Single-nucleus RNA-sequencing reveals oligodendrocytes and their progenitors as vulnerable cell types in prefrontal cortex and anterior cingulate of brains with Parkinson's disease

Mohammad Dehestani[1,2], Cornelis Blauwendraat[3], Thomas Gasser[1,2*], Vikas Bansal[2*]

1. Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany
2. German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany
3. Laboratory for Neurogenetics, National Institute of Health NIH, Bethesda, Maryland, USA

[*]**Correspondence**: thomas.gasser@uni-tuebingen.de, vikas.bansal@dzne.de

Abstract

Several prior studies have proposed the involvement of various brain regions and cell types in Parkinson's disease (PD) pathology. Here, we performed snRNA-seq on the prefrontal cortex and anterior cingulate regions from post-mortem control and PD brain tissue. We found a significant association of oligodendrocytes and oligodendrocyte precursor cells with PD-linked risk loci and report several dysregulated genes and pathways, including regulation of tau-protein kinase activity, regulation of inclusion body assembly and protein processing involved in protein targeting to mitochondria. In an independent PD cohort with clinical measures (681 cases and 549 controls), polygenic risk scores derived from the dysregulated genes significantly predicted Montreal Cognitive Assessment (MoCA)-, and Beck Depression Inventory-II (BDI-II)-scores as well as Age at Onset but not motor impairment (UPDRS-III). These results suggest that by expanding our focus to glial cells, we can uncover molecular pathways associated with non-motor symptoms that are frequently observed in PD patients, also prior to diagnosis.

Keywords: Parkinson's disease, oligodendrocytes, oligodendrocyte precursor cells, snRNA-seq, prefrontal cortex, anterior cingulate cortex, non-motor symptoms

Parkinson's disease (PD) is the second most common neurodegenerative disorder. It is characterized by the pathologic aggregation of alpha-synuclein and its pathology is known to progress in a predictable spatiotemporal manner. The spread of the disease pathology commences in the olfactory bulb and the lower brainstem and moves through the substantia nigra pars compacta (SN) in the midbrain, eventually reaching the meso- and neocortical areas[1,2]. PD arises from a complex interplay of various factors, such as aging, genetic predisposition, and environmental factors. While the cause of PD is unknown in most cases, specific mutations in genes such as *LRRK2* and *GBA1* have been identified to significantly increase the risk of developing the disease, although likely by affecting different molecular pathways. This may be one reason why the impact of *LRRK2* and *GBA1* mutations on clinical presentations may differ[3]. In recent decades, transcriptome profiling has emerged as a preeminent methodology for exploring human pathologies at the molecular and cellular level. In PD, alpha-synuclein pathology has been shown to be associated with the transcriptional programs of various brain cell types, including neurons and glial cells[4]. Furthermore, although the degeneration of dopaminergic neurons (DA) in PD primarily occurs in the SN region, Lewy bodies can form in other brain regions, such as the limbic system and the prefrontal cortex[5]. Moreover, there has been no investigation into the relative impacts of *LRRK2* and *GBA1* risk genotypes on transcriptional programs across several brain regions of PD patients.

In this study, we performed single-nucleus RNA-sequencing (snRNA-seq) on the prefrontal cortex (PFC) and anterior cingulate (ACC) brain regions from the same individuals (2 *LRRK2* PD, 2 *GBA1* PD and 2 Healthy Controls) to identify neuronal and non-neuronal cell-type differences (Figure 1A and Table S1). After data cleaning and quality control (see Methods), 88,876 high-quality single nuclei were retained. The clustering of these high-quality nuclei identified 13 clusters covering major broad cell types in the brain, i.e. excitatory neurons (ExN), inhibitory neurons (InN), oligodendrocytes (ODCs), oligodendrocyte precursor cells (OPCs), microglia (MG), astrocytes (Astro) and vascular cells (Vas) (Figure 1B). While ~50% of the total nuclei were annotated as ODCs, only 4.3% and 0.5% of nuclei accounted for OPCs and Vas cells, respectively (Figure 1C). On average, we obtained 7406 high-quality nuclei per sample ranging from 3.6% to 16.5% of the total nuclei (Figure 1C and Table S2). None of the clusters was significantly enriched for a particular brain region or mutation group (Figure 1D and Table S2). Although the percentage of each cell type varied between the samples, no consistent cell-type enrichment could be identified based on an individual sample, except for the ExN5 cluster, which was enriched in Sample 2 (Figure 1E). The clusters were annotated based on the expression of well-known cell-type markers (Figure 1F and Table S3). The top 5 markers of each cell type cluster are shown in Figure 1G.

Integration of snRNA-seq data with PD genome-wide association studies (GWAS)[6] showed the strongest association of ODCs and OPCs with PD-linked risk loci (Figure 2A and TableS4). Next, we performed differential expression gene analysis to identify dysregulated genes between PD and controls in ODCs and OPCs. While 1040 and 543 genes were differentially expressed between *GBA1* and control samples in ODCs and OPCs, respectively, only 278 genes in ODCs and 108 genes in OPCs were differentially expressed between *LRRK2* and control samples (Table S5). The fold change (FC) of these differentially expressed genes (DEGs) was significantly correlated with the bulk RNA-seq data (Figure 2B and Table S6). The correlation for *GBA1* DEGs was better than *LRRK2* DEGs owing to the higher number of genes. The FC of the top 10 up-regulated and top 10 down-regulated genes are shown in Figure 2C, along with the FC in bulk RNA-seq data. While 5 genes were up-regulated in all four comparisons, 11 genes were down-regulated in all four comparisons (Figure 2D). Earlier studies have found some of these genes to be dysregulated or at least associated with PD. For instance, recently *MAPT* was found to be up-regulated in oligodendrocytes of the substantia nigra in PD patients[7], and *SUMF1* has been associated with lysosomal storage disorders[8]. Enriched biological processes in up-regulated DEGs (Figure 2E and Table S7) exhibited negative regulation of inclusion body assembly (in *GBA1*/Control ODCs), regulation of microtubule nucleation (in *LRRK2*/Control ODCs), positive regulation of tau-protein kinase activity (in *GBA1*/Control OPCs) and regulation of protein polymerization (in *LRRK2*/Control OPCs). On the other hand, enriched biological processes in down-regulated DEGs exhibited protein processing involved in protein targeting to mitochondria (in *GBA1*/Control ODCs), regulation of potassium ion transport (in LRRK2/Control ODCs), modulation of chemical synaptic transmission (in *GBA1*/Control OPCs) and cholesterol biosynthetic process (in *LRRK2*/Control OPCs). Intriguingly, the pseudotime analysis indicated a significantly different pseudotime distribution of *GBA1* and *LRRK2* ODCs compared to controls (Figure 3A-B). Interestingly, co-expression network analysis in ODCs revealed 7 co-expression modules, out of which module 1 and 6 were enriched in *LRRK2* and *GBA1* samples, respectively (Figure 3C-D and Table S8). ClinVar phenotype enrichment analysis of module 1 genes exhibited diseases related to iron accumulation, while module 6 exhibited schizophrenia (Figure 3E-F).

Next, we performed polygenic prediction in an independent PD cohort from Tuebingen (681 cases and 549 controls), which has a uniquely rich and detailed set of PD clinical measures. We computed six different Polygenic Risk Scores (PRS) derived from four gene lists obtained through our analysis of differential expression in ODCs and OPCs, along with two gene lists derived from module 1 and module 6 (Table S9). Except module 1 PRS, all the scores significantly predicted case-control status. Subsequently, we conducted prediction of PD measures among patients. We focused on Unified Parkinson Disease Rating Scale-III (UPDRS-III), Montreal Cognitive Assessment (MoCA), Age at

Onset (AAO) and Beck Depression Inventory-II (BDI-II) among PD patients (*N* ranges from 379 to 636). Significant associations were observed between GBA1_OPC_DEG and LRRK2_ODC_DEG scores with MoCA, LRRK2_ODC_DEG and LRRK2_OPC_DEG scores with BDI-II, and GBA1_ODC_DEG, LRRK2_OPC_DEG, and ODC_Module6 scores with AAO. However, it is noteworthy that none of the scores were able to predict motor examination measures i.e. UPDRS-III (Table 1).

To summarize, in this short report, we found a significant enrichment of ODC and OPC expressed genes within PD GWAS risk loci and revealed several dysregulated genes and pathways. Although two decades ago, Wakabyashi et al., observed an abnormal accumulation of alpha-synuclein in the oligodendrocytes within the substantia nigra of PD patients[9], the role of oligodendrocytes in PD has been mostly neglected. Reynolds and colleagues found a significant enrichment of PD heritability in glial cell types like oligodendrocytes and astrocytes[10]. By integrating GWAS results with single-cell transcriptomic data, Bryois et al., and Agarwal et al., observed oligodendrocytes and oligodendrocyte precursor cells to be significantly associated with PD[11,12]. Moreover, an increasing amount of evidence points out that oligodendrocytes also provide support to neurons via mechanisms beyond the insulating function of myelin[13]. Therefore, it is tempting to speculate that the abnormally regulated pathways, which extend beyond myelination in ODCs, may contribute to the pathology of PD. While the loss of DA neurons in the SN region of the midbrain is a well-known pathological hallmark of PD closely associated with motor symptoms, it is important to note that PD patients also encounter various non-motor symptoms, including cognitive and psychopathological manifestations[14]. Szabolcs and colleagues found higher rates of psychiatric morbidity (especially mood disorders, cognitive impairment, anxiety disorders, schizophrenia) in the premotor phase of PD and were more common in PD patients before PD diagnosis[15]. In line with this, our weighted gene co-expression network analysis (WGCNA) revealed a gene module enriched in *GBA1* samples associated with schizophrenia. On the other hand, another co-expression gene module enriched in *LRRK2* samples exhibited diseases related to iron accumulation, aligning with previous observations[16,17]. Moreover, the polygenic predictions in this study showed notable associations with non-motor symptoms, suggesting a crucial involvement of glial cells in neuropsychiatric symptoms that may extend beyond the SN region of the midbrain. Altogether, we anticipate that our study will serve as a valuable resource and prompt further research into the involvement of oligodendrocytes in the pathology of PD.

Methods

Samples used in this study

The research was conducted using fresh-frozen postmortem brain tissues obtained from four PD patients, two of which had the *LRRK2* p.G2019S mutation and the other two had the *GBA1* mutation (one with p.L444P and one with p.E326K), along with two healthy controls. All donors were males, aged between 65 and 80, and PD patients had Lewy body Braak stages of 5-6. Two brain regions, namely the prefrontal cortex and anterior cingulate cortex, were investigated for each donor. All tissues were procured from the Netherlands brain bank except one *LRRK2* brain from UCL Queen Square Brain Bank for Neurological Disorders, following the policies and regulations of the institutional ethics board at the University Hospital Tuebingen in Germany. For transcriptome analysis, single nuclei and RNA were extracted from all samples, and snRNA-seq and bulk RNA-seq were performed. The clinical characteristics of the donors have been elaborated in Table S1.

Generation of single nuclei from postmortem human brains

The process of isolating nuclei entails the utilization of a detergent lysis technique, where a detergent is employed to break down the cellular membranes, followed by the centrifugal separation of the nuclei. In brief, 300 mg of post-mortem brain tissue was dounce-homogenized in 2 ml of Nuclei EZ Prep Lysis Buffer (Sigma Aldrich, MA, USA) spiked with 0.2 U μl^{-1} RNase inhibitor (Sigma Aldrich, MA, USA) , 3.3 µl DTT (Thermo Fisher Scientific, MA, USA) and 33 µl of 10% Triton X100 which were added before incubating on ice for 5 minutes in a final volume of 10 ml. Homogenized tissue was washed with 3-4 ml washing buffer which was fresh PBSB 1% and then was filtered through a 70-µm cell strainer (BD Bioscience, NY, USA). Then using a long tube, 10 ml of 1.8 M ice-cold sucrose cushion solution is added to each sample i.e the roughly 3 ml lysate. After carefully and completely discarding the supernatant and the sucrose cushion layer containing debris and myelin, 1 ml PBS buffer added to resuspend the nuclei and 4 ml nuclei suspension buffer (1% BSA-PBS solution). It was finally centrifuged on 500 g for 5 minutes. At the end, 2 µl DAPI (Sigma Aldrich, MA, USA) with the concentration of 1:100 was added to stain the nuclei. Final centrifugation step on 500 g for 5 minutes was preceded by incubating the DAPI added suspension for 15-20 minutes in the cold & dark room on the rotation wheel at 4 C. Sorting buffer which consists of 99 µl PBS and 1 µl RNase inhibitor (Sigma Aldrich, MA, USA) was added to resuspend the nuclei and make them ready for quality/quantity inspection and then to run on a 10x genomic chromium controller. Quality assessment was performed using fluorescence-activated cell sorting (FACS) to detect all DAPI-positive events, i.e. individual nuclei comprising more than 95 % of all events.

196 **Droplet-based snRNA-seq using 10x Genomics**
197 Single-nuclei suspension concentration was determined by automatic cell counting
198 (DeNovix CellDrop, DE, USA) using an AO/PI viability assay (DeNovix, DE, USA) and
199 counting nuclei as dead cells. Single-nucleus gene expression libraries were generated
200 using the 10x Chromium Next gel beads-in-emulsion (GEM) Single Cell 3' Reagent Kit
201 v3.1 (10x Genomics, CA, USA) according to manufacturer's instructions. In brief, cells
202 were loaded on the Chromium Next GEM Chip G, which was subsequently run on the
203 Chromium Controller (10x Genomics, CA, USA) to partition cells into GEMs. Cell lysis
204 and reverse transcription of poly-adenylated mRNA occurred within the GEMs and
205 resulted in cDNA with GEM-specific barcodes and transcript-specific unique molecular
206 identifiers (UMIs). After breaking the emulsion, cDNA was amplified by PCR,
207 enzymatically fragmented, end-repaired, extended with 3′ A-overhangs, and ligated to
208 adapters. P5 and P7 sequences, as well as sample indices (Chromium i7 Multiplex kit,
209 10x Genomics, CA, USA), were added during the final PCR amplification step. The
210 fragment size of the final libraries was determined using the Bioanalyzer High-Sensitivity
211 DNA Kit (Agilent, CA, USA). Library concentration was determined using the Qubit dsDNA
212 HS Assay Kit (Thermo Fisher Scientific, MA, USA). snRNA libraries were pooled and
213 paired-end-sequenced on the Illumina NovaSeq 6000 platform (Illumina, CA, USA).
214
215 **snRNA-seq quality control**
216 Samples were demultiplexed using Illumina's bcl2fastq conversion tool and the 10x
217 Genomics pipeline Cell Ranger count v6.0.1 to perform alignment against the 10x
218 Genomics pre-built Cell Ranger reference GRCh38-2020-A (introns included), filtering,
219 barcode counting, and UMI counting. As a default, a cut-off value of 200 unique molecular
220 identifiers expressed in at least 3 cells was used to select nuclei of sufficient complexity
221 for further analysis. Each sample's count was normalized by the SCTransform method in
222 Seurat v4.1.0[18] with mitochondrial reads regressed out. Two approaches were combined
223 for quality control: (1) Doublets and multiplets were filtered out using DoubletFinder
224 v2.0.3[19] for each individual sample; (2) outliers with a high ratio of mitochondrial and
225 ribosomal counts (each >10%) and cells with low a number of genes (N < 1000) were
226 removed. The core statistical parameters of DoubletFinder used to build artificial doublets
227 for true doublet classification were determined automatically using recommended
228 settings. After applying these filtering steps on 105,781 input nuclei, the dataset contained
229 88,876 high-quality single nuclei that were eligible for further analysis.
230
231 **Cell annotations and differential expression**
232 After combining all samples into a single Seurat object, genes were projected into
233 principal component space using the principal component analysis (RunPCA). Harmony
234 R package[20] was used for integration as well as for removing unwanted effects across
235 subjects. The first 12 PC dimensions of data processed with Harmony were used as

236 inputs into the FindNeighbours, FindClusters (at 0.1 resolution obtained out of a range of
237 tested resolutions (0.1, 0.2, 0.5, 1.0)) and RunUMAP functions of Seurat. In brief, a
238 shared-nearest-neighbor graph was constructed on the basis of the Euclidean distance
239 metric in principal component space, and cells were clustered using the Louvain
240 algorithm. The RunUMAP function with default settings was used to calculate 2D UMAP
241 coordinates and search for distinct cell populations. Cluster markers and differential
242 expression testing was performed on Seurat "RNA" assay containing seurat log-
243 normalized counts using default Wilcoxon method implemented in Seurat v4.1.0.
244 Differential gene expression test between cases and controls was performed for each cell
245 type using the Wilcoxon ranked sum method implemented within the FindMarkers
246 function. Gene ontology enrichment analysis for biological processes was performed
247 using EnrichR[21]. In addition, hierarchical clustering of enriched GO terms was performed
248 for a set of paired comparisons, including *LRRK2* vs. HC and *GBA1* vs. HC brains, in
249 which differentially expressed genes showed significant enrichment (Adjusted P-Value <
250 0.05).
251
252 **Bulk RNA-seq, data quality control and differential gene expression analysis**
253 Total RNA was extracted from brain tissue homogenate for all samples using RNeasy
254 plus mini kit (Qiagen, Hilden, Germany) with on-column DNase treatment according to
255 manufacturer's protocol. Final elution was made in 65 µl of dH_2O. RNA concentration was
256 measured using the Qubit Fluorometric Quantitation and RNA Broad-Range Assay
257 (Thermo Fisher Scientific, MA, USA). RNA Integrity Number (RIN) was determined using
258 the Fragment Analyzer 5300 and the Fragment Analyzer RNA kit (Agilent, CA, USA) and
259 presented a moderate degradation pattern with RIN value between 3.4 and 7. For library
260 preparation, ribosomal RNA fraction was depleted using the QIAseq FastSelect rRNA
261 removal (Qiagen, Hilden, Germany) from 100 ng of total RNA. Next, libraries were
262 prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB)
263 according to the manufacturer's instructions. Library molarity was determined by
264 measuring the library size (mean 340 bp) using the Fragment Analyzer 5300 and the
265 Fragment Analyzer DNA HS NGS fragment kit (Agilent, CA, USA) and the library
266 concentration (mean 30 ng/µl) using Qubit Fluorometric Quantitation and dsDNA high
267 sensitivity assay (Thermo Fisher Scientific, MA, USA). The libraries were denatured
268 according to the manufacturer's instructions, diluted to 270 pM and sequenced as paired-
269 end 100 bp reads on an Illumina NovaSeq 6000 (Illumina, CA, USA) with a mean
270 sequencing depth of 55 million clusters per sample. Read quality of RNA-seq data in fastq
271 files was assessed using ngs-bits to identify sequencing cycles with low average quality,
272 adaptor contamination, or repetitive sequences from PCR amplification. Reads were
273 aligned using STAR v2.7.9a[22] to the GRCh37. Both unsupervised principal component
274 analysis on normalized counts as well as differential gene expression analysis were
275 performed using the DESeq2 R package v1.22.2 with default parameters[23]. For each

gene, gene expression fold changes (log_2 fold change) were computed and FDR test was performed to assess the significance.

Cell-type association with genetic risk of PD

Association analysis of cell type-specific expressed genes with genetic risk of PD was performed as described previously[24], using Multi-marker Analysis of GenoMic Annotation (MAGMA) v1.0.0, in order to identify disease-relevant cell types in the data[25,26]. MAGMA, as a gene set enrichment analysis method, tests the joint association of all SNPs in a gene with the phenotype, while accounting for LD structure between SNPs. Competitive gene set analysis was performed on SNP p-values from the latest PD GWAS summary statistics including 23andMe data and the publicly available European subset of 1000 Genomes Phase 3 was used as a reference panel to estimate LD between SNPs. SNPs were mapped to genes using NCBI GRCh37 build (annotation release 105). Gene boundaries were defined as the transcribed region of each gene. An extended window of 35 kb upstream and 10 kb downstream of each gene was added to the gene boundaries. Gene sets used in MAGMA were filtered: FDR-corrected p-values < 0.05, percentage of cells of the cluster where the expression was detected > 0.5, and logFC > 0.25.

Trajectory analysis

Slingshot v2.2.0 was used to perform the trajectory analysis[27]. ODCs and OPCs as cell types expressing most genetic risk loci identified by cell type association analysis, were extracted together and were put through the slingshot for the pseudotime inference. According to the knowledge from biology, OPCs were set as root cells during the analysis. UMAP embeddings generated from Seurat were converted to a cell_data_set object using SeuratWrappers v0.2.0 and then used as input to perform graph learning and pseudotime measurement through reversed graph embedding with slingshot.

Weighted gene co-expression network analysis

Utilizing the hdWGCNA v0.2.16 software package[28], we conducted a single-cell level weighted gene coexpression network analysis to identify gene modules exhibiting comparable expression patterns in all ODCs. In consideration of their biological relevance and markedly substantial cell count, we performed the analysis only on ODCs. Therefore, note that the modules may not necessarily express specifically in ODCs. The initial phase involved the creation of metacells, i.e. aggregates of homogenous cells stemming from identical biological groups. The k-nearest neighbor (KNN) algorithm was used to identify groups of similar cells for aggregation, and then the average expression of these cells was calculated, resulting in a metacell gene expression matrix. A correlation adjacency matrix was created, which is essential for inferring co-expression relationships between genes. After calculating the appropriate soft power threshold, co-expression networks were then constructed using the ConstructNetwork function. Module eigengenes (MEs),

316 a commonly used metric to summarize the gene expression profile of an entire co-
317 expression module, were calculated by performing principal component analysis (PCA)
318 on the subset of the gene expression matrix that comprises each module. Next, a function
319 OverlapModulesDEGs was used to explore the overlap between modules and cell-type
320 marker genes that were distinctly expressed in each mutation group, here called as
321 marker genes, and were calculated using the Seurat embedded function of
322 FindAllMarkers using only ODCs.
323
324 **Polygenic Risk Scores (PRS)**
325 Using summary statistics data from PD meta GWAS (https://pdgenetics.org/resources),
326 we performed a comprehensive annotation using the Region Annotation function in
327 Annovar[29] to find all genes corresponding to the SNPs in the whole genome and then
328 generated the GWAS gene list. We performed an overlap analysis between the GWAS
329 annotated genes and four gene lists obtained from our differential expression
330 comparisons in ODCs and OPCs between *GBA1* or *LRRK2* and HC. Additionally, we
331 included two gene lists derived from informative modules identified through weighted
332 gene co-expression network analysis, specifically module 1 and module 6. Next, we
333 retrieved SNPs in each of the six gene lists, including MAF, beta, and p-value from the
334 base data summary statistics. For the prediction, we used imputed genotypes of 681
335 cases and 549 controls from the Tuebingen cohort[30], which were not included in the base
336 data. To construct polygenic risk score (PRS) models, we utilized the R package PRSice2
337 v2.3.5[31]. We applied clumping procedure using $r^2 > 0.1$ and 1000 kb as the clumping
338 parameters in PRSice2 and a p-value of 0.05 was chosen as the threshold to exclude
339 non-significant SNPs. In other words, lead SNPs with a p-value of 0.05 from the LD-
340 clumped list were included in the calculation of PRS used in the regression models. The
341 null model is a logistic regression model that measures the power of covariates including
342 age, sex, and genetic principal components (PC1-4) in the prediction of PD status,
343 whereas the full model adds the PRS to the null model, thereby isolating the additive
344 influence of the PRS on risk prediction. The R^2 was also adjusted for an estimated PD
345 prevalence of 0.005 on the liability scale. To predict the clinical measures, PRS from all
346 six gene lists were included in a linear regression model using the “lm” function in R.
347 Standardized beta was obtained using the “lm.beta” function.
348
349
350
351
352
353
354
355

Ethics approval and consent to participate

Our study was approved by the Ethics Committee of the University Hospital and Medical Faculty Tuebingen.

Availability of data and materials

A Seurat object of the processed snRNA-seq data has been uploaded to Zenodo (https://doi.org/10.5281/zenodo.7886802).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the Charitable Hertie Foundation and in part by the Intramural Research Programs of the National Institute on Aging (NIA). V.B. is supported by a Career Development Fellowship at DZNE Tuebingen.

Authors' contributions

M.D., T.G. and V.B. conceived the project. T.G. and V.B. supervised the study. M.D. performed all the computational analysis and nuclei isolation. C.B. performed the MAGMA analysis. M.D. and V.B. wrote the manuscript. All authors contributed to and reviewed the manuscript.

Acknowledgments

We thank the NGS Competence Center Tübingen core facility for library preparations and sequencing. We also thank the Jonas Neher research group at DZNE Tübingen for providing the protocols and FACS facility. We would like to express our gratitude to the brain banks and the donors. We also thank Manu Sharma, Claudia Schulte and Anastasia Illarionova for genotype data, clinical data and MAGMA script, respectively.

Main figure legends

Figure 1: Overview of snRNA-seq profiling in the human post-mortem brain tissues. (A) Schematic overview of the experimental plan. **(B)** Uniform manifold approximation and projection (UMAP) visualization of the snRNA-seq clusters from 88,876 high quality nuclei. **(C)** Percentage of nuclei for each cell type across samples, mutation group and brain regions. **(D)** UMAP embeddings of nuclei colored by mutation group and brain regions. **(E)** Barplot displaying the distribution of cell-type percentage sample-wise. **(F)** Violin plot illustrating the expression distribution of known gene markers. **(G)** Genes most up-regulated in identified cell types: excitatory neurons (ExN), inhibitory neurons (InN), oligodendrocytes (ODCs), oligodendrocyte precursor cells (OPCs), microglia (MG), astrocytes (Astro) and vascular cells (Vas).

Figure 2: Association of PD susceptibility with ODCs and OPCs. (A) Multi-marker analysis of genomic annotation (MAGMA) gene set enrichment based on the snRNA-seq data showed significant associations with oligodendrocytes (ODCs) and oligodendrocyte precursor cells (OPCs). **(B)** Spearman's correlation between the ODCs and OPCs DEGs (GBA1 or LRRK2 vs controls), and the bulk RNA-seq data, were calculated based on the logarithmic fold-change. **(C)** Top ten up- and down-regulated genes ranked by logarithmic fold-change in snRNA-seq. **(D)** Venn diagrams displaying the overlap between up-regulated (left) or down-regulated (right) genes identified in ODCs and OPCs. **(E)** Gene ontology enrichment analysis of up-regulated (left) or down-regulated (right) genes. Top five biological process terms for each gene list are indicated.

Figure 3: Pseudotime and weighted gene co-expression network analysis. (A) Trajectory and pseudotime inference by Slingshot. **(B)** Boxplots showing the distribution of pseudotime in ODCs and OPCs separated by mutation groups. The p-values are calculated using the Kolmogorov-Smirnov test. **(C)** WGCNA dendrogram showing the different co-expression modules resulting from the network analysis in ODCs. Each leaf on the dendrogram represents a single gene, and the color at the bottom indicates the co-expression module assignment. **(D)** Results from the OverlapBarPlot function implemented in the hdWGCNA R package demonstrating the overlap between ODCs positive markers in each mutation group and the modules. *denotes a significant association exclusively in the mutation group. **(E-F)** Disease enrichment analysis of module 1 and module 6 genes using ClinVar phenotypes provided via EnrichR.

Supplementary table legends

Table S1: Sample overview with meta-data.

Table S2: Number of nuclei for cell types across samples, mutation group and brain regions.

Table S3: Known and identified cluster markers.

Table S4: MAGMA snRNA-seq cell-type enrichment results for PD GWAS.

Table S5: Differentially expressed genes in snRNA-seq for ODCs and OPCs.

Table S6: Differentially expressed genes from bulk RNA-seq.

Table S7: Enriched gene ontology terms for DEGs in ODCs and OPCs.

Table S8: Module genes from weighted gene co-expression network analysis.

Table S9: Polygenic risk scores in Tuebingen PD cohort with case-control prediction.

Table 1: Polygenic prediction of Parkinson's measures in the Tuebingen patient cohort.

		UPDRS-III	Montreal Cognitive Assessment	Beck Depression Inventory-II	Age at Disease Onset
N		514	450	379	636
GBA1_ODC_DEG	*standardized beta*	-0.045	0.035	-0.087	**0.040****
	P value	0.345	0.474	0.116	0.037
GBA1_OPC_DEG	*standardized beta*	0.002	**0.093****	0.054	0.022
	P value	0.970	0.050	0.316	0.254
LRRK2_ODC_DEG	*standardized beta*	0.052	**-0.089***	**0.116****	0.006
	P value	0.266	0.060	0.032	0.746
LRRK2_OPC_DEG	*standardized beta*	0.005	0.003	**-0.165*****	**0.035***
	P value	0.907	0.946	0.001	0.056
ODC_Module1	*standardized beta*	-0.060	0.022	-0.001	-0.007
	P value	0.175	0.633	0.989	0.679
ODC_Module6	*standardized beta*	0.048	-0.010	-0.063	**0.030***
	P value	0.285	0.832	0.219	0.097
	Adj. R^2	0.039	0.145	0.053	0.804
	Model P value	0.001	2.4×10^{-12}	0.001	$<2.2 \times 10^{-16}$

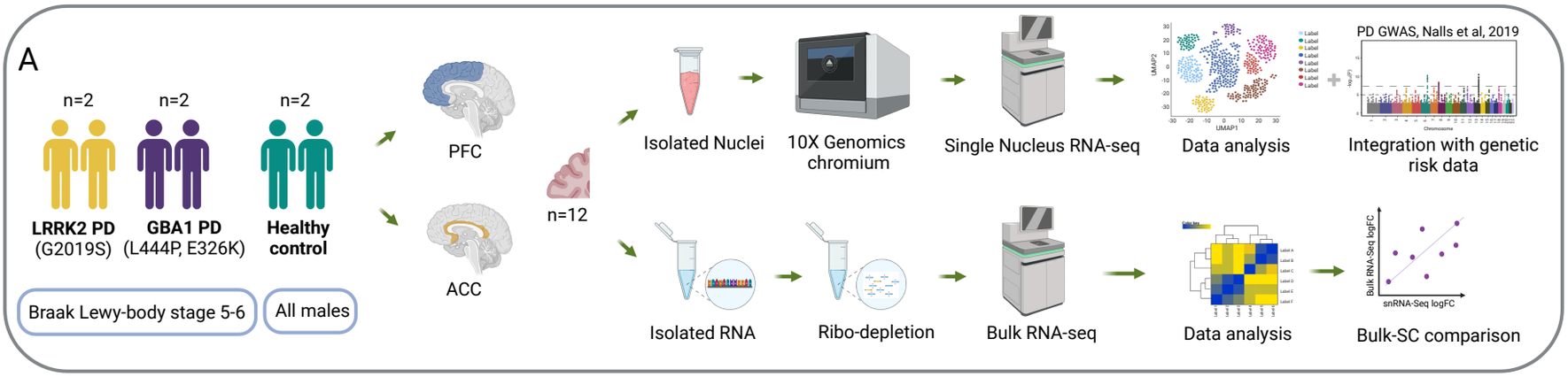
Notes: Linear regression using the sex, age and genetic principal components as control variables. *denotes significance at $P \leq 0.1$. **denotes significance at $P \leq 0.05$. ***denotes significance at $P \leq 0.01$. UPDRS, Unified Parkinson Disease Rating Scale; ODC, Oligodendrocytes; OPC, Oligodendrocyte precursor cells; DEG, differentially expressed genes.

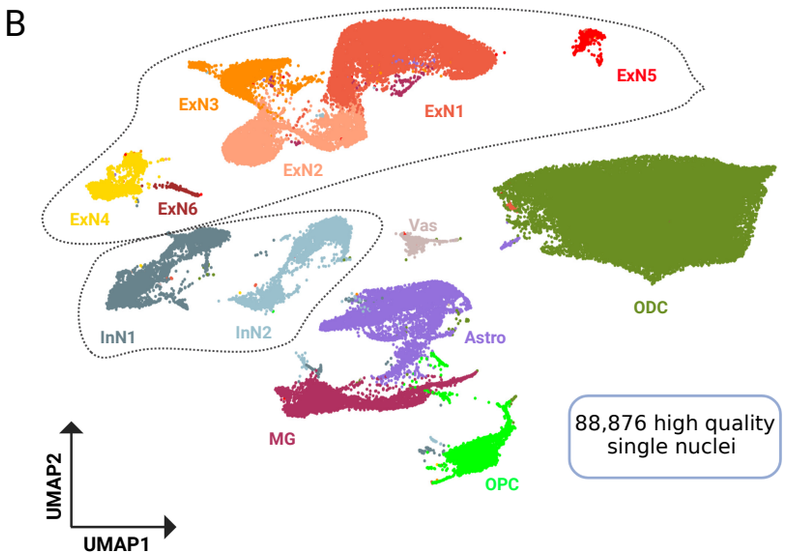
455 ## References
456

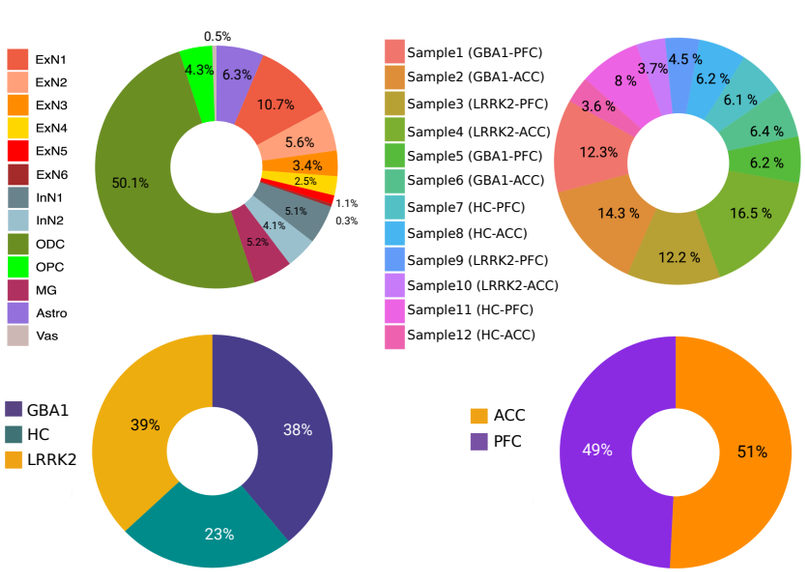
457 1. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease.
458 *Neurobiol. Aging* **24**, 197–211 (2003).
459 2. Obeso, J. A. *et al.* Past, present, and future of Parkinson's disease: A special essay
460 on the 200th Anniversary of the Shaking Palsy. *Mov. Disord.* **32**, 1264–1310 (2017).
461 3. Gan-Or, Z. *et al.* LRRK2 and GBA mutations differentially affect the initial
462 presentation of Parkinson disease. *Neurogenetics* **11**, 121–125 (2010).
463 4. Olsen, A. L. & Feany, M. B. Glial α-synuclein promotes neurodegeneration
464 characterized by a distinct transcriptional program in vivo. *Glia* **67**, 1933–1957
465 (2019).
466 5. Dumitriu, A. *et al.* Gene expression profiles in Parkinson disease prefrontal cortex
467 implicate FOXO1 and genes under its transcriptional regulation. *PLoS Genet.* **8**,
468 e1002794 (2012).
469 6. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk
470 for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*
471 *Neurol.* **18**, 1091–1102 (2019).
472 7. Lee, A. J. *et al.* Characterization of altered molecular mechanisms in Parkinson's
473 disease through cell type-resolved multiomics analyses. *Sci Adv* **9**, eabo2467
474 (2023).
475 8. Zhao, Y.-W. *et al.* The Association Between Lysosomal Storage Disorder Genes and
476 Parkinson's Disease: A Large Cohort Study in Chinese Mainland Population. *Front.*
477 *Aging Neurosci.* **13**, 749109 (2021).
478 9. Wakabayashi, K., Hayashi, S., Yoshimoto, M., Kudo, H. & Takahashi, H. NACP/alpha-
479 synuclein-positive filamentous inclusions in astrocytes and oligodendrocytes of
480 Parkinson's disease brains. *Acta Neuropathol.* **99**, 14–20 (2000).
481 10. Reynolds, R. H. *et al.* Moving beyond neurons: the role of cell type-specific gene
482 regulation in Parkinson's disease heritability. *NPJ Parkinsons Dis* **5**, 6 (2019).
483 11. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits
484 yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493
485 (2020).
486 12. Agarwal, D. *et al.* A single-cell atlas of the human substantia nigra reveals cell-
487 specific pathways associated with neurological disorders. *Nat. Commun.* **11**, 4183
488 (2020).
489 13. Pöyhönen, S., Er, S., Domanskyi, A. & Airavaara, M. Effects of Neurotrophic Factors
490 in Glial Cells in the Central Nervous System: Expression and Properties in
491 Neurodegeneration and Injury. *Front. Physiol.* **10**, 486 (2019).
492 14. Han, J. W. *et al.* Psychiatric Manifestation in Patients with Parkinson's Disease. *J.*
493 *Korean Med. Sci.* **33**, e300 (2018).

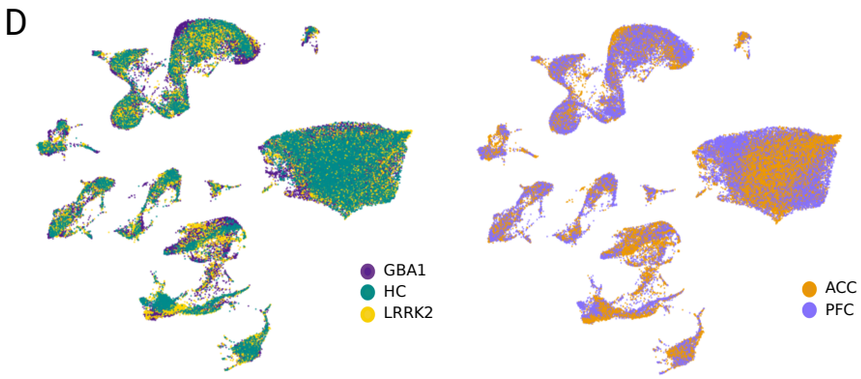
494 15. Szatmári, S., Jr, Ajtay, A., Oberfrank, F., Dobi, B. & Bereczki, D. The prevalence of
495 psychiatric symptoms before the diagnosis of Parkinson's disease in a nationwide
496 cohort: A comparison to patients with cerebral infarction. *PLoS One* **15**, e0236728
497 (2020).
498 16. Mamais, A. *et al.* Mutations in LRRK2 linked to Parkinson disease sequester Rab8a
499 to damaged lysosomes and regulate transferrin-mediated iron uptake in microglia.
500 *PLoS Biol.* **19**, e3001480 (2021).
501 17. Mochizuki, H. & Yasuda, T. Iron accumulation in Parkinson's disease. *J. Neural*
502 *Transm.* **119**, 1511–1514 (2012).
503 18. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–
504 3587.e29 (2021).
505 19. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in
506 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**,
507 329–337.e4 (2019).
508 20. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
509 Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
510 21. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment
511 analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
512 22. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
513 (2013).
514 23. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
515 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
516 24. Bressan, E. *et al.* The Foundational Data Initiative for Parkinson Disease: Enabling
517 efficient translation from genetic maps to mechanism. *Cell Genom* **3**, 100261
518 (2023).
519 25. Skene, N. G. *et al.* Genetic identification of brain cell types underlying
520 schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
521 26. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized
522 gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
523 27. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell
524 transcriptomics. *BMC Genomics* **19**, 477 (2018).
525 28. Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic
526 characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
527 29. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
528 variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
529 30. Dehestani, M. *et al.* Mitochondrial and autophagy-lysosomal pathway polygenic
530 risk scores predict Parkinson's disease. *Mol. Cell. Neurosci.* **121**, 103751 (2022).
531 31. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-
532 scale data. *Gigascience* **8**, (2019).

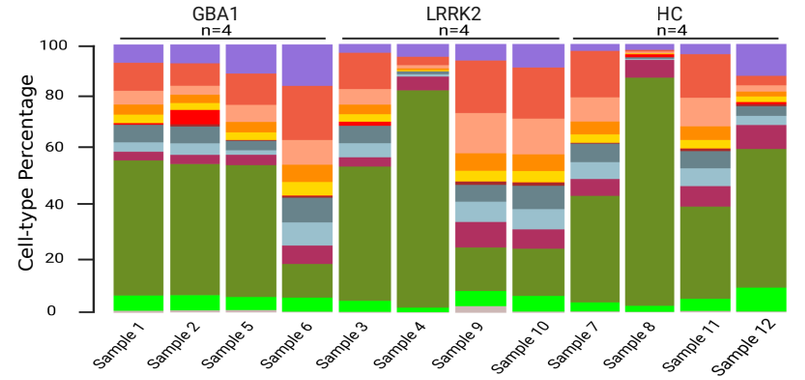
533

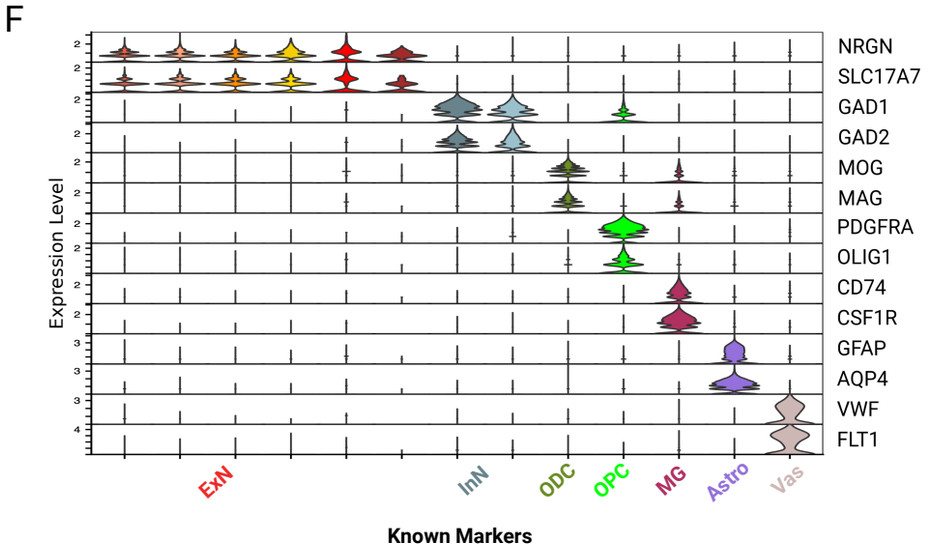

A
n=2
n=2
n=2
LRRK2 PD
(G2019S)
GBA1 PD
(L444P, E326K)
Healthy
control
Braak Lewy-body stage 5-6
All males
PFC
ACC
n=12
Isolated Nuclei
10X Genomics
chromium
Single Nucleus RNA-seq
Data analysis
PD GWAS, Nalls et al, 2019
Integration with genetic
risk data
Isolated RNA
Ribo-depletion
Bulk RNA-seq
Data analysis
Bulk-SC comparison

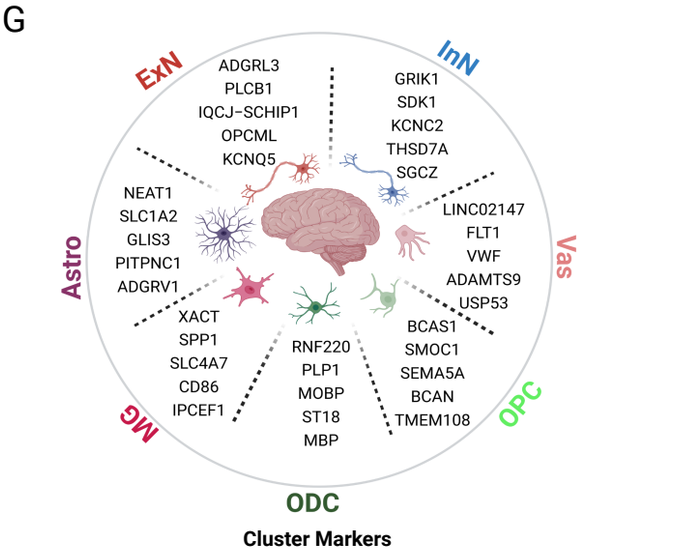

B
ExN3
ExN1
ExN5
ExN2
ExN4
ExN6
Vas
InN1
InN2
Astro
ODC
MG
OPC
88,876 high quality
single nuclei
UMAP2
UMAP1

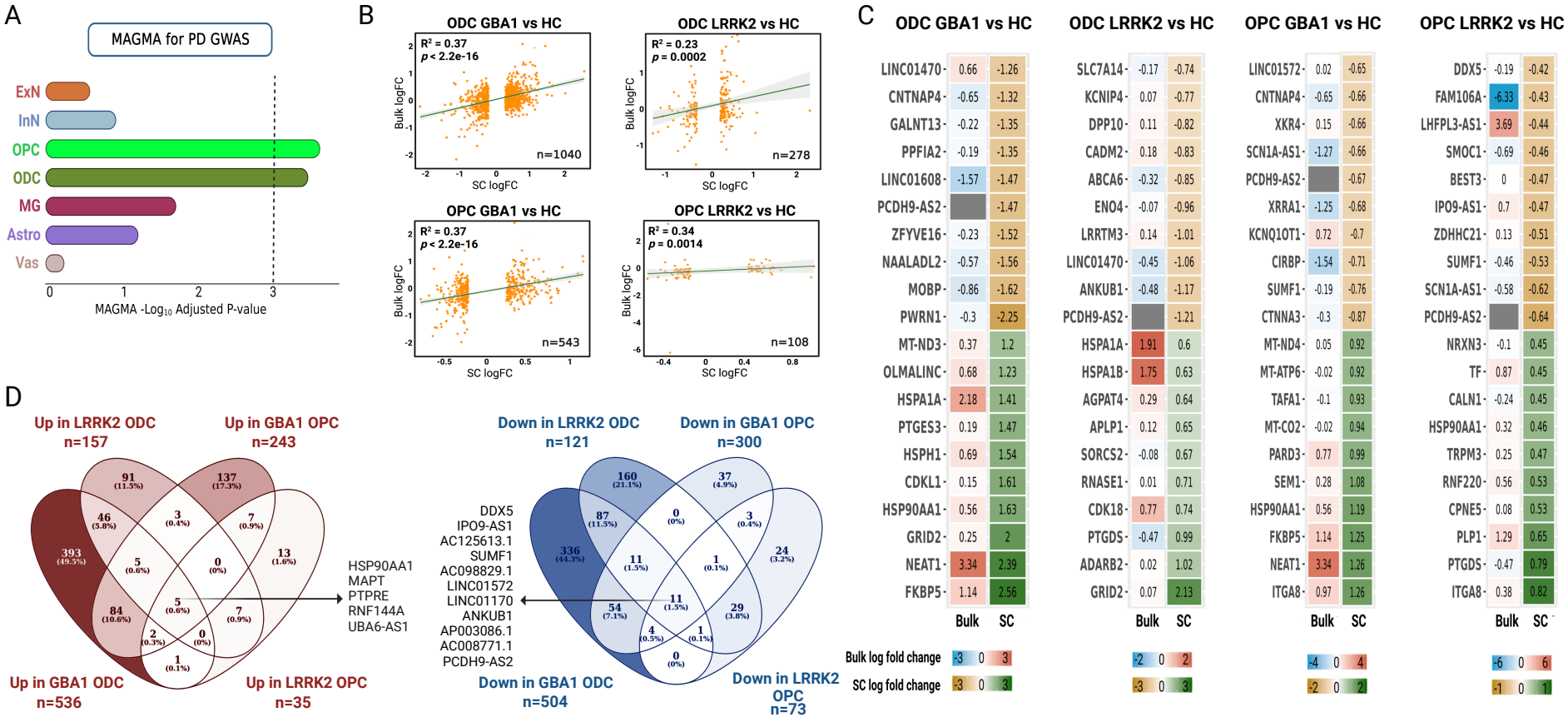

C
ExN1
ExN2
ExN3
ExN4
ExN5
ExN6
InN1
InN2
ODC
OPC
MG
Astro
Vas
0.5%
4.3%
6.3%
10.7%
5.6%
3.4%
2.5%
1.1%
0.3%
5.1%
4.1%
5.2%
50.1%
Sample1 (GBA1-PFC)
Sample2 (GBA1-ACC)
Sample3 (LRRK2-PFC)
Sample4 (LRRK2-ACC)
Sample5 (GBA1-PFC)
Sample6 (GBA1-ACC)
Sample7 (HC-PFC)
Sample8 (HC-ACC)
Sample9 (LRRK2-PFC)
Sample10 (LRRK2-ACC)
Sample11 (HC-PFC)
Sample12 (HC-ACC)
4.5 %
3.7%
8 %
6.2 %
6.1 %
3.6 %
6.4 %
12.3%
6.2 %
14.3 %
16.5 %
12.2 %
GBA1
HC
LRRK2
39%
38%
23%
ACC
PFC
49%
51%

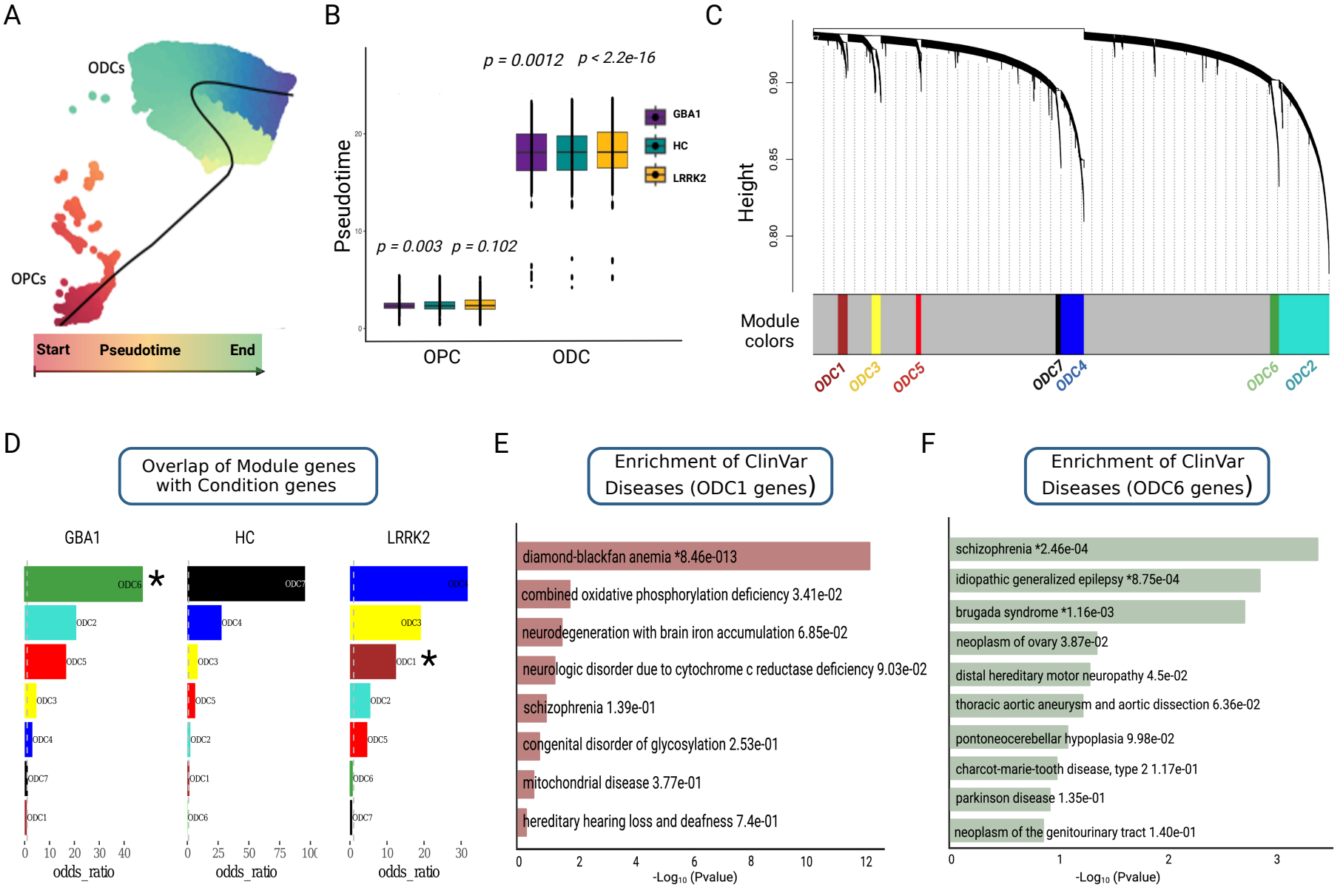

D
GBA1
HC
LRRK2
ACC
PFC


E
GBA1
n=4
LRRK2
n=4
HC
n=4
Cell-type Percentage
100
80
60
40
20
0
Sample 1
Sample 2
Sample 5
Sample 6
Sample 3
Sample 4
Sample 9
Sample 10
Sample 7
Sample 8
Sample 11
Sample 12


F
Expression Level
NRGN
SLC17A7
GAD1
GAD2
MOG
MAG
PDGFRA
OLIG1
CD74
CSF1R
GFAP
AQP4
VWF
FLT1
ExN
InN
ODC
OPC
MG
Astro
Vas
Known Markers


G
ExN
ADGRL3
PLCB1
IQCJ-SCHIP1
OPCML
KCNQ5
InN
GRIK1
SDK1
KCNC2
THSD7A
SGCZ
Astro
NEAT1
SLC1A2
GLIS3
PITPNC1
ADGRV1
Vas
LINC02147
FLT1
VWF
ADAMTS9
USP53
MG
XACT
SPP1
SLC4A7
CD86
IPCEF1
ODC
RNF220
PLP1
MOBP
ST18
MBP
OPC
BCAS1
SMOC1
SEMA5A
BCAN
TMEM108
Cluster Markers


A
MAGMA for PD GWAS
ExN
InN
OPC
ODC
MG
Astro
Vas
0
1
2
3
MAGMA -Log10 Adjusted P-value
B
ODC GBA1 vs HC
R² = 0.37
p < 2.2e-16
n=1040
ODC LRRK2 vs HC
R² = 0.23
p = 0.0002
n=278
OPC GBA1 vs HC
R² = 0.37
p < 2.2e-16
n=543
OPC LRRK2 vs HC
R² = 0.34
p = 0.0014
n=108
Bulk logFC
SC logFC
C
ODC GBA1 vs HC
LINC01470 0.66 -1.26
CNTNAP4 -0.65 -1.32
GALNT13 -0.22 -1.35
PPFIA2 -0.19 -1.35
LINC01608 -1.57 -1.47
PCDH9-AS2 -1.47
ZFYVE16 -0.23 -1.52
NAALADL2 -0.57 -1.56
MOBP -0.86 -1.62
PWRN1 -0.3 -2.25
MT-ND3 0.37 1.2
OLMALINC 0.68 1.23
HSPA1A 2.18 1.41
PTGES3 0.19 1.47
HSPH1 0.69 1.54
CDKL1 0.15 1.61
HSP90AA1 0.56 1.63
GRID2 0.25 2
NEAT1 3.34 2.39
FKBP5 1.14 2.56
ODC LRRK2 vs HC
SLC7A14 -0.17 -0.74
KCNIP4 0.07 -0.77
DPP10 0.11 -0.82
CADM2 0.18 -0.83
ABCA6 -0.32 -0.85
ENO4 -0.07 -0.96
LRRTM3 0.14 -1.01
LINC01470 -0.45 -1.06
ANKUB1 -0.48 -1.17
PCDH9-AS2 -1.21
HSPA1A 1.91 0.6
HSPA1B 1.75 0.63
AGPAT4 0.29 0.64
APLP1 0.12 0.65
SORCS2 -0.08 0.67
RNASE1 0.01 0.71
CDK18 0.77 0.74
PTGDS -0.47 0.99
ADARB2 0.02 1.02
GRID2 0.07 2.13
OPC GBA1 vs HC
LINC01572 0.02 -0.65
CNTNAP4 -0.65 -0.66
XKR4 0.15 -0.66
SCN1A-AS1 -1.27 -0.66
PCDH9-AS2 -0.67
XRRA1 -1.25 -0.68
KCNQ1OT1 0.72 -0.7
CIRBP -1.54 -0.71
SUMF1 -0.19 -0.76
CTNNA3 -0.3 -0.87
MT-ND4 0.05 0.92
MT-ATP6 -0.02 0.92
TAFA1 -0.1 0.93
MT-CO2 -0.02 0.94
PARD3 0.77 0.99
SEM1 0.28 1.08
HSP90AA1 0.56 1.19
FKBP5 1.14 1.25
NEAT1 3.34 1.26
ITGA8 0.97 1.26
OPC LRRK2 vs HC
DDX5 -0.19 -0.42
FAM106A -6.33 -0.43
LHFPL3-AS1 3.69 -0.44
SMOC1 -0.69 -0.46
BEST3 0 -0.47
IPO9-AS1 0.7 -0.47
ZDHHC21 0.13 -0.51
SUMF1 -0.46 -0.53
SCN1A-AS1 -0.58 -0.62
PCDH9-AS2 -0.64
NRXN3 -0.1 0.45
TF 0.87 0.45
CALN1 -0.24 0.45
HSP90AA1 0.32 0.46
TRPM3 0.25 0.47
RNF220 0.56 0.53
CPNE5 0.08 0.53
PLP1 1.29 0.65
PTGDS -0.47 0.79
ITGA8 0.38 0.82
Bulk
SC
Bulk log fold change
SC log fold change
D
Up in LRRK2 ODC n=157
Up in GBA1 OPC n=243
Up in GBA1 ODC n=536
Up in LRRK2 OPC n=35
HSP90AA1
MAPT
PTPRE
RNF144A
UBA6-AS1
Down in LRRK2 ODC n=121
Down in GBA1 OPC n=300
Down in GBA1 ODC n=504
Down in LRRK2 OPC n=73
DDX5
IPO9-AS1
AC125613.1
SUMF1
AC098829.1
LINC01572
LINC01170
ANKUB1
AP003086.1
AC008771.1
PCDH9-AS2
E
Up in GBA1 ODC
Up in LRRK2 ODC
Up in GBA1 OPC
Up in LRRK2 OPC
chondroitin sulfate proteoglycan biosynthetic process (GO:0050650)
chondroitin sulfate metabolic process (GO:0030204)
cell morphogenesis involved in neuron differentiation (GO:0048667)
chondroitin sulfate biosynthetic process (GO:0030206)
response to unfolded protein (GO:0006986)
regulation of cellular response to stress (GO:0080135)
regulation of insulin receptor signaling pathway (GO:0046626)
cellular protein-containing complex assembly (GO:0034622)
regulation of cellular response to heat (GO:1900034)
regulation of protein polymerization (GO:0032271)
ERBB2 signaling pathway (GO:0038128)
regulation of cellular protein localization (GO:1903827)
cellular response to heat (GO:0034605)
regulation of protein modification by small protein conjugation or removal (GO:1903320)
central nervous system neuron axonogenesis (GO:0021955)
regulation of microtubule nucleation (GO:0010968)
central nervous system neuron development (GO:0021954)
positive regulation of tau-protein kinase activity (GO:1902949)
negative regulation of inclusion body assembly (GO:0090084)
Combined score (scaled)
Down in GBA1 ODC
Down in LRRK2 ODC
Down in GBA1 OPC
Down in LRRK2 OPC
regulation of mRNA splicing, via spliceosome (GO:0048024)
cholesterol metabolic process (GO:0008203)
regulation of calcium−mediated signaling (GO:0050848)
regulation of stress fiber assembly (GO:0051492)
axonogenesis (GO:0007409)
positive regulation of cell−cell adhesion (GO:0022409)
negative regulation of stress fiber assembly (GO:0051497)
secondary alcohol biosynthetic process (GO:1902653)
cholesterol biosynthetic process (GO:0006695)
sterol biosynthetic process (GO:0016126)
synapse organization (GO:0050808)
regulation of synapse assembly (GO:0051963)
modulation of chemical synaptic transmission (GO:0050804)
regulation of nervous system process (GO:0031644)
modulation of excitatory postsynaptic potential (GO:0098815)
regulation of AMPA receptor activity (GO:2000311)
nucleic acid metabolic process (GO:0090304)
establishment of protein localization to membrane (GO:0090150)
protein localization to plasma membrane (GO:0072659)
positive regulation of epithelial cell migration (GO:0010634)
regulation of potassium ion transport (GO:0043266)
regulation of cation transmembrane transport (GO:1904062)
transmembrane receptor protein tyrosine phosphatase signaling pathway (GO:0007185)
protein processing involved in protein targeting to mitochondrion (GO:0006627)
Combined score (scaled)


A
ODCs
OPCs
Start
Pseudotime
End
B
p = 0.0012
p < 2.2e-16
p = 0.003
p = 0.102
Pseudotime
GBA1
HC
LRRK2
OPC
ODC
C
Height
0.90
0.85
0.80
Module colors
ODC1
ODC3
ODC5
ODC7
ODC4
ODC6
ODC2
D
Overlap of Module genes with Condition genes
GBA1
HC
LRRK2
ODC6
ODC2
ODC5
ODC3
ODC4
ODC7
ODC1
ODC7
ODC4
ODC3
ODC5
ODC2
ODC1
ODC6
ODC4
ODC3
ODC1
ODC2
ODC5
ODC6
ODC7
odds_ratio
odds_ratio
odds_ratio
E
Enrichment of ClinVar Diseases (ODC1 genes)
diamond-blackfan anemia *8.46e-013
combined oxidative phosphorylation deficiency 3.41e-02
neurodegeneration with brain iron accumulation 6.85e-02
neurologic disorder due to cytochrome c reductase deficiency 9.03e-02
schizophrenia 1.39e-01
congenital disorder of glycosylation 2.53e-01
mitochondrial disease 3.77e-01
hereditary hearing loss and deafness 7.4e-01
-Log10 (Pvalue)
F
Enrichment of ClinVar Diseases (ODC6 genes)
schizophrenia *2.46e-04
idiopathic generalized epilepsy *8.75e-04
brugada syndrome *1.16e-03
neoplasm of ovary 3.87e-02
distal hereditary motor neuropathy 4.5e-02
thoracic aortic aneurysm and aortic dissection 6.36e-02
pontoneocerebellar hypoplasia 9.98e-02
charcot-marie-tooth disease, type 2 1.17e-01
parkinson disease 1.35e-01
neoplasm of the genitourinary tract 1.40e-01
-Log10 (Pvalue)