scientific reports

Check for updates

OPEN

Single nucleus RNA sequencing profile analysis to reveal cell type specific common molecular drivers of Parkinson's disease and therapeutic agents

Md. Al Amin Pappu[1], Md. Alamin[2]✉, Most. Humaira Sultana[2], AKM Azad[3,4], Md. Rabiul Auwul[5], Sabkat Mahmud[1], Alvira Ajadee[1], Arnob Sarker[1], Salem A. Alyami[3,4] & Md. Nurul Haque Mollah[1]✉

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder, characterized by progressive motor and cognitive decline, leading to long-term disability and significantly impacting quality of life. While PD research has traditionally focused on dopaminergic neurons in the substantia nigra (SN), emerging evidence also suggests glial involvement in disease progression. So, this study explored PD-associated key genes from neuronal and glial cell types to uncover pathogenetic mechanisms and potential therapeutics by employing single-nucleus RNA sequencing (snRNA-seq) data from the accession number GSE184950. A total of 426,886 nuclei were analyzed, yielding 129,473 high-quality nuclei. Through rigorous quality control, clustering, and marker gene analysis using scVI and Scanpy, nine distinct cell types were delineated, including neurons, astrocytes, and microglia. 18 common differentially expressed genes (cDEGs) were identified across neuronal and glial cell types. Gene ontology (GO) and KEGG enrichment analyses revealed key terms associated with neurodegeneration, including PD. A total of six critical KGs, including *HSPA1A*, *DNAJB1*, *BAG3*, *SYN1*, *CALB2*, and *NEFL*, along with their key regulators, were identified by the protein–protein interaction network. Finally, three repurposed drugs (Celastrol, Withaferin-A, and Apomorphine) were suggested as the therapeutics agents for PD by molecular docking. In-silico ADME/T analyses were conducted using pkCSM and SwissADME to evaluate the pharmacokinetic properties of these compounds. These findings could shed light on PD mechanisms and hold promise for advancing diagnostics and therapies.

Keywords Parkinson's disease, Disability research, Single-nucleus RNA sequencing, Differential gene expression, Drug repurposing, Protein–protein interaction network

Parkinson's disease (PD) is a neurodegenerative condition associated with aging, affecting approximately 1–2% of individuals aged 60 and above[1,2]. As the global population ages, the prevalence of PD is anticipated to rise significantly in the forthcoming decades[3]. As of now, PD impacts more than 10 million people worldwide[4]. Despite its high prevalence, PD remains without a targeted cure, with current treatments limited to symptom management through various medications[5], and research indicates that PD arises from a complex interplay between genetic predispositions and environmental influences[6–8]. Despite the identification of numerous genes linked to PD risk, such as *LRRK2*[9] and *VPS35*[10], the precise mechanisms driving the disease remain elusive. The specific genetic markers that can reliably track the progression of PD remain poorly understood. Consequently, it is imperative to discover novel genes that could play significant roles in PD onset and advancement, as these could provide promising therapeutic targets for intervention and treatment.

[1]Bioinformatics Lab (Dry), Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh. [2]Department of Mathematics and Physics, School of Engineering & Physical Sciences, North South University, Dhaka 1229, Bangladesh. [3]Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), 13318 Riyadh, Saudi Arabia. [4]King Salman Center for Disability Research, 11614 Riyadh, Saudi Arabia. [5]Department of Statistics, Gazipur Agricultural University, Gazipur, Dhaka, Bangladesh. ✉email: md.alamin06@northsouth.edu; mollah.stat.bio@ru.ac.bd

The single-nuclei RNA sequencing (snRNA-seq) is a recently developed and transformative tool in genomic research that allows for an unprecedented resolution in dissecting the profiles of gene expression[11]. This technique provides clarity to researchers about cellular heterogeneity and the molecular basis of diseases like PD. This aspect is paramount since PD is an intricate illness affecting a variety of brain cell types. In cases of PD, snRNA-seq has the potential to reveal changes in cell-specific gene expressions that would be hidden under bulk analyses[12]. Data concerning gene expression at the single-nuclei level enables the identification of major affected cell types and their differentially expressed genes (DEGs) in relation to PD. Recognizing these changes at the cellular level is especially important to be able to identify critical genes and pathways involved in the pathology of PD.

There have been a number of snRNA-seq investigations to discover major genes and pathways implicated in PD. These investigations have reported some genes that show different expression levels towards healthy controls than in patients caused by PD. In 2022, Smajić and their team generated a dataset of 6 idiopathic PD patients with more than 41,000 single-nuclei transcriptomes of post-mortem midbrain. They produced the very first snRNA-seq datasets for idiopathic PD midbrain, which points to a disease-specific neuronal cell cluster and 'pan-glial' activation as key mechanisms of the pathology of the movement disorder[13]. A recent study identified that a certain group of glutamatergic neurons is involved in the development of heterogeneous alterations associated with depression and a full gene expression map for PD[14]. Glial pathology is thought to drive neuroinflammatory processes throughout the brain. These processes are believed to contribute to the degeneration of neurons in PD[13]. A study performed a transcriptomic analysis at the single-nucleus level on post-mortem substantia nigra pars compacta (SNpc) tissue derived from 15 randomly acquired PD cases and 14 controls, resulting in a dataset of around 84,000 nuclei derived from all primary brain cell types[15]. In PD samples, researchers found a marked decrease in the neuronal cell population. In contrast, they noted an increase in glial and T cell populations, with specific subtypes such as astrocytes, microglia, and oligodendrocytes showing clear reductions. They also identified 28 co-expressed genes related to dopamine metabolism, providing insights into the molecular basis of dopaminergic neuron (DA) degeneration and glial cell responses in PD[15]. Recent studies have employed integrative approaches combining snRNA-seq and bulk mRNA transcriptomics to compare molecular signatures between PD and related neuropsychiatric disorders, such as major depressive disorder (MDD), revealing distinct neuroinflammatory and synaptic transmission profiles in microglia and oligodendrocytes[16]. Besides, glial dysfunction contributes to this process by triggering glutamate-induced excitotoxicity, which accelerates neurodegeneration[17]. A recent molecular profiling study on both cultured organoids from human SN and midbrain uncovered a unique sub-cluster of neurons expressing the *RIT2* gene, known to confer risks for developing PD[18]. While neuronal populations are well-established contributors to PD pathogenesis, the potential involvement of non-neuronal cell populations remains poorly defined. This underscores the need for comprehensive molecular investigations across neuronal and non-neuronal cell populations. We have hypothesized that critical genes driving PD pathology might be present across both neuronal and non-neuronal cell types and that these key genes can be identified by analyzing publicly available single-nucleus RNA sequencing (snRNA-seq) data.

Therefore, we conducted a comprehensive investigation across multiple cell types by reanalyzing the publicly available data (accession number GSE184950) in this study. Our goal was to identify PD-relevant cell types by using common differentially expressed genes (cDEGs) that capture shared molecular mechanisms. We pinpointed cell-specific key genes (KGs) across glial and neuronal cell types, uncovering their distinct glial and neuronal responses in PD. Also, downstream analyses, including functional enrichment, gene regulatory networks, protein-protein interactions, and transcription factors enriched with the identified cDEGs, have been conducted to uncover valuable insights into the molecular mechanisms and biological significance of PD. Through this approach, we aimed to uncover therapeutic targets associated with PD that were used to repurpose drug molecules as potential treatments for PD in this study. The schematic view of the present study is given in Fig. 1.

Methods and materials

Data collection, processing, and integration

To investigate the cell-type-specific impacts of PD, we have conducted a comprehensive profiling of the SN samples from the Gene Expression Omnibus (GEO) under the accession number GSE184950, including 9 healthy and 23 idiopathic PD samples. The healthy or normal sample was defined as the term "control", which means no clinical symptoms related to PD (Table S1). The treatment samples were defined as those obtained from individuals diagnosed with idiopathic PD exhibiting clinical symptoms characteristic of the disease without a known genetic cause. These samples represent the disease-affected state, enabling us to explore the molecular and cellular alterations associated with PD compared to the control group. Across all samples, the average age was ~80 years. Of the 32 individuals, 21 were male and 11 were female, distributed across both control and PD groups.

First of all, we gathered all the data in the form of a count matrix containing cell barcodes and gene expression information. This matrix was then imported into the *Scanpy* tool in Python[19], a versatile tool for single-cell gene expression analysis, offering pre-processing, visualization, clustering, differential expression, and various other advantages. To identify and remove doublets we applied SOLO[20]. After excluding all doublets, we retained only singlets for further analysis in this study. Subsequently, we tailored Scanpy's quality control settings by excluding cells expressing < 200 genes, genes detected in < 3 cells, and cells with over 5% mitochondrial gene content to eliminate low-quality data[21]. Also, MALAT1 was excluded from the transcript list due to its high enrichment, which could introduce bias into the clustering analysis[15]. A gene was then eliminated if its expression levels in every cell exceeded 99%. After that 32 individual samples were merged to generate a unified snRNA-seq dataset for analysis.

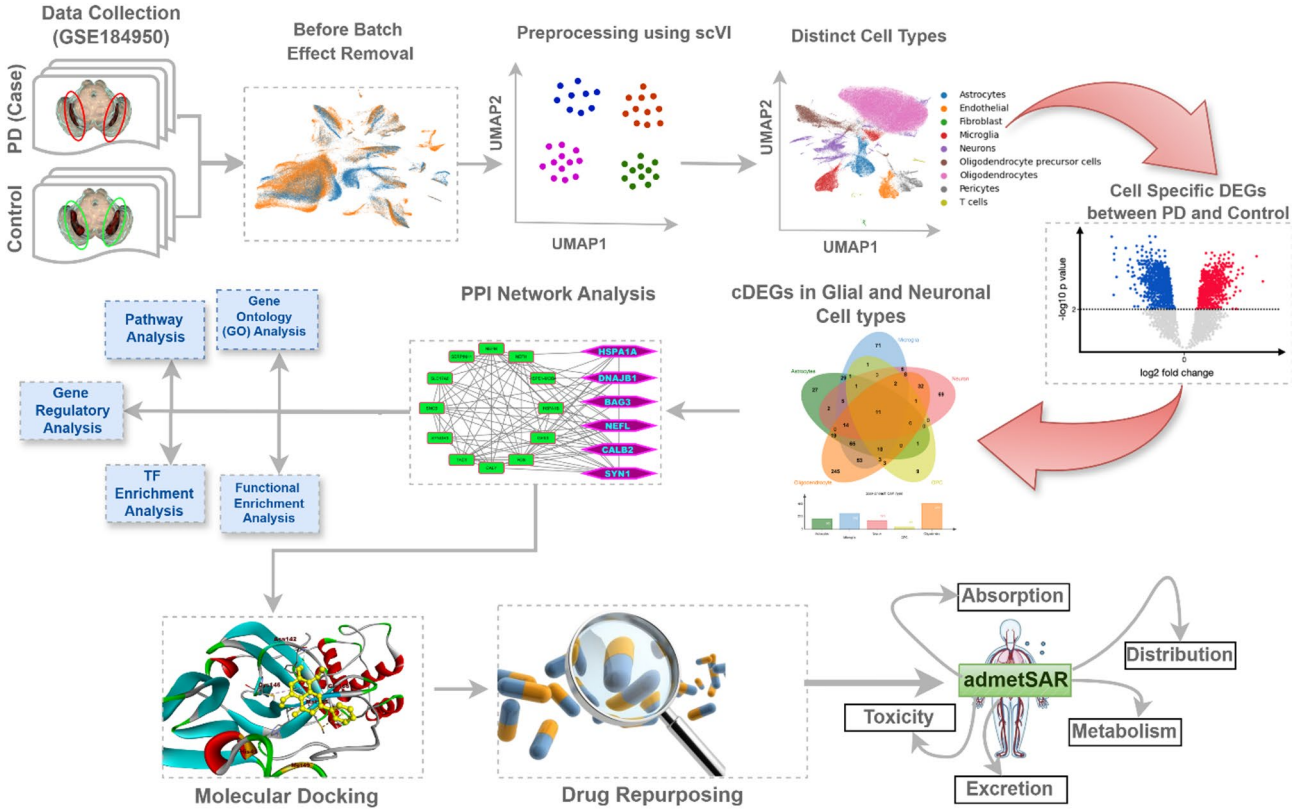


Fig. 1. A schematic outline and comprehensive computational analysis pipeline for identifying potential drug targets in PD. PD (case) and control datasets were downloaded from the public database Gene Expression Omnibus (accession number: GSE184950). This pipeline involves several key steps, including data preprocessing, single-nuclei analysis to identify distinct cell types, DEGs between PD and control groups, functional enrichment analysis, protein–protein interaction (PPI) network analysis, gene regulatory network analysis, and drug repurposing.

Following quality control, the remaining cells were normalized to a target total count of 1×10^4 using the '*scanpy.pp.normalize_total*' function, effectively standardizing sequencing depth across samples to ensure data consistency. To stabilize the variance across expression values, log normalization using the "*scanpy.pp.log1p*" transformation was used. The *scVI* package was used for data integration, which processed unnormalized data along with categorical and continuous covariates using a Variational Autoencoder approach in deep learning[22]. During integration, it removed unwanted variation, corrected for batch effects, and performed dimensionality reduction, resulting in a 10-dimensional representation for further downstream analysis[22,23]. A total of 6000 highly variable genes were identified using the '*scanpy.pp.highly_variable_genes*' function[21,24] and put them in a new layer called "counts". Finally, the transcriptomes of about ~ 130 K high-quality single nuclei were found in this study.

Cluster stability analysis

The *scanpy.pp.neighbors* function was utilized with default parameters to compute the nearest-neighbors distance matrix and construct a neighborhood graph for the cells. Community detection was conducted using the Leiden algorithm ('*scanpy.tl.leiden*'), which separates the given network into smaller groups at a resolution of 0.5[25], enabling the assignment of cells into distinct communities based on their expression profiles. Consequently, a total of 20 communities have been identified. Using the Scanpy function "*scanpy.get.rank_genes_groups_df*", the top 15 markers for each community have been identified. We then used the Human Protein Atlas[26], CellMarker 2.0[27], Panglaodb[28], and Enrichr annotation to annotate the major clusters. Clusters were visualized by applying non-linear dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP)[29] in the context of graph-based clustering using the 'scanpy.pl.umap' function.

Identification of cell-type markers

Marker genes were identified through differential expression analysis across clusters using the *scanpy.get.rank_genes_groups_df* function. This analysis focused on genes meeting stringent criteria, including a p-value < 0.01, |logFC| > 2, and t-score from the method "*t-test_overstim_var*". The cell-type specificity score quantitatively evaluates the exclusivity of a marker gene in identifying a particular cell type within the tissue, where higher scores denote highly specific markers, and lower scores indicate broader, less specific expression across multiple cell types. The top marker genes used for annotating each cell cluster included: Microglia (ARHGAP24),

Oligodendrocytes (ST18), Oligodendrocyte Precursor Cells (OPCs) (TNR), Astrocytes (TRPM3), Neurons (SYT1, SNAP25), Fibroblasts (COL1A2), Endothelial cells (FLT1), Pericytes (EBF1), and T-cells (PTPRC). These marker genes were used for initial cell-type labeling based on their relative expression patterns across clusters.

Cell-type specific DEGs and common DEGs (cDEGs)

Differential expression analysis for snRNA-seq data was conducted using the *t*-test, utilizing the '*de.test.t_test*' function from the *diffxpy* package in Python. We have considered those DEGs that were common among the major clusters, and we finally used them in our study for functional study. We have identified DEGs for each of the cell types independently in order to detect cDEGs. We looked for DEGs with |logFC| > 1 and adjusted *p*-value < 0.05[30]. The cDEGs were identified using the *JVenn* web-based tool[31].

PPI network construction

Using the STRING protein interactome database, the protein-protein interaction (PPI) network of cDEGs was built[32]. Cytoscape tool was used to enhance the PPI network's quality[33] and the key genes (KGs) from the PPI network were chosen using the Cytoscape plugin cytoHubba[34]. The highest-ranked KGs are nodes that have the most meaningful connections, interactions, and edges with other nodes. Six topological studies of the PPI network (Degree, Betweenness, Closeness, Edge Percolation Component (EPC), maximal clique centrality (MCC), and Stress) were used to choose the KGs. We have utilized the Enrichr web tool[35], incorporating the DisGeNET database[36], to conduct disease-key genes enrichment analysis. This approach allowed us to investigate the association of KGs with various neurological disorders, including PD.

Analysis of the regulatory factors of KGs

To identify the regulatory factors of the key KGs, we have followed the processes described by the previous study[37]. Using transcription factors (TFs) and microRNAs (miRNAs), we have investigated the transcriptional and post-transcriptional regulators of KGs. In order to determine the primary TFs connected to KGs, we analyzed the TFs-KGs links network using the JASPAR database[38]. By examining miRNA-KG associations with the TarBase[39] databases, the important miRNAs that affect KGs at the post-transcriptional stage were identified, and the relationships were replicated using NetworkAnalyst[40]. The post-transcriptional regulators of KGs were selected from among the top-ranked miRNAs. We have utilized Cytoscape[33] to visualize their interaction networks.

GO and KEGG pathway enrichment analysis

Gene ontology (GO) functional and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis is a widely used approach to identify significantly annotated terms and pathways by the KGs. It is crucial to understand the mechanisms of how proteins function at the molecular level The Gene Ontology (GO) concepts, which are broken down into three categories: Molecular Function (MF), Cellular Component (CC), and Biological Process (BP), are utilized to annotate the cDEGs[41]. Understanding the molecular activity, cellular function, and the location inside the cell where genes carry out their roles is the main driving force behind the identification of GO terms. Gene annotation is widely used by the "Kyoto Encyclopedia of Genes and Genomes" (KEGG) pathway, which is frequently utilized to comprehend metabolic pathways[42]. We took into consideration the Enrichr database web server for analyzing GO-terms and KEGG pathways in order to examine the biological importance of cDEGs highlighting KGs[35].

Molecular docking for exploring KGs-guided drugs

Molecular docking analysis was used to investigate potential drugs/ligands for treating PD. We considered the identified 6 kg and their regulatory 3 TF proteins as the receptors to investigate prospective ligands (drug molecules) for treating PD. The 3D structures of the receptors were retrieved from the Protein Data Bank (PDB), AlphaFold Protein Structure Database[43], and SWISS-MODEL[44]. The 3D structures of the target proteins were visualized with Discovery Studio Visualizer 2019[40], and non-essential protein chains were eliminated. Receptor proteins were pre-processed using AutoDock tools[45] by removing water molecules, assigning charges, and adding polar hydrogens. Drug agents or ligands underwent energy minimization via Avogadro[46] and were further pre-processed with AutoDock tools[45]. Next, molecular docking was conducted between receptors and ligands to determine their binding affinity scores (kcal/mol) using AutoDock Vina[47].

Let A_{ij} represents the binding affinity between the *i*th target protein $i = 1, 2, \ldots, x$ and the *j*th drug agent $j = 1, 2, \ldots, y$. The target proteins were ranked in descending order based on their row sums,

$$\frac{1}{x} \sum {}_{j=1}^{y} A_{ij};\ i = 1, 2, \ldots, x$$

while drug agents were ranked by column

$$\frac{1}{y} \sum {}_{i=1}^{x} A_{ij};\ j = 1, 2, \ldots, y$$

sum to identify the highest-ranked candidate drugs. We subsequently validated the proposed repurposed drugs through molecular docking studies with the top-ranked receptor proteins linked to PD, as determined through a comprehensive literature review.

In-silico validation of candidate drug molecules by ADME/T analysis

The ADMET (absorption, distribution, metabolism, excretion, and toxicity) characteristics and drug-like features of the top three compounds were evaluated to understand their structural and chemical properties better. Using the SCFBio tool (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp), we assessed their compliance with Lipinski's rule of five, examining essential factors like molecular weight, hydrogen bond donors and acceptors, rotatable bonds, and the octanol/water partition coefficient (LogP), among other key descriptors for drug-likeness[48]. ADMET properties were subsequently predicted using the online platforms SwissADME[49] and pkCSM[50]. The ADME/T evaluations of the medicinal compounds were conducted based on their optimized molecular structures in SMILES format[51].

Results

Sample clustering and batch effect correction

To investigate the cell-type-specific impacts of PD, we have conducted a comprehensive profiling of the SN samples collected from different ages and ethnic groups in this study. We have identified 426,886 droplet-based snRNA-seq profiles from the raw data of the individual brain samples (Fig. 2A). A total of 129,473 high-quality nuclei (79,514 of PD and 49,959 of the control) were obtained (Fig. 2B). Raw data was confounded with technical batch-effect, with cells from different samples forming distinct clusters (Fig. 2C). In contrast, the batch correction applied demonstrates a more homogenous distribution of cells, suggesting that the batch-related technical variability has been effectively mitigated (Fig. 2D). Without batch correction, there appears to be some separation between control and PD cells, but the clustering is not as distinct (Fig. 2E). However, with batch correction applied, the separation between control and PD cells becomes more noticeable (Fig. 2F). This suggests that batch correction has enhanced the ability to identify disease-specific differences in gene expression profiles.

Clustering and marker gene analysis reveal nine cell types

The initial grouping of data from control and PD samples revealed 20 distinct cell groups (Fig. 3A). To determine the cell types within each group, the top 15 marker genes per group were selected based on the highest |logFC| values (Fig. S1 and Table S2), and evaluated for cell specificity using comprehensive databases, including CellMarker 2.0[27], Human Protein Atlas[26], and PanglaoDB[28]. According to Khan et al., if 50% or more of these top marker genes were associated with a particular cell group, the group was assigned to that cell type[52]. This approach streamlined the groups into nine distinct cell types with the following distributions, including oligodendrocytes (56.6%), neurons (10.3%), microglia (8.8%), astrocytes (7.1%), endothelial cells (4.8%), OPCs (9.0%), pericytes (3.7%), fibroblasts (0.2%), and T cells (0.4%) (Fig. 3B,C). Oligodendrocytes emerged as the most prevalent cell type in the human SN, followed by neurons, OPCs, microglia, astrocytes, and the remaining cell types (Fig. 3C,D). The most significant marker genes were selected and plotted using the dot plot displaying mean gene expression and a stacked violin plot illustrating median gene expression (Fig. S2, and Table S3). The leading marker genes for each cluster are given within the first bracket after the cluster name, such as Microglia (ARHGAP24), Oligodendrocyte (ST18), OPC (TNR), Astrocyte (TRPM3), Neurons (SYT1, SNAP25), Fibroblast (COL1A2), Endothelial (FLT1), Pericytes (EBF1), T-cells (PTPRC).

Identification of common DEGs

We analyzed DEGs between PD and control samples for each major cell type individually to preserve the resolution of snRNA-seq analysis. DEGs were identified across all cell type clusters, using *t*-tests, applying thresholds of $|\log FC| > 1$ and adjusted p-value < 0.05. Specifically, Neurons exhibited 320 DEGs, with 169 up-regulated and 151 down-regulated. In astrocytes, 238 DEGs were identified, with 53 up-regulated and 185 down-regulated. In microglia, 319 DEGs were found, including 39 up-regulated and 280 down-regulated. OPCs showed 126 DEGs, comprising 83 up-regulated and 43 down-regulated. Oligodendrocytes had the highest number, with 488 DEGs (22 up-regulated and 466 down-regulated). Endothelial cells presented 121 DEGs (94 up-regulated and 27 down-regulated), and pericytes had 183 DEGs, where 77 were up-regulated and 106 were down-regulated (Table S4). However, no DEGs were detected in Fibroblasts and T cells under the applied statistical thresholds.

While snNA-seq allows cell type-specific insights, but we aimed to identify shared dysregulated genes across key PD-related cell types to highlight potential common molecular pathways involved in disease progression. Therefore, we focused on identifying cDEGs across all major brain cell types; however, only five cell types (astrocytes, microglia, neurons, OPCs, and oligodendrocytes) exhibited overlapping DEGs and are well-documented in the literature for their involvement in PD pathology. Other cell types, including T cells, fibroblasts, endothelial cells, and pericytes, were excluded from this specific analysis as they did not share any cDEGs with these key brain cell types.

Using the *JVenn* web tool[31], a total of 18 cDEGs were identified (7 up-regulated and 11 down-regulated) that were consistently present across the major glial (non-neuronal) and neuronal cell types (Fig. 5A,B and Table S5). These cDEGs were subsequently used for functional enrichment analysis to uncover shared biological processes potentially driving PD pathogenesis.

Identification of KGs from cDEGs using PPI network analysis with their regulatory factors and disease-KGs interaction analysis

Proteins interact with other proteins within the cell to perform their functions, and analyzing the Protein–Protein Interaction (PPI) network provides deeper insights into their coordinated functions[53]. The STRING[32] database was utilized to construct a PPI network of the cDEGs, comprising 18 nodes and 90 edges (Fig. 4A). Through the application of six topological measures (Degree, Betweenness, Closeness, EPC, MCC, and Stress) within the PPI network. We identified the top six KGs: *HSPA1A*, *DNAJB1*, *BAG3*, *SYN1*, *CALB2*, and *NEFL* (Table S7).

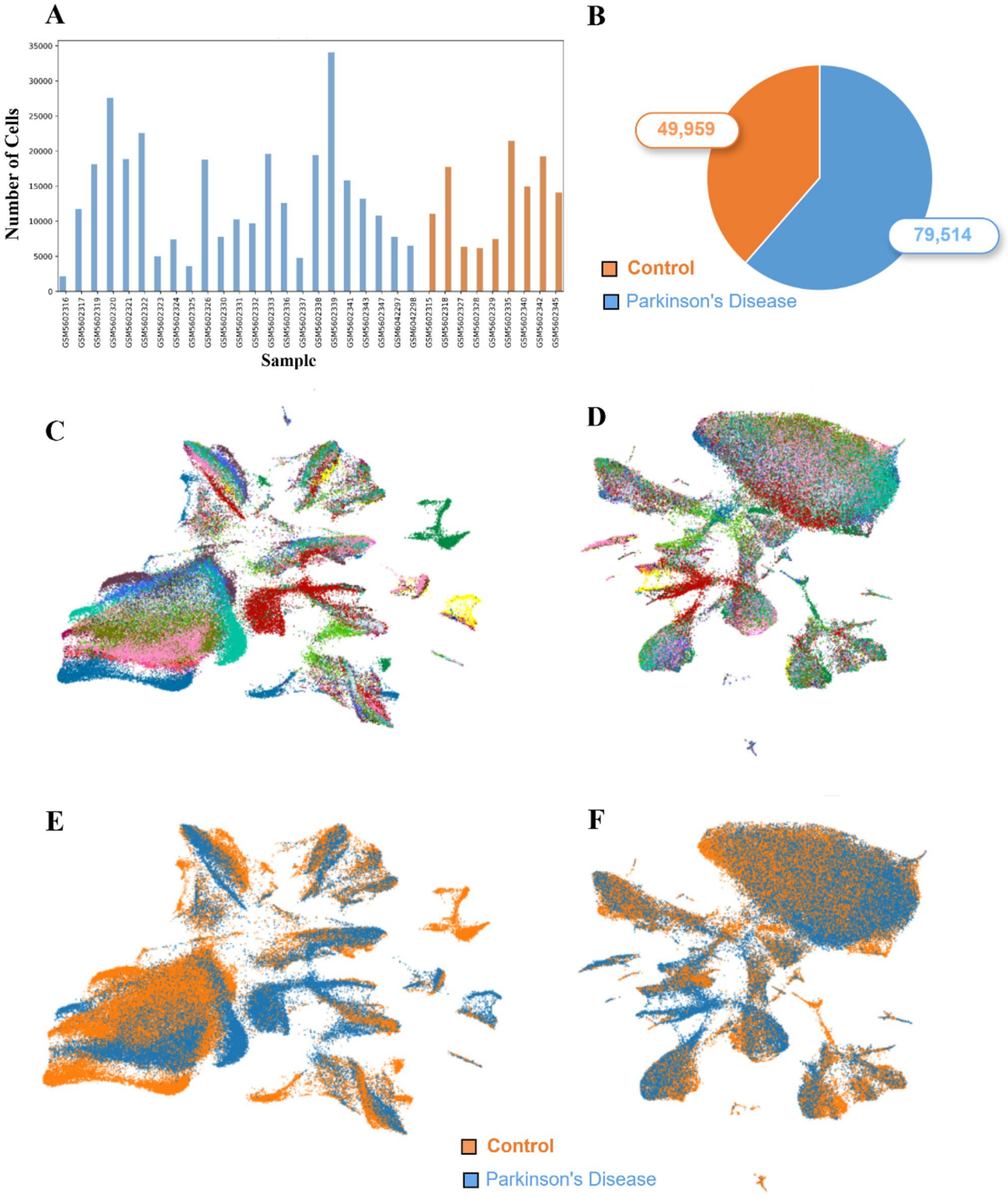


Fig. 2. UMAP visualization of snRNA-seq data. (**A**) Bar plot showing the number of nuclei per sample. (**B**) Pie chart representing the number of high-quality nuclei in PD and control samples. (**C**,**D**) UMAP plots by samples without and with batch correction, respectively. This UMAP plot shows the snRNA-seq data colored by individual sample IDs. Each color represents a different sample. This visualization highlights the distribution and clustering of cells from various samples. (**E**,**F**) UMAP plot without and with batch effect for control and PD samples, respectively. This UMAP plot shows the snRNA-seq data colored by case-control status. The yellow color represents a control sample, and the contrasting blue color represents a PD sample. This visualization highlights the distribution and clustering of cells between case-control status, providing insights into the heterogeneity and composition of the dataset across different conditions.

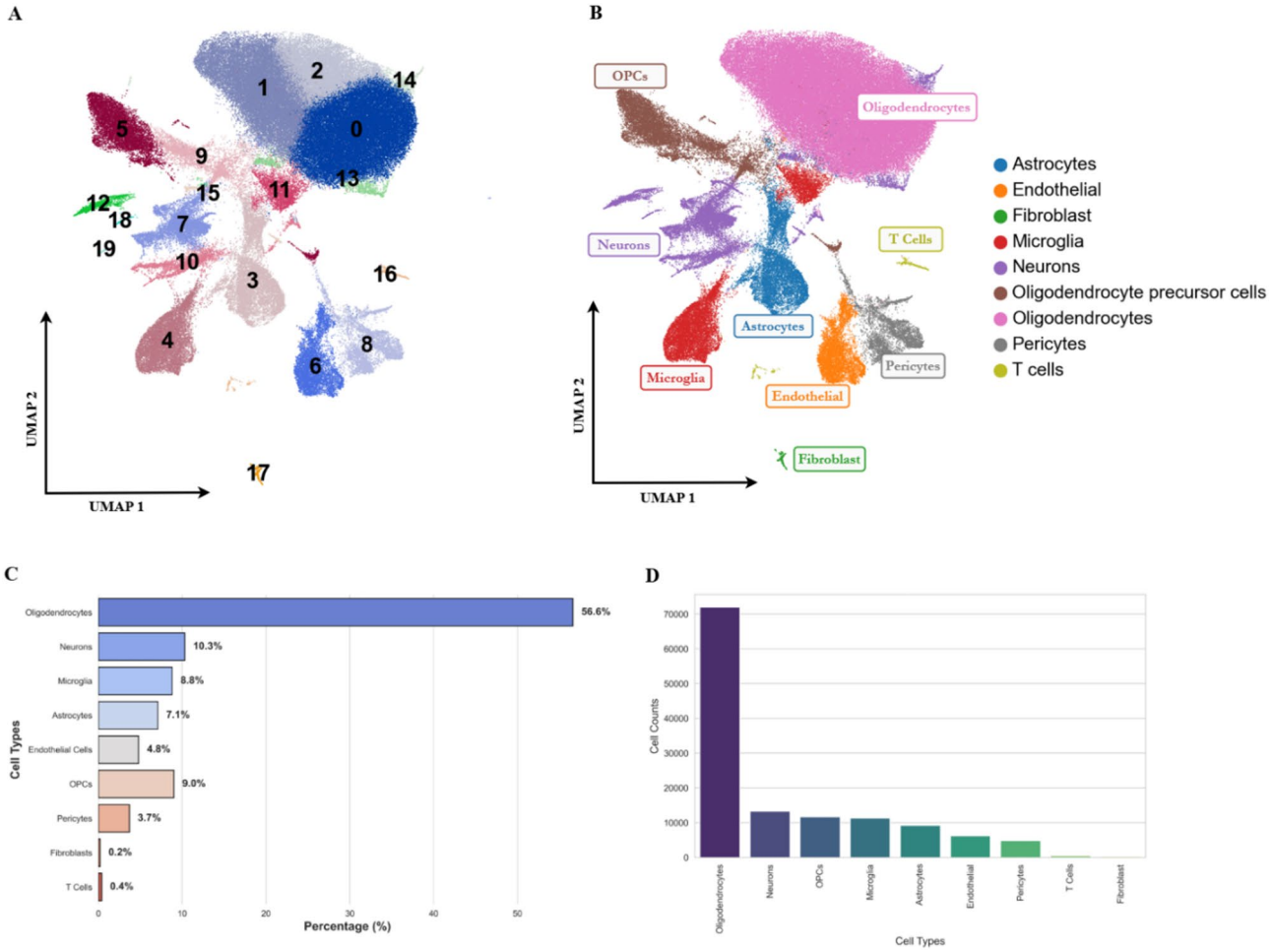


Fig. 3. (**A**,**B**) This UMAP plot shows the clustering of different cell types from a snRNA-seq dataset. Each dot represents an individual cell, and each color represents a particular cell type, color-coded by its cell type: Astrocytes (blue), Endothelial cells (orange), Fibroblasts (green), Microglia (red), Neurons (purple), OPC (brown), Oligodendrocytes (pink), Pericytes (gray), and T cells (yellow). (**C**) Percentage of each cell type. (**D**) Number of high-quality nuclei in each cell type.

To find the transcriptional and post-transcriptional regulators of KGs, the networks of miRNAs and TFs with KGs were verified in this study. Initially, we determined the transcriptional regulators of KGs by ranking the top three TFs (*HINFP*, *YY1*, *and SREBF1*) using two topological measures: degree and betweenness, with cutoffs of degree ≥ 2 and betweenness ≥ 29, respectively (Fig. 4B). Then, using the same topological measures with cutoff degree ≥ 6 and betweenness ≥ 660, respectively, the top two miRNAs (hsa-mir-34a-5p, hsa-mir-103a-3p) as the post-transcriptional regulators of KGs were chosen (Fig. 4C). Moreover, the analysis of disease-KG interactions using the 'DisGeNET-database' indicated that many significant diseases, such as PD and other neuronal disorders, are (Adj. p-value < 0.05) notably associated with KGs (Table S8).

GO and KEGG pathway enrichment analysis from cDEGs, highlighting KGs

To investigate the pathogenetic mechanisms of PD by including KGs, GO and KEGG pathway enrichment analysis for 18 cDEGs have been performed. To emphasize the KGs, the significant five GO terms were considered from each of the following: molecular function (MF), cellular component (CC), biological process (BP), and five KEGG pathways. The most 5 significantly enriched BPs were neurotransmitter transport, cellular response to heat, response to unfolded protein, cellular response to oxidative stress, and regulation of cell death (Fig. 5C and Table S6). The enriched MFs comprised ATP binding, cadherin binding, Hsp70 protein binding, calcium ion binding, and ubiquitin protein ligase binding (Fig. 5D and Table S6). The key CCs identified were the aggresome, synaptic vesicle membrane, neuron projection, axon, and neuromuscular junction (Fig. 5E and Table S6). Furthermore, KEGG pathway analysis revealed significant enrichment in the estrogen signaling pathway, protein processing in the endoplasmic reticulum, MAPK signaling pathway, pathways of neurodegeneration, and amyotrophic lateral sclerosis, highlighting the potential involvement of KGs in PD pathogenesis (Fig. 5F and Table S6).

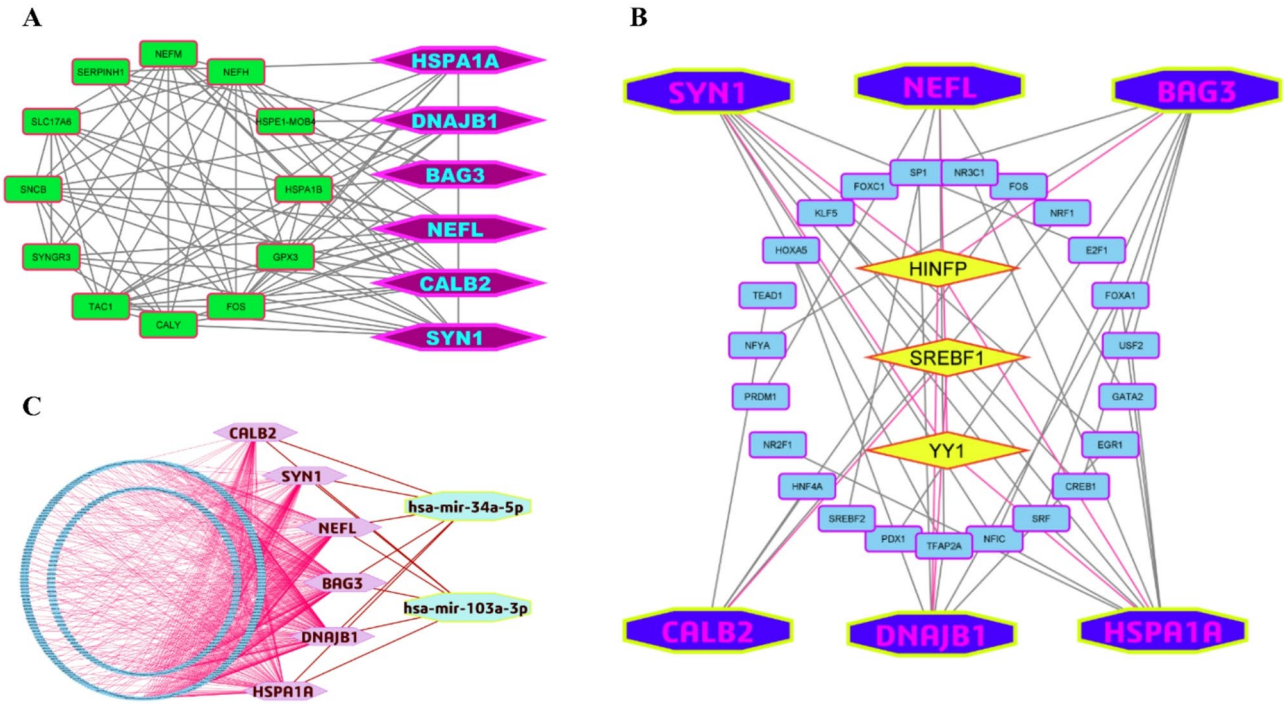


Fig. 4. PPI Network of cDEGs highlighting key genes, transcriptional regulators, and miRNA interactions. (**A**) Protein–protein interaction (PPI) network of cDEGs to identify key genes (KGs), where the purple color nodes indicate the KGs. KGs were marked as purple with hexagon shapes. (**B**) Transcriptional regulators of KGs were marked as yellow. (**C**) miRNAs-KGs interaction network. Here, KGs were marked as green with a rectangle shape and miRNAs were marked as red color with an octagon shape.

Drug repurposing by molecular docking study

To look for potential ligands (drug molecules) for PD treatment, we focused on our selected 6 kg and their 3 TFs proteins as target receptors. Data from two different sources was gathered to acquire 3D structures of these receptors. The 3D structures of the six target proteins (*BAG3*, *SYN1*, *CALB2*, *NEFL*, *HINFP*, and *SREBF1*) were retrieved from the "AlphaFold Protein Structure Database". The structural data of the remaining three receptors (*HSPA1A*, *DNAJB*, and *YY1*) were obtained from the PDB using the specific PDB codes 5GJJ, 6Z5N, and 1ZNM, respectively. We assessed the binding affinity scores (BAS) between the proposed receptors and the candidate drug molecules through molecular docking analysis, quantifying the potential interactions between them (Table S9). To prioritize the most promising therapeutic candidates, the drug molecules were ranked based on their average BAS across all receptors (Table S10). Additionally, receptors were ranked based on their mean BAS across all drug molecules, providing a comprehensive evaluation of receptor-drug interaction dynamics, which offers valuable insight into receptor-specific drug efficacy. The top 30 ranked drug molecules are mapped to their corresponding receptors (Fig. 6). Among these, celastrol, withaferin-A, and apomorphine emerged as leading candidates, displaying BAS < -7.0 kcal/mol, indicating strong receptor interaction. In contrast, the remaining molecules exhibited weaker binding, with BAS > -7.0 kcal/mol. Thus, the three proposed drug molecules could serve as potential inhibitors against targeting genes responsible for PD.

In-silico drug validation

The ADME characteristics of a drug molecule were assessed to determine its absorption, distribution, metabolism, excretion, and toxicity profiles. ADME/T properties were predicted using two well-established online tools: SwissADME[49] and pkCSM[50]. Specifically, SwissADME was used to compute physicochemical descriptors, lipophilicity (LogP), gastrointestinal absorption, and drug-likeness based on Lipinski's Rule of Five (ROF) and other criteria, while pkCSM was employed to predict toxicity-related endpoints and additional pharmacokinetic parameters. The drug-likeness properties were examined by evaluating physicochemical descriptors, providing a comprehensive understanding of each compound's chemical behavior (Table 1). According to ROF criteria, two drugs (Withaferin-A and Apomorphine) fully complied with the rule, while Celastrol violated one criterion (MLOGP)[54]. The lipophilicity (LogP value) of the three drugs falls within the acceptable range (1 to ≤ 5)[48] as outlined by Lipinski's rule. Based on their LogP values, these compounds are classified as lipophilic when compared to standard benchmarks. Consequently, our top three ranked drug candidates meet nearly all of the drug-likeness criteria (Table 1A). The ADME and toxicity profiles of the proposed compounds were assessed through several key parameters to gauge both their efficacy and safety. Predictions indicate that these compounds possess favorable gastrointestinal absorption, positioning them as strong candidates for oral drug administration[55]. A compound is regarded as highly absorbed in the human intestine if it achieves a Human Intestinal Absorption

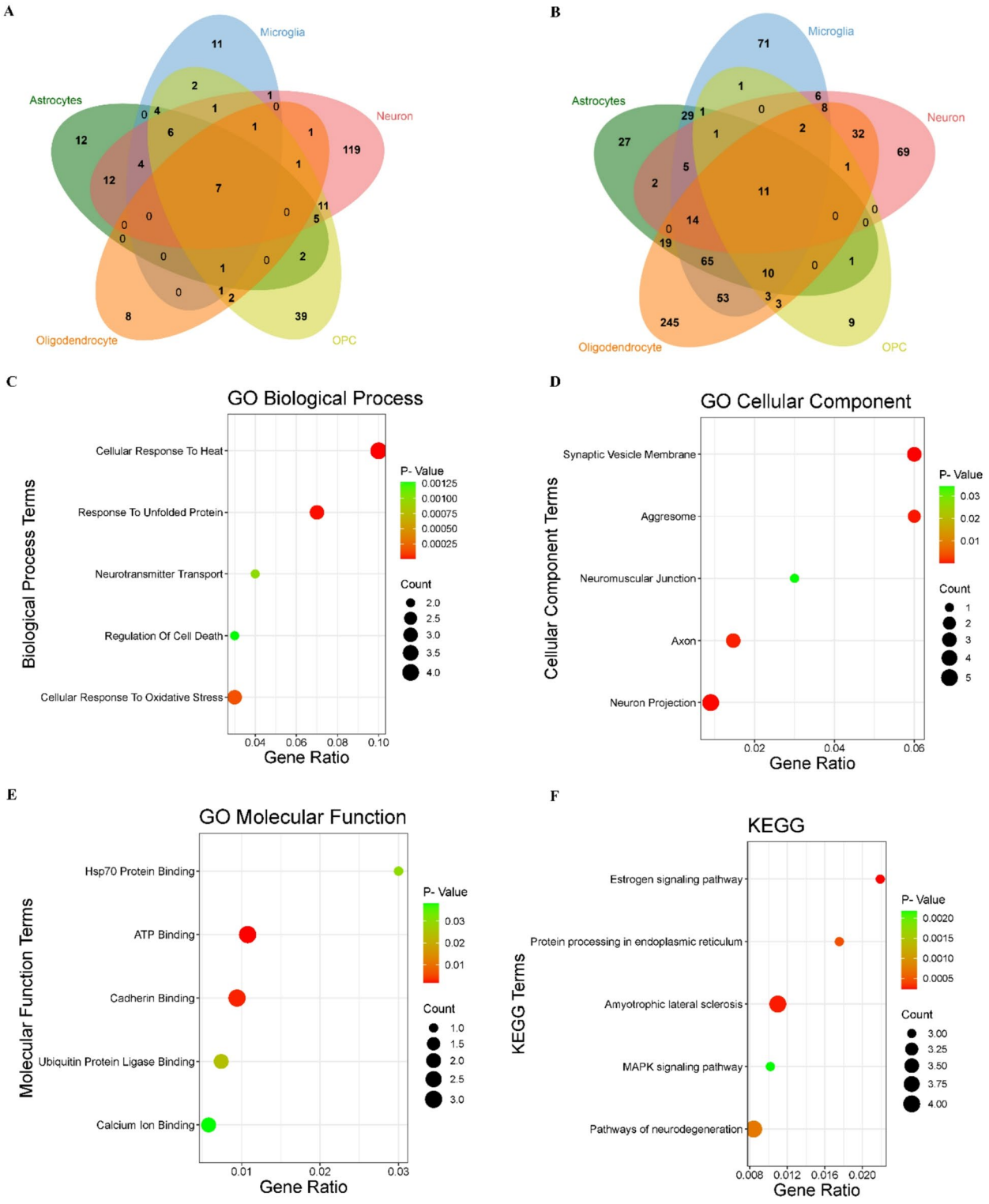


Fig. 5. Venn diagram, GO and KEGG analysis results of cDEGs. (**A**) Up-regulated cDEGs and (**B**) Down-regulated cDEGs. There were 7 up-regulated and 11 down-regulated cDEGs in Glial and Neuronal cell types of PD patients in this study. (**C**–**E**) Biological Process (BP), Cellular Components (CC), and Molecular Function (MF) function, respectively. (**F**) KEGG pathway.

(HIA) score of ≥ 30%, demonstrating favorable absorption properties. Among these three proposed drugs, all of them exhibit their HIA scores of ≥ 80%. High absorption reduces the variability in drug levels between patients, leading to more consistent therapeutic outcomes, indicating excellent absorption potential in the human body. Furthermore, Withaferin-A drug also possesses the capability to inhibit P-glycoprotein (P-gpl), an important

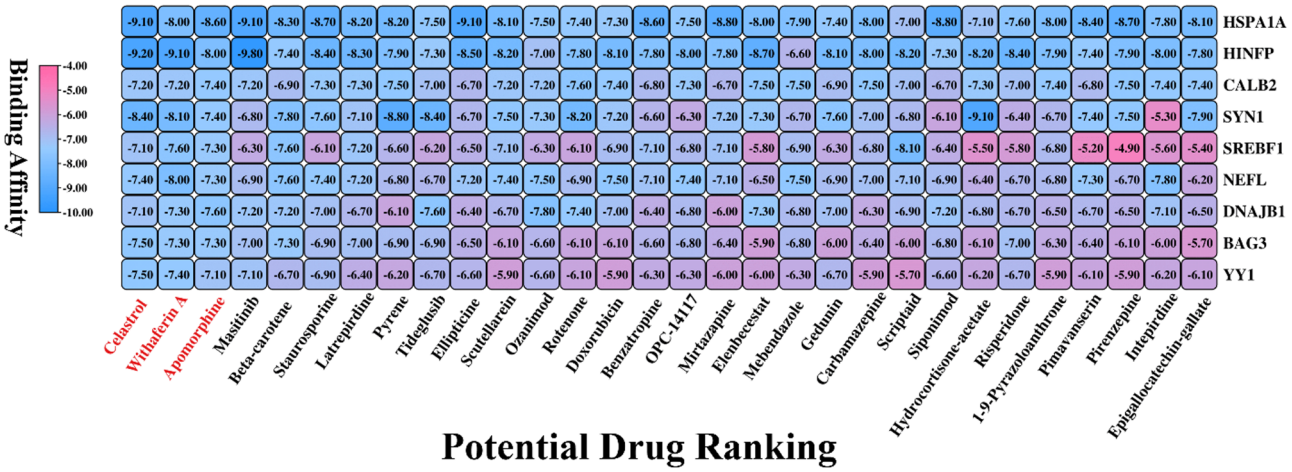


Fig. 6. Results from the molecular docking analysis, presenting the top 30 ranked drug molecules mapped to their corresponding receptors. Molecular docking scores are shown where green colors indicate the strong binding affinities between target proteins and drug agents, and pink colors indicate weak bindings. Also, a higher negative value represents stronger binding. The image depicts these affinities, with the X-axis showing the top 30 ranked drug agents (out of 170) and 9 target proteins (proposed) in the Y-axis.

Compounds	Molecular weight	Log P	H-bond acceptor (HBA)	H-bond donor (HBD)	Polar surface area (Å2)	No. of rotatable bond
(A) Drug likeness profile of candidate drug molecules						
Celastrol	450.61	6.6977	3	2	197.082	1
Withaferin A	470.6	3.3529	6	2	201.317	3
Apomorphine	267.32	2.8499	3	2	117.65	0

Compounds	Absorption	Distribution	Metabolism	Excretion	Toxicity	
	$Caco_2$ permeability	HIA (%)	P-gpI	BBB (LogBB)	CNS LogPS	CYP3A4 inhibitor
(B) ADME and toxicity (ADME/T) profile of top-ranked 3 drug molecules						
Celastrol	0.464	100	No	0.078	− 1.278	Yes
Withaferin A	0.829	85.345	Yes	− 0.03	− 2.72	No
Apomorphine	1.182	94.222	No	0.452	− 1.72	No

Table 1. Drug likeness and ADME/T analysis results. (A) drug likeness profile of candidate drug molecules. (B) ADME and toxicity (ADME/T) profile of top-ranked 3 drug molecules.

factor in drug efflux and bioavailability. In contrast, Apomorphine and Celastrol lack this inhibitory activity, highlighting a distinction in their pharmacokinetic profiles. The capacity of a compound to penetrate the blood-brain barrier (BBB) is quantified by its BBB-permeability index. Compounds with a LogBB ≥ 0.3 are generally considered to have good potential for efficient BBB traversal, whereas those with a LogBB < − 1 are likely to exhibit limited distribution through the BBB. Apomorphine is evaluated and found capable of crossing the BBB, a key factor for the advancement of successful neurotherapeutics. In contrast, the remaining compounds exhibit poor ability to traverse the BBB.

Moreover, the LogPS (CNS) values for these compounds suggest that they may fully penetrate the central nervous system (Table 1B). Human cytochrome P450 (CYP) enzymes, which are membrane-bound hemoproteins, are crucial for maintaining homeostasis, detoxifying drugs, and managing cellular metabolism. These enzymes are responsible for approximately 50% of the elimination of commonly prescribed medications and account for nearly 80% of oxidative metabolic processes. This extensive role involves multiple CYP enzymes from classes 1 to 3, highlighting their central importance in pharmacokinetics and metabolic regulation[56]. Among the three drugs, only Celastrol exhibits the property of inhibiting the CYP3A4 enzyme in the human body. The value of oral toxicity or lethal dose (LD50) was found to be 2.36, 2.78, and 2.57 mol/kg for Celastrol, Withaferin-A, and Apomorphine, respectively. A higher LC50 value corresponds to lower toxicity in a drug molecule, while a lower value indicates higher toxicity. Toxicity assessments, including AMES, Minnow toxicity (LC50), and lethal dose (LD50) tests, showed that the proposed compounds were inactive across all toxicity prediction parameters used, leading to the conclusion that they are likely non-toxic.

Literature-driven validation of KGs and proposed drugs

To substantiate our findings, we performed an extensive literature review to validate the identified KGs and proposed drug candidates for PD, as illustrated in Fig. 7A,B. This cross-verification not only reinforces the

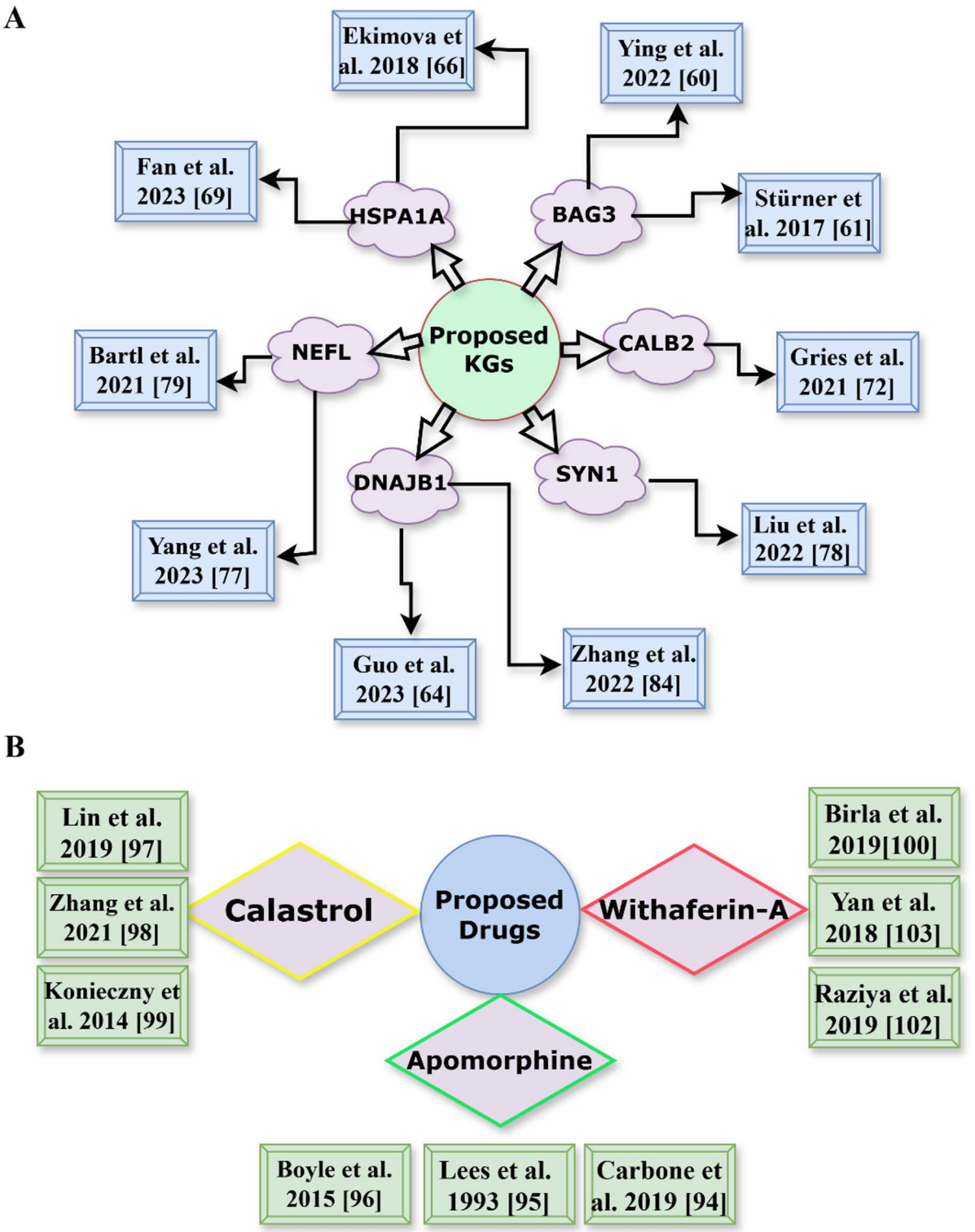


Fig. 7. Validation of the proposed KGs and drug molecules supporting PD was conducted through a literature review. The arrows highlight the alignment of our findings with those of previous studies. (**A**) Validation of the predicted KGs: Clouds in light purple represent KGs. (**B**) Validation of the suggested drugs: Diamond shapes in light purple represent FDA-approved drug molecules, the blue border denotes approved drugs, the yellow border introduces investigational drugs, and the red border indicates unapproved drugs.

biological relevance of the selected KGs but also highlights the therapeutic promise of the identified compounds. Our results are consistent with previous studies, further supporting the role of these genes in PD pathogenesis and endorsing the potential of these drug candidates as viable treatment options.

Discussion

Parkinson's disease (PD) is a chronic and progressive neurodegenerative condition with no definitive cure, characterized by gradual deterioration over time[57,58]. Although it has been widely studied, the exact molecular mechanisms driving PD are still unclear, highlighting the need for advanced single-nucleus analyses to uncover its underlying pathology, potential therapeutic targets, and treatment options. The snRNA-seq technology offers tremendous opportunities in the investigation of complicated diseases, such as PD, by providing unique insights into cellular heterogeneity and disease-specific molecular mechanisms[59]. In this study, we reanalyzed snRNA-seq data (GSE184950) to identify PD-relevant cell types by leveraging cDEGs that capture shared molecular mechanisms involved in PD pathogenesis. We introduced a novel approach and identified PD-relevant cell types across all major cell populations in the human SN affected by PD.

Analysis of snRNA-seq data revealed several cell populations, such as neurons, oligodendrocytes, OPC, microglia, astrocytes, fibroblasts, endothelial, pericytes, and T cells. However, analysis revealed 18 overlapping DEGs exclusively among neurons, astrocytes, microglia, OPCs, and oligodendrocytes. PPI network analysis revealed six key DEGs, some of which, include *BAG3* (BCL2-associated athanogene 3), a chaperone protein involved in protein quality control, responses to cellular stress, and regulation of apoptosis. This protein facilitates the stabilization and degradation of misfolded proteins. Mutations or dysregulation of *BAG3* may potentially result in the accumulation of toxic protein aggregates like alpha-synuclein, which is a signature of PD[60–64]. A recent study suggested that *HSPA1A*, like other heat-shock proteins, helps refold misfolded proteins or directs them for degradation[65,66]. Failure in this process due to unmodulated or dysfunction of *HSPA1A* could lead to increased alpha-synuclein aggregation and raised vulnerability of neurons to apoptosis, contributing to neurodegeneration[2,67,68]. Research suggests that *CALB2* (Calbindin 2), also known as Calretinin, plays a key role in calcium signaling regulation in both neurons and certain types of glial cells[69]. *CALB2* regulates intracellular calcium, which is crucial for cell survival, synaptic function, and neurotransmitter release[70–72]. Another study suggests that impairment of *CALB2* may disrupt calcium balance, affecting glial cells' capacity to regulate neuronal function, which could play a role in developing neurodegenerative diseases[73]. A study demonstrated that *NEFL* genes encode the light subunit of neurofilament, which is essential for maintaining neuronal structure and axonal transport[74]. Elevated levels of *NEFL* in cerebrospinal fluid and plasma are recognized as biomarkers for neuronal damage, which is linked to motor dysfunction in PD patients[75–77]. It can also be associated with glial responses, especially astrocytes, and microglia, which play a key role in responding to neuronal injury and degeneration[78]. A study reported that *SYN1* belongs to the synapsin gene family, which encodes neuronal phosphoproteins that associate with the cytoplasmic surface of synaptic vesicles, playing a critical role in synaptic function and neurotransmitter release[79–81]. Glial activation can also lead to neuroinflammation, which may alter neuronal pathways, potentially impacting proteins like *SYN1* involved in synaptic plasticity[82]. Research has shown that *DNAJB1*, a member of the DNAJ/Hsp40 family of co-chaperones, dysfunction or altered expression can impair the cell's ability to manage misfolded proteins, contributing to the accumulation of toxic protein aggregates like alpha-synuclein, a hallmark of the disease's pathology[6,83–86]. However, numerous studies have demonstrated that PD is strongly influenced by autosomal-dominant genes (*SNCA*, *LRRK2*, and *VPS35*) and autosomal-recessive genes (*PINK1*, *DJ-1*, and *Parkin*)[87–89], but several studies have also suggested these KGs as genes linked to the development of PD (Fig. 7A).

Among the identified GO enrichment terms, 'ATP binding', 'axon', and 'protein binding' are highly enriched in PD[90]. The Estrogen signaling pathway has been identified as the mechanism through which tocotrienol exerts its protective effects in a cellular model of PD[91]. Additionally, the amyotrophic lateral sclerosis (ALS) pathways are directly associated with PD, and one-third of ALS patients are directly affected by PD[92]. The identified GO and KEGG pathway enrichment terms could lead to the emergence of novel therapeutic approaches and vaccines to combat PD effectively.

To identify KG-guided drug candidates for PD, we have performed molecular docking analysis and discovered three top-ranking molecules (Celastrol, Withaferin A, and Apomorphine) demonstrating high binding affinities with KG-mediated target proteins. These molecules were subsequently validated as potential PD treatments through a comprehensive literature review (Fig. 7B).

We have performed an ADME/T analysis to assess the drug-likeness of the molecules computationally. Among the three evaluated drugs, Withaferin A and Apomorphine met al.l the criteria of Lipinski's rule of five, indicating drug-like properties, while Celastrol breached one of the criteria. The selected compounds exhibited desirable ADME/T properties, including adequate water solubility and high levels of Human Intestinal Absorption (HIA), ranging from 85 to 100%. These properties suggest strong bioavailability and potential effectiveness in oral administration, enhancing their viability as therapeutic candidates. Apomorphine, one of the identified candidate drugs, is an FDA-approved medication for PD, listed in the Drug Bank database with the accession number DB00714. Research has reported that Apomorphineis is a potent D1 and D2 dopamine receptor agonist[93–95], meaning it mimics the action of dopamine by stimulating dopamine receptors in the brain, which are often deficient in PD patients due to the degeneration of DA. In the PD mouse model, Celastrol alleviates motor impairments and prevents neurodegeneration in the SN[96]. An experimental study found that celastrol provides protection against the loss of dopamine-producing neurons, reduces neuroinflammation, and alleviates motor impairments[97]. However, Celastrol's protective effects are inconsistent under certain conditions, such as ubiquitin-proteasome system inhibition, which may accelerate disease progression[98]. Several studies have suggested that Withaferin A, a major bioactive compound for PD, has shown effectiveness in combating

brain aging, supporting mental well-being, and preventing motor function decline[99,100]. It has been noted for its effectiveness in addressing multiple neurological disorders[101–104].

While this study provides novel insights into PD mechanisms through snRNA-seq profiling, several limitations must be acknowledged. First, the analysis was based on a specific dataset with a relatively small number of individual samples, which may affect the generalizability of the findings. Additionally, the use of publicly available data introduces potential biases related to sample handling, processing, and the completeness of metadata. Our methodology also relied heavily on computational tools for drug repurposing analyses, which, although powerful, are inherently limited in capturing the full clinical support and may introduce algorithm-specific biases. Furthermore, no experimental validation was performed. Nevertheless, the findings will serve as a useful resource and hypothesis-generating framework for future wet lab investigations.

Conclusion

This study provides an in-depth exploration of cell-type-specific alterations in PD through snRNA-seq of SN samples. The reanalysis revealed 129,473 high-quality nuclei, leading to the identification of nine distinct cell types and highlighting significant gene expression dysregulation across these clusters. Eighteen cDEGs were pinpointed and enriched in pathways related to neurodegeneration and protein processing, shedding light on the mechanisms underlying PD. Furthermore, a PPI network analysis identified six critical KGs, which were assessed for their regulatory interactions. Disease association analysis confirmed these genes' significant links to PD and other neurological disorders. Moreover, drug repurposing efforts identified Celastrol, Withaferin-A, and Apomorphine as promising therapeutic candidates, demonstrating strong binding affinities and favorable ADME-Tox profiles. These findings underscore the potential of these compounds for PD treatment and highlight important targets for future research. We hope our findings may enhance our understanding of PD pathology while providing a foundation for the development of novel therapeutic strategies.

Data availability

The data used in the current study were obtained from the published articles as referenced in the Methods section and listed in Supplementary Table S1. The analyzed datasets generated during the study are listed in Supplementary Figs. S1–S2.

Received: 7 April 2025; Accepted: 27 June 2025
Published online: 25 July 2025

References

1. Reeve, A., Simcox, E. & Turnbull, D. Ageing and parkinson's disease: why is advancing age the biggest risk factor? *Ageing Res. Rev.* **14**, 19–30 (2014).
2. Huang, J. et al. Single-cell transcriptomics uncovers cellular heterogeneity, mechanisms, and therapeutic targets for parkinson's disease. *Front. Genet.* **13**, 686739 (2022).
3. Olanow, C. W., Stern, M. B. & Sethi, K. The scientific and clinical basis for the treatment of Parkinson disease (2009). *Neurology*. **72** (21 Suppl 4), S1–136 (2009).
4. Selvaraj, S. & Piramanayagam, S. Impact of gene mutation in the development of Parkinson's disease. *Genes Dis.* **6** (2), 120–128 (2019).
5. Kalia, L. V. & Lang, A. E. Parkinson's disease lancet. **386**(9996), 896–912 (2015).
6. Gao, X. et al. Human Hsp70 disaggregase reverses Parkinson's-linked α-synuclein amyloid fibrils. *Mol. Cell.* **59** (5), 781–793 (2015).
7. Onisiforou, A. & Spyrou, G. M. Immunomodulatory effects of microbiota-derived metabolites at the crossroad of neurodegenerative diseases and viral infection: network-based bioinformatics insights. *Front. Immunol.* **13-2022**. (2022).
8. Onisiforou, A. & Spyrou, G. M. Systems bioinformatics reveals possible relationship between COVID-19 and the development of neurological diseases and neuropsychiatric disorders. *Viruses*. **14** https://doi.org/10.3390/v14102270 (2022).
9. Tolosa, E. et al. LRRK2 in Parkinson disease: challenges of clinical trials. *Nat. Rev. Neurol.* **16** (2), 97–107 (2020).
10. Zimprich, A. et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am. J. Hum. Genet.* **89** (1), 168–175 (2011).
11. Chambers, D. C. et al. Transcriptomics and single-cell RNA-sequencing. *Respirology*. **24** (1), 29–36 (2019).
12. Volpato, V. Integration of functional genomics data to uncover cell type-specific pathways affected in Parkinson's disease. *Biochem. Soc. Trans.* **49** (5), 2091–2100 (2021).
13. Smajić, S. et al. Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. *Brain*. **145** (3), 964–978 (2021).
14. Weng, Y. et al. Single-cell RNA-sequencing of cellular heterogeneity and pathogenic mechanisms in paraquat-induced Parkinson's disease with depression. *Ecotoxicol. Environ. Saf.* **273**, 116169 (2024).
15. Martirosyan, A. et al. Unravelling cell type-specific responses to parkinson's disease at single cell resolution. *Mol. Neurodegener.* **19** (1), 7 (2024).
16. Christodoulou, C. C. et al. Unraveling the transcriptomic signatures of Oarkinson's disease and major depression using single-cell and bulk data. *Front. Aging Neurosci.* **15–2023** (2023).
17. Iovino, L., Tremblay, M. E. & Civiero, L. Glutamate-induced excitotoxicity in Parkinson's disease: the role of glial cells. *J. Pharmacol. Sci.* **144** (3), 151–164 (2020).
18. Wang, Q. et al. Molecular profiling of human substantia Nigra identifies diverse neuron types associated with vulnerability in Parkinson's disease. *Sci. Adv.* **10** (2), eadi8287 (2024).
19. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19** (1), 15 (2018).
20. Bernstein, N. J. et al. Solo: doublet identification in single-cell RNA-Seq via semi-supervised deep learning. *Cell. Syst.* **11** (1), 95–101e5 (2020).
21. Wu, B. et al. Single-cell analysis of the amphioxus hepatic caecum and vertebrate liver reveals genetic mechanisms of vertebrate liver evolution. *Nat. Ecol. Evol.* **8**(10), 1972–1990 (2024).
22. Lopez, R. et al. Deep generative modeling for single-cell transcriptomics. *Nat. Methods*. **15** (12), 1053–1058 (2018).
23. Lakkis, J. et al. A joint deep learning model enables simultaneous batch effect correction, denoising, and clustering in single-cell transcriptomics. *Genome Res.* **31** (10), 1753–1766 (2021).
24. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods*. **19** (1), 41–50 (2022).

25. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to leiden: guaranteeing well-connected communities. *Sci. Rep.* **9** (1), 5233 (2019).
26. Thul, P. J. & Lindskog, C. The human protein atlas: A Spatial map of the human proteome. *Protein Sci.* **27** (1), 233–244 (2018).
27. Hu, C. et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.* **51** (D1), D870–d876 (2023).
28. Franzén, O., Gan, L. M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)*. **2019**. (2019).
29. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37** (1), 38–44 (2019).
30. Zeng, X. et al. Single-cell transcriptomics dissects the transcriptome alterations of hematopoietic stem cells in myelodysplastic neoplasms. *J. Transl. Med.* **22** (1), 359 (2024).
31. Bardou, P. et al. Jvenn: an interactive Venn diagram viewer. *BMC Bioinform.* **15** (1), 293 (2014).
32. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47** (D1), D607–D613 (2018).
33. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13** (11), 2498–2504 (2003).
34. Chin, C. H. et al. *cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome.*
35. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44** (W1), W90–W97 (2016).
36. Piñero, J. et al. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48** (D1), D845–d855 (2020).
37. Alamin, M. et al. Single-cell RNA-seq data analysis reveals functionally relevant biomarkers of early brain development and their regulatory footprints in human embryonic stem cells (hESCs). *Brief. Bioinform.* **25**(3). (2024).
38. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48** (D1), D87–d92 (2020).
39. Sethupathy, P., Corda, B. & Hatzigeorgiou, A. G. TarBase: A comprehensive database of experimentally supported animal MicroRNA targets. *RNA*. **12** (2), 192–197 (2006).
40. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* **41** (Database issue), D991–D995 (2013).
41. Doms, A. & Schroeder, M. GoPubMed: exploring pubmed with the gene ontology. *Nucleic Acids Res.* **33**(Web Server issue), W783–W786 (2005).
42. Kanehisa, M. et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36** (suppl_1), D480–D484 (2007).
43. Varadi, M. et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50** (D1), D439–D444 (2021).
44. Schwede, T. et al. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31** (13), 3381–3385 (2003).
45. Morris, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30** (16), 2785–2791 (2009).
46. Hanwell, M. D. et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4** (1), 17 (2012).
47. Trott, O. & Olson, A. J. AutoDock vina: improving the speed and accuracy of Docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31** (2), 455–461 (2010).
48. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today Technol.* **1** (4), 337–341 (2004).
49. Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.
50. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. PkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **58** (9), 4066–4072 (2015).
51. Quirós, M. et al. Using SMILES strings for the description of chemical connectivity in the crystallography open database. *J. Cheminform.* **10** (1), 23 (2018).
52. Khan, A. H., Lee, L. K. & Smith, D. J. Single-cell analysis of gene expression in the substantia Nigra Pars compacta of a pesticide-induced mouse model of parkinson's disease. *Transl. Neurosci.* **13** (1), 255–269 (2022).
53. Ben-Hur, A. & Noble, W. S. Kernel methods for predicting protein-protein interactions. *Bioinformatics*. **21** (Suppl 1), i38–46 (2005).
54. Chen, X. et al. Analysis of the physicochemical properties of acaricides based on lipinski's rule of five. *J. Comput. Biol.* **27** (9), 1397–1406 (2020).
55. Durán-Iturbide, N. A., Díaz-Eufracio, B. I. & Medina-Franco, J. L. Silico adme/tox profiling of natural products: A focus on BIOFACQUIM. *ACS Omega*. **5** (26), 16076–16084 (2020).
56. Zanger, U. M. & Schwab, M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* **138** (1), 103–141 (2013).
57. Tori, K. L. & Eva, L. Y. A review on parkinson's disease treatment. *Neuroimmunol. Neuroinflamm.*. **8**, 222 (2021).
58. Wood-Kaczmar, A., Gandhi, S. & Wood, N. Understanding the molecular causes of Parkinson's disease. *Trends Mol. Med.* **12** (11), 521–528 (2006).
59. Ma, S. X. & Lim, S. B. Single-Cell RNA sequencing in Parkinson's disease. *Biomedicines*. **9** (4), 368 (2021).
60. Ying, Z. M. et al. BAG3 promotes autophagy and suppresses NLRP3 inflammasome activation in parkinson's disease. *Ann. Transl. Med.* **10** (22), 1218 (2022).
61. Stürner, E. & Behl, C. The role of the multifunctional BAG3 protein in cellular protein quality control and in disease. *Front. Mol. Neurosci.* **10**, 177 (2017).
62. Brenner, C. M. et al. BAG3: Nature's quintessential multi-functional protein functions as a ubiquitous intra-cellular glue. *Cells*. **12**(6). (2023).
63. Lin, H. et al. The role of BAG3 in health and disease: A magic BAG of tricks. *J. Cell. Biochem.* **123** (1), 4–21 (2022).
64. Guo, J. F. et al. Involvement of Bcl-2-associated athanogene (BAG)-family proteins in the neuroprotection by Rasagiline. *Int. J. Clin. Exp. Med.* **8** (10), 18158–18164 (2015).
65. Joshi, N., Raveendran, A. & Nagotu, S. Chaperones and proteostasis: Role in Parkinson's disease. *Diseases*. **8**(2). (2020).
66. Ekimova, I. V. et al. New HSF1 inducer as a therapeutic agent in a rodent model of parkinson's disease. *Exp. Neurol.* **306**, 199–208 (2018).
67. Silvestro, S., Raffaele, I. & Mazzon, E. Modulating stress proteins in response to therapeutic interventions for Parkinson's disease. *Int. J. Mol. Sci.* **24**(22). (2023).
68. Fan, L. Y. et al. Integrating single-nucleus sequence profiling to reveal the transcriptional dynamics of alzheimer's disease, parkinson's disease, and multiple sclerosis. *J. Transl. Med.* **21** (1), 649 (2023).
69. Perea, G. & Araque, A. Glial calcium signaling and neuron-glia communication. *Cell. Calcium*. **38** (3–4), 375–382 (2005).
70. König, N. et al. Selective knockout of the vesicular monoamine transporter 2 (Vmat2) gene in Calbindin2/Calretinin-Positive neurons results in profound changes in behavior and response to drugs of abuse. *Front. Behav. Neurosci.* **14**, 578443 (2020).
71. Gries, M. et al. Parkinson mice show functional and molecular changes in the gut long before motoric disease onset.

72. Pchitskaya, E., Popugaeva, E. & Bezprozvanny, I. Calcium signaling and molecular mechanisms underlying neurodegenerative diseases. *Cell. Calcium*. **70**, 87–94 (2018).
73. Ureshino, R. P. et al. The interplay between Ca(2+) signaling pathways and neurodegeneration. *Int. J. Mol. Sci.* **20**(23). (2019).
74. Lin, F. et al. Sequencing of neurofilament genes identified NEFH Ser787Arg as a novel risk variant of sporadic amyotrophic lateral sclerosis in Chinese subjects. *BMC Med. Genom.* **14** (1), 222 (2021).
75. Nakamura, T. et al. High levels of plasma neurofilament light chain correlated with brainstem and peripheral nerve damage. *J. Neurol. Sci.* **463**, 123137 (2024).
76. Yang, D. et al. Neurofilament light chain as a mediator between LRRK2 mutation and dementia in parkinson's disease. *NPJ Parkinsons Dis.* **9** (1), 132 (2023).
77. Liu, Y. et al. Neurofilament light as a biomarker for motor decline in Parkinson's disease. *Front. NeuroSci.* **16**. (2022).
78. Bartl, M. et al. Biomarkers of neurodegeneration and glial activation validated in alzheimer's disease assessed in longitudinal cerebrospinal fluid samples of parkinson's disease. *PLoS One*. **16** (10), e0257372 (2021).
79. Huttner, W. B. et al. Synapsin I (protein I), a nerve terminal-specific phosphoprotein. III. Its association with synaptic vesicles studied in a highly purified synaptic vesicle Preparation. *J. Cell. Biol.* **96** (5), 1374–1388 (1983).
80. Liu, S. et al. Gene expression profiling predicts pathways and genes associated with Parkinson's disease. *Neurol. Sci.* **37** (1), 73–79 (2016).
81. Chandrasekaran, S. & Bonchev, D. A network view on Parkinson's disease. *Comput. Struct. Biotechnol. J.* **7**, e201304004 (2013).
82. Nguyen, P. T. et al. Microglial remodeling of the extracellular matrix promotes synapse plasticity. *Cell.* **182** (2), 388–403e15 (2020).
83. Zhang, K. et al. Genetic analysis of HSP40/DNAJ family genes in Parkinson's disease: a large case-control study. *Mol. Neurobiol.* **59** (9), 5443–5451 (2022).
84. Hu, S. et al. Molecular chaperones and parkinson's disease. *Neurobiol. Dis.* **160**, 105527 (2021).
85. Wentink, A. S. et al. Molecular dissection of amyloid disaggregation by human HSP70. *Nature*. **587** (7834), 483–488 (2020).
86. Guo, H. et al. Potential application of heat shock proteins as therapeutic targets in Parkinson's disease. *Neurochem. Int.* **162**, 105453 (2023).
87. Li, W. et al. PARK genes link mitochondrial dysfunction and Alpha-Synuclein pathology in sporadic Parkinson's disease. *Front. Cell. Dev. Biol.* **9**. (2021).
88. Cherian, A. & Divya, K. P. Genetics of Parkinson's disease. *Acta Neurol. Belg.* **120** (6), 1297–1305 (2020).
89. Salles, P., Tirapegui, J. M. & Chaná-Cuevas, P. Genetics of Parkinson's disease: dominant forms and GBA. *Neurol. Perspect.* **4** (3), 100153 (2024).
90. Quan, W. et al. Identification of potential core genes in Parkinson's disease using bioinformatics analysis. *Parkinsons Dis.* **2021**, 1690341 (2021).
91. Nakaso, K. et al. The Estrogen receptor β-PI3K/Akt pathway mediates the cytoprotective effects of tocotrienol in a cellular parkinson's disease model. *Biochim. Et Biophys. Acta (BBA) - Mol. Basis Dis.* **1842** (9), 1303–1312 (2014).
92. Lualdi, M. et al. Shared and unique disease pathways in amyotrophic lateral sclerosis and Parkinson's disease unveiled in peripheral blood mononuclear cells. *ACS Chem. Neurosci.* **14** (23), 4240–4251 (2023).
93. Carbone, F. et al. Apomorphine for Parkinson's disease: efficacy and safety of current and new formulations. *CNS Drugs*. **33** (9), 905–918 (2019).
94. Lees, A. J. Dopamine agonists in Parkinson's disease: a look at apomorphine. *Fundam. Clin. Pharmacol.* **7** (3–4), 121–128 (1993).
95. Boyle, A. & Ondo, W. Role of apomorphine in the treatment of Parkinson's disease. *CNS Drugs*. **29** (2), 83–89 (2015).
96. Lin, M. W. et al. Celastrol inhibits dopaminergic neuronal death of Parkinson's disease through activating mitophagy. *Antioxid. (Basel)*. **9**(1). (2019).
97. Zhang, C. et al. The Nrf2-NLRP3-caspase-1 axis mediates the neuroprotective effects of Celastrol in Parkinson's disease. *Redox Biol.* **47**, 102134 (2021).
98. Konieczny, J. et al. Lack of neuroprotective effect of Celastrol under conditions of proteasome Inhibition by lactacystin in in vitro and in vivo studies: implications for Parkinson's disease. *Neurotox. Res.* **26** (3), 255–273 (2014).
99. Birla, H. et al. Neuroprotective effects of Withania somnifera in BPA induced-cognitive dysfunction and oxidative stress in mice. *Behav. Brain Funct.* **15** (1), 9 (2019).
100. Prakash, J. et al. Withania somnifera alleviates parkinsonian phenotypes by inhibiting apoptotic pathways in dopaminergic neurons. *Neurochem. Res.* **39** (12), 2527–2536 (2014).
101. Raziya Banu, M. et al. Ameliorative effect of Withaferin A on Ageing-Mediated impairment in the dopamine system and its associated behavior of Wistar albino rat. *Pharmacology*. **103** (3–4), 114–119 (2019).
102. Yan, Z. et al. Withaferin A inhibits apoptosis via activated Akt-mediated Inhibition of oxidative stress. *Life Sci.* **211**, 91–101 (2018).
103. Narayan, M. et al. Withaferin A regulates LRRK2 levels by interfering with the Hsp90- Cdc37 chaperone complex. *Curr. Aging Sci.* **8** (3), 259–265 (2015).
104. Maitra, R. et al. Inhibition of NFκB by the natural product Withaferin A in cellular models of cystic fibrosis inflammation. *J. Inflamm.* **6** (1), 15 (2009).

Author contributions

M.A. and M.N.H.M. conceived and designed the study. M.A.A.P., S.M., A.A., A.S., extracted and analyzed the data. M.A.A.P. and M.A. generated data visualization and data interpretation. M.A., M.H.S., M.R.A., and M.A.A.P. wrote the first draft of the manuscript. A.K.M. A., S.A.A., and M.N.H.M. revised and discussed the final edition. All authors declare that there is no interest and no relevant financial disclosure.

Funding

The authors extend their appreciation to the King Salman Center for Disability Research for funding this work through Research Group no KSRG-2024-089.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-09417-w.

Correspondence and requests for materials should be addressed to M.A. or M.N.H.M.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025