nature portfolio

https://doi.org/10.1038/s43587-021-00110-x

Supplementary information

Dysregulation of mitochondrial and proteolysosomal genes in Parkinson's disease myeloid cells

In the format provided by the authors and unedited

Dysregulation of mitochondrial and proteo-lysosomal genes in Parkinson's disease myeloid cells

Authors

Elisa Navarro[1,2,3,4,†], Evan Udine[1,2,3,4,†], Katia de Paiva Lopes[1,2,3,4], Madison Parks[1,2,3,4], Giulietta Riboldi[5,6], Brian M. Schilder[1,2,3,4], Jack Humphrey[1,2,3,4], Gijsje J. L. Snijders[7,8], Ricardo A. Vialle[1,2,3,4], Maojuan Zhuang[1,2,3,4], Tamjeed Sikder[1,2,9,10], Charalambos Argyrou[1,2,3,4], Amanda Allan[1,2,3,4], Michael Chao[1,2,3,4], Kurt Farrell[1,2,9,10], Brooklyn Henderson[5], Sarah Simon[4,11], Deborah Raymond[4,11], Sonya Elango[4,11], Roberto A. Ortega[4,11], Vicki Shanker[4,11], Matthew Swan[4,11], Carolyn W. Zhu[12,13,14], Ritesh Ramdhani[15], Ruth H. Walker[4,16], Winona Tse[4], Mary Sano[12,13,14], Ana C. Pereira[1,4], Tim Ahfeldt[1,2,4], Alison M. Goate[1,2,3,4], Susan Bressman[4,11], John F. Crary[1,2,9,10], Lotje de Witte[7,8], Steven Frucht[5], Rachel Saunders-Pullman[4,11], Towfique Raj[1,2,3,4*].

[1]Nash Family Department of Neuroscience & Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[2]Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[3]Department of Genetics and Genomic Sciences & Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[4]Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[5]The Marlene and Paolo Fresco Institute for Parkinson's Disease and Movement Disorders, New York University Langone Health, New York, NY, United States of America.

[6]Universita degli Studi di Milano, Molecular and Translational Medicine, Milan, Italy.

[7]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, United States of America

[8]Mental Illness Research Education Clinical, Centers of Excellence, VA, Mental Health, Veterans, Bronx, United States of America.

[9]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[10]Neuropathology Brain Bank & Research CoRE, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[11]Department of Neurology, Mount Sinai Beth Israel, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[12]Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[13]Geriatric Research, Education and Clinical Centers (GRECC), James J. Peters VA Medical Center, Bronx, New York, NY, United States of America.

[14]Alzheimer's Disease Research Center (ADRC), Icahn School of Medicine at Mount Sinai, New York, NY, United States of America.

[15]Department of Neurology, Zucker School of Medicine at Hofstra Northwell, New York, NY, United States of America.

[16]Department of Neurology, James J. Peters VA Medical Center, Bronx, NY, United States of America.

†These authors contributed equally

*Correspondence should be addressed to T.R. (towfique.raj@mssm.edu)

Supplementary Notes

Demographics and Clinical Overview of Study cohorts

In this study, 230 samples (including controls and PD patients) have been included. This study complied with all ethical regulations and was approved by the Institutional Review Board of each institution. Sample collection occurred during routine visits of the patients to the clinic, minimizing inconvenience to the patients and their families. No compensation was provided to participants. Written informed consent was obtained from all participants for the collection of samples and subsequent analysis. Samples were recruited from the following clinical cohorts: Movement Disorder Center at Mount Sinai Beth Israel (MSBI), Bendheim Parkinson and Movement Disorders Center at Mount Sinai (BPMD), Fresco Institute for Parkinson's and Movement Disorders at New York University (NYUMD), the Alzheimer's Research Center (ADRC) and Center for Cognitive Health (CCH) at Mount Sinai Hospital. Only subjects with diagnosis of Parkinson's disease following the UKBB criteria were included as cases (see below). 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.

Mount Sinai Movement Disorder Centers (BPMD and MSBI)

PD subjects were recruited from two Mount Sinai Movement Disorder Centers: The Robert and John M. Bendheim Parkinson and Movement Disorders Center (BPMD) and the Mount Sinai Beth Israel (MSBI). Study participants from BPMD and MSBI included PD cases and family and friend controls who were part of two genetic studies of PD, one focused on gene identification in PD, and a second on ascertainment of biologic markers of glucocerebrosidase mutations. Affected individuals met the UK Parkinson's disease Society Brain Bank (UKBB) criteria for probable PD [4], except that family history of PD was not an inclusion criteria. Controls did not have a history of a neurodegenerative disorder. All gave informed consent. Family history and pedigree were ascertained. A movement disorder trained neurologist assessed clinical features as well as performed the Unified Parkinson Disease Rating Scale (UPDRS)[5]. Medical history and medications were captured.

New York Movement Disorder (NYUMD)

Patients affected with PD and controls were enrolled at the Marlene and Paolo Fresco Institute for Parkinson's disease and Movement Disorders by Movement Disorder specialists between March 2018 and December 2019. Inclusion criteria were a diagnosis of PD according to the United Kingdom Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria and age between 18 and 100 years [6]. A population of aged and sex-matched non-affected subjects was enrolled among subjects who did not have a known diagnosis of PD at the time of evaluation and no history of other relevant neurological conditions. For each enrolled subject, both PD patients and controls, the following assessments were performed by qualified personnel and the following data were collected: demographic information (age, sex, self-reported ancestry), family history of PD, hand dominance, Montreal Cognitive Assessment (MoCA), UPDRS, H&Y rating scale, self-reported presence of the following motor and non-motor symptoms associated with PD (constipation, urinary symptoms, symptomatic orthostatic hypotension, subjective loss of sense of smell, REM sleep behavior disorders, hallucinations, anxiety, depression, motor fluctuations, dyskinesia, dopamine-related impulse control disorders), medication history, history of concomitant clinical conditions with particular attention of inflammatory diseases. For PD patients only the following information was also collected: age of onset of the disease (since onset of motor symptoms), symptoms at onset, and PD motor subtype (tremor dominant (TD) versus postural instability and gait difficulty (PIGD)). Data were collected in a password-protected database. Only subjects with the cognitive capacity to understand the study procedures, and the risks and benefits of the study, as assessed by licensed clinicians and established based on MoCA score greater or equal to 22, were enrolled.

Mount Sinai Alzheimer's Disease Research Center

The Alzheimer's Disease Research Center (ADRC) at the Icahn School of Medicine at Mount Sinai is a comprehensive research facility and clinical program dedicated to the study and treatment of normal aging and Alzheimer's disease. With research into the causes of dementia, diagnostic services, and caregiver programs, the ADRC seeks to improve diagnosis, delay disease progression, and enhance the well-being of those affected by AD. The ADRC recruits participants who are cognitively normal, MCI, AD

and other dementia into the National Alzheimer's Coordinating Center Uniform Data Set (NACC UDS). Participants are followed annually and provide permission to contact them as additional studies become available including studies for the contribution of biosamples for new or ongoing projects. DNA is banked both locally and through the National Cell Repository for Alzheimer's Disease (NCRAD). All subjects included have a Clinical Dementia Rating (CDR) of 0, thus only subjects cognitively normal were included in this study as controls.

Mount Sinai Center for Cognitive Health (CCH)

Through collaborations with Mount Sinai ADRC, the CCH evaluates individuals with concerns about cognition, such as memory, language and thinking difficulties. At the initial visit, a comprehensive neurological examination is performed which includes a thorough review of medical history, social history and detailed descriptions of cognitive complaints, and changes in behavior. The neurologists may do cognitive testing to assess memory, language, visual processing and other symptoms related to thinking. All subjects recruited from the CCH have a Mini-Mental State Exam (MMSE) test of cognitive function. Subjects that are cognitively normal based on cognitive testing were included in this study.

Genotype Quality Control and Imputation

We applied genotype quality control (QC) metrics such as SNP call rate > 95%, minor allele frequency (MAF) > 5%, Hardy-Weinberg equilibrium (HWE) *P-value* > 1 x 10^{-6}, sample call rate > 95% to prepare high quality genotype data. Duplicated/related samples were determined based on pairwise IBD (identity-by-descent) estimation using PLINK [7]. Duplicated samples with PLINK PI_HAT values between 0.99 to 1 were identified and files were converted to Variant Call Format (VCF) using VCFTools. Genetic ancestry of samples was confirmed by principal components analysis using PLINK; MDS values of study subjects were compared to those of 1000 Genome Project samples (Phase 3). An AJ only analysis was completed using a custom reference panel following that same protocol, and 82 samples (35.6%) were of AJ ancestry, 58 PD cases and 24 controls (**Extended Data Fig. 1B, Fig. S8**).

Genotype Imputation was done using the Michigan Imputation Server v1.0.4 (Minimac 3) [8] using the 1000 genomes phase 3 v5 mixed panel and eagle v2.3 phasing in QC and imputation mode. The following filtering was applied post imputation: SNPs that had imputation $R^2 > 0.3$, removing multi-allelic SNPs, filtering MAF > 5% and HWE *P-value* > 1×10^{-6}, and removing indels. We performed liftover of the imputed VCF files to hg38 using hg19toHg38 liftover chain file from UCSC Genome browser and liftoverVCF from Picard to match imputed genotypes to the hg38 reference used for RNA-seq. After imputation and liftover, a total of 5,951,770 variants were included in downstream analyses.

Understanding sources of expression variation and covariate selection

As highly correlated covariates cannot be included in the model, we selected covariates that were not very strongly correlated to run the variancePartition analysis (**Fig. S3A, B**). Gene counts were normalized using TMM values calculated from edgeR and *voom* transformed, [9] as input to variancePartition. We ran several differential expression (DE) models and tested the suitability of each of them. In all the models we included sex, age, RIN value and the four MDS from the genotypes (which are enough to separate different populations, see *Genotyping and QC* section) as covariates. For each model, we added different covariates ranging from the simplest (no additional covariates) to the most complex (all covariates from variancePartition). None of the covariates were correlated with diagnosis except an expected correlation with sex. In all cases, there was a high fold change correlation between DEGs generated from the different designs (>0.90) and high percentage of shared DEGs and similar pathways. We decided to use a design which includes those covariates that explained the most variance in gene expression (on average across genes) according to variancePartition results, which is as follows: *expression ~ rna_batch + age + sex + RIN + PCT_USABLE_BASES + PCT_RIBOSOMAL_RNA + MDS1 + MDS2 + MDS3 + MDS4*.

Splicing analysis

Differential splicing (DS) was assessed using Leafcutter [10]. Leafcutter pools splice junction-spanning reads from each sample together and clusters junctions that overlap at either end. Differential splicing is then defined in a shift of junction usage within a cluster between two groups. Firstly, splice junction reads were extracted from each BAM file using regtools [11] and any junction reads aligned to scaffold chromosomes were removed. All junction files were clustered using leafcutter_cluster_regtools.py, specifying for each junction in a cluster a maximum length of 100kb. This led to 194,127 junctions within 45,631 clusters. We used a custom script to restrict our analysis set to junctions present in at least 25% of samples contributing at least 5% of the total reads to their cluster (**Extended Data Fig. 3A**). Any cluster with only a single junction remaining after filtering or more than 10 junctions were removed. This led to a final set of 22,888 junctions within 8,882 clusters. Differential splicing between PD cases and controls was performed, testing each cluster if it was present in at least 50 samples per group with a minimum coverage of 20 reads in total. The same covariates were used in model fitting as for the differential expression analysis. Results were visualized using the LeafViz browser. Junction ratios were corrected for covariates using the "quantify_PSI.R" script found within the "psi_2019" branch of Leafcutter.

Parkinson's Disease Progressive Marker Initiative (PPMI) RNA-seq data analysis

RNA-seq counts and TPMs generated from whole blood from the PPMI (part of the AMP-PD cohorts) were downloaded from AMP-PD Knowledge Platform [12]. AMP-PD has RNA-seq FASTQ files and workflow products from Salmon, STAR, and featureCounts for the PPMI cohort. All RNA sequencing was performed by Hudson Alpha and is supplied along with corresponding clinical data. RNA-seq samples from the baseline visit were extracted for subjects with sporadic PD and controls. We followed a similar QC pipeline to our monocyte transcriptome analysis (**Fig. S5**). Differential expression analysis was performed with the read counts generated from rsubread featureCounts using the R package limma, adjusting for the following covariates: *% usable bases + % intergenic bases + sex + age + race + ethnicity + plate* (**Fig. S5**).

Heritability analysis

Stratified LD score regression (S-LDSC) [13] was used to partition SNP-based disease heritability within each co-expression module. Using GWAS summary statistics from PD and LD modeled from 1000 genomes reference panel of European ancestry, we calculated the proportion of genome-wide SNP-based heritability that be attributed to SNPs (by mapping SNPs within each gene plus 10 kb +/- from transcript start and stop sites) within each module. To improve model accuracy, the LD-scores from each co-expression module were added to the 'full baseline model' which included 53 functional categories capturing a broad set of functional and regulatory elements. Enrichment is defined as the proportion of SNP-heritability accounted for by each module divided by the proportion of total SNPs within the module. Modules with FDR-corrected enrichment *P-values* of less than 0.05 were considered significant heritability contributors.

Mediated expression score regression (MESC) method [14] was used to estimate disease heritability mediated by the *cis* genetic component of gene expression levels. The expression scores were estimated using eQTL summary statistics from monocytes (this study), microglia and DLPFC. The MESC python script also estimates the expression *cis*-heritability of each gene (using LDSC for eQTL summary statistics). The expression scores and GWAS summary statistics (PD, AD, Schizophrenia, and Height) [15–18] were then used to estimate expression-mediated heritability (h2med).

Single-cell RNA-seq data QC

Extensive QC and doublet removal was also performed by the NYGC according to their previously described protocol [19]. Briefly, barcoded reads were demultiplexed following the Cell Hashing technology [19]. Each cell barcode that had more than 200 nUMI were classified according to DigitalExpression. HTO raw counts were normalized using centered log-ratio (CLR) transformation. Then, we ran k-means clustering on these normalized counts using k = number of HTOs. Each cluster from the k-means was assigned to an HTO, and cells that were not in that cluster and were in the top 0.5% of the remaining cells were excluded. We then fitted a negative binomial distribution to the remaining HTO values (raw, not log-normalized), calculated the q=0.99 quantile of the fitted distribution, and thresholded each cell in the dataset based on this HTO-specific value. Barcodes that were positive for only one HTO were classified

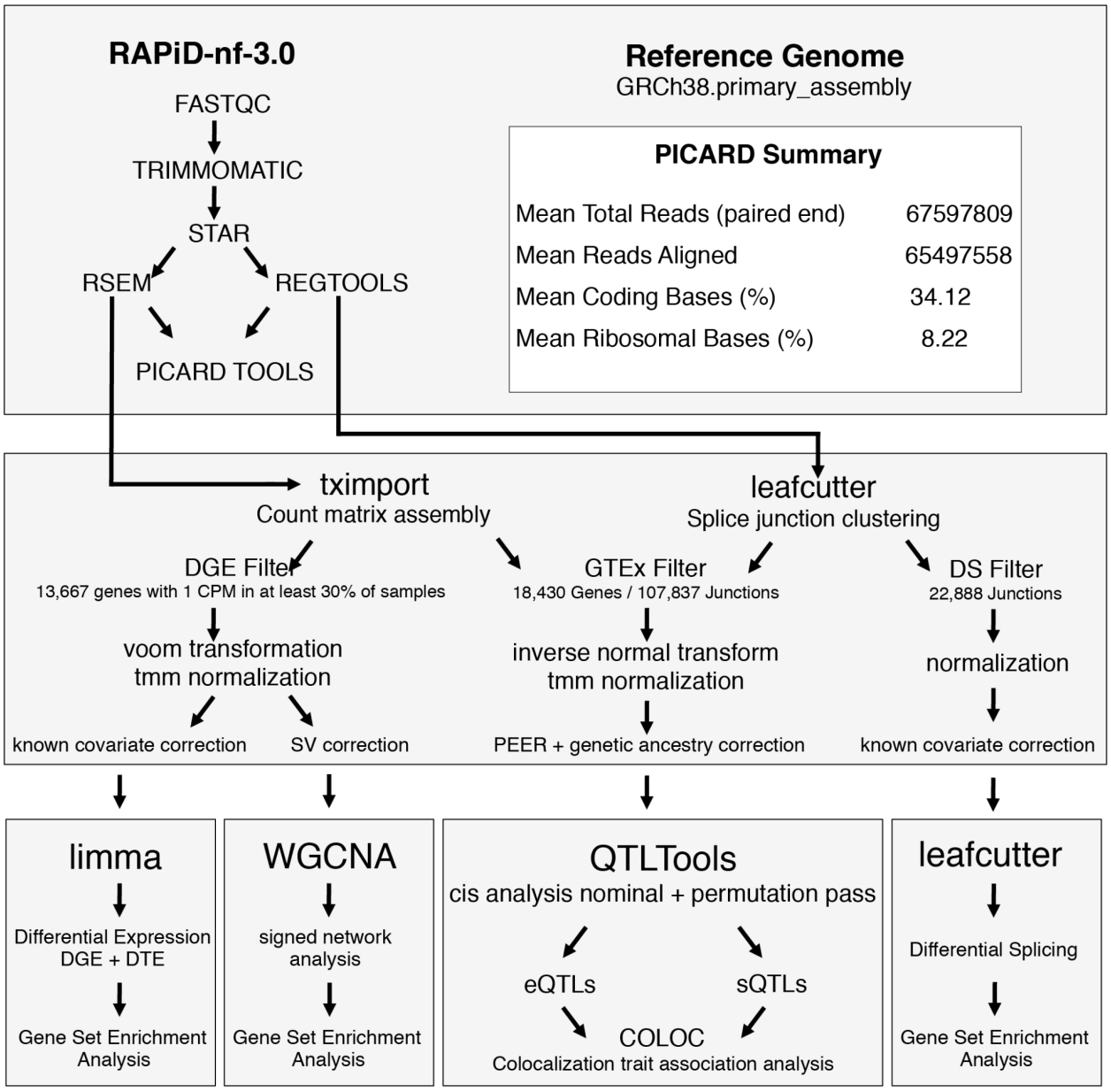
as singlets. Barcodes that were positive for two or more HTO were classified as multiplets. Barcodes that were negative for all HTOs were classified as negative. nly singlets were kept for downstream analysis. Number of cells per sample can be found in **Table S12** and **Fig. S7B**.

Flow cytometry

FACS was used to assess (i) the purity of the monocyte population and (ii) the proportion of monocyte subclusters. For the purity, monocytes were labeled with CD14-PE and the Live-dead marker (Far Red) (Thermofisher) for 15 mins in ice and no light. For assessing the proportion of monocyte subcusters, 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 . Biolegend antibodies were used at a concentration of of 2.5 µl in 50 µl and Live-dead 1:1000 Afterwards, samples were washed with PBS + BSA and data was acquired in LSRFortessa X-20 and analyzed using FCS Express 6 Flow. Compensation was made using beads and unlabeled cells. Gating strategy: only singlets were selected based on SCC-W and FSC-A. From there, live cells were selected based on Live-dead marker (using dead cells as positive control). From there, positive cells were selected based on the cellular markers.

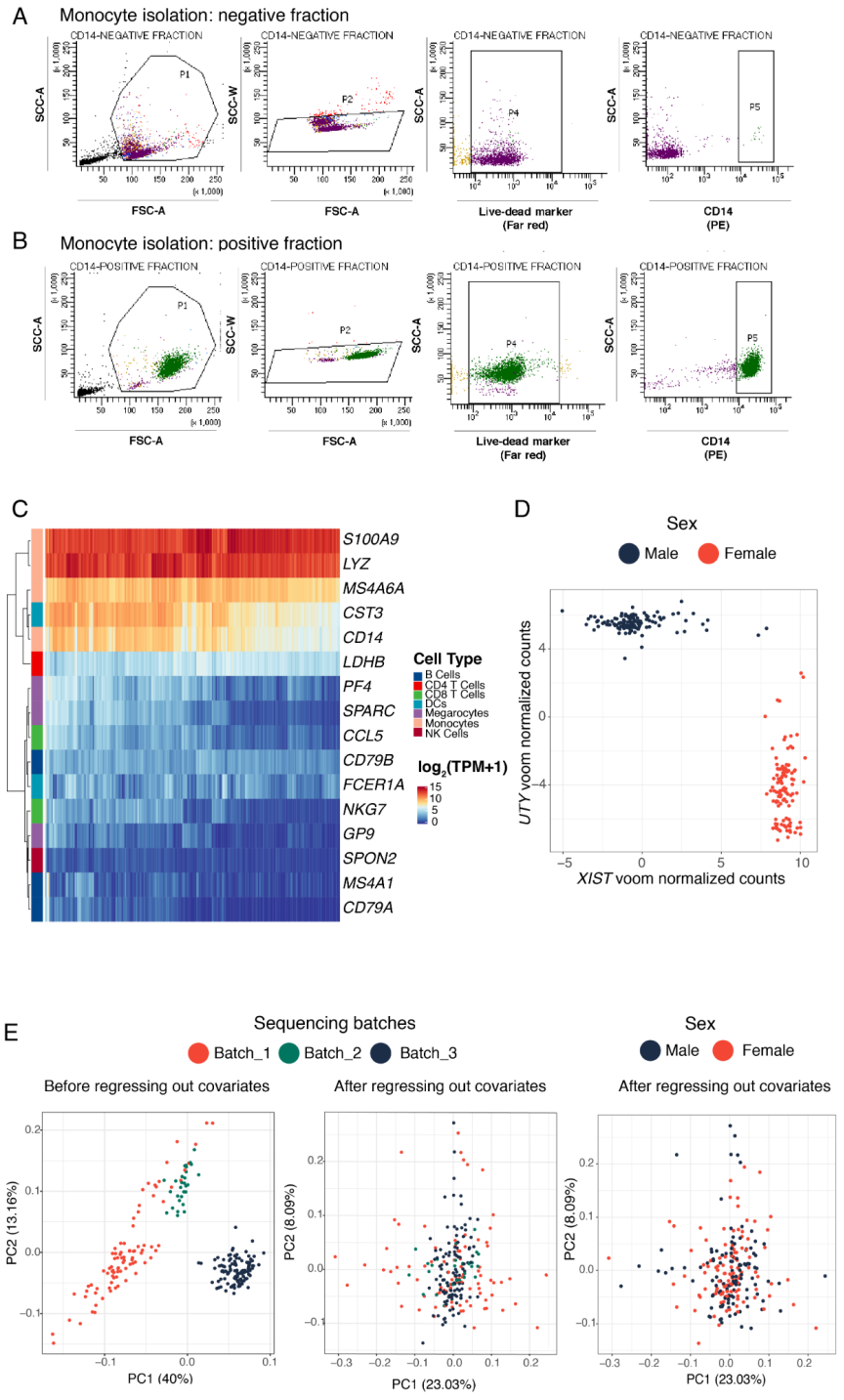
Supplementary Figures

Supplementary Figure 1



Supplementary Figure 1. Diagram representing the computational approach used in this study. (A) Overview of RNA-seq analysis pipeline. RAPiD pipeline of mapping, QC, and quantification. **(B)** Normalization, covariate selection, and outlier removal. **(C)** Downstream analyses plan for differential expression, splicing, network analysis, and quantitative trait loci analyses.

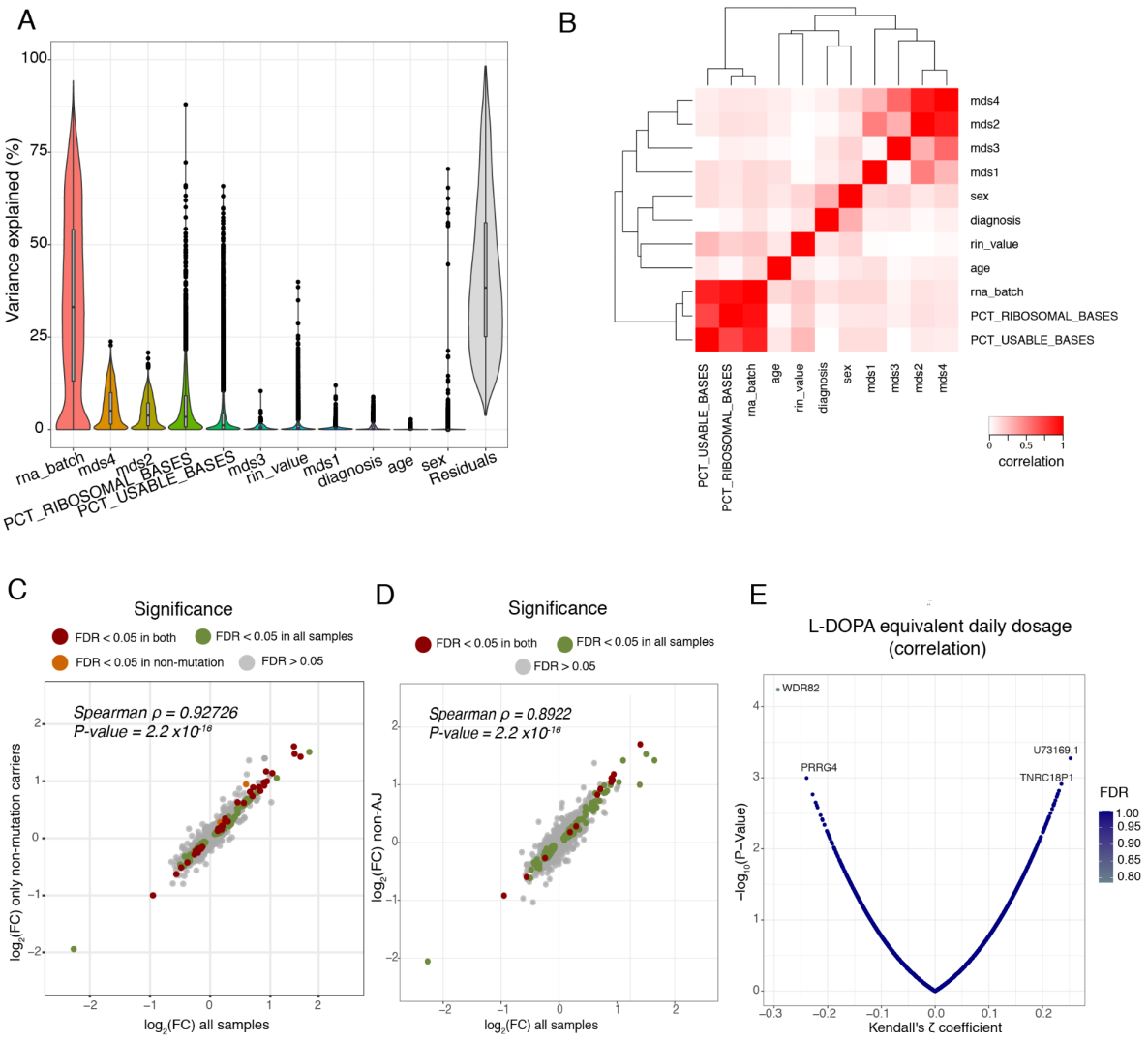
Supplementary Figure 2



Supplementary Figure 2. Purity of monocyte isolation and dimensionality reduction. FACS plots of the **(A)** negative and **(B)** positive fractions after CD14-magnetic selection. Purity

of the monocyte was 97.8%. **(C)** Heatmap representing in the y-axis the cell markers corresponding to different blood immune cells and the 230 samples in the x-axis. Color corresponds to the levels of expression (TPM + 1, $\log_2$ scale), blue representing low, and red meaning high. **(D)** Sex mismatch QC: scatter plot showing the voom transformed expression of *XIST* (X chromosome gene) in the x-axis and *UTY* (Y chromosome gene) in the y-axis. Each dot represents a sample and is colored by reported sex (red = female; blue = male). **(E)** PCA plots showing the voom transformed expression (TPMs) of the 230 samples before (left panel) and after (middle and right panels) regressing out known covariates. Each dot represents a sample. Left and middle panels are colored by sequencing batches (orange = batch_1, green = batch_2, blue = batch_3). Each batch has an equal proportion of case and control subjects. Right panel is colored by sex.

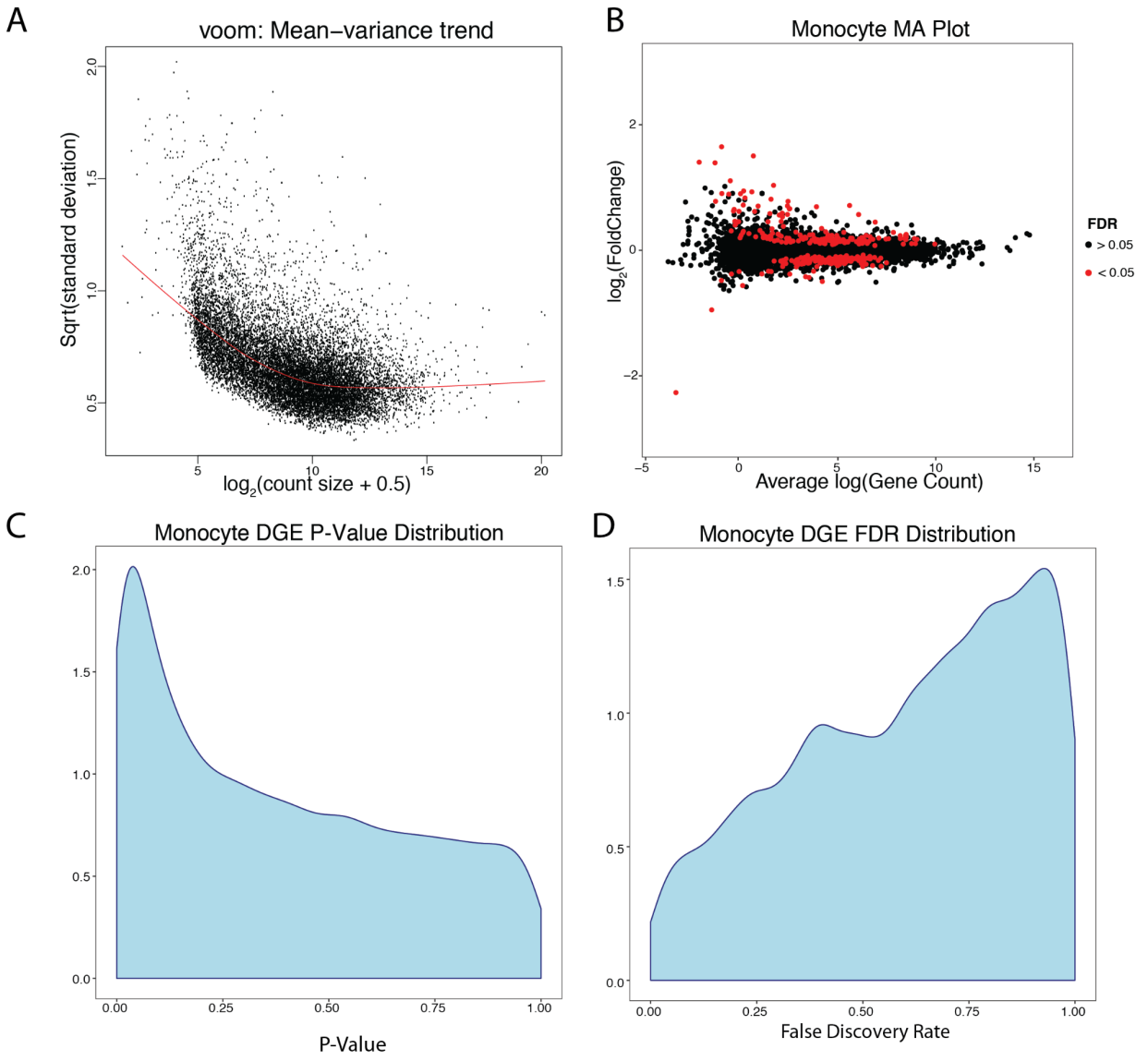
Supplementary Figure 3



Supplementary Figure 3. Covariate selection and correlation with possible confounders. Main sources of variation were analyzed using the variancePartition package [1]. **(A)** Violin plot showing the % of the variance (y-axis) explained by the covariates (x-axis) which contribute the most to the variability. Each dot represents a gene. n=230 independent samples. **(B)** Heatmap showing the correlation among the different covariates (red represents high correlation vs white low correlation). **(C)** Scatter plot representing fold-change correlation transcriptome-wide when

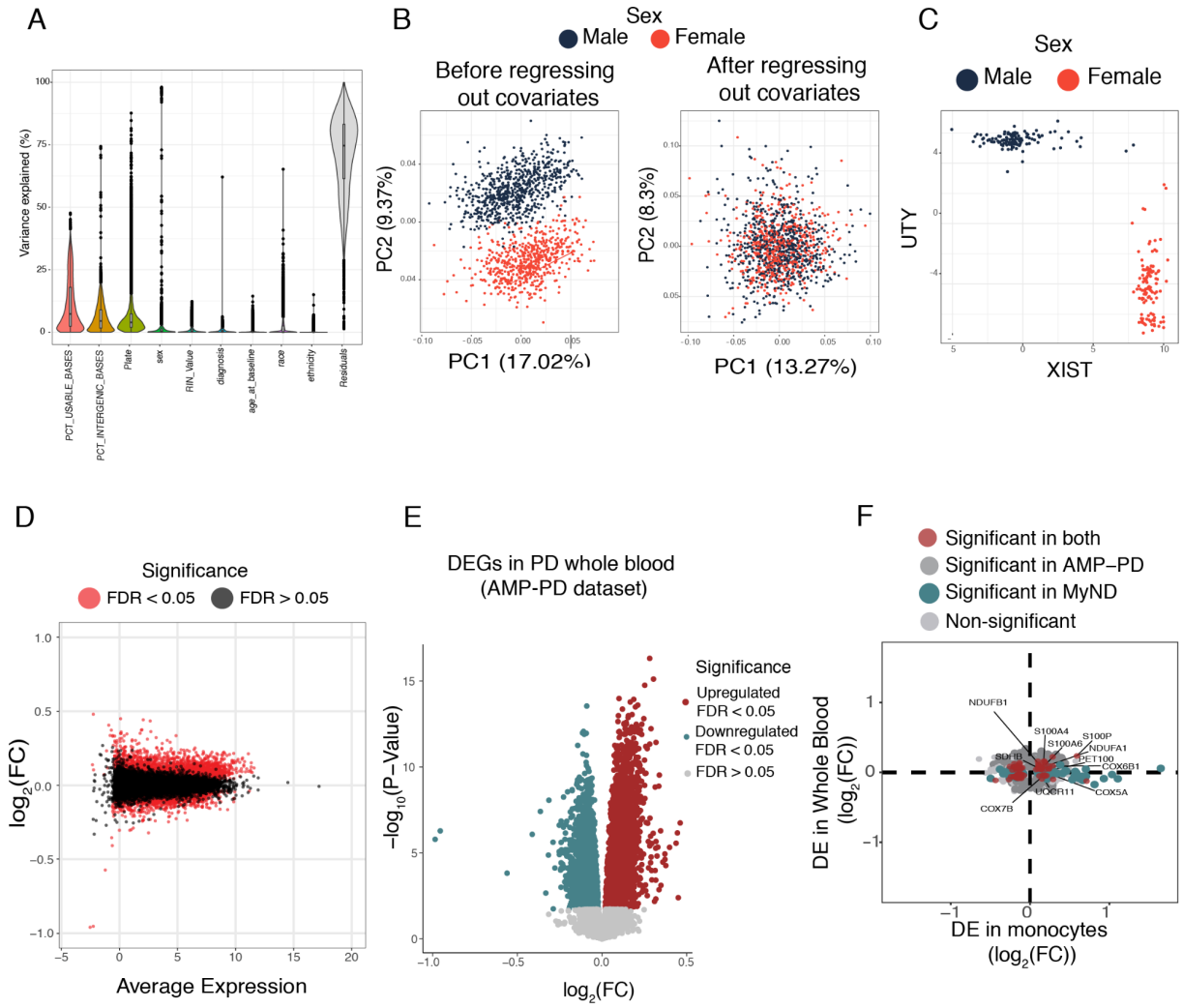
DE analysis was performed including all samples (n = 230) (x-axis) versus including only samples which are not *GBA* or *LRRK2* mutation carriers (n = 170) (y-axis). Spearman's Rank correlation. P-value < $2.2x10^{-6}$. **(D)** Scatter plot representing fold-change correlation transcriptome-wide when DE analysis was performed including all samples (n = 230) (x-axis) versus only samples with non-AJ ancestry (n = 148) (y-axis). Spearman's Rank correlation. P-value < $2.2x10^{-6}$. **(E)** Correlation of Levodopa equivalent daily dosage (LEDD) with gene expression levels for all genes tested in differential expression analysis. The four most significant genes are shown, although none met the FDR threshold. The FDR is indicated by the color of each dot. Kendall's rank correlation.

Supplementary Figure 4



Supplementary Figure 4. RNA-seq normalization and *P-value* distribution of differential expression. (A) Plots showing voom mean-variance trend with $\log_2$ (count size + 0.5) on the x-axis and square root of the standard deviation on the y-axis **(B)** MA plot showing fold-change ($\log_2$ scale) on the y-axis and mean of $\log_2$counts (x-axis). Genes differentially expressed at FDR < 0.05 are highlighted in red. **(C)** Distributions of uncorrected *P-values* (moderated t-statistic (two-sided)) and **(D)** FDR corrected *P-values* (moderated t-statistic (two-sided)).

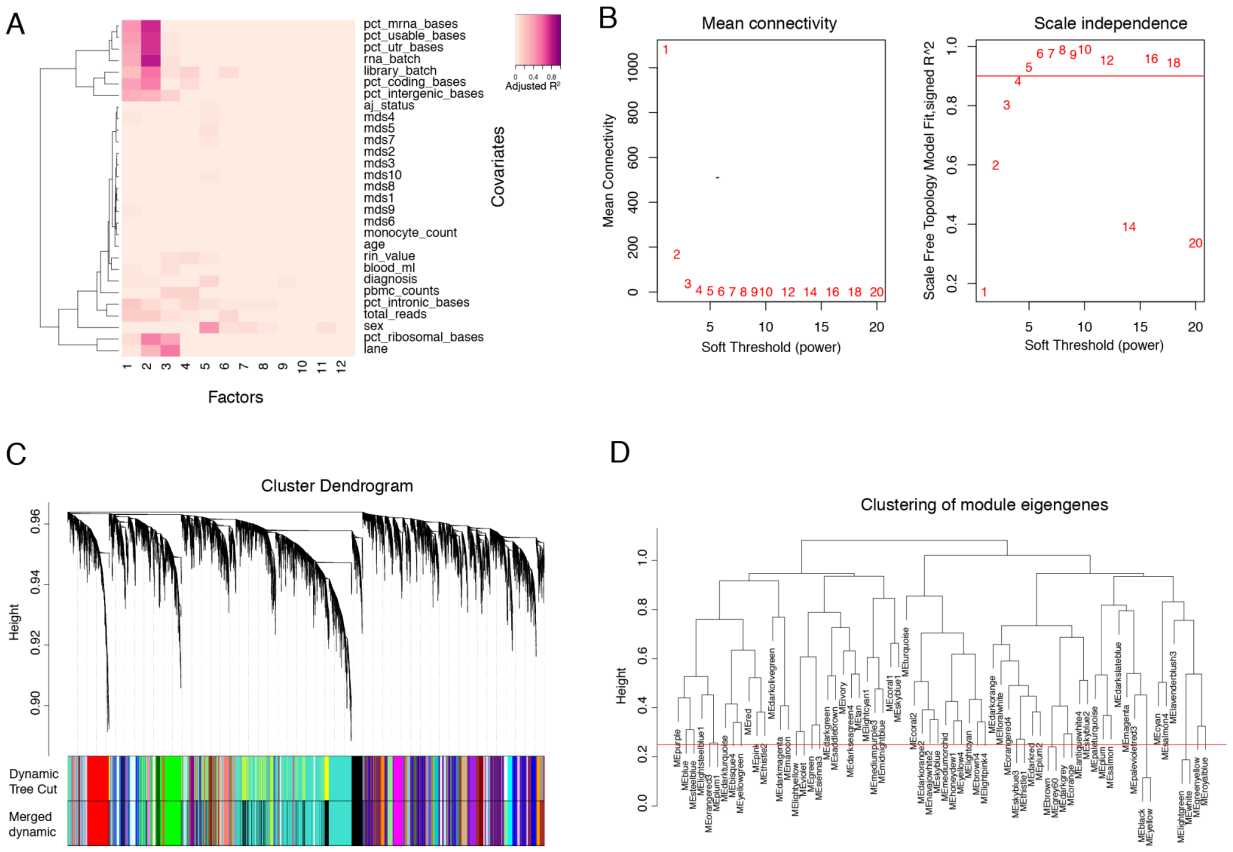
Supplementary Figure 5



Supplementary Figure 5. RNA-seq QC of the whole blood transcriptomic analysis from AMP-PD dataset. (A) Violin plot showing the % of variation (y-axis) explained by the covariates (x-axis). Each dot represents a gene. n=1284 independent samples. **(B)** Multidimensional reductionality using PCA of all the samples before (left panel) and after (right panel) regressing out the known covariates. Each dot represents a sample. Colored by sex (red = female, blue = male). **(C)** Sex mismatch QC: scatter plot showing the voom normalized expression of *XIST* (X-chromosome linked gene) in the x-axis and *UTY* (Y-chromosome linked gene) in the y-axis. Each dot represents a sample, colored by sex (red = female, blue = male). **(D)** MA plot showing

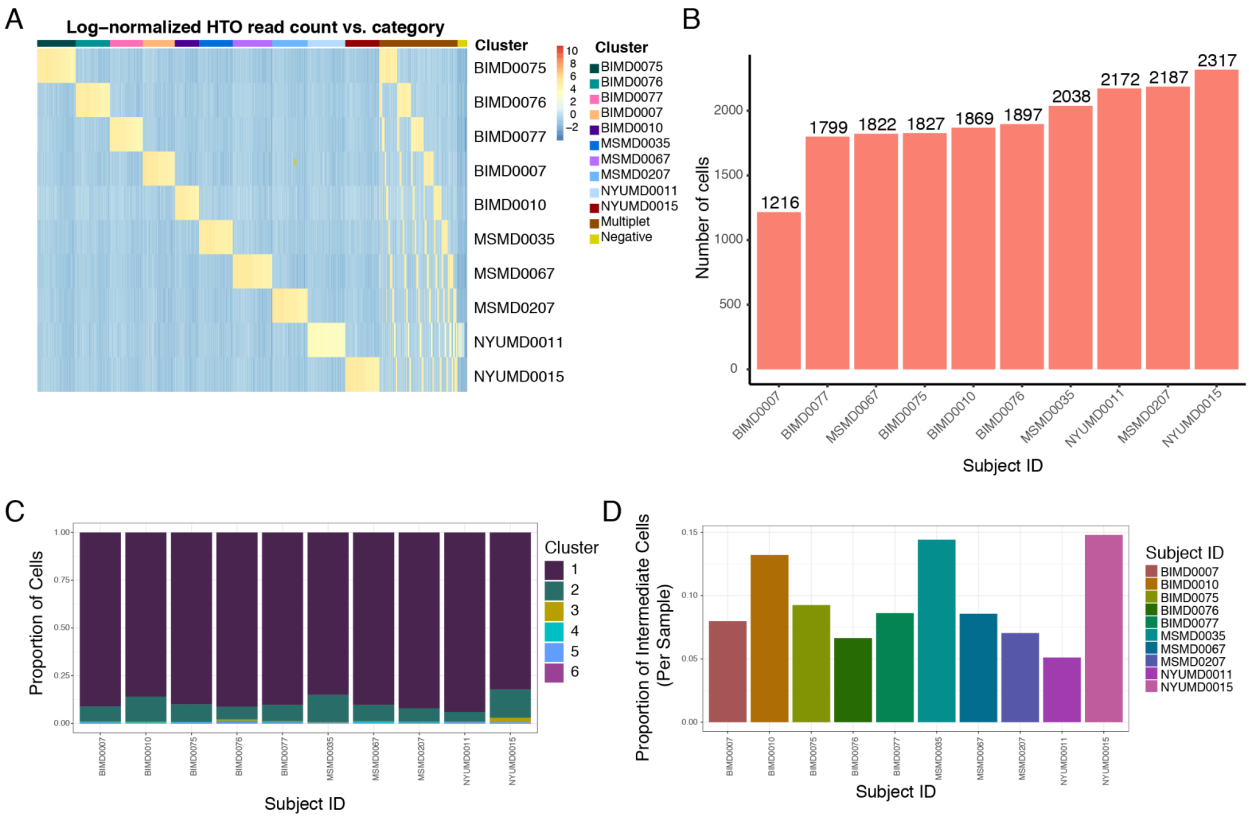
the fold-change (log_2 scale) of the DE cases vs controls in the y-axis and the mean of log_2counts (x-axis), highlighting the DEGs at FDR < 0.05 in red. **(E)** Volcano plot showing the fold-change (log_2 scale) of PD-whole blood (n = 780) and controls (n = 504) (x-axis) and their significance in the y-axis ($-log_{10}$ scale). Upregulated DEGs at FDR < 0.05 are highlighted in red and downregulated DEGs in blue. Moderated t-statistic (two sided) is used for statistical test (see R package limma). **(F)** Fold-change (log_2 scale) correlation of DEGs between MyND monocytes (x-axis) and AMP-PD whole blood (y-axis). Genes are colored by significance, considering significant DEGs at FDR < 0.05.

Supplementary Figure 6



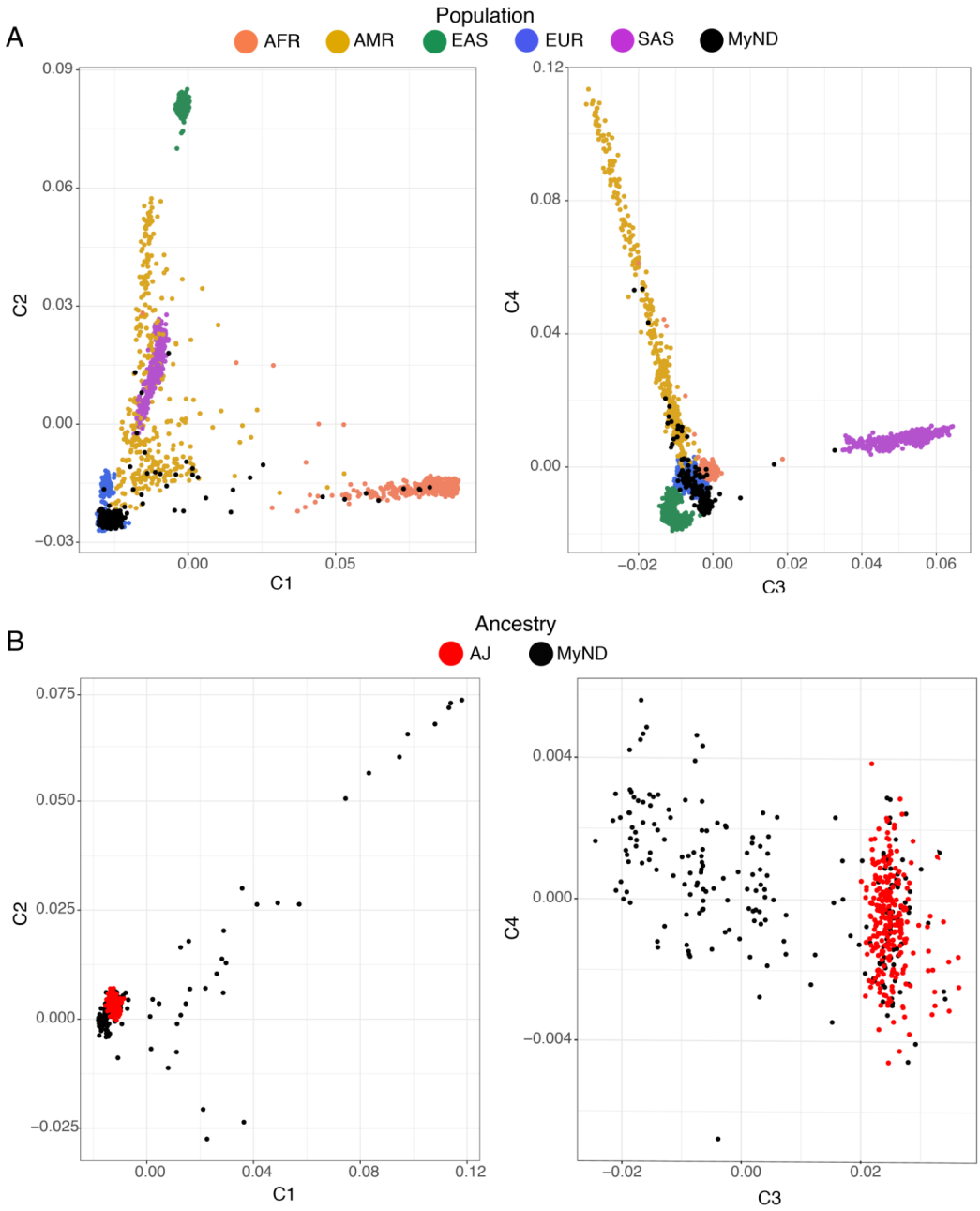
Supplementary Figure 6. Gene network construction in human monocytes using WGCNA. Using the 230 monocyte samples, 65 co-expression modules were obtained using WGCNA **(A)** Heatmap showing the correlation of the first 12 SVs (x-axis) and the known covariates (y-axis). **(B)** Right: Evaluation of network topology with different soft-thresholding powers. The y-axis represents the scale-free fit index as a function of the soft-thresholding power (x-axis). Left: The mean connectivity (y-axis) as a function of the soft-thresholding power. **(C)** Gene dendrogram using “Dynamic Tree Cut” to assign genes to different modules and modules to colors, showing before and after collapsing modules into 65 final networks. **(D)** Module eigengenes clustering dendrogram based on topological overlap. Modules below the threshold (Module Dissimilarity = 0.25) indicated by the red line were merged. These values correspond to a correlation of 0.75.

Supplementary Figure 7



Supplementary Figure 7. Single cell RNAseq QC. (A) Heatmap showing the HTO read counts (log transformed) for each of samples after demultiplexing. It includes the HTO read counts corresponding to Multiplets and Negative clusters that were removed for downstream analyses. **(B)** Number of cells in each sample that were used for subsequent analysis. **(C)** Cell proportions of the 6 sub-clusters obtained by unsupervised clustering with *monocle3* in scRNA-seq per individual (n = 3 controls and 7 PD). **(D)** Proportion of intermediate monocytes (cluster 2) per individual obtained by unsupervised clustering with *monocle3* in scRNA-seq (n = 3 controls and 7 PD).

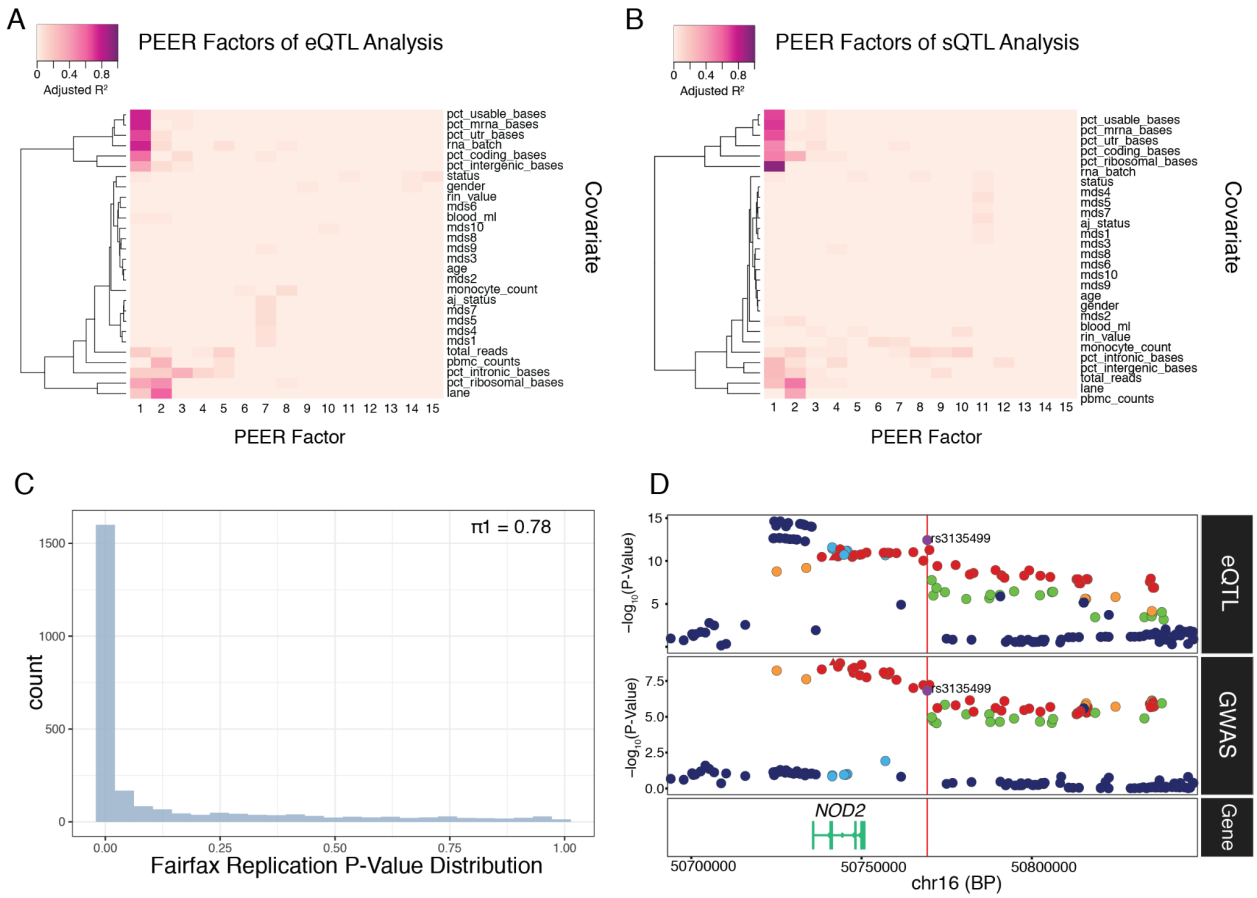
Supplementary Figure 8



Supplementary Figure 8. Analysis of population structure and ancestry estimates of subjects in MyND. (A) Ancestry of each subject was confirmed by PCA using PLINK (Purcell et al., 2007 Am J Hum Genet) and multidimensional scaling (MDS) values of study subjects were compared to those of 1000 Genome Project samples (Phase 3). AFR = African, AMR =

Admixed America, EAS = East Asian, EUR = European, SAS = South Asian. Samples of the MyND project are represented in black. **(B)** Multidimensional scaling values of study subjects were compared to those of Ashkenazi Jewish descent to determine ancestry.

Supplementary Figure 9



Supplementary Figure 9. Using PEER to regress out technical confounders in eQTL and sQTL analysis. Heatmap showing the correlation of the first 15 PEER factors (x-axis) and known covariates (y-axis) for the **(A)** eQTL analysis and **(B)** sQTL analysis**. (C)** Histogram showing replication of MyND eQTLs with Fairfax [2] eQTLs. **(D)** Regional association plot of monocyte eQTL (top panel) and PD GWAS (bottom panel) at the *NOD2* locus. The lead PD GWAS SNP rs6500328 is shown in red triangle. The lead eQTL SNP rs3135499 (in LD with rs6500328; $r^2 > 0.8$) is shown in purple. P-Values determined by linear regression, see QTLtools for details.

Supplementary Tables

Table 1: Metadata for monocyte samples. This includes RNA-seq sample ID, the sample ID, ancestry (based on genetics), age, sample collection date, sex, and diagnosis

Table 2: Metadata for microglia samples. This includes tissue sample ID, donor ID, donor diagnosis, main diagnosis, tissue region, donor sex, donor age, post mortem delay, and tissue pH.

Table 3: Full summary statistics from monocyte differential gene expression between PD cases and control. This table includes $\log_2$(FC), average expression, t-statistic (two-sided), *P-value*, adjusted *P-value* and B (posterior probability of DE).

Table 4: Pathway analysis results for all monocyte differential gene expression analyses. This table includes gene set, number of genes in the set, description of the gene set, number of genes that overlap, ratio of overlapping genes with number of genes in the set, *P-value*, and adjusted *P-value* (see GSEA for statistical methods).

Table 5: Full summary statistics from monocyte differential transcript expression summary. This table includes $\log_2$(FC), average expression, t-statistic (two-sided), *P-value*, adjusted *P-value* and B (posterior probability of DE).

Table 6: Full summary statistics from monocyte differential splicing analysis. This table includes the cluster ID, the change in percent spliced in, number of junctions in that cluster***,*** the genomics coordinates of the cluster, the gene the cluster is in, the cluster annotation, FDR, the coordinates of the most significant intron junction, the log effect size, the estimated mean junction contribution from control subjects, the estimated mean junction contribution from PD subjects, the *P-value*, the adjusted *P-value*. Annotations are included only for significant clusters (FDR < 0.05).

Table 7: Pathway analysis results for all monocyte differential transcript expression analyses. Analysis was performed independently for upregulated and downregulated genes. This table

includes gene set, number of genes in the set, description of the gene set, number of genes that overlap, ratio of overlapping genes with number of genes in the set, *P-value*, and adjusted *P-value* (see GSEA for statistical methods).

Table 8: Pathway analysis results for all monocyte differential splicing analyses. This table includes gene set, number of genes in the set, description of the gene set, number of genes that overlap, ratio of overlapping genes with number of genes in the set, *P-value*, and adjusted *P-value* (see GSEA for statistical methods).

Table 9: Full summary statistics from whole blood (AMP-PD) differential gene expression between PD cases and control. This table includes $\log_2$(FC), average expression, t-statistic (two-sided), *P-value*, adjusted *P-value* and B (posterior probability of DE).

Table 10: Full summary statistics from differential gene expression from monocyte scRNA-seq across-clusters between PD cases and controls. This includes the gene name, estimate, the standard error, the Wald statistic, *P-value*, normalized effect size, and q-value.

Table 11: List of top 10 hub genes from WGCNA co-expression networks based on module membership values. This table includes the gene, the module to which the gene is a member, and the module membership score.

Table 12: Metadata for monocyte scRNA-seq samples. This includes the sample ID, donor diagnosis, mutation status, reported ethnicity, donor sex, donor age, the propotion of cells per cluster from each sample, and the total number of cells per sample included in the analysis.

Table 13: Full summary statistics from monocyte differential gene expression from scRNA-seq comparing intermediate vs classical cluster. This includes the gene name, estimate, the standard error, the Wald statistic, *P-value*, normalized effect size, and q-value.

Table 14: Full summary statistics from monocyte differential gene expression from scRNA-seq between PD cases and controls in the intermediate cluster. This includes the gene name, estimate, the standard error, the Wald statistic, *P-value*, normalized effect size, and q-value.

Table 15: Full summary statistics from scRNA-seq marker gene identification for all 6 clusters

using monocle3. This table includes gene name, cluster number, marker score, mean gene expression in the cluster, fraction of cells expressing that gene in the cluster, the specificity to the cluster, the pseudo R^2, *P-value*, and q-value. Wald statistic is used for statistical test (see *monocle3*).

Table 16: Full summary statistics from microglia differential gene expression between PD cases and control. This table includes $\log_2$(FC), average expression, t-statistic (two-sided), *P-value*, adjusted *P-value* and z.std (*P-value* transformed signed z-score).

Table 17: Summary statistics for *cis*-eQTL permutation analysis between genes and the top variant in *cis* for all associations with FDR < 0.05*.* This table includes the Ensembl gene ID, the gene symbol, chromosome that the gene is located on, the gene start position, the gene position, strand orientation, number of variants tested in *cis,* distance between the lead variant and the gene, the ID of the lead variant, the chromosome of the lead variant, the variant start position, the variant end position, the number of degrees of freedom, a dummy column, the first parameter value of the fitted beta distribution, the second parameter value of the fitted beta distribution, the nominal *P-value*, the regression slope, the empirical *P-value*, the beta fitted *P-value*, and the FDR corrected beta fitted *P-value*. P-Values determined by linear regression, see QTLtools for details.

Table 18: Summary statistics for *cis*-sQTL permutation analysis between genes and the top variant in *cis* for all associations with FDR < 0.05*.* This table includes the Ensembl gene ID, the gene symbol, chromosome that the gene is located on, the gene start position, the gene position, strand orientation, the lead intron cluster, the number of clusters in the gene, the number of number of variants tested in *cis,* distance between the lead variant and the lead cluster, the ID of the lead variant, the chromosome of the lead variant, the variant start position, the variant end position, the number of degrees of freedom, a dummy column, the first parameter value of the fitted beta distribution, the second parameter value of the fitted beta distribution, the nominal *P-value*, the regression slope, the empirical *P-value*, the beta fitted

P-value, and the FDR corrected beta fitted *P-value*, and the nominal *P-value* threshold per phenotype. P-Values determined by linear regression, see QTLtools for details.

Table 19: Summary of PD GWAS and monocyte QTL (expression and splicing) colocalization analysis. The table includes the most recent PD GWAS reported locus, the lead SNP from GWAS in that locus, the colocalized QTL gene, the QTL lead SNP in LD with the lead GWAS SNP ($R^2 > 0.8$), the QTL lead SNP-Gene *P-value* and beta, the lead QTL reference allele, colocalization statistics for inclusion, and overlap of CD14 monocyte H3K27ac marks and microglia H3K27ac marks, ATAC-seq peaks, and PU.1 enhancer annotations.

Table 20: Summary of differential gene expression results from monocytes, whole blood (AMP-PD), microglia, and substantia nigra tissue [3]. The table includes the gene name, adjusted *P-value* t-statistic (two-sided) in monocytes, $\log_2$(FC) in monocytes, adjusted *P-value* t-statistic (two-sided) in whole blood, $\log_2$(FC) in whole blood, adjusted *P-value* t-statistic (two-sided) in microglia, $\log_2$(FC) in microglia, adjusted *P-value* in substantia nigra tissue, beta in substantia nigra tissue (standardized mean difference, see Wang et al.), and a column of Yes/No/NA for each cell type/tissue indicating if a gene is significant in that analysis. NA is given if the gene was not tested in that analysis. *Note that FDR < 0.15 is used in microglia to indicate significance.

REFERENCES

1. Hoffman, G. E. & Schadt, E. E. variancePartition: Interpreting drivers of variation in complex gene expression studies. doi:10.1101/040170.
2. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
3. Wang, Q. *et al.* The landscape of multiscale transcriptomic networks and key regulators in Parkinson's disease. *Nat. Commun.* **10**, 5234 (2019).
4. Daniel, S. E. & Lees, A. J. Parkinson's Disease Society Brain Bank, London: overview and research. *J. Neural Transm. Suppl.* **39**, 165–172 (1993).
5. FAHN & S. Members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. *Recent developments in Parkinson's disease* **2**, 293–304 (1987).
6. Hughes, A. J., Ben-Shlomo, Y., Daniel, S. E. & Lees, A. J. UK Parkinson's disease society brain bank clinical diagnostic criteria. *J. Neurol. Neurosurg. Psychiatry* **55**, e4 (1992).
7. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
8. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
9. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
10. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
11. Feng, Y.-Y. *et al.* RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. *bioRxiv* 436634 (2018) doi:10.1101/436634.

12. Home | AMP-PD. https://amp-pd.org/.
13. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
14. Young, A., Kumasaka, N., Calvert, F. & Hammond, T. R. A map of transcriptional heterogeneity and regulatory variation in human microglia. *bioRxiv* (2019).
15. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
16. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
17. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
18. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
19. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).