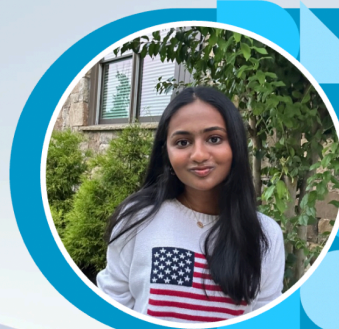# Siri Chivukula

Alliance Academy for Innovation
Cumming, Georgia 30040

Investigating Potential Biomarkers for Early Intervention in Parkinson's Disease using Bioinformatics

ABSTRACT

Background

Parkinson's disease is a progressive neurodegenerative disorder caused by the loss of dopamine-producing neurons, leading to motor symptoms such as tremors and rigidity. A significant challenge is the lack of understanding of the disease's early molecular changes, which hinders early detection and therapeutic development. This study aims to identify key genes and biological pathways associated with Parkinson's disease using a bioinformatics approach.

Methods

We utilized the GSE156926 dataset from the Gene Expression Omnibus and the GEO2R tool to identify differentially expressed genes. Functional analysis was conducted using gene ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses to predict gene functions and pathways.

Results

Our analysis identified 30 differentially expressed genes, suggesting significant molecular changes. Gene ontology analysis revealed enrichment in terms related to G-protein complexes, GTPase activity, and hormone-related processes. Kyoto Encyclopedia of Genes and Genomes

pathway analysis highlighted the P53 signaling pathway and the dopaminergic synapse as being significantly altered. These findings suggest that cellular stress, programmed cell death, and defects in cellular transport are central to the disease.

Conclusion

The genes and pathways identified in this study provide a foundation for further investigation and may serve as novel targets for the development of therapeutic drugs to slow the progression of Parkinson's disease.

INTRODUCTION

Parkinson's Disease (PD) is a complex and progressive neurological disorder that impacts millions of individuals globally. It involves the degeneration of specific nerve cells in the brain, primarily those that produce a crucial chemical messenger called dopamine [1]. Dopamine is essential for smooth and coordinated movements, so its depletion leads to the characteristic motor symptoms of PD, such as tremors, rigidity, bradykinesia (slowness of movement), and balance problems [2]. Beyond these visible motor symptoms, PD also presents with a wide array of non-motor symptoms like sleep disturbances, depression, anxiety, and cognitive changes, which can manifest many years before the motor signs and significantly diminish a patient's quality of life [3]. The global prevalence of PD is steadily increasing due to the aging global population, highlighting its growing significance as a major public health challenge [4].

A primary challenge in managing Parkinson's Disease is the lack of research on early diagnosis. By the time visible symptoms become apparent, a substantial loss of dopamine-producing neurons has already occurred in the brain, often an irreversible process [1]. This late diagnosis limits the effectiveness of current treatments, which are largely symptomatic and do not halt or reverse the underlying neurodegeneration. Therefore, a critical need exists for reliable biomarkers that can detect PD in its pre-symptomatic stages. Addressing this problem is crucial because early intervention can potentially slow disease progression and significantly improve long-term outcomes for patients.

This research is driven by the question: "What biomarkers are most commonly noticed among infected gene samples?" As Parkinson's is a neurodegenerative disorder, "infected" here is used to represent gene samples that are affected by the processes of Parkinson's Disease. More specifically, this research seeks to identify what particular biomarkers (primarily focusing on gene expression changes) are consistently and significantly observed within gene expression datasets derived from individuals with Parkinson's Disease, distinguishing them from healthy controls [5]. This detailed investigation involves the analysis of public genetic data using

bioinformatics tools to pinpoint specific genes whose altered activity could serve as reliable indicators of the disease, particularly in its early stages.

Parkinson's Disease is characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta region of the brain. While the exact cause is unknown, it generally manifests from a complex blend of genetic and environmental factors. Its impact is profound, including a spectrum of even non-motor symptoms such as cognitive impairment, depression, sleep disorders, and autonomic dysfunction, which often precede motor symptoms by years and significantly contribute to patient disability [3]. Globally, the number of people living with PD has more than doubled in the last 25 years, with over 8.5 million individuals affected in 2019, highlighting its increasing worldwide burden [4].

Bioinformatics has become an important tool in Parkinson's Disease research, enabling the analysis of large datasets to research disease mechanisms and identify potential biomarkers. Researchers have used bioinformatics to pinpoint differentially expressed genes, identify perturbed biological pathways, and explore genetic risk factors associated with PD. Public databases like the Gene Expression Omnibus (GEO) house numerous datasets, which have been extensively studied to compare gene expression profiles between PD patients and healthy controls, as well as across different stages of the disease [6]. Tools like NCBI GEO2R facilitate these comparisons, allowing for the rapid identification of genes and pathways. These studies have contributed to a deeper understanding of the mechanisms of PD and have highlighted various genes and pathways involved.

Despite significant research, several challenges persist. The mechanisms that initiate neurodegeneration in PD are still not fully understood, and the variability in how the disease progresses across individuals makes both clinical diagnosis and biomarker discovery difficult [8]. From a bioinformatics point of view, the sheer volume and diversity of biological data are challenges. Furthermore, the absence of easily accessible early diagnostic biomarkers continues to be a major obstacle to developing and implementing effective therapies for PD.

The goal of this research is to utilize advanced bioinformatics tools to analyze publicly available gene expression datasets associated with Parkinson's Disease, with the specific aim of identifying novel biomarkers that are indicative of early-stage PD.

Our hypothesis is that if the development of early-onset Parkinson's Disease involves alterations in key biological pathways, then specific genes related to these pathways will show statistically significant upregulation or downregulation in gene samples from individuals with early-onset Parkinson's Disease compared to gene samples from healthy control individuals.

This research is crucial because it addresses the need for early Parkinson's Disease diagnosis, which currently occurs too late to fully halt the progress of neurodegeneration. By identifying key biomarkers from "infected" (affected) gene samples, we aim to enable predictive diagnosis, allowing earlier intervention with therapy. This will significantly improve patient outcomes, provide insights into the disease's origins, and encourage the development of new treatments, ultimately reducing the immense global impact of Parkinson's Disease.

METHODS

Data Collection for Gene Expression Data

In this study, the dataset for Parkinson's disease was collected from NCBI GEO2R [7] by using the keywords "Parkinson's disease" and "gene expression". NCBI GEO is defined as a public repository that archives and freely distributes high-throughput functional genomics data, including gene expression data from microarray and sequencing experiments. It serves as a comprehensive resource for researchers to access, explore, and analyze data to support their studies. [7] The specific dataset utilized was GSE156926. Then the dataset was categorized into two groups: "Parkinson's Disease patients" and "healthy controls." The data was analyzed using the no-code GEO2R bioinformatics tool, which utilizes the R programming language to identify differentially expressed genes. **Figure 1** shows an overview of the methodology used in this study and includes the bioinformatics tools and databases used.
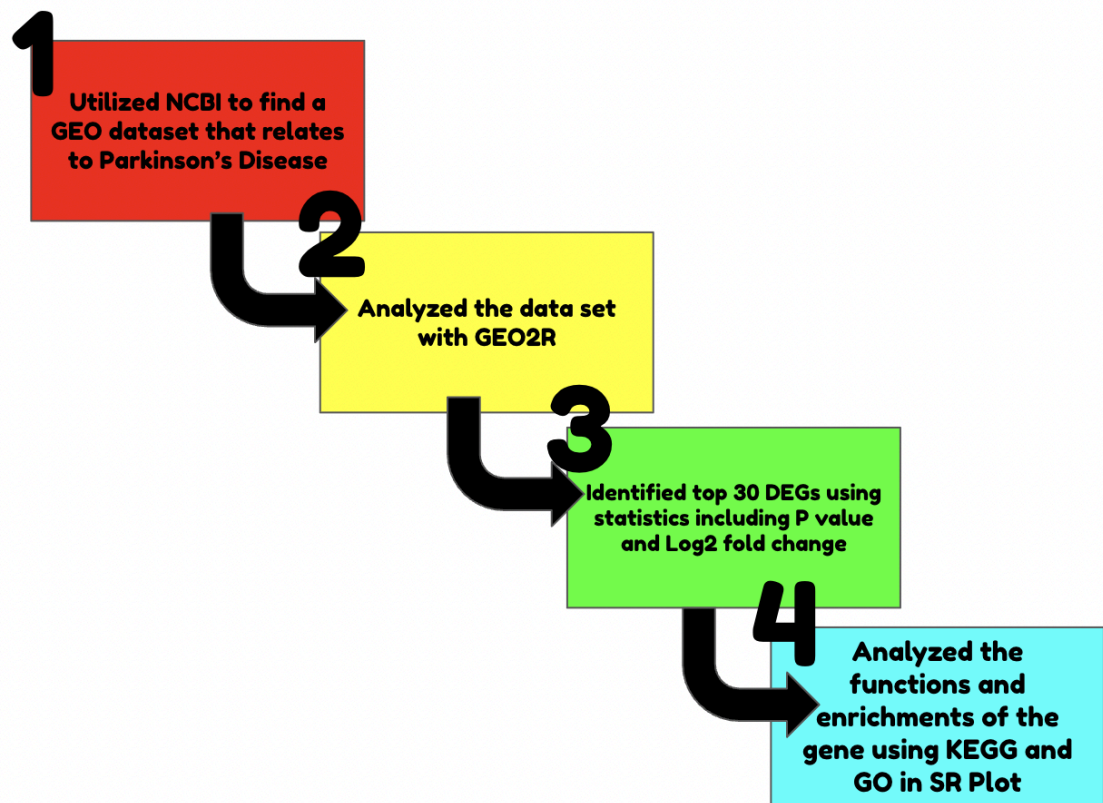
Figure 1: Summary of the methods and bioinformatics tools and databases used in this study. The methods used in the study were data collection, gene expression analysis, statistical analysis, and biological function analysis. The bioinformatics tools and databases used were GEO2R, SR Plot, KEGG and GO.

Identification of the Top Differentially Expressed Genes

To identify the most significant differentially expressed genes (the top 30), statistical analysis was applied. This process used a p-value cutoff of p <0.05 to prioritize the most important genes based on their differential expression across samples. A volcano plot and Venn diagram were generated to visualize the differentially expressed genes and their overlaps

Data Analysis Using SRPlot, KEGG, and GO Bioinformatics Tools

Following the identification of the top differentially expressed genes, SRPlot, KEGG (Kyoto Encyclopedia of Genes and Genomes), and GO (Gene Ontology) bioinformatics tools and databases were utilized to analyze the functions and enrichment of these top genes. These tools helped uncover the potential roles of the genes in the pathogenesis and progression of

Parkinson's disease, as well as identify relevant biological processes, molecular functions, and cellular components. Specifically, GO enrichment analysis was performed across three ontologies: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), to understand the high-level functions of the genes. KEGG pathway analysis was conducted to identify the most significant pathways in which these genes are involved, providing insights into the molecular mechanisms underlying Parkinson's disease. The top significant 3 to 5 pathways/genes were further investigated to elucidate their specific contributions.

RESULTS

The primary purpose of our analysis with GEO2R was to specifically identify genes that showed significant differential expression between the two sample groups. The initial results confirmed that a substantial number of genes were expressed differently in the Parkinson's disease patients compared to the healthy controls (Figure 2), providing a foundation for our subsequent functional and enrichment analyses.

In the volcano plot, the red dots represent the genes that were significantly upregulated in the Parkinson's disease group compared to the control group (Figure 2A),. These genes have a high positive fold change and a low p-value. The blue dots represent the genes that were significantly downregulated, having a high negative fold change and a low p-value. The black dots represent genes that were not considered statistically significant, as they did not meet our predetermined fold change and p-value thresholds (Figure 2A). Based on the analysis, a total of 30 differentially expressed genes were identified(⊞ PD Gene Dataset Top 30 DEGs).
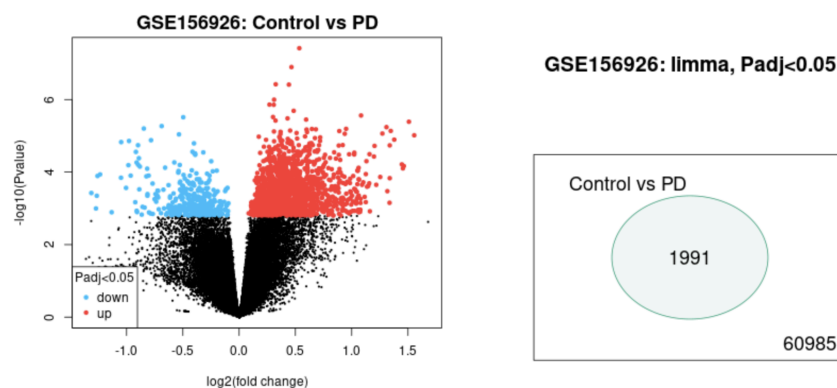

Figure 2. (a) Volcano Plot, (b) Venn Diagram illustrates gene expression differences across disease samples based on data from GSE156926. (a) Genes with substantial upregulation or downregulation in Control (not infected) and PD (infected) (based on log2 fold change thresholds) are shown in red and blue. Black dots illustrate genes with little to no expression

change. (b) The Venn diagram summarizes how from 60,985 genes analyzed, 1991 were found to be differentially expressed based on log2FC cutoffs, with very limited overlap among the conditions.

The Venn diagram, having a total of 60,985 genes, was generated to visualize the overlap between upregulated and downregulated genes within the Parkinson's disease group. Zero genes overlapped between the upregulated and downregulated groups, as these are mutually exclusive sets. While the study's core comparison was between the disease and control groups, the diagram showed that the 1991 genes were a unified set (Figure 2B), representing genes commonly and significantly altered in the disease state. Our focus was on this total set of 30 genes (gathered using statistics) that were consistently identified as differentially expressed.

Following the identification of the differentially expressed genes (DEGs), further analysis was conducted using the SR Plot bioinformatics tool. The Gene Ontology (GO) analysis revealed significant enrichment in terms related to G-protein complexes, GTPase activity, and hormone-related processes (Figure 4B). In parallel, the KEGG pathway analysis highlighted several key pathways, including the P53 signaling pathway and the dopaminergic synapse, as being significantly altered in the disease state (Figure 3A). These analyses provided critical insights into the biological functions and pathways affected by Parkinson's disease.
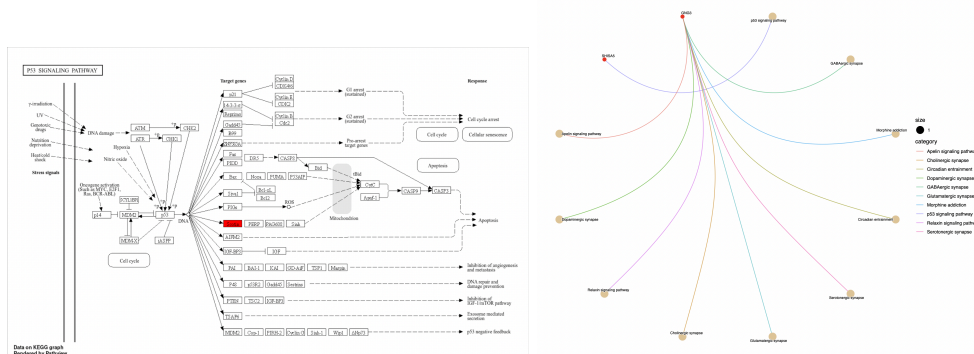

Figure 3A and B: a) Detailed visualization of the P53 signaling pathway, rendered by Pathview. This network diagram shows the complex interactions and downstream effects of the p53 protein, including its role in cellular stress responses, DNA damage repair, and apoptosis. The figure highlights the specific components and interactions within this pathway that are relevant to the pathogenesis of Parkinson's disease. b) The Cnet plot is a display of important BP genes and their relation to specific functions and activities. The red dots represent significant, upregulated genes. In this case, the two most upregulated genes are SHISA5 and GNG3.
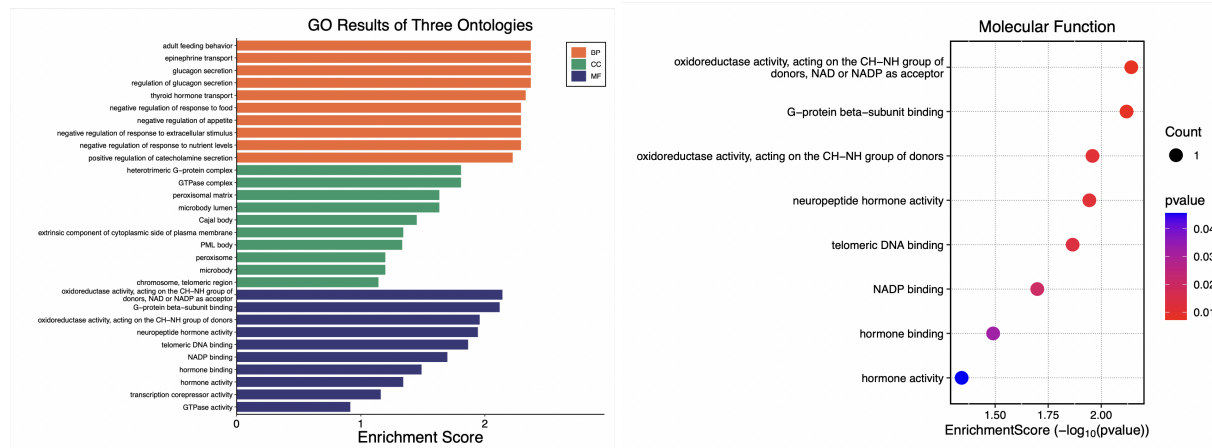
Figure 4A and B: a) BP, CC, and MF are the three ontologies that are pictured in this graph. The graph reveals that the top 30 DEGS are most enriched in BP and MF. Overall, the enrichment scores were greater than or around 1.5. b) Dot plot illustrating the enrichment score and p-value for various molecular functions. The plot shows that genes involved in oxidoreductase activity, G-protein beta-subunit binding, and neuropeptide hormone activity are among the most significantly enriched functions. The color and size of the dots correspond to the p-value and the number of genes associated with each term, respectively.

DISCUSSION

The primary goal of this research was to identify potential biomarkers for the early detection of Parkinson's disease by analyzing gene expression data. The study's main findings were derived from the GEO2R analysis of a Parkinson's disease dataset, followed by functional and enrichment analysis using GO and KEGG. The GEO2R analysis identified a list of top differentially expressed genes (DEGs) between the Parkinson's disease patient group and the healthy control group (Figure 2A). Subsequent Gene Ontology (GO) analysis revealed significant enrichment in terms related to G-protein complexes, GTPase activity, and hormone-related processes (Figure 4A). The KEGG pathway analysis highlighted several key pathways, including the P53 signaling pathway and the dopaminergic synapse, as being significantly altered in the disease state (Figure 3A).

Interpretation of Results

The differentially expressed genes and their associated pathways provide critical insights into the underlying molecular mechanisms of Parkinson's disease. The enrichment of genes associated with the GTPase complex and heterotrimeric G-protein complex (Figure 4B) is particularly significant. GTPases, such as Rab GTPases, are known to be key regulators of vesicle-mediated transport within cells. Defects in this transport system are strongly implicated in the pathogenesis

of Parkinson's disease, as they can lead to the impaired degradation and aggregation of proteins like α-synuclein, a hallmark of the disease [15].

Furthermore, mutations in the LRRK2 gene, a major genetic cause of familial Parkinson's disease, directly affect GTPase activity, highlighting this pathway's central role [9]. The identification of the P53 signaling pathway as a top enriched KEGG pathway (Figure 3A) is consistent with the current understanding of neuronal cell death in Parkinson's disease [16]. The p53 protein is a well-established regulator of cellular stress responses, including apoptosis (programmed cell death), which is a key process in the progressive degeneration of dopaminergic neurons in the substantia nigra [17]. Studies have shown that p53 levels and activity are substantially increased in the brains of Parkinson's disease patients, and its activation is closely linked to mitochondrial dysfunction, oxidative stress, and protein aggregation [10]. Therefore, the differential expression of genes within this pathway suggests a heightened state of cellular stress and apoptosis in the brains of Parkinson's disease patients, which directly relates to the loss of dopaminergic neurons and the clinical symptoms of the disease. The prominence of the dopaminergic synapse pathway (Figure 3A) is a direct and expected finding given the nature of the disease. Dopamine is the primary neurotransmitter involved in motor control, and the motor symptoms of Parkinson's disease are a direct consequence of the loss of dopamine-producing neurons [3]. The altered gene expression within this pathway reflects the widespread dysfunction of dopamine synthesis, transport, and signaling that defines the disease [3]. The results confirm the central hypothesis that synaptic dysfunction is a key event in Parkinson's disease and may even precede the widespread death of neurons.

Comparison with Previous Studies

The findings from this study are in strong agreement with a growing body of literature on the molecular pathology of Parkinson's disease. The implication of the P53 signaling pathway is supported by numerous research papers that have linked p53-mediated apoptosis to the neurodegeneration observed in both human patients and animal models of the disease [11]. Similarly, the involvement of the dopaminergic synapse (Figure 3A) is a foundational concept in Parkinson's disease research, and our results corroborate the findings of many studies that have focused on the failure of this pathway as a key event in disease progression. While this study provides an overview of several affected pathways, other research has delved into the specifics of these mechanisms. For example, some studies have shown that the G-protein-coupled receptors (GPCRs) and the associated heterotrimeric G-proteins play a crucial role in regulating dopaminergic neuronal survival and neuroinflammation, which is a major contributor to the disease's progression [12]. The overlap between our findings and these studies reinforces the validity of using bioinformatics tools like GEO2R, GO, and KEGG to uncover core pathological mechanisms.

Implications

The identification of differentially expressed genes and significantly altered pathways, such as those related to GTPase activity, p53 signaling, and dopaminergic synapses (Figure 3A), holds substantial implications for Parkinson's disease research and potential clinical applications. These findings suggest that the identified genes could serve as promising biomarkers for the early detection of Parkinson's disease, potentially enabling earlier diagnosis and intervention. Furthermore, a deeper understanding of these dysregulated pathways could pave the way for the development of novel therapeutic strategies. For instance, targeting specific components within the GTPase or p53 signaling pathways might offer new avenues for drug development aimed at slowing or halting neurodegeneration. These insights could also contribute to personalized medicine approaches, allowing for more tailored treatments based on an individual's specific molecular profile.

Limitations

A significant limitation of this study stems from its reliance on publicly available bioinformatics datasets from microarray experiments conducted by other researchers. While computational analysis provides valuable insights into gene expression patterns, it does not involve direct sample collection from human subjects or animal models. Consequently, the identified differentially expressed genes and pathways require rigorous experimental validation in

laboratory settings and clinical trials. This crucial step is necessary to confirm their biological relevance and functional impact before these findings can be translated into practical applications, such as the development of novel diagnostic tools, therapeutic drugs, or interventional strategies for Parkinson's disease. Additionally, the inherent variability and potential biases within publicly available datasets, including sample size and diversity, may influence the generalizability of the findings [14].

Future Directions

Building upon the current findings, future research should focus on experimentally validating the identified differentially expressed genes and their roles in Parkinson's disease pathogenesis. This could involve in vitro studies using cell culture models of Parkinson's disease to confirm gene expression changes and assess their functional consequences on neuronal survival and dopamine metabolism. Subsequently, in vivo studies using animal models of Parkinson's disease could be employed to further elucidate the mechanisms by which these genes contribute to neurodegeneration and to test potential therapeutic interventions. The identified genes could also be investigated as targets for drug development, with a focus on compounds that modulate their activity or expression to ameliorate disease symptoms or progression. Ultimately, these findings could progress towards clinical trials to determine if modulating these specific genes or pathways can lead to effective treatments or preventative measures for Parkinson's disease.

Table 1: Summary of Key Pathways and Genes Identified in this Study

Key Pathway/Gene Identified in this Study	Role in Parkinson's Disease
P53 signaling pathway	A well-established regulator of cellular stress responses and apoptosis (programmed cell death), which leads to the progressive degeneration of dopaminergic neurons.
Dopaminergic synapse	The primary site of dopamine communication. Its dysfunction is a direct consequence of the loss of dopamine-producing neurons, leading to the motor symptoms of the disease.
GTPase complex / *LRRK2* gene	GTPases regulate vesicle-mediated transport, and defects can lead to impaired protein degradation and the aggregation of proteins like -synuclein. Mutations in the *LRRK2* gene directly affect GTPase activity.
Heterotrimeric G-protein complex	These proteins are involved in cellular signaling and communication. Their

	dysregulation can contribute to neurodegeneration and neuroinflammation.

References

1. Kalia L. Parkinson's disease - the lancet [Internet]. [cited 2025 Aug 7]. Available from: https://www.thelancet.com/article/S0140-6736(14)61393-3/abstract

2. Klein MO, Battagello DS, Cardoso AR, Hauser DN, Bittencourt JC, Correa RG. Dopamine: functions, signaling, and association with neurological diseases. Cellular and molecular neurobiology. 2019 Jan 31;39(1):31-59.

3. Sveinbjornsdottir S. The clinical symptoms of Parkinson's disease. Journal of neurochemistry. 2016 Oct;139:318-24.

4. Rocca WA. The burden of Parkinson's disease: a worldwide perspective. The Lancet Neurology. 2018 Nov 1;17(11):928-9.

5. Emamzadeh FN, Surguchov A. Parkinson's disease: biomarkers, treatment, and risk factors. Frontiers in neuroscience. 2018 Aug 30;12:612.

6. Jiang F, Wu Q, Sun S, Bi G, Guo L. Identification of potential diagnostic biomarkers for Parkinson's disease. FEBS open bio. 2019 Aug;9(8):1460-8.

7. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Research [Internet]. 2008.

8. Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leimpeter A, Bloch DA, Nelson LM. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. American journal of epidemiology. 2003 Jun 1;157(11):1015-22.

9. Tsika E, Moore DJ. Contribution of GTPase activity to LRRK2-associated Parkinson disease. Small GTPases. 2013 Jul 1;4(3):164-70.

10. Szybińska A, Leśniak W. P53 dysfunction in neurodegenerative diseases-the cause or effect of pathological changes?. Aging and disease. 2017 Jul 21;8(4):506.

11. Chang JR, Ghafouri M, Mukerjee R, Bagashev A, Chabrashvili T, Sawaya BE. Role of p53 in neurodegenerative diseases. Neurodegenerative Diseases. 2012 Oct 28;9(2):68-80.

12. Jing Y, Yao P, Zhu H, Yu L, Lin Y, Kang D. The Role and Mechanisms of G protein-coupled receptors in Parkinson's disease. Neurological Sciences. 2025 Jun 11:1-5.

13. Advani D, Kumar P. Uncovering cell cycle dysregulations and associated mechanisms in cancer and neurodegenerative disorders: A glimpse of hope for repurposed drugs. Molecular Neurobiology. 2024 Nov;61(11):8600-30.

14. Grisci BI, Feltes BC, de Faria Poloni J, Narloch PH, Dorn M. The use of gene expression datasets in feature selection research: 20 years of inherent bias?. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2024 Mar;14(2):e1523.

15. Srinivasan E, Chandrasekhar G, Chandrasekar P, Anbarasu K, Vickram AS, Karunakaran R, Rajasekaran R, Srikumar PS. Alpha-synuclein aggregation in Parkinson's disease. Frontiers in medicine. 2021 Oct 18;8:736978.
16. Lei C, Zhongyan Z, Wenting S, Jing Z, Liyun Q, Hongyi H, Juntao Y, Qing Y. Identification of necroptosis-related genes in Parkinson's disease by integrated bioinformatics analysis and experimental validation. Frontiers in Neuroscience. 2023 May 22;17:1097293.
17. Li DW, Li GR, Zhang BL, Feng JJ, Zhao H. Damage to dopaminergic neurons is mediated by proliferating cell nuclear antigen through the p53 pathway under conditions of oxidative stress in a cell model of Parkinson's disease. International journal of molecular medicine. 2016 Feb;37(2):429-35.