IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A Multi-Omics Network Approach to Uncover Novel Therapeutic Targets in Parkinson's Disease

Anshuman Roy

Bridgewater-Raritan Regional High School, 600 Garretson Rd, Bridgewater, NJ 08807, USA; anroy2209@gmail.com
Mentor: Jobin Varkey, Virgel Torremocha, Jothsna Kethar

ABSTRACT: Parkinson's Disease (PD) is a progressive neurodegenerative disorder that impacts motor and cognitive abilities. It is predicted that more than 1.2 million people will be diagnosed with PD in the United States by 2030. There are ~20,000 genes in the human genome; however, only a tiny proportion of them have been linked to PD and related mechanisms, even though ~85-90% of cases are idiopathic or sporadic, occurring due to a combination of numerous genetic and environmental factors. This study sought to identify potential novel idiopathic candidates via network analysis in a biological context, where 27 monogenic PD-related genes were input from the OMIM Database. Through a combination of the STRING, GTEx, and GWAS Databases, after data cleaning scripts were finalized, a network proximity calculation was executed utilizing Dijkstra's algorithm, finding the shortest distance between nodes, which represented genes. An expression filter was also defined of greater than 1 transcripts per million (TPM) in the substantia nigra (SN), where the loss of dopaminergic neurons is most impactful to PD development. The top twenty candidates that fulfilled both criteria were displayed, and the top three genes, *EIF4A2, EIF4A1,* and *PABPC1*, were further analyzed regarding potential for therapeutic targeting if identified in PD. A statistically significant correlation was found between network proximity and SN expression, with a p-value of 3.92e-57.

KEYWORDS: Computational Biology and Bioinformatics, Computational Neuroscience, Parkinson's Disease (PD), Network Analysis, CRISPR-Cas9 Gene Editing.

Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that manifests heavily through motor symptoms such as tremors, muscle stiffness, and slowed physical movement, as described by the Mayo Clinic.[1] The Parkinson's Foundation in 2018 reports that it is predicted that 1.2 million people will be diagnosed with PD in the United States by 2030, with more than 90,000 people diagnosed per year in the nation.[2]

Surmeier describes PD as marked by a selective loss of dopaminergic neurons (DaNs) in the substantia nigra pars compacta (SNpc) and the formation of Lewy Bodies.[3] This SN-specific degeneration is responsible for the motor symptoms associated with PD. Brichta and Greengard share that very similar DaNs in the adjacent ventral tegmental area (VTA), a pivotal facilitator in the reward system, demonstrate a substantially reduced degree of degeneration.[4] This finding highlights the importance of selectively analyzing SNpc-specific gene expression to identify PD-related genes.

Clinically, PD is classified into two types: idiopathic, which occurs without a known cause and accounts for roughly 85-90% of cases, and familial, which is genetic in origin and comprises the remaining 10-15%. Tran *et al.* state that in familial cases, mutations in genes such as *SNCA* and *LRRK2* are often responsible, with prior research more indicative of gene targets in these cases.[5]

Wang *et al.* indicate that Parkin-mediated mitophagy, a process where the protein Parkin tags damaged mitochondria for removal by the cell's autophagy pathway, becomes exacerbated in PD.[6] This mitochondrial dysfunction is linked to the expression of vital mediating proteins, including alpha-synuclein, expressed by the *SNCA* gene; DJ-1, expressed by the *PARK7* gene; LRRK2, expressed by the *LRRK2* gene; and numerous others. Essentially, damaging genes responsible for producing essential mitophagy pathway proteins are hallmarks in monogenic PD. There is potential to further assess non-hereditary altered genes, but rather epigenetically impacted genes in idiopathic cases.

Dorsey and Bloem present that PD heavily occurs due to environmental toxicants, indicating a linkage to the disease via epidemiological and pre-clinical studies.[7] Notably, industrial solvents, including trichloroethylene (TCE), a contaminant present in groundwater most prevalent in factories and occupational environments, and air pollution, are shown to correlate with the disease. A study conducted by Goldman *et al.* demonstrates the impacts of TCE. The risk of developing PD for individuals who served at Camp Lejeune, where chemical levels exceeded 3000 times permitted by safety standards, was 70% higher in comparison to others residing in a less polluted base.

Through experimentation on transgenic mice after exposure to fine particulate matter, particles with diameters smaller than 2.5 μm were determined to indicate mitochondrial dysfunction and the formation of *SNCA* fibrils, a hallmark of PD. Pesticides and pollutants, specifically in agricultural areas, once again pinpoint the primary causes of PD to be linked to environmental reasons, as indicated in the proportion of idiopathic cases.

Subsequently, this study aimed to integrate multi-omics expression data, gene co-expression modules, and

DOI: 10.36838/IJHSR815.19

network-proximity-based scoring to identify novel therapeutic targets, drawing information from a combination of varying sources. Although the majority of idiopathic cases involve a combination of factors, recent advances in network-analysis methods and CRISPR gene editing opened opportunities to discover and manipulate novel PD-related genes. These are potentially specifically influenced by environmental toxicants, pollutants, and other factors. Previous studies included machine-learning biomarker discovery, data extraction from the GEO database, repurposing drugs using network biology, the utilization of computational methods, and other pivotal information, which have advanced research on PD. However, none identified novel gene targets for RNA-based gene therapy across both sporadic and familial PD populations, which this study aims to do by simultaneously filtering for SN expression. This study prioritized bridging these research gaps, specifically through further research in this field, which was essential for the discovery of genes previously unlinked to Parkinson's disease, as there are over 20,000 genes within the human genome.

Dias *et al.* expressed the role of oxidative stress (OS) in PD, specifically the degeneration of dopaminergic neurons.8 Reactive oxygen species (ROS) production also increases in this scenario, which subsequently results in a buildup of unstable molecules that can damage DNA, RNA, proteins, and lipids, seen in PD through neurodegeneration. Amongst several molecules, including neuromelanin and glutathione, which increase ROS production in the brain, dopamine plays a substantial role. Specifically, ROS production occurs due to dopamine oxidation products, dopamine quinones, which are accompanied by level regulation from monoamine oxidase-B (MAO-B). Evidence also suggests mitochondrial dysfunction as an indicator of OS within PD.

Methodology

The objective of this computational, in-field study was to identify novel idiopathic PD candidates via network analysis, where monogenic seeds from the OMIM Database and related literature served as the inputs. Data obtained from GWAS, GTEx, and STRING Databases represented the nodes of the network in a biological context, with weights indicating the strength of each candidate's relationship to the imported monogenic genes, on a scale of 0 to 1. Following, a network proximity score was computed using Dijkstra's algorithm to map the biological distance, or edge score. The top twenty genes with minimal distance to an identified monogenic input, alongside an expression greater than 1 transcript per million (TPM) in the SN, were displayed as the results. Further analysis was conducted, assessing the relationship between SN expression and computed network proximity, alongside a PubMed novelty assessment to determine prior PD-mechanisms-related research for the top twenty genes. The top three potential idiopathic genes, considering SN expression and network proximity score, were researched for involvement in neurodegeneration and protein synthesis pathways, followed by feasibility for therapeutic targeting if identified in PD. No human subjects were involved, and to ensure ethical data use, all data acquired were publicly available and remained anonymized on an individual patient basis.

Network Analysis and Protein-Protein Interaction Overview:

Before proceeding with detailed methodological procedures, it is crucial to outline the fundamental principles of network analysis and their application to protein-protein interaction data. This accentuates the involvement of numerous genes and components in the human interactome, a detailed map of protein-protein interactions (PPIs), involved in disease pathology. Barabási *et al.* detail critical aspects of a network in a biological context, where nodes represent proteins and edges represent interactions that connect these proteins.[9] In parallel, co-expression networks are also outlined, in which two genes possessing similar expression patterns are linked if they modify the phenotype when both are mutated, contrasting with a single mutation.

In this study, weighted networks were used to display the strength of interactions between protein expression of genes through confidence scores. These relationships were captured through network proximity, which captured the minimal distance between gene interactions. Büttner and Krieter exemplify this as a defining characteristic compared to unweighted networks, which display the existence of an interaction as binary, given that it exceeds a specific threshold.[10]

Network Medicine and Gene Prioritization:

Wu *et al.* investigated potential biomarkers of PD by analyzing microarray gene expression data from patients. Utilizing a computational approach, specifically Weighted Gene Co-expression Network Analysis (WGCNA), data were extracted from the GSE99039 dataset of NCBI's Gene Expression Omnibus (GEO) database.[11] After preprocessing the data, WCGNA was used to create a co-expression network, with genes in the top 25% variance, to examine how a model's predictions change depending on training data. Afterward, Pearson correlation matrices were used to transform into a weighted adjacency matrix for the genes. Module membership and gene significance were used to calculate the biological proximity of the genes detected to PD, with the ClusterProfiler Package responsible for gene ontology, determining the functions of the genes before undergoing calculation. To construct the PD prediction model, 7 genes were selected, all related to inflammation and immunity. This study demonstrated the effective use of the GEO database, along with essential preprocessing steps. In addition, it utilized network analysis and created a PD prediction model closely analyzing the genes, information aligning with the usage of RNA-based gene targeting.

Using network proximity and node similarity for drug-gene connections, Dey *et al.* built DTI-prox to find early-onset Parkinson's disease (EOPD) markers and possible drug repurposing candidates.[12] After preprocessing the data, 55 disease-specific genes, 806 drug targets, 417 novel drug-target pairs, and six candidate biomarkers were discovered, based on the proximity scores with the biomarkers and drug targets.

3 different databases, including the GEO and UniProt, were utilized as publicly available data. Protein-protein interactions (PPI) were also established from the STRING database, with only high-confidence interactions retained for detecting meaningful models. Figure 1 shows the shortest path algorithm used to find the distances between nodes, depicted as circles.

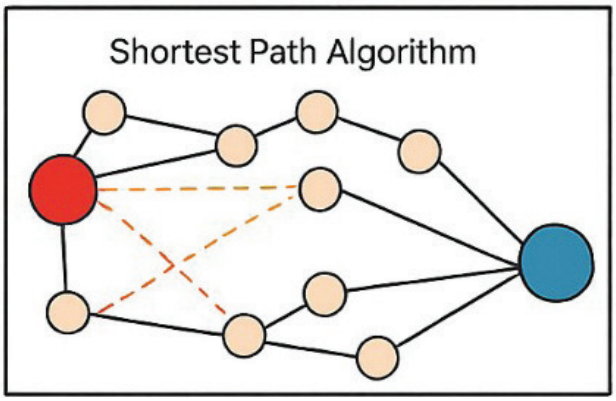

Figure 1: Schematic representation of the DTI-Prox workflow. This figure illustrates the network construction, proximity measurement, and drug-target identification steps. This visualization was implemented as a workflow to calculate the distance from monogenic PD genes to potential candidates. Created using BioRender.

These studies showcase the importance of multiple proximity parameters when calculating confidence scores, including the usage of nodes, edges, and other representations. They also demonstrate the effective use of the GEO database, along with essential preprocessing steps. In addition, they utilize network analysis and create a PD prediction model, closely analyzing the genes, information aligning with the usage of RNA-based gene targeting as a therapeutic intervention.

Termine *et al.* identified PD subtypes using post-mortem RNA-Sequencing (RNA-seq) data for precision medicine, specifically identifying the targeted genes within the subtypes.[13] Targeting the issues of any a priori conceptions and the flaws present within empirical clustering, scientists transformed data analysis in PD research through unsupervised machine learning and network analysis, effectively displaying the complex heterogeneity present. The pipeline yielded 2 clusters: PD Cluster 1 and PD Cluster 2. The gene expression profiles were compared between the groups via differential gene analysis (DGA), resulting in Differentially Expressed Genes (DEG). The Wnt/β-catenin pathway, regulating adult neurogenesis, largely differed in both groups, and *NEUROD1* was a key regulator in gene networks. As a result, PDC1 had a lower life expectancy on average in comparison to PCD2. Afterward, a drug repurposing pipeline was used to target pivotal genes in PCD1 that were downregulated, with many filtered and removed.

Dou *et al.* developed a network-based genetic framework for identifying disease-associated genes in PD, utilizing five types of brain-active xQTLs, a locus where genetic variation determines phenotypes. PD risk genes (pdRGs) underwent multi-omics validation, examining the pathobiological pathways, including DaN-specific genes, as well as gene expression in the substantia nigra, the most likely place of expression.[14] Next, using an snRNA-seq database, the risk genes were analyzed across seven different cell types, and found to be more differentially expressed in human PD brain cells. After the 175 disease-gene associations were established, network proximity scores were calculated between the encoded proteins and the human PPI drug target network. After evaluation, Simvastatin was found to be associated with a reduced incidence of PD. In addition, Simvastatin was found to target a protein product, ITGB2, of a mutated gene playing a role in neurodegenerative diseases and PD, aligning with a potential gene therapy solution.

These studies highlight the importance of unsupervised learning-based data clustering, which can provide crucial insights into the reasons behind a gene's ineffectiveness in a given scenario. Through these studies, the usage of xQTLs and multi-omics to identify disease-associated genes has potential in a network analysis model, alongside other data points.

Moreira and Jan demonstrate the accessibility and utility of numerous publicly available datasets for neurodegenerative diseases (NDDs), including those from four cohorts of PD, covering bulk RNA-seq, CSF proteomics, and microarray studies.[15] These account for standardized pipelines (in Python and R) for downloading, quality controlling, preprocessing, and combining these heterogeneous data types. An example analysis is the construction of co-expression modules from GEO transcriptomes, showing the widespread accessibility of the data even to non-computational researchers. In addition, crucial data considerations, including age, disease stage, and medication status, are highlighted. This guide is often utilized to receive and preprocess GEO and proteomics datasets before displaying network-proximity scores, ensuring targeted genes are sufficiently evaluated.

Across these five studies, the usage of NCBI's Gene Expression Omnibus (GEO) database is profuse, providing the necessary information for data clustering and preprocessing for creating PD subgroups. Specifically, microarray data, variations of RNA-seq data, and proteomics are used in prior research. Furthermore, multi-omics, the integration of varying data, as well as network analysis, are used to determine novel gene targets, PPIs, and drug repurposing candidates. Wu *et al.* utilized co-expression network analysis to discover novel genes and the plausibility of prediction analysis in the scope of gene therapy.[11] In contrast, Dey *et al.* identified PD genes through unsupervised learning data clustering, and then compared the genes between the two groups via DGA.[12]

The following study brought together a growing body of research in Parkinson's Disease network analysis, emphasizing the intricate relationships between data retrieval, preprocessing, and clustering methods. It brought to light the necessity of more targeted studies on weighted gene co-expression and multi-omics approaches, which are vital for enhancing the precision of gene identification in both idiopathic and familial forms of PD. This project also incorporated multiple elements for accurate network proximity calculations, such as protein-specific information obtained from STRING. Addressing these gaps may pave the way for refining disease

DOI: 10.36838/IJHSR815.19

models, uncovering previously overlooked genetic contributors, and ultimately accelerating the development of innovative therapeutic strategies.

Substantia Nigra-Specific Gene Expression:

Martirosyan *et al.* investigated human post-mortem SNpc samples from 29 patients, divided into two groups: one consisting of 15 sporadic PD patients, and the other 14 serving as control samples. 28 dopaminergic and overlapping disease-associated genetic markers, including *ALDH1A1, SLC6A3*, and *SLC18A2*, were identified.[16] The researchers compiled an snRNA seq dataset, reporting analysis of the transcriptomes of approximately 84,000 high-quality nuclei, and exceeding 2,000 DaN neuron nuclei, one of the most prevalent cell types in the substantia nigra. Association of other cells within the SN, including tyrosine hydroxylase (TH) and glial populations, was depleted, linked via a pathway analyzed in unfolded protein response (UPR), a pathway activated due to the accumulation of unfolded proteins in the endoplasmic reticulum (ER). As UPR is consistent with dysfunction in vesicle tracking, it displays a higher prevalence in PD cases from genes involved in multiple processes in the SN. After preprocessing the dataset, scientists analyzed GWAS data for cell types present within the SN, comparing them to high-confidence genes associated with monogenic PD. To identify potentially overrepresented cell types, a Binomial test was performed, assessing the cell-type-specific enrichment, a computational method using the two-sided R function 'binom.test()'. This research demonstrates the variance of PD-related cells present in the substantia nigra, including DaNs, TH neurons, glial cells, and numerous others. It also showcases the applicability of selectively filtering to select relevant genes, crucial in the pipeline this study developed. Similarly, clustering is used to successfully eliminate various genes in Seurat V4.4.0, while the known SN markers and provided monogenic seeds serve as nodes in the network analysis model.

The loss of DaNs in the SN has been indicated to be the foremost contributor to the pathology of PD, including the loss of fine motor control, movement, cognition, and motivation; however, other cell types, including oligodendrocytes (ODC), also display impacts. A large-scale assessment considering these factors was necessary to accurately identify PD-associated genes, as cell-type-specific gene expression varies within the SN, uniquely contributing to the disease. Agarwal *et al.* established the significance of the SN through sequencing and comparing the transcriptomic profiles of nuclei across ten distinct cell populations from the cortex or the SN.[17] The scientists determined 95.5% of nuclei in the SN, represented predominantly by ODCS, were from glia, exceeding the 12% of nuclei in the cortex from glia. An increase in glia can serve either as a neuroprotective mechanism by clearing debris or as an inflammatory response by contributing to oxidative stress, which can lead to neuronal death. Furthermore, conditional analysis employed LD score regression (LDSC) to evaluate that ODC-specific expression patterns were distinct from DaN patterns, underscoring the dispersion of known-PD risk genes across varying SN neuronal cell types. Cell-type-specific PPI networks were also utilized to determine relations between genes, specifically the enrichment of metabolic processes, gene regulation, protein phosphorylation, and kinase activity involved in ODCs associated with PD. Similarly, this study used SN expression and cell-type-specific gene expression as parameters to identify potential idiopathic-PD genes via PPI, comparing the expression of established monogenic genes to novel discoveries.

Therapeutic Targeting and Gene Therapy Readiness:

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) gene therapy is described as a recently developed genetic engineering technology, involving a Cas9 nuclease guided by Guide RNA (gRNA), according to CRISPR Therapeutics.[18] Undergoing rapid improvement, the therapeutic strategy has shown promise for numerous diseases, as it can either silence, delete, or correct genes through cutting DNA segments. Ntetsika *et al.* highlight the potential of CRISPR-Cas9 genome editing in modulating disease-related pathways, demonstrating how targeted correction of PD-associated mutations can restore DaN functionality within preclinical models.[19] Moreover, Pinjala *et al.* reviewed CRISPR/Cas9-based stem cell therapy as a potential disease-modifying, therapeutic strategy in PD.[20] In this mechanism, currently experimented on via disease models, dysfunctional DaNs are restored to recover motor functions, substantially initiated by gene recovery. The CRISPR/Cas9 system is preferred over other gene editing techniques due to its off-target site predictions, rapid editing efficacy, and adaptation to the target genomic sequence. The scientists also expressed the *in vivo* applications of the tool, citing a previous study conducted by Yoon *et al.*, where an adeno-associated virus (AAV) vector was designed to delete the A535-SNCA PD mutation.[21] In PD, this has been applied to deliver therapeutic genes to the SN and striatum and traverse the Blood-Brain Barrier (BBB), crucial for modifying the expression of disease-involved genes, and their respective protein expression and transcriptomic data.

Environment Setup and Libraries:

To start the model development process, Anaconda PowerShell was used to deploy and activate the Microsoft Visual Studio (VS) Code environment: pd-netprox.

The following key Python libraries were imported across the preprocessing and network analysis scripts. The pandas library, a frequently used library for data manipulation, analysis, and handling large datasets, is imported for reading and cleaning the OMIM, STRING, GWAS, and GTEx files. Open Chromosome Collective (Open2C) *et al.* underscored the applicability of pandas in manipulating genomic intervals through developing a novel Python library, named bioframe, which maintained the API and several other principal components of pandas.[22] The NumPy library possesses mathematical functions and multidimensional arrays, alongside logarithmic capabilities, which are used to convert STRING database confidence scores, initially ranging from 1 to 1000. These confidence scores were later transformed into final edge weights from 0 to 1 and used to calculate the network proximity. NetWorkX is a Python

package used to represent the human protein-protein interaction (PPI) network based on data from the STRING database, holding nodes, which represent the individual proteins, and edges representing the interactions. Zhou and Lauschke utilized the NetWorkX package to analyze the assortativity of pharmacogenomics gene-drug links to determine the impact on epistatic interactions.[23] They discovered that the prevalence of complex gene-gene-drug interactions was increased when the nodes were in closer proximity, a principle that is applied in this application. Other libraries used included the os and pickle libraries, which formed the backbone of data retrieval and management infrastructure. They ensure cross-platform compatibility to enhance reproducibility across different operating systems, and effectively construct object structures as well as reuse preprocessed networks without reconstruction once the user reloads the application.

For reproducibility purposes, all aspects of the project can be found in the following GitHub Repository: https://github.com/Anshuman-Roy-22/pd-netprox.

Monogenic Seeds from the OMIM Database and Literature:

Confirmed Monogenic PD Genes (n = 15)	Literature-Supported Candidates (n = 10)	Lower-Confidence / Controversial (n = 2)
SNCA	DCTN1	EIF4G1
LRRK2	ATP6AP2	LRRK1
PRKN (PARK2)	MAPT	
PINK1 (PARK6)	GRN	
PARK7 (DJ-1)	PSEN1	
VPS35 (PARK17)	PSEN2	
GBA (GBA1)	HTRA2	
ATP13A2 (PARK9)	AFG3L2	
PLA2G6 (PARK14)	COQ2	
FBXO7 (PARK15)	TARDBP	
DNAJC6 (PARK19)		
SYNJ1 (PARK20)		
RAB39B		
CHCHD2		
TMEM230		

Figure 2: Monogenic Parkinson's disease (PD) gene set used for network construction, stratified by evidence level. Fifteen confirmed genes from OMIM and literature, ten literature-supported candidates, and two lower-confidence genes (EIF4G1, LRRK1) were included to capture both established and exploratory PD-associated network nodes.

The data collection and preprocessing steps were initialized when monogenic PD seeds were extracted from the Online Mendelian Inheritance in Man (OMIM) Database of Johns Hopkins Medicine.[24] Further genes were extracted upon examination of the literature, indicating their involvement within PD mechanisms. Genes specific, causative, and autosomally linked to PD, including the *PARK7* and *PARK6* genes, were filtered for usage in the protein-protein interaction comparisons. Subsequently, the following 15 genes were directly input into the monogenic PD genes TSV file: *SNCA, LRRK2, PRKN (PARK2), PINK1 (PARK6), PARK7 (DJ-1), VPS35 (PARK17), GBA (GBA1), ATP13A2 (PARK9), PLA2G6 (PARK14), FBXO7 (PARK15), DNAJC6 (PARK19), SYNJ1 (PARK20), RAB39B, CHCHD2, TMEM230*. Another ten genes present in the OMIM Database, yet not confirmed to be associated with monogenic PD, were also included in the list based on literature indicating their association in monogenic cases. These were also identified in the existing literature, and they were as follows: *DCTN1, ATP6AP2, MAPT, GRN, PSEN1, PSEN2, HTRA2, AFG3L2, COQ2*, and *TARDBP*. Two more genes whose role remains controversial in monogenic PD were included as lower-confidence seeds to account for potentially relevant nodes: *EIF4G1* and *LRRK1*.

STRING Database:

Data from the human PPI network from the STRING v11.0 database were utilized to compare the expression of monogenic genes with other potential PD candidates. The individual PPI interactions, which represent edges in a network, were filtered to retain a confidence score of ≥ 700, indicating high confidence in the interaction between protein expression. A detailed dataset for Homo sapiens (9606) was extracted for both protein links (interaction data), alongside protein info (provided gene symbols for comparison to monogenic seeds). Heterogeneous evidence sources across multiple studies and methodologies were integrated, introducing variability and limiting direct comparability of interaction confidence. Consequently, proximity scores may be biased by underlying network construction and should be interpreted as measures of topological association rather than definitive evidence of a gene's involvement in Parkinson's disease.

NHGRI-EBI GWAS Catalog:

Data from the National Human Genome Research Institute - European Bioinformatics Institute Genome-Wide Association Study Catalogue was obtained, specifically genes categorized under "*Parkinson's Disease*." The catalogue was manually downloaded as a TSV file and placed within the pd-netprox VS Code environment. To identify known associations, the GWAS gene names were standardized to network gene symbols for verification.

GTEx v8 RNA-seq Substantia Nigra Expression Data:

Gene expression data for human substantia nigra (SN) were obtained from the GTEx project. GTEx collected RNA-sequencing data, which provided the median Transcripts Per Million (TPM) for each gene. This was utilized to quantify expression of candidate genes in the SN for later filtration, previously established to be the foremost region affected in PD. A threshold of >1 TPM was established to sort for moderately expressed genes in the SN at a minimum, with genes not detected filtered out.

Dijkstra's Algorithm / Shortest Path Computation:

The network graph was constructed to account for weighted interactions and compute proximities. After construction, Dijkstra's Algorithm was used to calculate the lengths between the monogenic PD inputs and viable sporadic genes, based upon the weights of the STRING database interactions. The weights were calculated by dividing 1000.0 by the confidence score, which would result in higher weights classified as maintaining "shorter paths" in terms of biological distance, compared to lower weights. Accounting for every gene, the *nx.multi_source_dijkstra_path_length* function in the NetworkX library effectively algorithmically served as Dijkstra's. Figure 2 shows the implementation of the aforementioned function in the model.

DOI: 10.36838/IJHSR815.19

```
# ------------------------------------------------------
# Multi-source Dijkstra's Algorithm to Compute Shortest Path
# ------------------------------------------------------
print("[4/7] Running multi-source Dijkstra...")
distances = nx.multi_source_dijkstra_path_length(G, seeds, weight="weight")
print("Finished Dijkstra.")
```

Figure 3: Code within the compute_proximity.py file using multi-source Dijkstra's Algorithm to compute the shortest path between all Homo sapiens protein links (STRING Database) and GWAS genes to the monogenic seeds. This program was used to implement the network proximity calculation and yielded the top PD-associated candidates for further assessment. The g parameter was initially defined as infinity for all the potential candidates, and then was continuously updated across nodes. Also, multi_source_dijkstra_path_length determined genes associated with all of the original monogenic input genes, finding the total cost and the distance.

Candidate Prioritization and Analysis:

In addition to the network proximity calculation, a further candidate prioritization criterion was established to select the top twenty priority candidates. In accordance with the GTEx v8 RNA-seq Substantia Nigra Expression Data, candidates were filtered based on expression in the SN, whereas previously stated, the required threshold was > 1 TPM to ensure the candidate list would have a plausible involvement in PD. The genes with the smallest distance that passed this required threshold would be added to the final candidate list until it contained twenty genes. A detailed description and protein-interaction overview were constructed for each gene using the Homo sapiens (9606) protein info dataset from STRING v11. Afterward, a graph was constructed to compare the correlation between network proximity and SN expression. The outliers, namely mitochondrial genes, were previously extracted due to their substantially higher SN expression levels.

A detailed analysis was performed regarding involvement in PD mechanisms, alongside the most prominent gene involved in mitochondrial dysfunction in PD. Three genes were selected for further validation for therapeutic targeting, where CRISPR feasibility was evaluated through review of prior literature and experimentation.

Following this step, a manual PubMed novelty assessment was conducted on each of the genes. Within PubMed, the following would be entered into the search bar: *[Gene Name][tiab] AND Parkinson[tiab]*, where [tiab] represented title/abstract, allowing the search to focus on specific PD mentions. The results for each gene were obtained to showcase which candidates had either an abundance or a lack of prior experimentation. Lastly, the limitations of this study were highlighted.

Results and Discussion

Identified Idiopathic PD Candidates:

Potential idiopathic PD genes were identified based on their network proximity scores and placed in descending order, indicated by the values in yellow in Figure 3. The original monogenic PD-related seeds, which included *SNCA* and *LRRK2*, yielded a proximity score of 0, as they were mapped to themselves. Subsequently, after running Dijkstra's algorithm, the genes that held proximity scores closer to 0 were prioritized. SN expression was obtained from the GTEx database and showcased in the rightmost green column in Figure 3, measured in TPM. When selecting the top three candidates for therapeutic targeting analysis, genes with the lowest proximity score, yet the highest SN expression, were selected.

SNX3	9606.ENSP00000230085	0.0010010010010010	141.838
VPS26B	9606.ENSP00000281187	0.0010010010010010	20.9821
PSENEN	9606.ENSP00000468411	0.0010010010010010	34.675
APH1A	9606.ENSP00000358105	0.0010010010010010	39.9957
SNX2	9606.ENSP00000368831	0.0010010010010010	17.8205
APH1B	9606.ENSP00000261879	0.0010010010010010	7.57099
EIF4E	9606.ENSP00000425561	0.0010010010010010	4.83596
EIF4B	9606.ENSP00000388806	0.0010010010010010	62.5892
EIF4A2	9606.ENSP00000326381	0.0010010010010010	238.014
NCSTN	9606.ENSP00000294785	0.0010010010010010	21.3908
PABPC1	9606.ENSP00000313007	0.0010010010010010	66.9184
DCTN2	9606.ENSP00000408910	0.0010010010010010	66.5155
EIF4A1	9606.ENSP00000293831	0.0010010010010010	72.4078
VPS26A	9606.ENSP00000362480	0.0010010010010010	21.1793
VPS29	9606.ENSP00000480853	0.0010010010010010	24.2028
APP	9606.ENSP00000284981	0.001002004008016032	289.376
CLIP1	9606.ENSP00000479322	0.001002004008016032	9.69188
DYNC1H1	9606.ENSP00000348965	0.001002004008016032	44.8053
ACTR1B	9606.ENSP00000289228	0.001002004008016032	43.0907
DCTN3	9606.ENSP00000259632	0.001002004008016032	20.4078

Figure 4: Top 20 gene candidates selected following the columns' gene_symbol, ensp_id, proximity, sn_tpm in respective order, and positioned directly after monogenic PD candidates. These 20 genes were further assessed for prior research conducted on them, based on a PubMed novelty assessment. Top candidates were selected based on proximity score and SN expression.

The correlation between proximity and SN expression was also identified in Figure 4, with an r-value of -0.156 (excluding outliers), highlighting a weak to nonsubstantial relation between increased distance to SN expression. The most notable outliers, however, were mitochondrial (MT) genes, notably the *MT-co2* gene, which had a computed proximity score of ~0.002120, although an SN expression of 77608.2 TPM, far exceeding the threshold limit of 1 TPM defined in this study to ensure minimal expression. It also far exceeded the threshold classified as high expression, which is 1000 TPM.

Naren *et al.* identified the involvement of the *MT-Co2* gene in PD pathology, specifically mitochondrial dysfunction, which was previously referenced as a crucial hallmark.[25] The *MT-Co2* gene encodes subunit 2 of Cytochrome c oxidase (Complex IV) in the mitochondrial electron-transport chain, which is critical in transferring electrons to oxygen and maintaining Adenosine triphosphate (ATP) generation via oxidative phosphorylation. Dysfunction of this critical gene could disrupt Complex IV activity, leading to impaired mitochondrial respiration, alongside apoptosis of dopaminergic neurons, both of which are prominent in PD. The reasoning behind this gene's exclusion from the nearest candidates based on distance was explained further in the limitations section of this study, representative of the numerous factors involved in accurately identifying the best candidates.

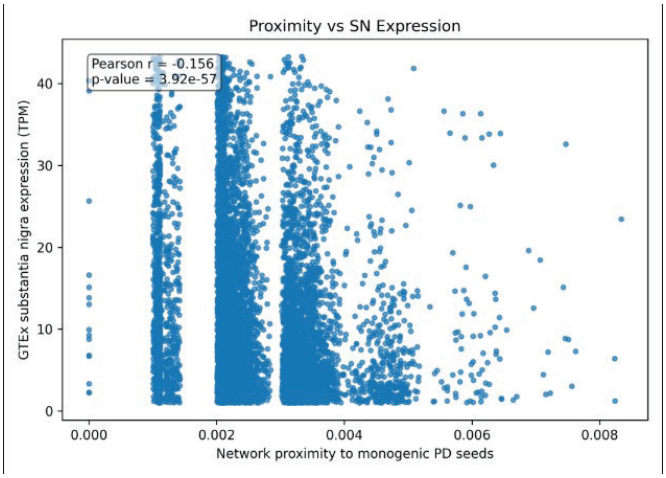

Figure 5: Across potential candidates (non-outliers) linked to PD from the GWAS and the STRING Database, a scatterplot was constructed to showcase the relation between network proximity and SN expression. Numerous MT genes, with a network proximity score of ~0.002, previously heavily skewed the graph, maintaining a significantly higher SN expression as previously defined. The proximity score is a weighted shortest-path distance in the STRING network, where 0 is reserved for the monogenic PD seed genes themselves, and values increase upward without a fixed upper bound as genes become further away. So, a score of ~0.002 in this project is extremely small and indicates that a gene lies only about two near-max confidence interaction steps from a known PD seed. This indicates the role MT variants play in the SN, the highest PD-involved brain region, and ultimately in PD mechanisms. The p-value of 3.92e-57 showcases a high statistical significance between network proximity and SN expression, despite the moderately weak linear relationship (r-value) of –0.156. This occurred since the p-value was below 0.05, leading to a high probability that the null hypothesis (indicating no association between two variables) could be rejected, except for potential errors. Accordingly, the weak negative correlation means that genes closer to the PD seed network tend to have somewhat higher SN expression on average, but the relationship is not strong or uniform, so proximity is useful as a closeness signal for candidate prioritization, while SN expression serves better as a supporting biological context than as a dominant standalone ranking metric.

Further Analysis and CRISPR Feasibility of the Top Three Genes:

The *EIF4A2*, *EIF4A1*, and *PABPC1* genes were selected for further analysis, resulting from a combination of their proximity scores and high SN expression.

Harrer *et al.* characterized the involvement of variants of the *EIF4A2* gene in dystonic conditions, defined as a phenotypically heterogeneous group of movement disorders resulting in involuntary muscle contractions with neurodevelopmental abnormalities.[26] Previously, missense and deletion mutations of the gene led to intellectual disability stemming from the eukaryotic translation initiation factor 4A isoform 2 (eIF4A2) in the protein synthetic pathway. The study conducted an integrated analysis of rare variants of the *EIF4A2* gene across independent patients from 1100 families who suffered from dystonic conditions. A single heterozygous 2-bp deletion (c.896_897del) in exon 8 of *EIF4A2* was identified as a top disease-causing candidate, in which the frameshift mutation was predicted to trigger nonsense-mediated mRNA decay (NMD) to an overactive state, leading to the degradation of protein development. This mutation contributed to a ~50% reduction of eIF4A2 protein, which is disease-causing with the potential for severe encephalopathic recessive inherited disease, causing developmental and intellectual delay. Shetty *et al.* noted the overlap between key attributes of dystonia and PD, notably the motor control and cognitive abnormalities, with dystonia observed in over 30% of PD cases.[27] Mutations in various DYT genes blocking the dopamine synthesis pathway were confirmed as hallmarks of PD, mechanisms that could be integrated into the findings of *EIF4A2* implicated in movement disorders.

EIF4A1 and *PABPC1* have not yet been genetically linked to PD, although their functions within initiating protein translation and mRNA stability machinery serve as crucial indicators. Xue *et al.* confirmed the functions of protein expression of the *EIF4A1* gene on the ATP-dependent RNA helicase, eIF4F, which plays a significant role in human cancers.[28] Specifically, eIF4F translation was found to be a crucial step, and when dysregulated, it increased the chances of tumorigenesis and malignant progression, as found in gastric cancer. Furthermore, mRNA expression levels of the eIF4A1 protein were obtained from the GEO Database, where it was overexpressed in colorectal cancer, cervical cancer, breast cancer, and various other cancers. Surguchov and Surguchev referenced epidemiological evidence that indicated an inverse relationship between PD and all cancers except melanoma.[29] This could potentially indicate a relationship between a decrease in eIF4A1 SN expression and the development of PD, though further research is required.

Qi *et al.* assessed the expression of the Cytoplasmic poly A binding protein (PABPC1), which was similarly highly expressed in a variety of carcinomas.[30] They defined the protein as a central regulator of cytoplasmic mRNA, which binds to the poly(A) tail during translation. Furthermore, they identified it to localize to stress granules under cellular stress, an important mechanism for cellular defense, yet also a viable contributor to neurodegenerative diseases. Yuan *et al.* highlighted that the aggregation of alpha-synuclein, an unstructured protein that accumulates during PD pathogenesis, is caused by stress granule assembly from chronic stress driven by mitochondrial-impairing neurotoxins.[31] ATP production is subsequently impaired, which cyclically induces stress granule production and contributes to neurodegeneration.

Targeting eIF4A helicases has been validated as a feasible therapeutic strategy. Chu *et al.* implemented CRISPR/Cas9-mediated approaches to validate targeting of eIF4F by small molecule inhibitors, namely rocaglate family members, promoting antineoplastic activity, effectively destroying cancerous cells.[32]

Rocaglates have been studied for their biological capabilities of inhibiting translation; however, they were demonstrated to be resistant to eIF4A1-mutant cell lines generated through CRISPR/Cas9 editing, thereby confirming the involvement of these compounds through eIF4A helicase inhibition. The authors further tested this by introducing an eIF4A1 mutation directly into the endogenous *EIF4A1* locus, where they engineered a retroviral complementation vector (RCV) that used small hairpin RNA (shRNA) to suppress the gene. To further strengthen results, Cas9 was used for two single guide RNAs (sgRNAs) designed to target exon 5, confirming eIF4A1 as the foremost inhibiting target. These findings highlight the plau-

DOI: 10.36838/IJHSR815.19

sibility of therapeutic targeting methods to silence or mutate both *EIF4A1* and *EIF4A2* in PD, including shRNA and sgRNA to cross the BBB to target the SN.

Malerba *et al.* experimented *in vivo* on the A17 mouse model, where serotypes 8 and 9 AAV vectors were injected to silence the mutated *PABPC1* gene with a trinucleotide repeat expansion responsible for oropharyngeal muscular dystrophy (OPMD), an autosomal dominant, late-onset muscle disorder.[33] The study showed that when PABPN1 protein aggregates were successfully significantly degraded, however, muscle depletion and degeneration occurred, with 60% of centrally nucleated fibers three months after injection, highlighting the negative impacts arising from the complexity of treating the nervous system.

Collectively, these findings underscore the therapeutic promise of gene-based therapeutic interventions. As research advances, particularly given the majority of PD cases arising sporadically from a combination of epigenetic factors, such approaches, as depicted in Figure 5, hold significant potential for broad impact across numerous disease subtypes.

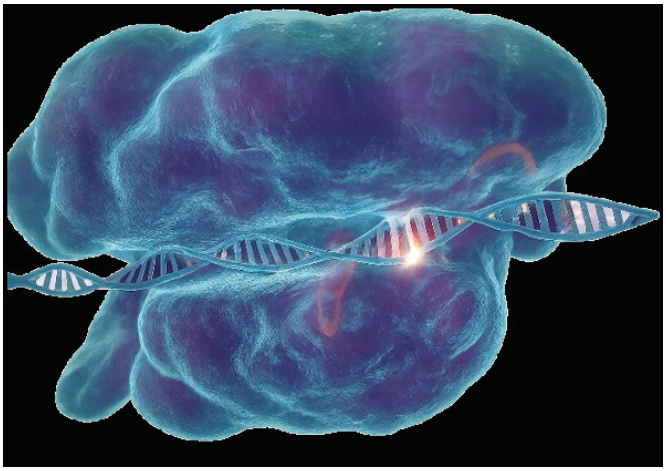

Figure 6: Image representing genome editing using CRISPR-based tools. The diagram illustrates the targeted cleavage of DNA by the Cas9 enzyme guided by RNA, enabling precise modification of specific genetic sequences. Furthermore, it promises high therapeutic potential for PD. Source: (MIT McGovern Institute 2014). Created using BioRender.

PubMed Novelty Assessment:

The results of the PubMed Novelty Assessment conducted were depicted in Figure 6 as follows, which excluded the APP gene as it was an outlier, yielding 490 identified results. Furthermore, the *MT-Co2* was similarly analyzed, yet not included in the graph, citing two PubMed results. The results are critical for analyzing conducted research targeted to specific genes and their relation to sporadic PD, notably *PABPC1* and *APH1A*, which both had moderate TPM expression values of ~30 and ~67, respectively. However, when the novelty search was conducted, neither gene was identified in any studies or research papers. In contrast, the *EIF4E* gene, which is closely linked to the *EIF4G1* gene associated with cases of monogenic familial PD, yielded more search results due to the closer proximity scores. Genes with the highest search results, such as *APP*, however, were well-studied, as referenced. Schulte *et al.* cited variants in the beta-amyloid precursor protein, expressed by the *APP* gene, that were more common in PD cases compared to their involvement in Alzheimer's Disease.[34] The former is a strong and established link.

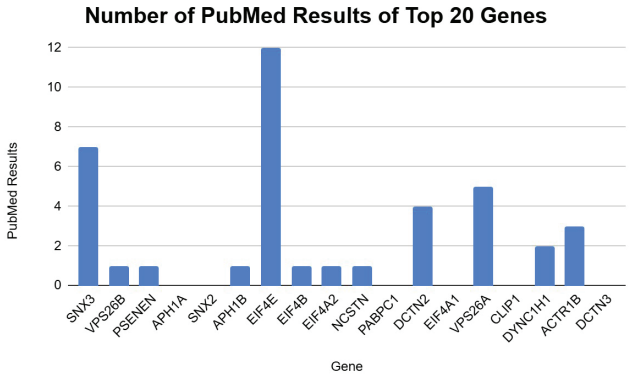

Figure 7: Results of PubMed Novelty Assessment, with every top 20 gene pictured, excluding the *APP* gene, which had 490 results. In contrast, genes such as the *EIF4A2* gene, with a proximity score difference of ~50 lower compared to *APP*, yielded one result. Furthermore, genes such as *PABPC1* yielded zero PubMed search results, despite their strong proximity scores. This figure highlights the potential for future research in this field, including the mechanisms of the identified genes.

Gene	PubMed Results
SNX3	7
VPS26B	1
PSENEN	1
APH1A	0
SNX2	0
APH1B	1
EIF4E	12
EIF4B	1
EIF4A2	1
NCSTN	1
PABPC1	0
DCTN2	4
EIF4A1	0
VPS26A	5
APP	490
CLIP1	0
DYNC1H1	2
ACTR1B	3
DCTN3	0

Figure 8: Summary table of PubMed Novelty Assessment, with every top 20 gene pictured in tabular format. Numerous genes potentially involved in underlying PD mechanisms are currently not adequately researched and require more attention from the scientific community.

■ Conclusion

Overall, this study aimed to identify twenty high-ranking novel idiopathic candidates by conducting a detailed approach that involved monogenic PD seeds as inputs. These candidates were identified by comparing protein expression interactions to form network proximity scores and filtered by SN expression to ensure relevance in disease mechanisms. Furthermore, the top three candidates were assessed for therapeutic and CRISPR feasibility via AAV9 viral vectors alongside other delivery methods. To further contextualize the top twenty genes, protein information data were obtained from the STRING

Database, followed by a PubMed novelty assessment of each gene in relation to PD.

There are limitations to this study's findings, however. For example, the study heavily relies on STRING v11 to construct the human PPI network, which integrates heterogeneous evidence, data extracted from multiple studies, leading to difficulties in direct comparison. This could have deflated or inflated proximity scores, which are also not a complete representation of whether a gene is involved in PD, but rather a measure of association. Furthermore, specific genomics of populations may have been overrepresented through data acquisition methods, such as the GWAS associations. The GTEx Expression data also come from healthy donors, rather than from PD-specific pathology. Most substantially, Dijkstra's algorithm does not account for complex biological pathways that may be involved in PD, in which computing the shortest distance may not be the most appropriate measure.

Regardless, this research, built upon prior studies, can be further developed, leading to potential for optimization and addressing the aforementioned limitations. The genes identified have a large scope for further research within PD itself. Experimentation with therapeutic strategies leveraging developing technologies such as CRISPR could increase the eight-year survival rate from 55.7%, affecting millions of Americans every day.[35]

Acknowledgments

I would like to express my sincere gratitude to Dr. Jobin Varkey and Professor Virgel Torremocha for their guidance and support throughout this project, whether through deepening my understanding of neuroscience, assisting with data acquisition, or providing invaluable feedback that strengthened the clarity and rigor of this research. I also thank Coach Jothsna Kethar for establishing and leading this outstanding research program through Gifted Gabber. Finally, I am deeply grateful to my family for their constant encouragement and unwavering support, without which this work would not have been possible.

References

1. Mayo Clinic. *Parkinson's Disease*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055.
2. *New Study Shows 1.2 Million People in the United States Estimated to be Living with Parkinson's Disease by 2030 | Parkinson's Foundation*. www.parkinson.org. https://www.parkinson.org/about-us/news/parkinsons-prevalence-project.
3. Surmeier, D. J. Determinants of Dopaminergic Neuron Loss in Parkinson's Disease. *The FEBS Journal* **2018**, *285* (19), 3657–3668. https://doi.org/10.1111/febs.14607.
4. Brichta, L.; Greengard, P. Molecular Determinants of Selective Dopaminergic Vulnerability in Parkinsonâ€TMS Disease: An Update. *Frontiers in Neuroanatomy* **2014**, *8*. https://doi.org/10.3389/fnana.2014.00152.
5. Tran, J.; Anastacio, H.; Bardy, C. Genetic Predispositions of Parkinson's Disease Revealed in Patient-Derived Brain Cells. *npj Parkinson's Disease* **2020**, *6* (1), 1–18. https://doi.org/10.1038/s41531-020-0110-8.
6. Wang, S.; Long, H.; Hou, L.; Feng, B.; Ma, Z.; Wang, Y.; Yu, Z.; Cai, J.; Zhang, D. W.; Zhao, G.-J. The Mitophagy Pathway and Its Implications in Human Diseases. *Signal Transduction and Targeted Therapy* **2023**, *8* (1). https://doi.org/10.1038/s41392-023-01503-7.
7. Dorsey, E. R.; Bloem, B. R. Parkinson's Disease Is Predominantly an Environmental Disease. *Journal of Parkinson's Disease* **2024**, *Preprint* (Preprint), 1–15. https://doi.org/10.3233/JPD-230357.
8. Dias, V.; Junn, E.; Mouradian, M. M. The Role of Oxidative Stress in Parkinson's Disease. *Journal of Parkinson's Disease* **2013**, *3* (4), 461–491. https://doi.org/10.3233/jpd-130230.
9. Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network Medicine: A Network-Based Approach to Human Disease. *Nature Reviews. Genetics* **2011**, *12* (1), 56–68. https://doi.org/10.1038/nrg2918.
10. Büttner, K.; Krieter, J. Comparison of Weighted and Unweighted Network Analysis in the Case of a Pig Trade Network in Northern Germany. *Preventive Veterinary Medicine* **2018**, *156*, 49–57. https://doi.org/10.1016/j.prevetmed.2018.05.008.
11. Wu, Z.; Hu, Z.; Gao, Y.; Xia, Y.; Zhang, X.; Jiang, Z. A Computational Approach Based on Weighted Gene Co-Expression Network Analysis for Biomarkers Analysis of Parkinson's Disease and Construction of Diagnostic Model. *Frontiers in Computational Neuroscience* **2023**, *16*. https://doi.org/10.3389/fncom.2022.1095676.
12. Dey, A.; Chakraborty, M.; Ujjwal Maulik; Bandyopadhyay, S. Network Based Approach for Drug Target Identification in Early Onset Parkinson's Disease. *Scientific Reports* **2025**, *15* (1). https://doi.org/10.1038/s41598-024-83178-w.
13. Termine, A.; Fabrizio, C.; Strafella, C.; Caputo, V.; Petrosini, L.; Caltagirone, C.; Cascella, R.; Giardina, E. A Hybrid Machine Learning and Network Analysis Approach Reveals Two Parkinson's Disease Subtypes from 115 RNA-Seq Post-Mortem Brain Samples. *International Journal of Molecular Sciences* **2022**, *23* (5), 2557. https://doi.org/10.3390/ijms23052557.
14. Dou, L.; Xu, Z.; Xu, J.; Zang, C.; Su, C.; Pieper, A. A.; Leverenz, J. B.; Wang, F.; Zhu, X.; Cummings, J.; Cheng, F. A Network-Based Systems Genetics Framework Identifies Pathobiology and Drug Repurposing in Parkinson's Disease. *npj Parkinson's Disease* **2025**, *11* (1). https://doi.org/10.1038/s41531-025-00870-y.
15. Gomes Moreira, D.; Jan, A. A Beginner's Guide into Curated Analyses of Open Access Datasets for Biomarker Discovery in Neurodegeneration. *Scientific Data* **2023**, *10* (1). https://doi.org/10.1038/s41597-023-02338-1.
16. Martirosyan, A.; Ansari, R.; Pestana, F.; Katja Hebestreit; Gasparyan, H.; Aleksanyan, R.; Hnatova, S.; Suresh Poovathingal; Marneffe, C.; Thal, D. R.; Kottick, A.; Hanson-Smith, V. J.; Guelfi, S.; Plumbly, W.; T. Grant Belgard; Emmanouil Metzakopian; Holt, M. G. Unravelling Cell Type-Specific Responses to Parkinson's Disease at Single Cell Resolution. *Molecular neurodegeneration* **2024**, *19* (1). https://doi.org/10.1186/s13024-023-00699-0.
17. Agarwal, D.; Sandor, C.; Volpato, V.; Caffrey, T. M.; Monzón-Sandoval, J.; Bowden, R.; Alegre-Abarrategui, J.; Wade-Martins, R.; Webber, C. A Single-Cell Atlas of the Human Substantia Nigra Reveals Cell-Specific Pathways Associated with Neurological Disorders. *Nature Communications* **2020**, *11* (1). https://doi.org/10.1038/s41467-020-17876-0.
18. CRISPR Therapeutics. *Gene Editing*. CRISPR Therapeutics. https://crisprtx.com/gene-editing.
19. Ntetsika, T.; Papathoma, P.-E.; Markaki, I. Novel Targeted Therapies for Parkinson's Disease. *Molecular Medicine* **2021**, *27* (1). https://doi.org/10.1186/s10020-021-00279-2.
20. Pinjala, P.; Tryphena, K. P.; Prasad, R.; Khatri, D. K.; Sun, W.; Singh, S. B.; Gugulothu, D.; Srivastava, S.; Vora, L. CRISPR/Cas9 Assisted Stem Cell Therapy in Parkinson's Disease. *Biomaterials research* **2023**, *27* (1), 46. https://doi.org/10.1186/s40824-023-00381-y.

DOI: 10.36838/IJHSR815.19

21. Yoon, H. H.; Ye, S.; Lim, S.; Jo, A.; Lee, H.; Hong, F.; Lee, S. E.; Oh, S.-J.; Kim, N.-R.; Kim, K.; Kim, B.-J.; Kim, H.; Lee, C. J.; Nam, M.-H.; Hur, J. W.; Jeon, S. R. CRISPR-Cas9 Gene Editing Protects from the A53T-SNCA Overexpression-Induced Pathology of Parkinson's Disease *in Vivo*. *The CRISPR Journal* **2022**, *5* (1), 95–108. https://doi.org/10.1089/crispr.2021.0025.
22. Nezar Abdennur; Fudenberg, G.; Flyamer, I. M.; Galitsyna, A. A.; Goloborodko, A.; Maxim Imakaev; Sergey Venev. Bioframe: Operations on Genomic Intervals in Pandas Dataframes. *Bioinformatics* **2024**. https://doi.org/10.1093/bioinformatics/btae088.
23. Zhou, Y.; Lauschke, V. M. Pharmacogenomic Network Analysis of the Gene-Drug Interaction Landscape Underlying Drug Disposition. *Computational and Structural Biotechnology Journal* **2020**, *18*, 52–58. https://doi.org/10.1016/j.csbj.2019.11.010.
24. *Gene Map Search - parkinsons disease autosomal - OMIM - (OMIM. ORG)*. Omim.org. https://www.omim.org/search?index=entry&search=parkinsons+disease+autosomal&start=1&limit=10&retrieve=geneMap&genemap_exists=true (accessed 2025-12-02).
25. Naren, P.; Cholkar, A.; Kamble, S.; Samim, K. S.; Srivastava, S.; Madan, J.; Mehra, N.; Tiwari, V.; Singh, S. B.; Khatri, D. K. Pathological and Therapeutic Advances in Parkinson's Disease: Mitochondria in the Interplay. *Journal of Alzheimer's Disease* **2022**, 1–30. https://doi.org/10.3233/jad-220682.
26. Harrer, P.; Matej Škorvánek; Volker Kittke; Dzinovic, I.; Friederike Borngräber; Thomsen, M.; Mandel, V.; Svorenova, T.; Ostrozovicova, M.; Kulcsarova, K.; Berutti, R.; Busch, H.; Ott, F.; Kopajtich, R.; Holger Prokisch; Kumar, K. R.; Mencacci, N. E.; Kurian, M. A.; Fonzo, A. D.; Boesch, S. Dystonia Linked to EIF4A2 Haploinsufficiency: A Disorder of Protein Translation Dysfunction. *Movement Disorders* **2023**, *38* (10), 1914–1924. https://doi.org/10.1002/mds.29562.
27. Shetty, A. S.; Bhatia, K. P.; Lang, A. E. Dystonia and Parkinson's Disease: What Is the Relationship? *Neurobiology of Disease* **2019**, *132*, 104462. https://doi.org/10.1016/j.nbd.2019.05.001.
28. Xue, C.; Gu, X.; Li, G.; Bao, Z.; Li, L. Expression and Functional Roles of Eukaryotic Initiation Factor 4A Family Proteins in Human Cancers. *Frontiers in Cell and Developmental Biology* **2021**, *9*. https://doi.org/10.3389/fcell.2021.711965.
29. Andrei Surguchov; Surguchev, A. A. Association between Parkinson's Disease and Cancer: New Findings and Possible Mediators. *International Journal of Molecular Sciences* **2024**, *25* (7), 3899–3899. https://doi.org/10.3390/ijms25073899.
30. Qi, Y.; Wang, M.; Jiang, Q. PABPC1——MRNA Stability, Protein Translation and Tumorigenesis. *Frontiers in Oncology* **2022**, *12*. https://doi.org/10.3389/fonc.2022.1025291.
31. Yuan, L.; Mao, L.-H.; Huang, Y.-Y.; Outeiro, T. F.; Li, W.; Vieira, T. C. R. G.; Li, J.-Y. Stress Granules: Emerging Players in Neurodegenerative Diseases. *Translational Neurodegeneration* **2025**, *14* (1). https://doi.org/10.1186/s40035-025-00482-9.
32. Chu, J.; Galicia-Vázquez, G.; Cencic, R.; Mills, J.; Katigbak, A.; Porco, J.; Pelletier, J. CRISPR-Mediated Drug-Target Validation Reveals Selective Pharmacological Inhibition of the RNA Helicase, EIF4A. *Cell Reports* **2016**, *15* (11), 2340–2347. https://doi.org/10.1016/j.celrep.2016.05.005.
33. Malerba, A.; Klein, P.; Bachtarzi, H.; Jarmin, S. A.; Cordova, G.; Ferry, A.; Strings, V.; Espinoza, M. P.; Mamchaoui, K.; Blumen, S. C.; St Guily, J. L.; Mouly, V.; Graham, M.; Butler-Browne, G.; Suhy, D. A.; Trollet, C.; Dickson, G. PABPN1 Gene Therapy for Oculopharyngeal Muscular Dystrophy. *Nature Communications* **2017**, *8*, 14848. https://doi.org/10.1038/ncomms14848.
34. Schulte, E. C.; Fukumori, A.; Mollenhauer, B.; Hor, H.; Arzberger, T.; Perneczky, R.; Kurz, A.; Diehl-Schmid, J.; Hüll, M.; Lichtner, P.; Eckstein, G.; Zimprich, A.; Haubenberger, D.; Pirker, W.; Brücke, T.; Bereznai, B.; Molnar, M. J.; Lorenzo-Betancor, O.; Pastor, P.; Peters, A. Rare Variants in β-Amyloid Precursor Protein (APP) and Parkinson's Disease. *European journal of human genetics: EJHG* **2015**, *23* (10), 1328–1333. https://doi.org/10.1038/ejhg.2014.300.
35. Joo, Y., Jo, Y.-I., & Park, J. H. Long-Term Survival Rates in Peritoneal Dialysis Initiated Immediately After Catheter Insertion: Comparative Study With Hemodialysis. *Journal of the American Society of Nephrology* **2022**, *33*(11S), 467–467. https://doi.org/10.1681/asn.20223311s1467a

Author

Anshuman Roy is a student at Bridgewater-Raritan High School, deeply interested in medicine and neuroscience, intersected with data analysis and computation. He hopes to major in neuroscience at university and pursue a career in medicine that heavily integrates research alongside patient care.