nature research

Corresponding author(s): Towfique Raj

Last updated by author(s): May 17, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	All custom code is available in the GitHub repository: https://github.com/RajLabMSSM/MyND-Analysis RNASeq was processed using the RAPiD 3.0 pipeline. RAPiD aligns samples to the hg38 genome build using STAR (v2.7.2a) software with GENCODE v30 transcriptome reference and calculates quality control metrics using FASTQC (0.11.8), Picard (v2.20), samtools (v1.9). Estimated transcript abundance was obtained using RSEM (v1.3.1) and transcripts were summed to the gene level with tximport (v1.20.0) # Softwares used for downstream analysis FCS Express 6 Flow Sources of variation from RNA-seq data: variancePartition (v1.17.7) Differential Expression Analysis: Monocyte and PPMI Samples: R package limma (v3.38.3); Microglia Samples: R package "Differential expression for repeated measures" (DREAM) from the variancePartition (v1.17.7). Pathway analysis: Gene Set Enrichment Analysis (GSEA) Co-expression Network Analysis: Weighted Gene Correlation Network Analysis (WGCNA v1.68) Calling local splicing events or intronic excision: Regtools (v0.5.1) and Leafcutter (0.2.9) Genotype, genetic association and QC: PLINK (v1.9), vcftools (v0.1.14) Imputing data: Michigan Imputation Server v1.0.4 (Minimac 3); 1000 genomes phase 3 v5 mixed panel and eagle v2.3 phasing; https://imputationserver.sph.umich.edu/index.html Mapping QTL: PEER (v1.3); QTLtools (v1.1) Colocalization: R package Coloc (v3.2) Fine-mapping: echolocatoR https://rajlab.shinyapps.io/Fine_Mapping_Shiny Co-expression Heritability: Stratified LD score regression (S-LDSC v1.0.1) eQTL Heritability: mediated expression score regression (MESC) eQTL Replication: R package q-value (v2.15.0)

Single Cell RNAseq: Seurat (v3.1.0); monocle3 (v0.1.2)

All software is freely available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Bulk and single cell RNA-seq raw data and genotypes from this study have been submitted to the NCBI database of Genotypes and Phenotypes (dbGaP; https://www.ncbi.nlm.nih.gov/gap) under accession number phs002400.v1.p1. Processed read counts and full eQTL summary statistics are available from Zenodo online data sharing portal (in progress). The link to download the data: https://zenodo.org/record/4715907. The raw RNA-seq and genotype data are available via dbGAP (Accession:phs002400.v1.p1; data will be available to download upon publication).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed for this study. The number of samples after quality control (Monocytes RNA-seq n = 230; Monocytes single-cell RNA-seq n = 10; Microglia RNA-seq n = 128, N = 55; Monocytes eQTLs n = 180), was determined by the availability of high quality samples.
Data exclusions	For monocyte RNA-seq samples, exclusion criteria was to remove sampl s with < 20% reads reads mapping coding rgions and/or <20 million passed reads and/or > 30% ribosomal rate. Also, samples were excluded if there were sex mismatches. All monocyte RNA-seq samples that passed QC and has European origin were used for eQTL analysis. No sample exclusion was applied to monocyte single cell RNA-seq For microglia RNA-seq samples the exclusion criteria was: samples with <10 million reads aligned to the reference genome and/or samples with <10% reads mapping to coding rgions and/or samples with >20% reads aligned to ribosomal regions
Replication	Our results were successfully replicated in several external datasets: 1. For monocyte RNA-seq data we replicated: - AMP-PD dataset: whole blood transcriptomic data from Control and Parkinson's disease donors. Data was obtained from the AMP-PD Knowledge Platform (https://amp-pd.org/), and filtered for unique donors. It is no a direct replication as our data has been generated in pure monocyte samples whereas AMP-PD is whole blood transcriptomics. Nevertheless, we observe a positive correlation (Fig. 2D) as well as a significant overlap of the differentially expressed genes in our dataset and the AMP-PD dataset (P-Value = 0.041, Fisher-exact test) - We also validated bulk RNA-seq results in the single cell RNA-seq data that we generated with independent donors. We analyzed the single-cell RNA-seq data without considering clusters and we observed a positive correlation of the differentially expressed genes in bulk and single cell data (Spearman p=0.59, P-value=5.04x10-6) (Fig 2.E) - We validated targeted differentially expressed genes by qPCR (Fig. 5C) 2. For microglia RNA-seq data we replicated: - We assessed an an independent dataset obtained from meta-analysis of 8 studies from bulk brain substantia nigra (Wang et al. 2019). It is not a direct replication (isolated microglia in our dataset vs bulk tissue for replication) due to the absence of purified human microglia data from Parkinson;s disease. Nevertheless, we observed that the directionality for the significant pathways (mitochondria, proteasome and S100 genes) was replicated in the substantia nigra dataset - We also replicated in a dataset of single nuclei from Parkinson's disease sample, recently posted in biorxiv work (Smajic et al. 2020). Again, it is not a direct replication (isolated microglia vs single nuclei). Nevertheless, we observed a significant overlap between the differentially expressed genes at FDR < 0.1 that we found and the genes associated to disease-activated microglia that the authors describe (Fisher's exact test, P-value = 0.041, OR = 6.56) - We differentiated monocytes to monocyte-derived macrophages and performed targeted qPCR, validating the directionality in gene expression. 3- For eQTL analysis: We estimated replication of MyND monocyte cis-eQTLs (discovery) using CD14+ eQTL data set from Fairfax et al. (replication) using the q-value R package to estimate π1.

Randomization	Samples were sequenced in different batches were diagnosis, sex, and other technical and biological covariates were evenly distributed to avoid batch effects. For other than sequencing, samples were randomly distrubuted maintaining same number of PD and Controls
Blinding	Cells were isolated (monocytes and microglia) in a blinded manner as clinical data was obtained subsequently

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Primary human CD14+ monocytes were freshly isolated from 230 donors, including Controls (n=95) and Parkinson's disease (n=135) patients. The mean of age is 67 years old. The average in age of diagnosis is 57 years old with a mean in disease duration of 8.3 years. Primary human CD11b+ microglia were isolated at autopsy from 55 donors with clinical diagnosis as Parkinson's disease (N=13), as well as unaffected subjects (Controls, N=42), generating a total of 128 samples (Controls n=106; Parkinson's disease n= 22) from the following brain regions: corpus callosum (CC; 13 samples), medial frontal gyrus (MFG; 40 samples), superior temporal gyrus (STG; 30 samples), thalamus (THA; 23 samples), sub-ventricular zone (SVZ; 18 samples) and substantia nigra (SN; 1 sample) The average age of death is 80 years old for Control samples and 78.5 for Parkinson's samples. Mean of disease duration is 13.5 years old. All analyses were adjusted for age, sex, RIN, and other covariates. The details are described in the Methods section. All biological (sex, diagnosis, ethnicity, etc) and technical covariates (RNA concentration, RIN value, million reads, etc) are included in the Supplementary tables: Table S1: metadata for monocyte samples Table S2: metadata for microglia samples Table S12: metadata for single-cell RNA-seq monocyte samples
Recruitment	Recruitment was approved by the Institutional Review Board of all the institutions implicated in the study. For isolation of primary monocytes, participants were asked to donate blood (maximum 40 mL). Sample collection occurred during routine visits of the patients and families to the clinic in order to minimize inconvenience. Only subjects with diagnosis of Parkinson's disease following the UKBB criteria (Fahn et al, 1987) were included as cases. The aged-matched healthy controls were subjects that did not carry a diagnosis of neurodegenerative disease (such as Parkinson's disease, atypical Parkinsonism or Alzheimer disease, among others) or chronic autoimmune disorder as obtained from their medical history. Subjects who did not fulfill these inclusion criteria were not included in the study. Post-mortem brain samples were obtained from the Netherlands Brain Bank (NBB) and the Neuropathology Brain Bank and Research CoRE at Mount Sinai Hospital. We have tried to avoid any bias. All details for controlling confounders are outlined in the Methods section.
Ethics oversight	The permission to obtain blood was approved by the Institutional Review Board of all the clinics that participated in the study. The permission to collect human brain material was obtained from the Ethical Committee of the VU University Medical Center, Amsterdam, The Netherlands, and the Mount Sinai Institutional Review Board. In all cases, donors were informed and consented to participate in the study. In all cases, procedures tried to minimize inconvenience to donors and families.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	FACS was used to assess (i) the purity of the monocyte population and (ii) the proportion of monocyte subclusters. For the purity, briefly, monocytes were isolated following the CD14+ selection described in the Methods section. Cells were labeled with CD14-PE (Biolegend) and the Live-dead marker (Far Red) (Thermofisher) for 15 mins in ice and no light at the dilution recommended by manufacturer. For the proportion of cells, the protocol was the same although samples were labeled with CD14-PE (Biolegend), CD45-FITC (Biolegend), CD16-APC (Biolegend) and Live-dead marker (Violet) (Thermofisher), 15 mins in ice and no light at the dilution recommended by manufacturer. Afterwards, samples were washed with PBS+BSA and data was acquired
Instrument	LSRFortessa X-20
Software	FCS Express 6 Flow
Cell population abundance	Gates were designed using beads and unlabeled cells. Based on those gates, abundance was determined as percentage of cells falling on those gates.
Gating strategy	Beads were used for compensation and designment of the gates Gates were confirmed using unlabeled cells to avoid any autofluorescence issues. Gating strategy is summarized in Fig. S3A.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.