

RESEARCH ARTICLE

Open Access



# Transcriptome deregulation of peripheral monocytes and whole blood in *GBA*-related Parkinson's disease

Giulietta Maria Riboldi<sup>1</sup> , Ricardo A. Vialle<sup>2,3,4,5,6</sup>, Elisa Navarro<sup>2,3,4,5,7</sup>, Evan Udine<sup>2,3,4,5</sup>, Katia de Paiva Lopes<sup>2,3,4,5,6</sup>, Jack Humphrey<sup>2,3,4,5</sup>, Amanda Allan<sup>2,3,4,5</sup>, Madison Parks<sup>2,3,4,5</sup>, Brooklyn Henderson<sup>1</sup>, Kelly Astudillo<sup>1</sup>, Charalambos Argyrou<sup>2,3,4,5</sup>, Maojuan Zhuang<sup>2,3,4,5</sup>, Tamjeed Sikder<sup>2,3,8,9</sup>, J. Oriol Narcis<sup>2,3,4,5</sup>, Shilpa Dilip Kumar<sup>10</sup>, William Janssen<sup>10</sup>, Allison Sowa<sup>8</sup>, Giacomo P. Comi<sup>11,12</sup>, Alessio Di Fonzo<sup>11,12</sup>, John F. Crary<sup>2,3,8,9</sup>, Steven J. Frucht<sup>1</sup> and Towfique Raj<sup>2,3,4,5,13\*</sup>

## Abstract

**Background:** Genetic mutations in beta-glucocerebrosidase (*GBA*) represent the major genetic risk factor for Parkinson's disease (PD). *GBA* participates in both the endo-lysosomal pathway and the immune response, two important mechanisms involved in the pathogenesis of PD. However, modifiers of *GBA* penetrance have not yet been fully elucidated.

**Methods:** We characterized the transcriptomic profiles of circulating monocytes in a population of patients with PD and healthy controls (CTRL) with and without *GBA* variants ( $n = 23$  PD/*GBA*, 13 CTRL/*GBA*, 56 PD, 66 CTRL) and whole blood ( $n = 616$  PD, 362 CTRL, 127 PD/*GBA*, 165 CTRL/*GBA*). Differential expression analysis, pathway enrichment analysis, and outlier detection were performed. Ultrastructural characterization of isolated CD14+ monocytes in the four groups was also performed through electron microscopy.

**Results:** We observed hundreds of differentially expressed genes and dysregulated pathways when comparing manifesting and non-manifesting *GBA* mutation carriers. Specifically, when compared to idiopathic PD, PD/*GBA* showed dysregulation in genes involved in alpha-synuclein degradation, aging and amyloid processing. Gene-based outlier analysis confirmed the involvement of lysosomal, membrane trafficking, and mitochondrial processing in manifesting compared to non-manifesting *GBA*-carriers, as also observed at the ultrastructural levels. Transcriptomic results were only partially replicated in an independent cohort of whole blood samples, suggesting cell-type specific changes.

**Conclusions:** Overall, our transcriptomic analysis of primary monocytes identified gene targets and biological processes that can help in understanding the pathogenic mechanisms associated with *GBA* mutations in the context of PD.

**Keywords:** Parkinson's disease, Monocytes, *GBA*, beta-glucocerebrosidase, Transcriptomic analysis

## Background

Mutations of the *GBA* gene, encoding beta-glucocerebrosidase (GCCase), have long been recognized as the major genetic risk factor for Parkinson's disease (PD) [1–4]. Mono- and biallelic mutations of *GBA* can increase the risk of developing PD up to 10 times compared to the

\*Correspondence: towfique.raj@mssm.edu

<sup>13</sup> Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, ICAHN 10-70E, New York, NY 10029–6574, USA  
Full list of author information is available at the end of the article



general population, with an incidence ranging from 2 to 30% across different ancestries [5]. More than 60 pathogenic variants of *GBA* have been identified in PD, where the N370S and L444P mutations account for up to 70–80% of those [6–8]. The molecular mechanisms mainly affected in *GBA*-PD that have been reported so far encompass the endo-lysosomal pathways, vesicular trafficking, lipid metabolism, and the cell stress response [9–17]. Different hypotheses have also been suggested to explain the relationship between *GBA* variants, reduced GCase activity, and alpha-synuclein accumulation, one of the hallmarks of PD [10, 14, 16]. However, because of the incomplete penetrance of the variants of this gene, it is still not clear whether additional modifiers are responsible for the onset of PD in some of the carriers. The identification of these possible modifiers is crucial for targeted therapeutic interventions or to be leveraged as diagnostic biomarkers. In the literature, a possible modifier effect on *GBA* has been reported for variants in the cathepsin B (*CTSB*) and alpha-synuclein (*SNCA*) genes, common variants in the proximity of the *GBA* gene, *GBA* pseudogene 1 (*GBAP1*), Metaxin 1 (*MTX1*) and Bridging Integrator 1 (*BINI*) genes, as well as the leucine repeat rich kinase 2 (*LRRK2*), the other major genetic risk factor of PD [18–24].

*GBA* is ubiquitously expressed across tissues. Mutations of this gene have been associated with aberrant monocyte/macrophage-mediated inflammatory response in the periphery, and with microglia activation in the brain of transgenic animal models [25–27]. Glial activation appears to start early in the presence of *GBA* variants, as it was reported in asymptomatic carriers of the mutation [28]. Since the initial report in 1998 of activated microglial cells in brains of subjects affected with PD, further evidence has stressed a possible role of inflammation in PD pathogenesis involving the peripheral and the central immune response [3, 29–35]. Genetic studies also suggest a link between PD and the immune response [36–40]. Indeed, recent genome-wide association studies (GWAS) identified a number of loci associated with PD which are in close proximity with genes related to the immune and inflammatory response [38]. In adjunct, a polarization of the cis-regulatory effect of common variants associated with PD was identified in the innate immune compartment compared to the adaptive response [39]. By assessing the transcriptomic profiles of CD14+ monocytes and microglia cells from a large cohort of subjects with PD and healthy controls we were able to describe a distinctive expression profile in these cells, identifying dysregulation of genes in the lysosomal and mitochondrial pathways [41].

Because of the described role of the immune system in the pathogenesis of PD and the important role of *GBA*

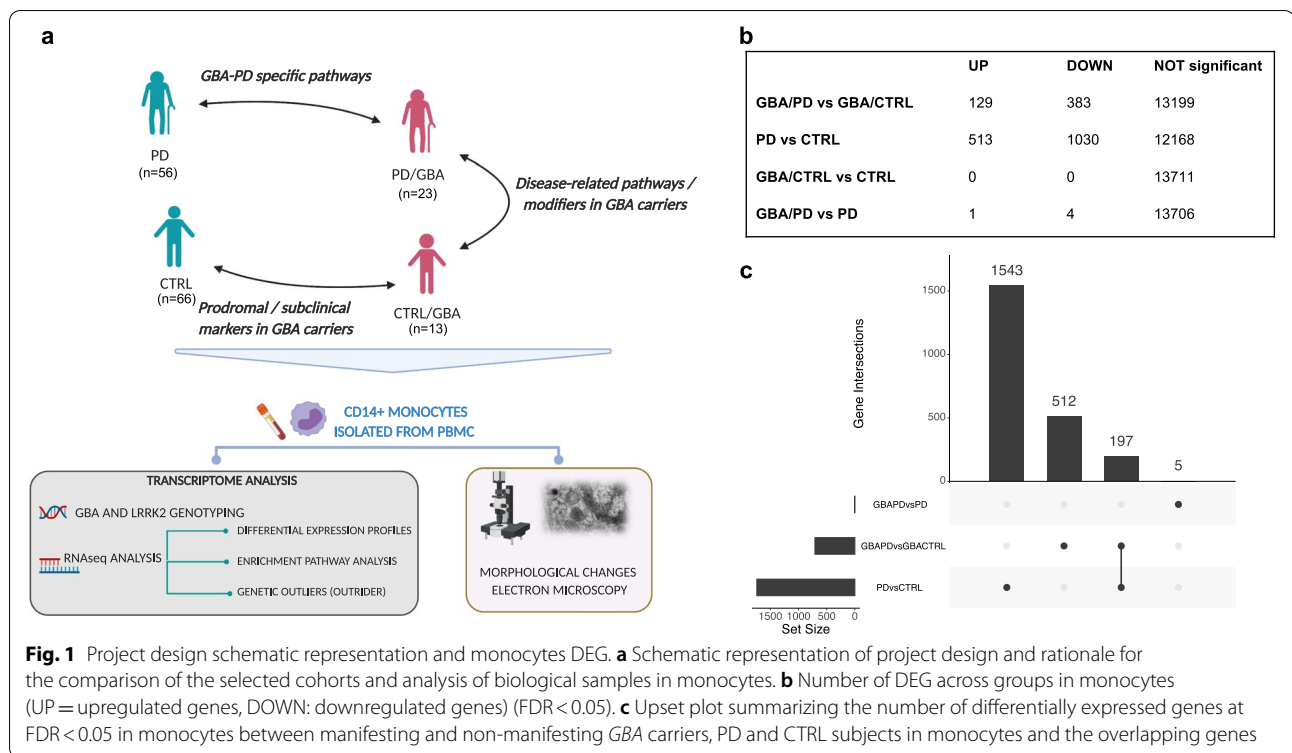
in these cells, monocytes can represent an informative cell-type to assess the role of this mutation in PD [42, 43]. With the present work we characterized the transcriptomic profiles of isolated CD14+CD16- monocytes in a cohort of PD patients and controls with and without *GBA* mutations and compared its results with expression data from whole blood generated from a validation cohort (Parkinson's Progression Markers Initiative - PPMI) (Fig. 1a). We reported a prominent involvement of lysosomal, membrane trafficking, and mitochondrial targets in manifesting vs non-manifesting *GBA*-carriers. Distinctive profiles related to alpha-synuclein-, amyloid- and aging-related processes were detected in manifesting *GBA*-carriers compared to subjects with PD without *GBA* variants. Electron microscopy analysis supported the presence of ultrastructural changes in monocytes from manifesting *GBA*-carriers.

## Material and methods

### Clinical centers and recruitment strategies

Subjects participating in the study were enrolled for the New York Movement Disorder (NYMD) cohort at The Marlene and Paolo Fresco Institute for Parkinson's and Movement Disorders at the New York University Langone Health (New York), the Bendheim Parkinson Movement Disorders Center at Mount Sinai (BPMD), the Alzheimer's Research Center (ADRC) and at the Center for Cognitive Health (CCH) at Mount Sinai Hospital (New York). Each Institution's Institutional Review Board approved the study protocol and the related procedures for subject recruitment, as well as data and samples collection. Subjects were enrolled only upon signing IRB approved informed consent. Enrolled subjects were between the age of 18 and 100 years. The diagnosis of PD was established by qualified movement disorder specialists according to the United Kingdom Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria [44]. Healthy controls (CTRL) were defined as aged and gender-matched non-affected subjects, who didn't have a known diagnosis or evidence of PD or other neurological conditions at the time of enrollment. Non-affected subjects were enrolled mostly among participants' partners and family members.

The study population of PD and CTRL subjects with and without *GBA* mutations was then selected for the study of the transcriptomic profiles of CD14+ monocytes, based on the availability of blood samples, good quality of extracted RNA and sequencing, according to procedures detailed above, as well as self-reported Northern European ancestry, in order to limit variability due to genetic background architecture related to ancestry. The final population consisted of 56 idiopathic PD (PD - subjects with PD and without *GBA* mutations),



**Fig. 1** Project design schematic representation and monocytes DEG. **a** Schematic representation of project design and rationale for the comparison of the selected cohorts and analysis of biological samples in monocytes. **b** Number of DEG across groups in monocytes (UP = upregulated genes, DOWN: downregulated genes) (FDR < 0.05). **c** Upset plot summarizing the number of differentially expressed genes at FDR < 0.05 in monocytes between manifesting and non-manifesting *GBA* carriers, PD and CTRL subjects in monocytes and the overlapping genes

66 CTRL (non-manifesting subjects without *GBA* pathogenic variants), 23 PD/*GBA* (subjects with PD and *GBA* pathogenic variants), and 13 CTRL/*GBA* subjects (non-manifesting subjects with *GBA* pathogenic variants) (Supplementary Table 1). Demographic variables are reported in Supplementary Table 1 and they were accounted for in the downstream normalization of the expression data. The great majority of the selected subjects presented overlap with European ancestry, equally distributed between Ashkenazi Jewish (AJ) and non-AJ ancestry (Supplementary Fig. 1a-b).

### Parkinson progressive markers initiative cohort

Data for the Parkinson Progressive Markers Initiative (PPMI) were downloaded from LONI in January 2021. The PPMI is an international, multicenter, observational study collecting clinical and biological data from subjects with PD, non-affected subjects and cohorts at risk for PD aiming to identify clinically significant biomarkers for the care and diagnosis of PD. Enrolled subjects with PD are de novo patients, initially recruited within 2 years from diagnosis. Subjects with PD and controls and known genetic mutations associated with PD, such as *GBA*, *LRRK2*, and *SNCA* are enrolled in the genetic cohort or genetic registry. An extensive description of the PPMI study and collected data and information can be found on [www.ppmi-info.org](http://www.ppmi-info.org). For this study we selected only

subjects with a diagnosis of Parkinson’s disease and non-affected control subjects, with and without mutations of the *GBA* gene. After removing subjects with missing data that would interfere with downstream analysis and considering only baseline visit, the final cohort consisted of 127 PD/*GBA*, 165 CTRL/*GBA*, 616 PD and 362 CTRLs.

### Sample collection and processing

Blood samples were collected fresh on the day of the research visit, in the morning, to reduce the variability of sample components and cell activation. Samples were collected in Vacutainer blood collection tubes with acid citrate dextrose (ACD) (BD Biosciences). Samples were processed within 2–3 hours from collection.

DNA was extracted from whole blood (0.5 ml) using the QiAamp DNA Blood Midi kit (Qiagen) according to the manufacturer’s instructions. Nanodrop was utilized to assess DNA quality and concentration.

Sample processing consisted in isolation of peripheral blood mononuclear cells (PBMC) and subsequent CD14+ monocytes purification. For PBMC isolation, SepMate tubes (StemCell Technologies) were used. After dilution in 2-fold PBS (Gibco) tubes were filled with 15 ml of Ficoll-Plaque PLUS (GE Healthcare) and centrifuged at 1200g for 10 mins, followed by wash with PBS. Monocyte isolation was performed through sorting of 5 million PBMCs utilizing the AutoMacs sorter with CD14+

magnetic beads (Miltenyi) according to manufacturer's instructions. Sorted monocytes were stored at  $-80^{\circ}\text{C}$  in RLT buffer (Qiagen) + 1% 2-Mercaptoethanol (Sigma Aldrich). Isolated monocytes stored in RLT buffer were first thawed on ice. RNA was isolated with the RNeasy Mini kit (Qiagen) according to manufacturer's instructions, including the DNase I optional step. RNA was then stored at  $-80^{\circ}\text{C}$  until library preparation. RNA integrity number (RIN) was assessed with TapeStation using Agilent RNA ScreenTape System (Agilent Technologies). RNA concentration was obtained with Qubit.

For the PPMI cohort, blood samples were centrally collected and processed as reported by the study's related documentation ([www.ppmi-info.org](http://www.ppmi-info.org)) and as previously reported [45]. Cell type composition (percentage of neutrophils, lymphocytes, monocytes, basophils, eosinophils) was provided as well with metadata.

### Sample genotyping

Genotyping of the DNA of the samples was performed with the Illumina Infinium Global Screening Array (GSA). This consists of a genome-wide backbone of 642,824 common variants and custom disease SNP content of about 60,000 SNPs. Screening for the most common genetic mutations of the *GBA* and *LRRK2* genes associated with Parkinson's disease and more frequent among the Ashkenazi Jewish ancestry was performed through targeted genotyping at Dr. William Nichols' laboratory at the Cincinnati Children's Hospital. In particular, for the *LRRK2* gene the G2019S variant was screened; for the *GBA* the following 11 variants were analyzed: IVS2 + 1, 84GG, E326K, T369M, N370S, V394L, D409G, L444P, A456P, R496H, RecNcil. The percentage of each mutation across the entire population and within manifesting and non-manifesting carriers was calculated for both *LRRK2* and *GBA* mutations.

### RNA sequencing

Part of the cohort were processed in house for RNA library preparation. TruSeq Stranded Total RNA Sample Preparation kit (Illumina), with the Low Sample (LS) protocol, was utilized for library preparation according to the manufacturer's instructions. For the rest of the samples RNA-seq libraries was prepared by a commercial service (Genewiz Inc.). RNA was shipped and processed according to a standard RNA-seq protocol. The Ribo-depletion strategy to remove rRNA was utilized for both samples processed in house and at Genewiz Inc. All samples were sequenced at Genewiz Inc. on an Illumina HiSeq 4000 platform with 150-bp paired-end reads and an average of 60 million reads per sample. Sequencing was performed in four independent batches. For isolated CD14+

monocytes, RNA-seq data were obtained from 56 PD, 23 PD/GBA, 66 CTRL, and 13 CTRL/GBA subjects.

For the PPMI cohort, RNA was sequenced at Hudson Alpha's Genomic Services Lab on an Illumina NovaSeq6000. As reported by the PPMI consortium, rRNA+globin reduction and directional cDNA synthesis using the NEB kit were performed. Samples were prepped using the NEB/Kapa (NEBKAP) based library prep, following second-strand synthesis. Sequencing was performed on the Illumina 6000 platform, generating on average 100 million 125 bp paired reads per sample.

### Genotyping and ancestry analysis

Global Screening Array (GSA) was used to genotype each individual. The following quality control metrics were applied: minor allele frequency (MAF)  $>5\%$ , SNP and samples call rate  $>95\%$ , Hardy-Weinberg equilibrium (HWE)  $P$ -Value  $>1 \times 10^{-6}$ . PLINK was utilized to identify and remove duplicated/related samples using pairwise IBD (identity-by-descent) estimation (PLINK PI\_HAT values 0.99–1).

Genetic ancestry of samples were confirmed through principal component analysis [46] and comparing multidimensional scaling (MDS) of the values of the study cohort with data from the Phase 3 of 1000 Genome Project samples. For the Ashkenazi Jewish (AJ) only, analyses were repeated using a custom panel as a reference (based on [47]).

### Expression data normalization

FASTQ files were processed with the RAPiD pipeline [48] implemented in the NextFlow framework (RAPiD-nf) ("Nextflow - A DSL for Parallel and Scalable Computational Pipelines" n.d.) providing automated alignment, quantification, and quality control of each RNA-seq sample. To assess quality of the sequences and technical metrics, SAMtools (v1.9) and Picard (2.20) ("Picard Tools - By Broad Institute" n.d.) were utilized prior to and after alignment with FASTQC (0.11.8) ("Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data" n.d.) [49]. Then, reads were processed with trimmomatic (v0.36) for adapter trimming [50]. Afterwards, upon creating indexes from GENCODE (v30) ("GENCODE - Human Release 30" n.d.), STAR (2.7.2a) was utilized for aligning the samples to the human reference genome hg38 build (GRCh38.primary\_assembly) [51]. Quantification of gene expression was obtained with RSEM (1.3.1) [52]. Quality control metrics were visualized with MultiQC and gene expression results were generated as raw counts, Counts Per Million (CPM), Transcripts Per Million (TPM), and TMM-voom. The following thresholds were used for initial filtering of the data at the sample level:  $>20\%$  of reads mapping to

coding regions, >20 Million aligned reads, and ribosomal rate < 30%. Sex mismatch was assessed by comparing reported sex with the expression of genes *UTY* and *XIST*, which didn't identify any sex mismatch in our cohort. At the gene level, genes with < 1 count per million in at least 30% of the samples were considered low expression genes and were excluded from the downstream analysis. The above processing led to a total of 13,711 genes that were used in all downstream analyses. Variance partitioning of each including surrogate variable (SVs) (explained in paragraph "Linear models for data regression") showed an average residual less than 25% (Supplementary Fig. 2a). Batch effects strongly associated with gene expression variation (Supplementary Fig. 2c), that was mitigated after SVs regression (Supplementary Fig. 2b, d).

For the PPMI cohort, transcriptome alignment and quantification were performed by the Accelerating Medicines Partnership (AMP)-PD consortium. FASTQ files were aligned using STAR v2.6.1d to the GRCh38 human genome build. Gene quantification was performed using Salmon v0.11.3 using GENCODE v29 annotations. Count per million (CPM) were calculated using edgeR package, and genes with CPM lower than 1 in more than 70% (of a total of 8322 samples from PPMI, PDBP, and BioFIND studies) were removed (total of 18,123 remaining genes). Next, only baseline PPMI samples ( $n=1270$ ) were selected for downstream analysis (only considering patients with PD and control subjects).

#### Data normalization and covariate selection

For differential expression analysis, gene counts were scale-normalized by the TMM method using edgeR R package and voom transformed using limma R package [53]. Different designs accounting for the majority of available technical and phenotypic variables (rna\_batch + Sex + PCT\_USABLE\_BASES + PCT\_RIBOSOMAL\_BASES + AJ\_gsa\_assignment + RIN+PCT\_CODING\_BASES + PCT\_INTERGENIC\_BASES + MEDIAN\_5PRIME\_BIAS + TOTAL\_READS + PF\_ALIGNED\_BASES + PF\_MISMATCH\_RATE + C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10 + PCT\_INTRONIC\_BASES + PCT\_ADAPTER + next\_day + C1\_AJ + C2\_AJ + C3\_AJ + C4\_AJ + C5\_AJ + C6\_AJ + C7\_AJ + C8\_AJ + C9\_AJ + C10\_AJ + MEDIAN\_CV\_COVERAGE + PCT\_ADAPTER + Diagnosis) were tested. However, principal component and MDS analysis showed the persistence of samples outliers with an impact on the downstream analysis. Therefore, to reduce error rate and increase reproducibility of the data, these were then processed with the 'sva' R package for Surrogate Variable Analysis [54]. This package allows the identification of surrogate variables to be built directly from a high-dimensional dataset. We estimated 13 surrogate variables while preserving effects from genetic status (presence

of *GBA* mutations) and phenotype (subjects with a diagnosis of Parkinson's disease or control groups). Surrogate variables were built in the design for linear regression. The contribution of known technical and phenotypic variables to the surrogate variables was obtained by linear regression between the surrogate variables and the covariates file and visualized in a heatmap (Supplementary Fig. 2c).

For the PPMI cohort, the following variables were regressed for downstream analysis: Interaction + age\_at\_baseline + sex + rin\_value + pct\_mrna\_bases\_picard + pct\_intergenic\_bases\_picard + PC1 + PC2 + PC3 + PC4 + PC5 + median\_insert\_size\_picard + plate. Interaction term accounted for genetic status (*GBA*) and disease status (PD vs CTRL).

#### Differential expression analysis based on interaction between genetic and disease status

A list of differentially expressed genes was obtained with the limma package in R by combining expression data (after TMM normalization and voom transformation in R) and surrogate variables. R package limma version 3.38.3 was used to fit a linear model and provide *P*-value upon performing Bayesian moderated t-test. Multiple testing correction with Benjamini-Hochberg False Discovery Rate (FDR) was obtained leveraging the function in the limma package. The cohort of subjects was subdivided into subgroups based on the disease status (subjects with PD vs controls (CTRL)) and *GBA* genetic mutation status (subjects carrying at least one *GBA* variant (PD/*GBA*, CTRL/*GBA*) and subjects with no *GBA* variants (PD, CTRL)).

We utilized a nested model to analyze expression data in reference to the variable of interest, which consisted in disease status and *GBA*-mutation status. In addition, to explore the effects of the *GBA* mutation on the disease status, an interaction term was added to the design as follows: [(PD/*GBA* – CTRL/*GBA*) – (PD – CTRL)]. Results from the comparison of each pair of groups were then extracted. A threshold of  $FDR < 0.05$  was utilized as a threshold of significance.

#### Gene set enrichment analysis

Gene set enrichment analysis was performed utilizing the set of differentially expressed genes from the nested interaction model analysis considering genes with  $FDR < 0.05$ . Enrichment analysis was performed separately for upregulated and downregulated genes in order to better characterize our set of differentially expressed genes. Gene Set Enrichment Analysis (GSEA) was used for the analysis of different terms from the Gene Ontology (GO) list (specifically: Cellular Component (CC), Molecular Function (MF), and Biological Processes (BP)) [55]. Gene-set enrichment with  $FDR < 0.01$  or 0.05 (as

specified in the results) were considered. Filters were set for gene-sets with less than 2000 genes. We analyzed up to the first 20 significant enriched terms.

For the dataset obtained from analysis with OUTRIDER [56] tool (see below) the additional following tools were utilized:

1. g-profiler (<https://biit.cs.ut.ee/gprofiler/gost>), a web server for functional enrichment analysis. Input data were the list of up and downregulated genes separately.
2. Ingenuity pathway analysis (IPA). Canonical data analysis for gene-set enrichment was performed. Statistically significant enriched terms with  $P$ -value  $< 0.05$  were accounted for in the final results.

The results from the different tools were then combined together based on  $P$ -values after multiple corrections.

#### Curated gene-set analysis

Gene-set enrichment analysis was also performed to assess the enrichment in our sets of differentially expressed genes of curated pathways and gene-sets relevant to previously reported pathways related with GBA functions and PD mechanisms. These encompassed genes involved in the following pathways (Supplementary Table 2): lysosomal database: 435 genes from The Human Lysosome Gene Dataset; lysosomal storage disease causative gene (LSD list) (54 genes) classified as sphingolipidoses, neuronal ceroid lipofuscinosis, mucopolipidosis/oligosaccharides diseases; mitochondrial gene list from [57] (315 genes), classified in distinct mitochondrial pathways as reported in the cited paper, such as mitonuclear cross-talk, mitochondrial dynamics, and OXPHOS; ubiquitin-related gene list (428 genes) from ubiquitin-like modifier activating enzymes and ubiquitin conjugating enzymes E2 (HUGO Gene Nomenclature Committee (HGNC) dataset), and ubiquitin ligase E3. We also assessed targeted enrichment for pathways identified by specific GO and involved in vesicle trafficking and the endolysosomal pathways (based on terms “membrane”, “lysosome”, “endocytosis”, “exocytosis”, that identified 32 pathways, (based on terms “membrane”, “lysosome”, “endocytosis”, “exocytosis”, that identified 32 pathways, Supplementary Table 3). and the following additional pathways: NOTCH1 signaling pathway GO:0007219; senescence associated vacuoles: GO:0010282 (plant); cell signaling via exosome: GO:0099156; cellular senescence: GO:0090398; lipid storage: GO:0019915, GO:0006869; lipid transport GO:0032594; tau protein binding GO:0048156; regulation Tau kinase activity GO:1902947, GO:1902949, GO:1902948; Golgi related pathways: GO:0048211, GO:0005795, GO:0005794, GO:0005796,

GO:0051645, GO:0006895, GO:0035621, GO:0055107, GO:0006888. Fisher exact test was run to assess the enrichment of curated terms in the differential expressed gene lists.

#### Genetic outliers

RNA-seq data can be also used to identify expression outliers within each single sample that may be expression of underlying genetic mutations, especially in regulatory regions, or compensatory/deregulated mechanisms. Different tools have been reported in the literature to explore this approach, based on Z-score distribution or a combination of Z-scores and the negative binomial distribution, respectively [58, 59]. These tools presented some limitations such as the lack of specific statistical tests to compare the expression data and the lack of regression for known and unknown covariates that can greatly affect gene expression profiles. OUTRIDER is an additional tool that, instead, utilizes autoencoders to control for variation linked to unknown factors for data normalization. Single genes and single individuals outliers are then detected by comparing univariate cases with the distribution of each gene across the population, by calculating the negative binomial distribution of each single sample compared to all samples [56]. Autoencoders are also discharging samples with an excess of outlier genes that may be related to other causes than having a biological relevance [56].

Count per million  $> 1$  in more than 30% of the samples were implemented in the tools. Data were normalized leveraging autoencoders (“OUTRIDER - OUTlier in RNA-Seq fInDER”). Normalized dispersion and mean were then fitted in a binomial model followed by computation of two-sided  $p$ -value. The significance threshold was set at an FDR adjusted  $P$ -value cut-off of 0.05 and  $z$ -score threshold of 2.

#### Ultrastructural characterization of CD14+ monocytes

CD14+ monocytes were isolated fresh, within 3 hours from collection, from PBMCs as described above. Isolated monocytes were pelleted ( $n=2$  per group) from manifesting (PD-GBA), non-manifesting (CTRL-GBA) GBA carriers, non-carrier controls (CTRL) and idiopathic PD (PD). Pellets were fixed (2% glutaraldehyde and 2% paraformaldehyde/ 0.1M Sodium Cacodylate buffer) for 1 week and Epon resin embedded. Ultrathin sections (75 nm) were collected onto 200 mesh copper grids using a Leica Em UCT ultramicrotome, counterstained with uranyl acetate/lead citrate, and imaged using a Hitachi 7500 TEM equipped with an AMT digital camera. Images were sized and adjusted for brightness and contrast. Transmission electron microscopy

images were acquired for 2–3 cells per field for a total of 15 cells for each sample. Sections were evaluated blinded. Qualitative assessment of the following parameters was assessed and compared among the four different groups: mitochondrial shape, endoplasmic reticulum, lysosomes, nucleus, chromatin, vacuoles, vesicles and inclusions.

### Statistical analysis and graphic representation

For comparison of the different parameters of the ultrastructural characterization between the four groups all analyses were performed with GraphPad Prism version 9.1.0 (GraphPad Inc. La Jolla CA, USA). One-way ANOVA between mean of counts from each sample was performed. Data are presented as Mean and SEM.

For transcriptomic analysis, sample size and statistical methods are reported in the legends of each figure. In the figures, asterisks indicate significant adjusted *p* value (\*=adjusted *p* value <0.05, \*\*=adjusted *p* value <0.01, \*\*\*=adjusted *p* value <0.001).

Bioinformatic analysis were performed with R (version 3.6.0) and graphic representation with R Studio version 1.2.1335. Figure 1a was created with [BioRender.com](https://BioRender.com).

## Results

### Phenotype and genetics drive different transcriptomic profiles in monocytes

We collected a cohort of subject with PD/GBA (*n*=23), PD (*n*=56), CTRL/GBA (*n*=13), and CTRL (*n*=66) (Fig. 1a), of European ancestry (Supplementary Fig. 1a, Supplementary Table 1). Differentially expression analysis of isolated CD14+CD16- monocytes showed a high number of DEG in subjects with GBA/PD compared to GBA/CTRL (*n*=512) and in PD compared to CTRL (*n*=1543) (FDR<0.05) (Fig. 1b,c). We found a more limited number of DEG between subjects with PD (with and without GBA variants, *n*=5) and no genes between CTRL/GBA and CTRL (Fig. 1b, c). The comparison of the transcriptomic profiles in monocytes from PD vs CTRL was previously explored in our recent work [41]. Here we will focus on the analysis of the subgroup of patients stratified by GBA variants.

Interestingly, expression levels of the GBA gene didn't show any significant differences across the four groups in monocytes (Supplementary Fig. 3a-b). However, lower expression levels of GBA were observed in carriers of GBA variants considered severe, such as the 84GG and V394L, although considering the limited number of subjects per each group (84GG: PD=2, CTRL=2; V394L: PD=1, CTRL =1; L444P/A456P/RecNcil: PD=1, CTRL=0) (Supplementary Fig. 3a, Supplementary Table 4).

### Exocytosis- and myeloid cell activation related genes are impaired in monocytes of manifesting vs non-manifesting GBA carriers

We first characterized the expression profiles of isolated monocytes from manifesting and non-manifesting GBA mutation carriers to elucidate disease mechanisms of GBA-related PD. Among the 512 DEG, 197 genes overlapped with genes differentially expressed between PD and CTRL subjects with no GBA mutations in monocytes (FDR<0.05) with significance concordance of directionality ( $R=0.92$ ,  $p<0.001$ ) (Fig. 1c, Supplementary Fig. 10a).

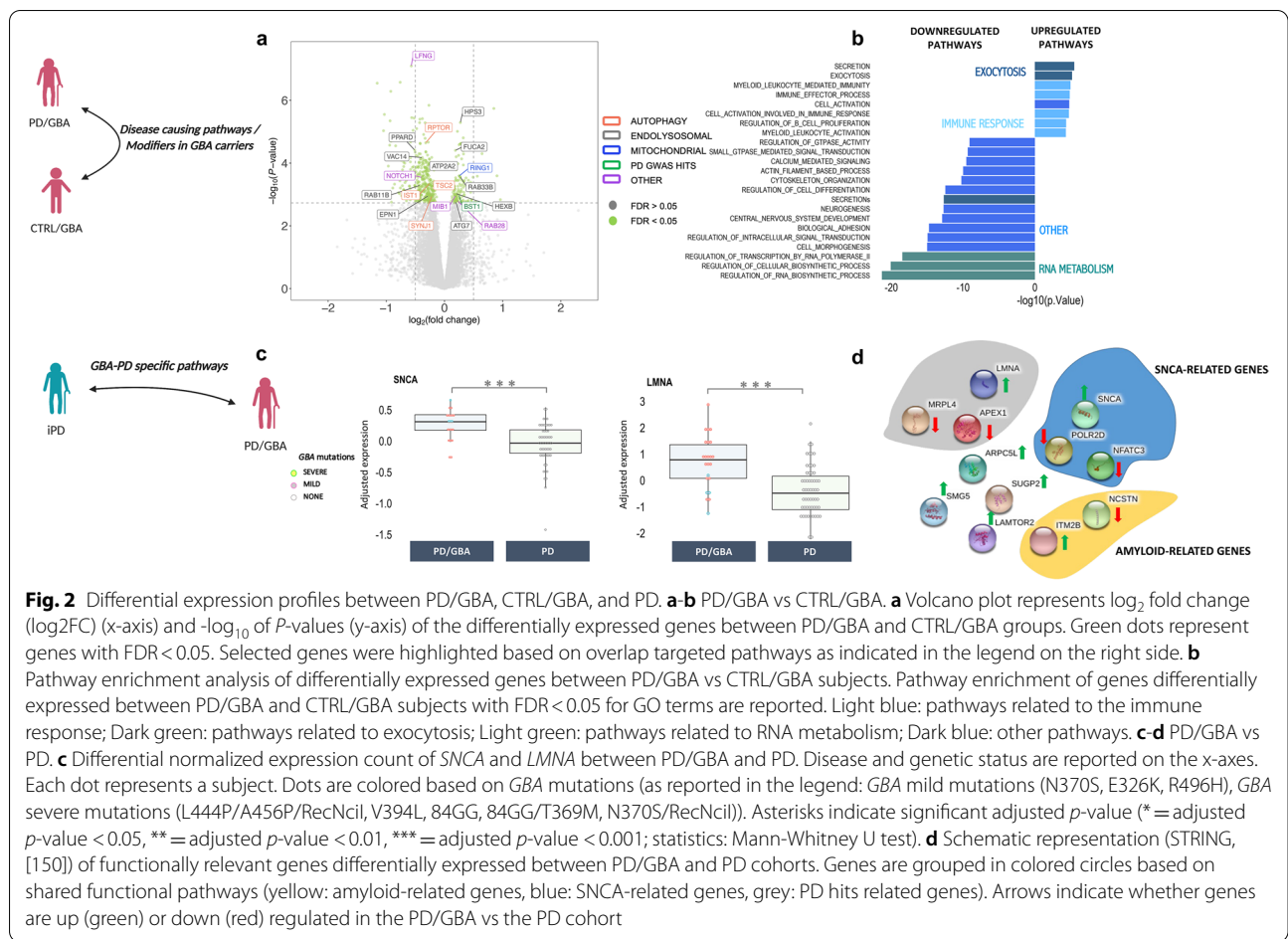
Gene-set enrichment analysis (GSEA) of differentially upregulated genes showed enrichment of pathways related to the myeloid compartment activation, exocytosis, and secretion (FDR<0.05) (Fig. 2b). Downregulated genes were enriched for transcription/RNA-metabolism related pathways, signal transduction (synapses and calcium mediated signal transmission), kinase activity, and vesicle secretion (Fig. 2b).

Taking a set of curated gene lists and gene ontology (GO) pathways related to membrane trafficking, exocytosis, lysosomal, ubiquitin, and proteasomal pathways, we observed a significant enrichment only for the vesicle membrane pathway within the set of significantly downregulated genes (Fisher's exact test; adjusted *P*-value <0.05), which included the *ATP2A2*, *RIPOR1*, *SYNJI* genes (Supplementary Table 2–3). No significant enrichment was identified for the membrane trafficking-related genes from previously prioritized genes in PD, as reported in Bandres-Ciga et al., 2019 [9].

Interestingly, significant DEG included targets implicated in monogenic forms of PD (such as *ATP13A2* and *LRKK2*), Tau-related pathways, senescence (*FUCA2* and *HEXB*), *NOTCH1*-related genes (a highly conserved transmembrane domain protein that is involved in different cellular processes such as cell proliferation, differentiation, apoptotic processes, and modulation of the secretome dynamics in cells), autophagy (such as *RPTOR*), the endolysosomal and mitochondrial pathways, as well as genes associated with PD-GWAS loci (Fig. 2a, Supplementary Fig. 4a-b).

### Aging- and alpha-synuclein-related pathways are impaired in PD/GBA vs PD

When comparing expression profiles in isolated monocytes of subjects with PD with vs without GBA-variants, the number of significantly DEG was more limited, including 5 genes at FDR<0.05 (1 upregulated and 4 downregulated) (Supplementary Fig. 5a). Pathway enrichment analysis was not possible due to the small number of DEGs between these two groups. Considering



the small sample size of these two groups, we also looked at genes with sub-threshold expression levels ( $FDR < 0.15$ ; 44 genes). Interestingly, in PD/GBA subjects there was an upregulation of the alpha-synuclein gene (*SNCA*) and related genes (i.e. *POLR2D* and *NFATC3*) compared to PD ( $FDR < 0.15$ ) (Fig. 2c-d, Supplementary Fig. 5b) [60]. We also identified genes related to the amyloid pathways (*ITM2B* and *NCSTN* genes), and aging process, such as *LMNA*, which is responsible for the progeria syndrome [61] (Fig. 2b). Finally, we found a deregulation of *LAMTOR2*, an amino acid sensing and activator of mTORC1 through its recruitment in the lysosomes (Fig. 2b).

Taken together, these observations suggest a deregulation of pathways associated with alpha-synuclein, aging and PD-related genetics in the PD/GBA vs PD group.

Finally, a small set of genes was significantly differentially expressed in the PD/GBA group compared to both PD and CTRL/GBA subjects ( $n=6$ ,  $FDR < 0.15$ ), and included *FILIP1L* [62], and *HPS3* (Supplementary Fig. 6). *HPS3* is particularly interesting in this context since this gene encodes for a subunit of the lysosome-related organelles complex-2. Mutations of this gene are responsible

for the Hermansky-Pudlak Syndrome 3, a systemic condition affecting the immune system and melanin, similarly to Chediak-Higashi syndrome (CHS). CHS has been previously associated with PD [63].

**Gene expression outliers highlight targeted pathways in manifesting and non-manifesting carriers**

We then performed gene expression outlier analysis. This tests for genes with significantly different levels of expression in a single individual compared to the rest of our cohort. Gene expression outlier individuals may harbor rare genetic variants in those genes [56]. In our cohort, we detected 493 outlier genes and 125 subjects (out of 158) with at least one outlier gene after normalization (Supplementary Fig. 7a-c).

Consistent with previous analysis, pathway enrichment analysis of the genetic outliers of subjects in the PD/GBA group confirmed a significant enrichment of pathways related to neuroinflammation (including *ICAM1*, *NFATC1*, *IRAK4*), membrane bounded organelle (such as *VPS41*), ERK/MAPK signaling, and autophagy (such as *DOCK1* - involved in cytoskeletal rearrangement for



phagocytosis of apoptotic cells and mobility of the cells) (Fig. 3a-b).

Instead, outlier genes in the CTRL/GBA group were mostly related to the mitochondrial functions: oxidative phosphorylation, inner mitochondrial membrane protein complex, thermogenesis, NADH dehydrogenase complex, and mitochondrial respiratory chain complex I. Outliers genes in the enriched pathways included: *MT-ATP* (mitochondrial complex 5), *MT-ND*, *MT-ND2*, *MT-ND4* (components of the NADH dehydrogenase or mitochondrial complex I) which are all genes encoded by the mitochondrial genome (Fig. 3a, Supplementary Fig. 7). We found 9 outlier genes in CTRL/GBA subjects and one outlier gene in PD/GBA (*NFATC1*) subjects that overlapped with the DEG between PD/GBA and CTRL/GBA (Fig. 3c, Supplementary Fig. 7).

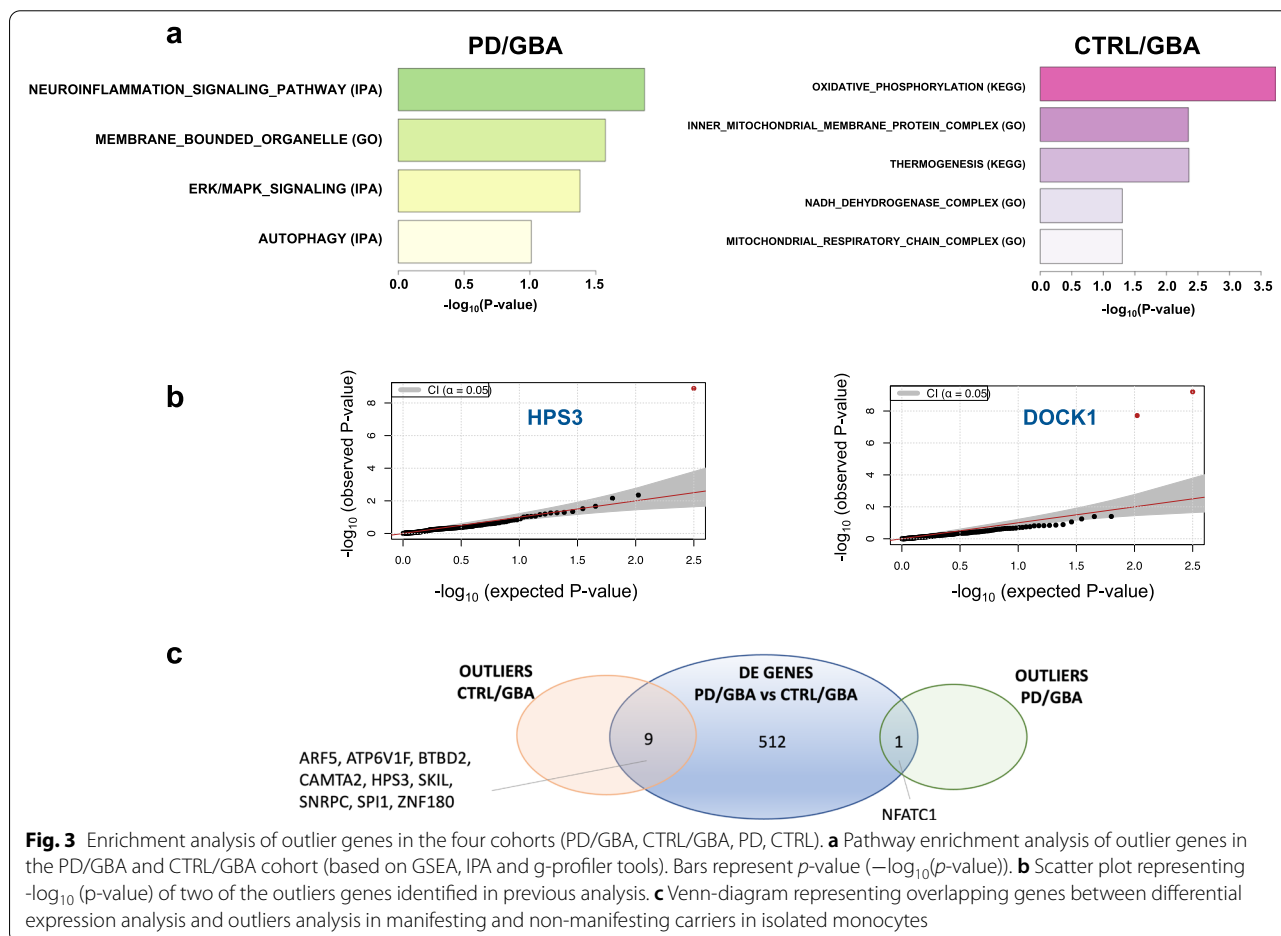
**Whole blood transcriptomic profiles only partially reflect the findings in isolated CD14+ monocytes**

We then assessed the overlap between the transcriptomic profiles in our monocytes cohort with an independent cohort of whole blood (Fig. 4a, Supplementary Table 5). Monocyte counts and RNA concentration in

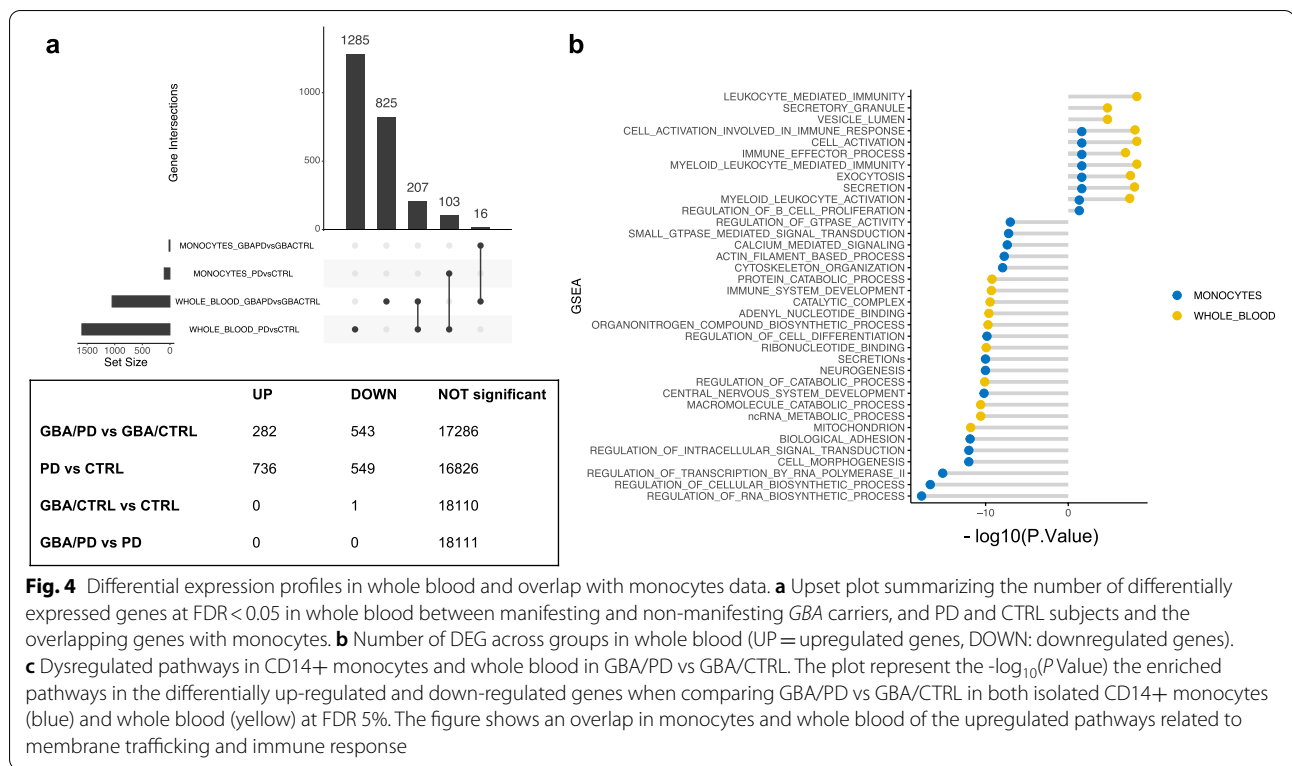
our cohort and in the PPMI study were not significantly different across the four groups, and therefore should not be accountable for the differences of the transcriptomic profiles (Supplementary Fig. 2 and 9a-b). Instead, whole blood composition showed a significant difference in the number of neutrophils and lymphocytes in *GBA* manifesting vs non-manifesting carriers, which may impact differential analyses of transcriptomic data (Supplementary Fig. 9b).

Although we found a high correlation of mean gene expression levels between monocyte and whole blood (Supplementary Fig. 11), the reproducibility of DEG in the two cohorts was only partial (Fig. 4a).

Interestingly, when comparing PD/GBA vs CTRL/GBA we found overlaps of enriched upregulated pathways related to vesicle trafficking and activation of the myeloid compartment in monocytes and whole blood (Fig. 4b). The number of overlapping DEG between GBA/PD and GBA/CTRL in monocytes and whole blood was limited ( $n = 16$ ), although with a high degree of concordance in direction (Fig. 4a, Supplementary Fig. 10c, Supplementary Table 6, Supplementary Table 7). Target genes that were found to be significantly differentially expressed in



**Fig. 3** Enrichment analysis of outlier genes in the four cohorts (PD/GBA, CTRL/GBA, PD, CTRL). **a** Pathway enrichment analysis of outlier genes in the PD/GBA and CTRL/GBA cohort (based on GSEA, IPA and g-profiler tools). Bars represent  $p$ -value ( $-\log_{10}(p\text{-value})$ ). **b** Scatter plot representing  $-\log_{10}(p\text{-value})$  of two of the outliers genes identified in previous analysis. **c** Venn-diagram representing overlapping genes between differential expression analysis and outliers analysis in manifesting and non-manifesting carriers in isolated monocytes



**Fig. 4** Differential expression profiles in whole blood and overlap with monocytes data. **a** Upset plot summarizing the number of differentially expressed genes at FDR < 0.05 in whole blood between manifesting and non-manifesting *GBA* carriers, and PD and CTRL subjects and the overlapping genes with monocytes. **b** Number of DEG across groups in whole blood (UP = upregulated genes, DOWN: downregulated genes). **c** Dysregulated pathways in CD14+ monocytes and whole blood in GBA/PD vs GBA/CTRL. The plot represent the  $-\log_{10}(P\text{ Value})$  the enriched pathways in the differentially up-regulated and down-regulated genes when comparing GBA/PD vs GBA/CTRL in both isolated CD14+ monocytes (blue) and whole blood (yellow) at FDR 5%. The figure shows an overlap in monocytes and whole blood of the upregulated pathways related to membrane trafficking and immune response

monocytes (such as *ATP13A2*, *LRRK2*, and *NOTCH1*) were not differentially expressed in whole blood, suggesting that purified cell population may increase the power for detecting important changes despite smaller cohorts (Supplementary Fig. 12a). Similarly to monocytes, *GBA* was not differentially expressed in whole blood across our four groups (Supplementary Fig. 12b).

When we compared the expression profiles of PD subjects with vs without *GBA*-variants in whole blood, no significantly differentially expressed genes at FDR < 0.15 were detected (Fig. 4a). At a nominal *p*-value < 0.001, some overlapping DEG between monocytes and whole blood were found, including *LMNA* (Supplementary Fig. 12b). Enrichment analysis for differentially expressed genes with nominal *p*-value < 0.01 highlighted pathways related to the immune response and cellular transport, but not to mitochondria and vesicular trafficking (Supplementary Fig. 12c). The number of DEG in subjects with PD with and without *GBA* variants was too limited to perform a pathway analysis.

**Ultrastructural characterization of CD14+ monocytes confirms membrane vesicle impairment in PD and GBA carriers**

Differentially expressed gene profiles from monocytes and whole blood both suggested a dysregulation of genes involved in the endo-lysosomal pathways and mitochondria.

To assess whether this had an impact at a molecular level, the ultrastructure analysis of CD14+ monocytes from our four cohorts were performed (Supplementary Table 7).

Qualitative assessments showed that in the CTRL/*GBA* and CTRL groups samples showed normal cell size (12–20 μm), with intact cytoplasm organelles (Golgi (G), endoplasmic reticulum (ER)/rough ER, ribosomes, pinocytotic vesicles, lysosomal granules, phagosomes, mitochondria, and microtubules), indented nuclei, and pseudopodia extend from cell surfaces (Fig. 5a, i/i'; ii/ii').

Images from the PD group showed that the majority of cells were relatively intact. However, subpopulations presenting distinct characteristics were identified. These cells showed impairment of cell membranes, manifesting with enhancing of the cytoplasmic membrane (in the Golgi and ER) and lacking external membranous encapsulation and swelling of internal swollen cristae of the mitochondria (MT) (Fig. 5a, iii/iii').

In the PD/*GBA* group, distinct subpopulations of monocytes showed small and large vacuoles, but normal cell membranes, pseudopodal extensions, and nuclei intact. Abnormal membrane-organelles, such as Golgi/ER, RER were severely affected in these cells as well. However, quantification of the number of mitochondria, lysosome, vesicles, and vacuolar inclusions found no statistical difference between the four groups (Fig. 5b).

## Discussion

The involvement of the immune system in the pathogenesis of PD, both of the innate immune system in the periphery and in the brain and the adaptive immune system, has attracted growing attention over the last few years [29–31, 33–35, 39, 64–66]. GCase, the enzyme encoded by the PD-genetic risk factor *GBA*, is important for the metabolic processes of scavenger cells such as monocytes and macrophages. Compared to previous works that performed unbiased transcriptomic analysis comparing PD and CTRL subjects, we report here the first transcriptomic analysis in whole blood and CD14+ monocytes stratified by *GBA* mutation status [41, 45]. We found that expression profiles of patients with PD and *GBA* variants were significantly different from both idiopathic PD patients and non-manifesting carriers (Fig. 1b-c) with only a partial overlap with genes differentially expressed between PD and CTRL subjects without *GBA* mutations, suggesting that specific pathways are altered in the presence of variants of this gene (Fig. 1c).

Deregulated genes and pathways in manifesting *GBA*-carriers highlighted molecular mechanisms previously associated with PD pathogenesis in other cellular or animal models, such as the endo-lysosomal pathway, the mitochondrial pathway, inflammatory markers as well as genes related to monogenic forms of PD and PD GWAS (such as *BST1* and *SNCA*) (Fig. 2a) [38, 67–69]. This suggests that monocytes can mirror many processes reported in cellular and animal models of PD as well as dopaminergic neurons. Interestingly, ultrastructural changes in monocytes may reflect transcriptomic signatures. In subpopulations of cells from *GBA*/PD subjects, we found impairment of the membrane compartment, including a poor organization of the endoplasmic reticulum and Golgi (Fig. 5). Whether these changes are consistent with an active role of the innate immune compartment in causing the disease or represent just a reactive response cannot yet be inferred from our data, but it would be worth further exploring. Nevertheless, monocytes may represent a good platform to recapitulate and study PD-associated pathogenic mechanisms. Even more

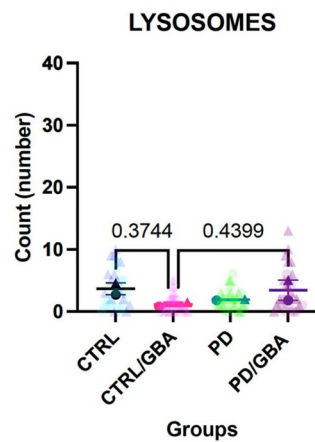
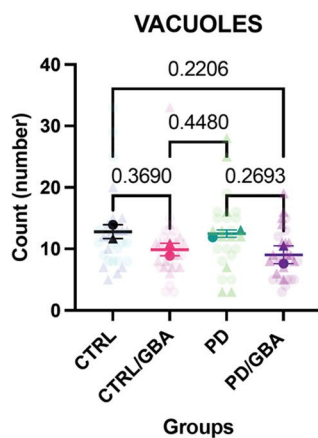
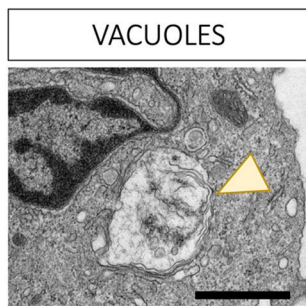
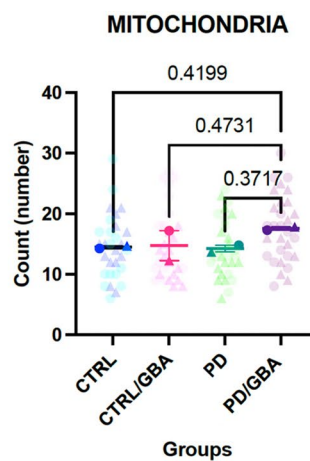
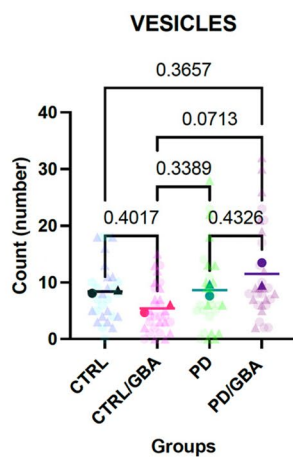
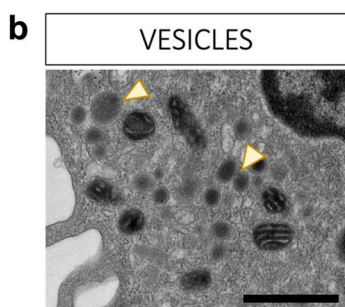
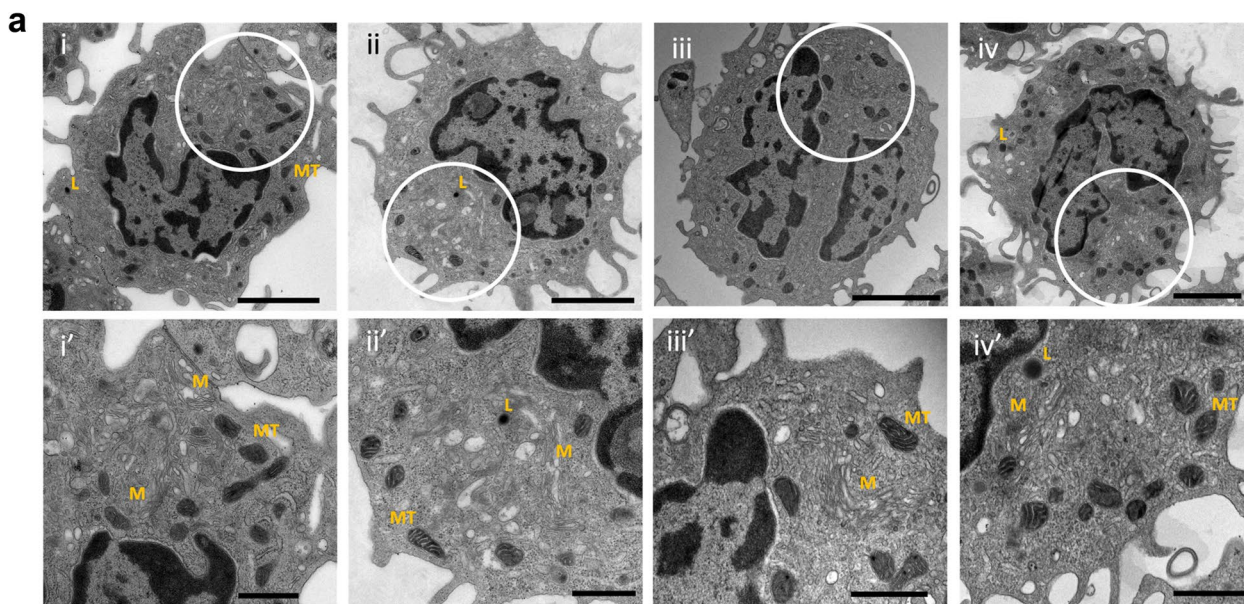
so, despite the small number of subjects in our cohorts – which, to the best of our knowledge, still represents the largest transcriptomic profiles analysis in patients with PD carrying mutations of the *GBA* gene – we were able to identify a large number of differentially expressed genes. This suggests the importance of considering purified cell types, such as isolated CD14+ monocytes, to dramatically reduce the variability due to background noise signals, such as in whole blood.

When we compared the expression profiles of subjects with PD with and without mutations of the gene *GBA*, we found an interesting dysregulation of aging-related targets (Fig. 2c-d). The question remains whether aging processes of PD can be further accelerated in carriers of *GBA* mutations or whether the pathogenic variants of this gene are instead responsible in the first place for the activation of pathways that can cause accelerating aging. Interestingly, we also reported that monocytes of subjects with PD and *GBA* mutations showed an increased expression of the *SNCA* gene, the hallmark protein of PD (Fig. 2c). We know that in brains of subjects with *GBA*-related PD there is a robust deposition of this protein, which could explain the more aggressive phenotype in carriers of these mutations compared to idiopathic PD in terms of an earlier age of onset and increased frequency of cognitive impairment and non-motor symptoms [70]. It can be speculated that increased expression levels of *SNCA* can represent a compensatory mechanism to its aberrant accumulation due to decreased GCase activity, as previously described [71, 72]. However, *SNCA* upregulation could also represent a precipitating factor for disease onset in predisposed subjects, such as carriers of mutations of *GBA*.

Only a small subset of genes were differentially expressed in both monocytes and whole blood from manifesting vs non-manifesting *GBA*-carriers (Supplementary Table 6). We hypothesized that monocytes-specific changes can be missed when analyzing multiple cell types at once in whole blood. Indeed, cell type composition in whole blood showed a significant difference in the number of neutrophils and lymphocytes in manifesting

(See figure on next page.)

**Fig. 5** Morphological characterization of CD14+ monocytes. **a** Low magnification (1800X) images (i, ii, iii, iv) and high magnification (4000x) images (i', ii', iii', iv'). Location of the high magnification images within low magnification pictures are highlighted by the white circle in the first row. Scale bar: 2  $\mu$ m and 800nm respectively. i/i') Cells from CTRL and ii/ii') CTR/*GBA* subjects showing normal membrane compartment (M, Golgi and Endoplasmic reticulum), lysosomes (L), mitochondria (MT), cell-to-cell adhesences, multiple pseudopodal extensions from the cell membrane. iii/iii') Cells from PD subjects showing highly thickened, and distorted cell membrane compartment (M, Golgi and Endoplasmic reticulum); decreased pseudopodia, RER and free ribosomes; mitochondrial (MT) membranes are severely affected, often lacking external membranous encapsulation and with internally swollen cristae. iv/iv') Cells from PD/*GBA* subjects showing small and large vacuoles, normal cell membranes, pseudopodal extensions appear normal, and nuclei. Abnormal membrane assembly (M), ultrastructure, and free ribosomes. **b** Quantification of mitochondria, lysosome, vesicles and vacuoles in four groups ( $n = 2$  subjects per group, 15 cells per sample). Example of vacuole and vesicles reported in the figures on the left highlighted by yellow arrowhead (scale bar 800nm). Data represents count from each cell from the two samples per group, and mean and SEM of the two replicates.  $p$ -value from one-way ANOVA analysis is reported above each comparison ( $p$ -value  $< 0.5$  were reported)



**Fig. 5** (See legend on previous page.)

vs non-manifesting *GBA* carriers, that could affect the overall transcription signal (Supplementary Fig. 9b). On the contrary, the expression of other gene sets may be more consistent across cell types, and thus overlap between monocytes and whole blood. Previous works from the literature showed that additional cell types of the innate immune compartment as well as T-cells, are involved in PD (65,81). To further explore these hypotheses it will be worth analyzing the expression profiles in PD and CTRL subjects in different purified cell types.

In our previous work assessing expression profiles of circulating monocytes and microglia in idiopathic PD patients, we identified an opposite deregulation of the mitochondrial signature in the immune cells in the periphery compared to the central nervous system in PD patients [41]. Impaired mitochondrial genes were also detected as outliers among CTRL/*GBA* subjects, suggesting an involvement and a possible modulatory effect of these genes in the *GBA*-related pathogenic mechanisms as previously suggested (Fig. 3) [73].

One limitation of our study, which is shared with many of the works comparing profiles of manifesting vs non-manifesting carriers of mutations with a reduced penetrance, is the fact that we cannot rule out whether some of the non-manifesting carriers of *GBA* mutations will eventually manifest PD at the end of their life. However, because of our large sample size (considering both monocytes and whole blood) and the reduced penetrance of *GBA*, even if some of the subjects who are currently considered non-manifest carriers will eventually phenocconvert to PD, this will not significantly impact the results of our analysis. In the future, it will be interesting to look for the possible prodromal changes in these subjects through longitudinal analysis.

Our analyzes aimed to leverage immune cells as a platform to identify in a systematic and rigorous manner dysregulated pathways and target genes that can represent potential modifiers of *GBA*-driven PD. Future validation analysis in experimental models (cellular or animal models) will help clarify the contribution of these modifiers in the disease mechanisms. In addition, integration with genome sequencing data may allow us to identify genetic variants that can modulate *GBA* penetrance.

## Conclusions

In conclusion, our results showed that peripheral innate immune cells can be informative in the assessment of disease mechanisms associated with PD. We identified a set of genes and molecular pathways that are specific for *GBA*-related PD, such as a deregulation of *SNCA* and pathways related to metabolism of beta-amyloid and aging compared to PD, as well as dysregulation of

the lysosomal, membrane trafficking, and mitochondrial processing in manifesting compared to non-manifesting *GBA*-carriers (including genes such as *LRRK2*, *ATP13A2*, *BST1* and *NOTCH1*), also confirmed at the ultrastructural level in monocytes. Further investigation will clarify the possible role as disease biomarkers and of these hits.

## Abbreviations

ACD: Acid citrate dextrose; ADRC: Alzheimer's Research Center; AJ: Ashkenazy Jewish; ANGPT1: Angiopoietin 1; AP2A1: Adaptor Related Protein Complex 2 subunit alpha-1; APEX1: Apurinic/Apyrimidinic Endonuclease 1; ARHGAP1: Rho GTPase activating protein; ATG7: Autophagy Related 7; ATP13A2: ATPase Cation Transporting 13A2; ATP2A2: ATPase Sarcoplasmic/Endoplasmic Reticulum Ca<sup>2+</sup> Transporting 2; BIN1: Bridging Integrator 1; BP: Biological Processes; BPMD: Bendheim Parkinson and Movement Disorders (Center at Mount Sinai School of Medicine (NY, US)); BST1: Bone marrow stromal cell antigen 1, ADP-ribosyl cyclase 2, CD157; CC: Cellular Component; CCH: Center for Cognitive Health; CHML: CHM Like Rab Escort Protein; CPM: Count per million; CTRL: Control; CTSB: Cathepsin B; DOCK1: Dedicator of Cytokinesis 1; ER: Endoplasmic reticulum; ERLIN1: ER lipid raft associated 1; FDR: False Discovery Rate; FILIP1L: Filamin A Interacting Protein 1 Like; FUCA2: Alpha-L-Fucosidase 2; *GBA*: Glucocerebrosidase; *GBAP*: Glucocerebrosidase pseudogene; *GCASE*: Beta-glucocerebrosidase; GO: Gene Ontology; GSEA: Gene Set Enrichment Analysis; GWAS: Genome wide association study; HEXB: Hexosaminidase Subunit Beta; HPS3: Hermansky-Pudlak Syndrome 3 Protein; IRAK4: Interleukin 1 Receptor Associated Kinase 4; IRB: Institutional Review Board; KMT2D: lysine methyltransferase 2D; ICAM1: Intracellular Adhesion Molecule 1; IL: Interleukin-; IPA: Ingenuity pathway analysis; iPD: Idiopathic Parkinson's disease; ITM2B: Integral membrane protein 2B; *GSA*: Global Screening Array; HGNC: HUGO Gene Nomenclature Committee; LAMTOR2: Late Endosomal/Lysosomal Adaptor, MAPK And MTOR Activator 2; LFNG: O-Fucosylpeptide 3-Beta-N-Acetylglucosaminyltransferase; LMNA: Lamin A/C; *LRRK2*: Leucine-Rich Repeat Kinase 2; LSD: lysosomal storage disorders; MAF: Minor allele frequency; MAP 4: Microtubule Associated Protein 4; MF: Molecular Function; MIB1: Mindbomb E3 Ubiquitin Protein Ligase 1; MRPL4: Mitochondrial Ribosomal Protein L4; mRNA: Micro RNA; MT-ND (2, 4): Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit (2, 4); mTORC1: Mechanistic target of rapamycin complex 1; MyND: Myeloid cells in Neurodegenerative Disease; NFATC (1, 3): Nuclear Factor of Activated T-cells, cytoplasmic (1, 3); NIH: National Institutes of Health; *NOTCH1*: Notch receptor 1; OUTRIDER: OUTlier in RNA-seq flnDER; PBMC: Peripheral blood mononuclear cells; PCA: Principal component analysis; PD: Parkinson's disease; PDHX: Propionyl-CoA Carboxylase Subunit Beta; PIK3R5: Phosphoinositide 3-kinase regulatory subunit 5; POLR2D: RNA polymerase II subunit D; QC: Quality control; RIN: RNA integrity number; RNA-seq: RNA-sequencing; RPTOR: Regulatory associated protein of MTOR complex 1; RSRP1: Arginine and Serine Rich Protein 1; *SNCA*: Alpha-synuclein; SNP: Single nucleotide polymorphism; SRGAP1: SLIT-ROBO Rho GTPase activating protein 1; SVA: Surrogate variable analysis; SYNJ1: Synaptojanin 1; TMM: Trimmed Mean of M-values; TPM: Transcripts Per Million; TSC2: Tuberous Sclerosis Complex 2; UTY: Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked; XIST: X-inactive specific transcript.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13024-022-00554-8>.

**Additional file 1: Supplementary Table 1.** Clinical characterization of study cohort. Summary of demographic, clinical and genetic features of the cohort of subjects (PD and CTRL) whose purified CD14<sup>+</sup> monocytes were used for integrated genomic analysis.

**Additional file 2: Supplementary Table 2.** Targeted pathway enrichment in PD/*GBA* vs CTRL/*GBA*. List of endolysosomal pathways (from GO terms and curated pathways, i.e. ubiquitin pathway) reported in Fig. 5. *p*-value of enrichment as per Fisher exact test of each pathway within the

set of up-regulated (UP) and down-regulated (DOWN) genes. Significant enriched pathways (at  $p$ -value < 0.15) are highlighted in red.

**Additional file 3: Supplementary Table 3.** GO terms pathways enrichment analysis. The table reports the GO terms and ID for the pathways that were tested for enrichment analysis of differentially expressed genes between the CTRL/GBA and PD/GBA cohorts.

**Additional file 4: Supplementary Table 4.** Number of subjects (PD and CTRL) with different *GBA* variants, reported in the text.

**Additional file 5: Supplementary Table 5.** Summary of demographic, clinical and genetic features of the PPMI cohort (whole blood) of subjects (PD and CTRL) in this study.

**Additional file 6: Supplementary Table 6.** Overlapping genes between monocytes and PPMI cohort (whole blood) in manifesting vs non-manifesting *GBA*-carriers. List of overlapping genes differentially expressed in monocytes and whole blood in manifesting vs non-manifesting carriers. Gene name, description of gene function and logFC and FDR in monocytes and whole blood are reported. In "grey" genes that represent PD-GWAS hits are highlighted.

**Additional file 7: Supplementary Table 7.** Up-regulated and down-regulated pathway related to vesicle and membrane trafficking in isolated CD14+ monocytes and whole blood. The table shows the list of pathways that were found to be deregulated in both isolated CD14+ monocytes and whole blood in the PD/GBA vs CTRL/GBA groups (top), the deregulated genes in each pathway (first column), and the directionality of dysregulation (second and third column). Red "x" indicates genes dysregulated in whole blood, black "x" represents genes dysregulated in isolated CD14+ monocytes.

**Additional file 8: Supplementary Fig. 1.** Characterization of genetic background of donor population. a) Representation of PCA analysis of ancestry of MDS values from the cohort of 158 subjects (PD/GBA, PD, CTRL/GBA, CTRL) compared to 1000 Genome Project samples (Phase 3). The different ancestry are represented in distinct colors (Orange: African; Gold: Ad Mixed American; Green: East Asian; Blue: European; Purple: South Asian; Black: study cohort). b) PCA considering only overlap of MDS values of donor cohort (black) with European ancestry (blue) and AJ ancestry (light blue). **Supplementary Figure 2.** Normalization and quality control of RNA-seq data from isolated CD14+ monocytes. a) Violin plot representing the contribution of each of the surrogate variables (as explained in the text) to the variability of expression data of the study cohort and residual (158 subjects). b) Violin plot representing the contribution of technical, demographical and clinical variables to the variability of expression data of the study cohort and residual (158 subjects). c) Heatmap representing the results of linear regression between the surrogate variables utilized for data normalization and technical variables (from RNA-seq analysis) and metadata. Coefficient of linear regression is reported in the heatmap for each correlation pair. d) Distribution of MDS values of study cohort identified a clear clustering based on batches used for RNA-seq analysis (batches 1 to 4). e) After regression of SVs, variability of MDS values is significantly reduced, with no significant outliers and no clustering based on experimental batches.

**Supplementary Fig. 3.** Differential expression of *GBA* in CD14+ isolated monocytes. Box plot representing differential expression levels (normalized expression count) of *GBA* in isolated CD14+ monocytes (a and b). In b) data from isolated CD14+ monocytes of *GBA*-carriers and non-carriers within PD and CTRL subjects were combined and compared. Each dot represents a subject. Dots are colored based on *GBA* mutations (as reported in the legend: *GBA* mild mutations (N370S, E326K, R496H), *GBA* severe mutations (L444P/A456P/RecNcil, V394L, 84GG, 84GG/T369M, N370S/RecNcil)).  $p$ -value of differential expression levels is reported on top (statistics: Mann-Whitney U test). *GBA* variants different from N370S are labeled in the boxplot. **Supplementary Fig. 4.** Differential expression of target genes in manifesting and non-manifesting carriers in monocytes. a) Volcano plot showing differentially expressed genes with FDR < 0.05 (green dots) in isolated monocytes from manifesting vs non-manifesting carriers. Genes related to targeted pathways (*LRKK2*, *ATP13A2*, *NOTCH1* and *TAU*) are highlighted. b) Box plots of differential levels of expression of the targeted genes (*ATP13A2*, *LRKK2*, *NOTCH1*), between manifesting and

non-manifesting carriers in isolated monocytes. Each dot represents a subject. Dots are colored based on *GBA* mutations (as reported in the legend: *GBA* mild mutations (N370S, E326K, R496H), *GBA* severe mutations (L444P/A456P/RecNcil, V394L, 84GG, 84GG/T369M, N370S/RecNcil)).

Asterisks indicate significant  $p$ -value (\* =  $p$ -value < 0.05, \*\* =  $p$ -value < 0.01, \*\*\* =  $p$ -value < 0.001, statistics: Mann-Whitney U test). **Supplementary Fig. 5.** Differential expression analysis in monocytes in PD patients with and without *GBA*-variants. a) Volcano-plot representing logFC (x-axes) and  $p$ -value (y-axes,  $-\log_{10}$   $p$ -value) of differential expressed genes between PD/GBA and PD as per nested interaction model. Highlighted in yellow are genes with FDR < 0.15 (44 total genes). ID labels of functionally relevant genes and of genes differentially expressed in whole blood (PPMI overlap of genes at nominal  $p$ -value < 0.001) are reported in the plot. b) Differential normalized expression count of *SNCA* and *LMNA* between PD/GBA and PD, compared to CTRL/GBA and CTRL subjects in isolated CD14+ monocytes. Asterisks indicate significant  $p$ -value (\* =  $p$ -value < 0.05, \*\* =  $p$ -value < 0.01, \*\*\* =  $p$ -value < 0.001, statistics: Mann-Whitney U test). Disease and genetic status are reported on the x-axes. Each dot represents a subject. Dots are colored based on *GBA* mutations (as reported in the legend: *GBA* mild mutations (N370S, E326K, R496H), *GBA* severe mutations (L444P/A456P/RecNcil, V394L, 84GG, 84GG/T369M, N370S/RecNcil)). *GBA* variants different from N370S are labeled in the boxplot. **Supplementary Fig. 6.** Differential expression profiles across the four cohorts based on diagnosis and genetic status interaction. a) Volcano-plot representing  $\log_2$  fold change (x-axes) and  $P$ -Value (y-axes,  $-\log_{10}$   $P$ -Value) of differential expressed genes based on diagnosis and genetics interaction between the four cohorts (PD/GBA, CTRL/GBA, PD, CTRL). Genes with FDR < 0.05 are highlighted in blue and labeled with their IDs. Differentially expressed genes encompassed: *ANGPT1* (angiotensinogen converting enzyme 1), *FILIP1L* (Filamin A Interacting Protein 1 Like), *AC138028.2* (novel transcript), *SRGAP1* (Slit-Robo GTPase-activating protein 1), *PIK3R5* (Phosphoinositide 3-kinase regulatory subunit 5, responsible for Ataxia with Oculomotor-Apraxia type 3), *HPS3* (Hermansky-Pudlak Syndrome 3 Protein, biogenesis of lysosomal organelle complex 2 subunit 1). b) Box plots representing expression levels (normalized expression count) of one differentially expressed targeted gene according to interaction term (diagnosis and genetics interaction). Disease and genetic status are labeled on the x-axes. Boxes are colored based on disease and genetic status. Each dot represents a subject. Dots are colored based on *GBA* mutations (as reported in the legend: *GBA* mild mutations (N370S, E326K, R496H), *GBA* severe mutations (L444P/A456P/RecNcil, V394L, 84GG, 84GG/T369M, N370S/RecNcil)). Asterisks indicate significant  $p$ -value (\* =  $p$ -value < 0.05, \*\* =  $p$ -value < 0.01, \*\*\* =  $p$ -value < 0.001; statistics: Mann-Whitney U test). **Supplementary Fig. 7.** QC for analysis of outlier genes. a)

Normalization based on surrogate variables, as provided by the OUTRIDER script, of a total of 13711 genes (considering only genes with > 30% of total expression). Discrete relevant variables (Diagnosis, batches of RNAseq analysis (rna\_batch), *GBA* mutations and *GBA*-related genetic status (carriers or non-carriers of *GBA* mutations), gender (Sex: male (M) and female (F)) are labeled at the top of the heatmap per each subject. b) Bar-plot reporting number of outlier genes per each subject (out of 158 subjects). Highlighted in orange samples with outlier gene count above 0.1%. c) Summary tables: on the left: number of outliers genes per cohort (PD/GBA, CTRL/GBA, PD, CTRL) (493 pairs); on the right: number of subjects per each cohort with at least one outlier gene (125 unique subjects total with at least one outlier gene). **Supplementary Fig. 8.** Scatter plot of selected genes identified through the OUTRIDER method. Scatter plot representing  $-\log_{10}$  ( $p$ -value) of outliers genes identified in previous analysis. **Supplementary Fig. 9.** Cell proportion across patient and control cohorts. a) Monocyte counts ( $\times 10^6$ ) in the GBA/PD, PD, GBA/CTRL, and CTRL groups in the cohort of isolated monocytes (NYMD cohort). Each dot represents the cell count of one subject. Cell count was available for 57 subjects in our cohort. Statistical analysis according to the Wilcoxon test is reported for the comparison between the four groups. b) Cell proportion (neutrophils, monocytes, lymphocytes, eosinophils, basophils) within whole blood samples in the four groups (GBA/PD, PD, GBA/CTRL, and CTRL) in the PPMI cohort. Comparison of each cell type percent across the four groups is reported on the box plots on the right side of the figure (statistical analysis: Wilcoxon test). **Supplementary Fig. 10.** Correlation between genes differentially expressed in isolated

CD14+ monocytes and whole blood. We compared the directionality of differentially expressed genes between a) GBA/PD vs GBA/CTRL and PD vs CTRL in CD14+ monocytes ( $n = 197$ ); b) GBA/PD vs GBA/CTRL and PD vs CTRL in whole blood ( $n = 207$ ); c) GBA/PD vs GBA/CTRL in CD14+ monocytes and whole blood ( $n = 16$ ); d) PD vs CTRL in CD14+ monocytes and whole blood ( $n = 103$ ). **Supplementary Fig. 11.** Correlation between gene expression levels in isolated CD14+ monocytes and whole blood. Genes with expression with more than 1 CPM in 30% of the samples were considered from both cohorts (discovery cohort: isolated CD14+ monocytes (total number of genes: 13711), validation cohort: whole blood - PPMI cohort (total number of genes: 18111)). Spearman correlation between levels of normalized mean gene expression across subjects within each cohort per sub-group of subjects was calculated ( $R = 0.78$   $p < 0.001$ ). Genes were normalized with TMM and voom, as detailed in the main text. **Supplementary Fig. 12.** Validation in whole blood of differentially expressed genes in monocytes. a) Differential levels of expression of the targeted genes (*ATP13A2*, *LRRK2*, *NOTCH1*, between manifesting and non-manifesting carriers in whole blood from manifesting and non-manifesting GBA-mutation carriers. b) Differential normalized expression count of *SNCA*, *LMNA*, and *GBA* between PD/GBA and PD, compared to CTRL/GBA and CTRL subjects in whole blood. In a) and b) each dot represents a subject. Dots are colored based on *GBA* mutations (as reported in the legend: *GBA* mild mutations (N370S, E326K, R496H), *GBA* severe mutations (L444P/A456P/RecNcil, V394L, 84GG, 84GG/T369M, N370S/RecNcil)).  $p$ -value of different expression levels is reported on top (statistics: Mann-Whitney U test). c) Pathway enrichment analysis of differentially expressed genes in whole blood between PD/GBA vs PD subjects with  $p$ -value  $< 0.01$  for GO terms are reported. Dark blue: pathways related to cell transport; Green: pathways related to immune response; Light blue: other pathways.

### Acknowledgements

We thank the study participants at the Marlene and Paolo Fresco Institute for Parkinson's and Other Movement Disorders for providing their time and participation in the study. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). For up-to-date information on the study, visit [www.ppmi-info.org](http://www.ppmi-info.org). PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including [list the full names of all of the PPMI funding partners found at [www.ppmi-info.org/fundingpartners](http://www.ppmi-info.org/fundingpartners)]. G.R. was supported by the American Parkinson's Disease Association Post-Doctoral Fellowship 2018, the Parkinson's Foundation Clinical Research Award PF-CRA-1940, and The Fresco Institute for Parkinson's and Movement Disorders Clinical Fellowship. T.R. supported by grants from the Michael J. Fox Foundation (Grant #14899 and #16743), US National Institutes of Health NIH NINDS R01-NS116006, NINDS U01-NS120256, NIA R01-AG054005, NIA R21-AG063130, and NIA U01 P50-AG005138.

### Authors' contributions

Conceptualization: TR, GMR; Methodology: GMR, TR, RAV, EN, EU, KPL, ONJ, SDK, WJ, AS; Formal analysis and investigation: GMR, RAV, EN, EU, KPL, AA, MP, BH, KA, CA, MZ, ST, ONJ, SDK, WJ, AS; Writing - original draft preparation: GMR, TR; Writing - review and editing: GMR, TR, RAV, EN, EU, KPL, AA, MP, BH, KA, CA, MZ, ST, ONJ, SDK, WJ, AA, JFC, ADF, GPC, SJF; Funding acquisition: TR, GMR; Resources: TR, SJF; Supervision: TR. All authors read and approved the final manuscript.

### Funding

G.R. was supported by the American Parkinson's Disease Association Post-Doctoral Fellowship 2018 and the The Marlene and Paolo Fresco Clinical Fellowship; T.R. supported by grants from the Michael J. Fox Foundation (Grant #14899 and #16743), US National Institutes of Health NIH NINDS R01-NS116006, NINDS U01-NS120256, NIA R01-AG054005, NIA R21-AG063130, and NIA U01 P50-AG005138.

### Availability of data and materials

The datasets supporting the conclusions of this article (Raw RNA-seq data) are available as part of the Myeloid cells in Neurodegenerative Disease (MyND)

study via dbGAP (study accession ID: phs002400.v1.p1) at [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs002400.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002400.v1.p1). RNA-seq data for Parkinson's Progression Markers Initiative (PPMI) cohort were obtained from the Accelerating Medicines Partnership program for Parkinson's disease (AMP-PD) Knowledge Platform. For up-to-date information on the study, <https://www.amp-pd.org>.

### Declarations

#### Ethics approval and consent to participate

All the procedures involving human subjects were performed upon written informed consent, approval from the institutional review board and in accord with the Helsinki Declaration of 1975. Informed consent was obtained from all individual participants included in the study.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>The Marlene and Paolo Fresco Institute for Parkinson's Disease and Movement Disorders, New York University Langone Health, 222 East 41st street, New York, NY 10017, USA. <sup>2</sup>Nash Family Department of Neuroscience & Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. <sup>3</sup>Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. <sup>4</sup>Department of Genetics and Genomic Sciences & Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1498, New York, NY 10029, USA. <sup>5</sup>Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1137, New York, NY 10029, USA. <sup>6</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. <sup>7</sup>Department of Biochemistry and Molecular Biology (Universidad Complutense de Madrid) & Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain. <sup>8</sup>Department of Pathology, Icahn School of Medicine at Mount Sinai, 1468 Madison Avenue, Annenberg Building, 15th Floor, New York, NY 10029, USA. <sup>9</sup>Neuropathology Brain Bank & Research CoRE, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, Room 9-22, New York, NY 10029, USA. <sup>10</sup>Microscopy Core and Advanced Bioimaging Center at the Icahn School of Medicine at Mount Sinai Center, 1468 Madison Avenue, Room 18-250, New York, NY 10029, USA. <sup>11</sup>Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milan, Italy. <sup>12</sup>Dino Ferrari Center, Neuroscience Section, Department of Pathophysiology and Transplantation, University of Milan, Via Francesco Sforza, 35, 20122 Milano, MI, Italy. <sup>13</sup>Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, ICAHN 10-70E, New York, NY 10029-6574, USA.

Received: 22 October 2021 Accepted: 29 June 2022

Published online: 17 August 2022

### References

- Gan-Or Z, Giladi N, Rozovski U, Shifrin C, Rosner S, Gurevich T, et al. Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. *Neurology*. 2008;70(24):2277–83.
- Lesage S, Anheim M, Condroyer C, Pollak P, Durif F, Dupuits C, et al. Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. *Hum Mol Genet*. 2011;20(1):202–10.
- Li Y, Sekine T, Funayama M, Li L, Yoshino H, Nishioka K, et al. Clinicogenetic study of GBA mutations in patients with familial Parkinson's disease. *Neurobiol Aging*. 2014;35(4):935.e3–8.
- Sidransky E, Samaddar T, Tayebi N. Mutations in GBA are associated with familial Parkinson disease susceptibility and age at onset. *Neurology*. 2009;73(17):1424–5 author reply 1425–1426.
- Neumann J, Bras J, Deas E, O'Sullivan SS, Parkkinen L, Lachmann RH, et al. Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. *Brain J Neurol*. 2009;132(Pt 7):1783–94.

6. Riboldi GM, Di Fonzo AB. GBA, Gaucher disease, and Parkinson's disease: from genetic to clinic to new therapeutic approaches. *Cells*. 2019;8(4):8(4):364.
7. Velez-Pardo C, Lorenzo-Betancor O, Jimenez-Del-Rio M, Moreno S, Lopera F, Cornejo-Olivas M, et al. The distribution and risk effect of GBA variants in a large cohort of PD patients from Colombia and Peru. *Parkinsonism Relat Disord*. 2019;63:204–8.
8. Zhang Y, Shu L, Zhou X, Pan H, Xu Q, Guo J, et al. A meta-analysis of GBA-related clinical symptoms in Parkinson's disease. *Park Dis*. 2018;2018:3136415.
9. Bandres-Ciga S, Saez-Atienzar S, Bonet-Ponce L, Billingsley K, Vitale D, Blauwendraat C, et al. The endocytic membrane trafficking pathway plays a major role in the risk of Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2019;34(4):460–8.
10. Henderson MX, Sedor S, McGeary I, Cornblath EJ, Peng C, Riddle DM, et al. Glucocerebrosidase activity modulates neuronal susceptibility to pathological  $\alpha$ -Synuclein insult. *Neuron*. 2020;105(5):822–836.e7.
11. Kilpatrick BS, Magalhaes J, Beavan MS, McNeill A, Gegg ME, Cleeter MWJ, et al. Endoplasmic reticulum and lysosomal  $\text{Ca}^{2+}$  stores are remodelled in GBA1-linked Parkinson disease patient fibroblasts. *Cell Calcium*. 2016 Jan;59(1):12–20.
12. Maor G, Cabasso O, Krivoruk O, Rodriguez J, Steller H, Segal D, et al. The contribution of mutant GBA to the development of Parkinson disease in *Drosophila*. *Hum Mol Genet*. 2016;25(13):2712–27.
13. Mazzulli JR, Xu YH, Sun Y, Knight AL, McLean PJ, Caldwell GA, et al. Gaucher disease glucocerebrosidase and  $\alpha$ -synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell*. 2011;146(1):37–52.
14. Papadopoulos VE, Nikolopoulou G, Antoniadou I, Karachaliou A, Ariano-glou G, Emmanouilidou E, et al. Modulation of  $\beta$ -glucocerebrosidase increases  $\alpha$ -synuclein secretion and exosome release in mouse models of Parkinson's disease. *Hum Mol Genet*. 2018;27(10):1696–710.
15. Ron I, Horowitz M. ER retention and degradation as the molecular basis underlying Gaucher disease heterogeneity. *Hum Mol Genet*. 2005;14(16):2387–98.
16. Sanchez-Martinez A, Beavan M, Gegg ME, Chau KY, Whitworth AJ, Schapira AHV. Parkinson disease-linked GBA mutation effects reversed by molecular chaperones in human cell and fly models. *Sci Rep*. 2016 Aug;19(6):31380.
17. Suzuki T, Shimoda M, Ito K, Hanai S, Aizawa H, Kato T, et al. Expression of human Gaucher disease gene GBA generates neurodevelopmental defects and ER stress in *Drosophila* eye. *PLoS One*. 2013;8(8):e69147.
18. Blauwendraat C, Reed X, Krohn L, Heilbron K, Bandres-Ciga S, Tan M, et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain J Neurol*. 2020;143(1):234–48.
19. Alcalay RN, Levy OA, Waters CC, Fahn S, Ford B, Kuo SH, et al. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. *Brain J Neurol*. 2015;138(Pt 9):2648–58.
20. Gan-Or Z, Bar-Shira A, Gurevich T, Giladi N, Orr-Urtreger A. Homozygosity for the MTX1 c.184T>a (p.S63T) alteration modifies the age of onset in GBA-associated Parkinson's disease. *Neurogenetics*. 2011;12(4):325–32.
21. Omer N, Giladi N, Gurevich T, Bar-Shira A, Gana-Weisz M, Goldstein O, et al. A possible modifying effect of the G2019S mutation in the LRRK2 gene on GBA Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2020;35(7):1249–53.
22. Sanyal A, DeAndrade MP, Novis HS, Lin S, Chang J, Lengacher N, et al. Lysosome and inflammatory defects in GBA1-mutant astrocytes are normalized by LRRK2 inhibition. *Mov Disord Off J Mov Disord Soc*. 2020;35(5):760–73.
23. Schierding W, Farrow S, Fadason T, Graham OEE, Pitcher TL, Qubisi S, et al. Common variants Coregulate expression of GBA and modifier genes to delay Parkinson's disease onset. *Mov Disord Off J Mov Disord Soc*. 2020;35(8):1346–56.
24. Straniero L, Rimoldi V, Samarani M, Goldwurm S, Di Fonzo A, Krüger R, et al. The GBAP1 pseudogene acts as a ceRNA for the glucocerebrosidase gene GBA by sponging miR-22-3p. *Sci Rep*. 2017;7(1):12702.
25. Chahine LM, Qiang J, Ashbridge E, Minger J, Yearout D, Horn S, et al. Clinical and biochemical differences in patients having Parkinson disease with vs without GBA mutations. *JAMA Neurol*. 2013;70(7):852–8.
26. Dzamko N, Gysbers A, Perera G, Bahar A, Shankar A, Gao J, et al. Toll-like receptor 2 is increased in neurons in Parkinson's disease brain and may contribute to alpha-synuclein pathology. *Acta Neuropathol (Berl)*. 2017;133(2):303–19.
27. Keatinge M, Bui H, Menke A, Chen YC, Sokol AM, Bai Q, et al. Glucocerebrosidase 1 deficient Danio rerio mirror key pathological aspects of human Gaucher disease and provide evidence of early microglial activation preceding alpha-synuclein-independent neuronal cell death. *Hum Mol Genet*. 2015;24(23):6640–52.
28. Mullin S, Stockholm MG, Hughes D, Mehta A, Parbo P, Hinz R, et al. Brain microglial activation increased in Glucocerebrosidase (GBA) mutation carriers without Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2021;36(3):774–9.
29. Gerhard A, Pavese N, Hottot G, Turkheimer F, Es M, Hammers A, et al. In vivo imaging of microglial activation with [<sup>11</sup>C](R)-PK11195 PET in idiopathic Parkinson's disease. *Neurobiol Dis*. 2006;21(2):404–12.
30. Grozdanov V, Bliederhaeuser C, Ruf WP, Roth V, Fundel-Clemens K, Zondler L, et al. Inflammatory dysregulation of blood monocytes in Parkinson's disease patients. *Acta Neuropathol (Berl)*. 2014;128(5):651–63.
31. Harms AS, Thome AD, Yan Z, Schonhoff AM, Williams GP, Li X, et al. Peripheral monocyte entry is required for alpha-Synuclein induced inflammation and neurodegeneration in a model of Parkinson disease. *Exp Neurol*. 2018;300:179–87.
32. McGeer PL, Itagaki S, Boyes BE, McGeer EG. Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains. *Neurology*. 1988;38(8):1285–91.
33. Mogi M, Harada M, Narabayashi H, Inagaki H, Minami M, Nagatsu T. Interleukin (IL)-1 beta, IL-2, IL-4, IL-6 and transforming growth factor-alpha levels are elevated in ventricular cerebrospinal fluid in juvenile parkinsonism and Parkinson's disease. *Neurosci Lett*. 1996;211(1):13–6.
34. Reale M, Iarlori C, Thomas A, Gambi D, Perfetti B, Di Nicola M, et al. Peripheral cytokines profile in Parkinson's disease. *Brain Behav Immun*. 2009;23(1):55–63.
35. Terada T, Yokokura M, Yoshikawa E, Futatsubashi M, Kono S, Konishi T, et al. Extrastriatal spreading of microglial activation in Parkinson's disease: a positron emission tomography study. *Ann Nucl Med*. 2016;30(8):579–87.
36. Funk N, Wieghofer P, Grimm S, Schaefer R, Bühring HJ, Gasser T, et al. Characterization of peripheral hematopoietic stem cells and monocytes in Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2013;28(3):392–5.
37. Holmans P, Moskva V, Jones L, Sharma M, International Parkinson's Disease Genomics Consortium, Vedernikov A, et al. A pathway-based analysis provides additional support for an immune-related genetic susceptibility to Parkinson's disease. *Hum Mol Genet*. 2013;22(5):1039–49.
38. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46(9):989–93.
39. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. 2014;344(6183):519–23.
40. Wissemann WT, Hill-Burns EM, Zabetian CP, Factor SA, Patsopoulos N, Hoglund B, et al. Association of Parkinson disease with structural and regulatory variants in the HLA region. *Am J Hum Genet*. 2013;93(5):984–93.
41. Navarro E, Udine E, de Paiva Lopes K, Parks M, Riboldi G, Schilder BM, et al. Dysregulation of mitochondrial and proteolysosomal genes in Parkinson's disease myeloid cells. *Nat Aging*. 2021;1(9):850–63. <https://doi.org/10.1038/s43587-021-00110-x>.
42. Atashrazm F, Hammond D, Perera G, Dobson-Stone C, Mueller N, Pickford R, et al. Reduced glucocerebrosidase activity in monocytes from patients with Parkinson's disease. *Sci Rep*. 2018;8(1):15446.
43. Hughes LP, Pereira MMM, Hammond DA, Kwok JB, Halliday GM, Lewis SJG, et al. Glucocerebrosidase activity is reduced in cryopreserved Parkinson's disease patient monocytes and inversely correlates with motor severity. *J Parkinsons Dis*. 2021;11(3):1157–65.
44. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992;55(3):181–4.
45. Craig DW, Hutchins E, Violich I, Alsop E, Gibbs JR, Levy S, et al. RNA sequencing of whole blood reveals early alterations in immune cells and gene expression in Parkinson's disease. *Nat Aging*. 2021;1(1):734–47.
46. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.



47. Carmi S, Hui KY, Kochav E, Liu X, Xue J, Grady F, et al. Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat Commun*. 2014;9(5):4835.
48. Shah, H. PgmNr 1856: RAPiD—An Agile and Dependable RNA-Seq Framework (American Society of Human Genetics, 2015).
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinforma Oxf Engl*. 2009;25(16):2078–9.
50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl*. 2014;30(15):2114–20.
51. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl*. 2013;29(1):15–21.
52. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
53. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1–2–3 with limma, Glimma and edgeR. *F1000Research*. 2018;5:ISCB Comm J-1408.
54. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma Oxf Engl*. 2012;28(6):882–3.
55. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
56. Brechtmann F, Mertens C, Matusevičiūtė A, Yépez VA, Avsec Ž, Herzog M, et al. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am J Hum Genet*. 2018;103(6):907–17.
57. Cuperfain AB, Zhang ZL, Kennedy JL, Gonçalves VF. The complex interaction of mitochondrial genetics and mitochondrial pathways in psychiatric disease. *Mol Neuropsychiatry*. 2018;4(1):52–69.
58. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9(386):eaal5209.
59. Kremer LS, Bader DM, Mertens C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun*. 2017;8:15824.
60. Luo J, Sun L, Lin X, Liu G, Yu J, Parisiadou L, et al. A calcineurin- and NFAT-dependent pathway is involved in  $\alpha$ -synuclein-induced degeneration of midbrain dopaminergic neurons. *Hum Mol Genet*. 2014;23(24):6567–74.
61. Gonzalo S, Kreienkamp R, Askjaer P. Hutchinson-Gilford progeria syndrome: a premature aging disease caused by LMNA gene mutations. *Ageing Res Rev*. 2017;33:18–29.
62. Pan Z, Li GF, Sun ML, Xie L, Liu D, Zhang Q, et al. MicroRNA-1224 splicing CircularRNA-Filip11 in an Ago2-dependent manner regulates chronic inflammatory pain via targeting Ubr5. *J Neurosci*. 2019;39(11):2125–43.
63. Bhamhani V, Introne WJ, Lungu C, Cullinane A, Toro C. Chediak-Higashi syndrome presenting as young-onset levodopa-responsive parkinsonism. *Mov Disord Off J Mov Disord Soc*. 2013;28(2):127–9.
64. Cook DA, Kannarkat GT, Cintron AF, Butkovich LM, Fraser KB, Chang J, et al. LRRK2 levels in immune cells are increased in Parkinson's disease. *NPJ Park Dis*. 2017;3:11.
65. Li X, Sundquist J, Sundquist K. Subsequent risks of Parkinson disease in patients with autoimmune and related disorders: a nationwide epidemiological study from Sweden. *Neurodegener Dis*. 2012;10(1–4):277–84.
66. Sulzer D, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Liebes J, et al. T cells from patients with Parkinson's disease recognize  $\alpha$ -synuclein peptides. *Nature*. 2017;546(7660):656–61.
67. Chang D, Nalls MA, Hallgrímsson IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet*. 2017;49(10):1511–6.
68. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, 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*. 2019;18(12):1091–102.
69. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkman J, et al. Parkinson disease. *Nat Rev Dis Primer*. 2017;3:17013.
70. Blandini F, Cilia R, Cerri S, Pezzoli G, Schapira AHV, Mullin S, et al. Glucocerebrosidase mutations and synucleinopathies: toward a model of precision medicine. *Mov Disord Off J Mov Disord Soc*. 2019;34(1):9–21.
71. Fishbein I, Kuo YM, Giasson BI, Nussbaum RL. Augmentation of phenotype in a transgenic Parkinson mouse heterozygous for a Gaucher mutation. *Brain J Neurol*. 2014;137(Pt 12):3235–47.
72. Schöndorf DC, Aureli M, McAllister FE, Hindley CJ, Mayer F, Schmid B, et al. iPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. *Nat Commun*. 2014;5:4028.
73. Kim S, Wong YC, Gao F, Krainc D. Dysregulation of mitochondria-lysosome contacts by GBA1 dysfunction in dopaminergic neuronal models of Parkinson's disease. *Nat Commun*. 2021;12(1):1807.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

