**Volume 11, Issue 03, March 2025,**
**Publish Date: 01-03-2025**
**Doi https://doi.org/10.55640/ijmsdh-11-03-01**

International Journal of Medical Science and Dental Health (Open Access)

# IN SILICO STUDY OF DELETERIOUS SINGLE NUCLEOTIDE POLYMORPHISMS IN HUMAN HTRA2 GENE WITH THE PARKINSON'S DISEASE

**OMAR QAHTAN YASEEN**

*Department of Heet Education, General Directorate of Education in Anbar, Ministry of Education, Hit, Anbar 31007, Iraq.*

## ABSTRACT

One of the most prevalent neurodegenerative illnesses that affects the elderly and is incurable is Parkinson's disease. Effective remedies for these illnesses are sometimes expensive, time-consuming, and coincidental to find. Prior research comparing several model organisms has shown that the majority of animals have comparable cellular and molecular traits. The Bioinformatics tools were used to find non-synonymous single nucleotide polymorphisms SIFTand Mutpred. The HTRA2 protein, a protein linked to Parkinson's disease, was used to predict nsSNPs.

This study is based on the analysis of single nucleotide polymorphisms in the human HTRA2 gene that cause Parkinson's disease. By determining the most effective prediagnosis techniques and Biological markers for illness prediction.

Our goal is to identify the most effective prediagnosis method for Parkinson's disease. A specific SNP is to serve as a molecular marker for the illness's prediagnosis. The analysis of non-synonymous SNPs (nsSNPs) in several Bioinformatics methods may facilitate the identification of critical diagnoses for the disease and enhance the treatment process for Parkinson's disease (PD).

## INTRODUCTION

Neurological illnesses known as neurodegenerative diseases (NDs) are brought on by a progressive loss of brain function brought on by slow neuronal death. [1] They primarily affect the elderly population and are incurable. Their term "incurability" alludes to both the late diagnosis, when most signs manifest in the latter steps of the disease, and brain death, which is the primary cause of these illnesses. Globally, the incidence of age-related neurodegenerative illnesses is rising as people age. Parkinson's disease (PD) was the most well-known [2]. Its symptoms develop slowly and ultimately result in death since they are incurable. They have a detrimental social and economic impact on patients' and their families' quality of life[3].

Most, like Parkinson's disease (PD), are the consequence of a hereditary and environmental cocktail. In comparison to vertebrate models, insects are recommended for usage as study animals due to their ease of handling, small size, ease of raising locations, ease of rearing costs, small life cycles, great fecundity, ease of gene manipulation, and lack of ethical permissions[4,5]. Mutations are the main source

of dysfunctional gene behavior and are ultimately accountable for the onset of diseases. Numerous mutations that cause illness have been found in the genome; around 0.5 million of them are SNPs.[19]

In other words, one base gets swapped out for another base. These mutations can include single nucleotide variations (SNPs) that are synonymous or nonsynonymous, or SNPs that are located in intergenic regions, non-coding areas, or coding sequences of genes[20]. SNPs are important and raise the risk of developing a variety of illnesses. In coding areas, synonymous SNPs (sSNPs) had no influence on translated proteins[21]. They may, however, also have an impact on translation rate and mRNA stability. Amino acid alterations brought on by nonsynonymous SNPs (nsSNPs) directly affect the structure and function of proteins. SNPs in non-coding areas may have an impact on transcription[22], RNA degradation, and gene splicing, among other biological processes.

The consequences of mutations on the structure and function of proteins are predicted using computational techniques. They are crucial for SNV analysis and prioritizing them for characterization through experimentation. Computational methods, such as the SIFT tool [23,24], can discover highly harmful variants by aligning them with known pathogenic mutations using a sequence homology algorithm. Other computational methods, such as Mutpred, INPS, and I-Mutant 2.0 programs, categorize the nsSNVs into sick or neutral substitutions using artificial neural networks and support vector machines. [25, 26]. Consensus-based methods use a combination of (SIFT, Mutpred) [27] tools to determine the pathogenicity of nsSNPs as a bioinformatics tool.

It takes money and effort to establish a gene-disease causal relationship. As a result,, thorough prioritization of potential SNPs and identification of the most accurate model to replicate the illness prior to experimental testing significantly lower related costs, save time, and speed up the drug development process.

Our goal is to identify the most effective prediagnosis method for Parkinson's disease, even if a specific SNP is chosen to serve as a molecular marker for the illness's prediagnosis. The anticipated non-synonymous SNPs (nsSNPs) in several bioinformatics methods may facilitate the identification of single nucleotide polymorphisms and enhance the process of drug discovery for Parkinson's disease (PD).

## MATERIALS AND METHODS

### *Data Collection*

The protein sequences of the Parkinson's disease were retrieved from UniProt sequence database (www.uniprot.org). SNPs of the Parkinson's disease proteins were obtained from UniProt database. Finally, two of Bioinformatics tools were useful for defining hurtful SNPs in HTRA2 gene for predicting the structure of protein.

### *SNP Analysis*

SNPs retrieved from UniProt database were subjected for analysis using

***Table 1*** *(In silico approaches available as online tools.)*

| Algorithm | web |
|---|---|
| **SIFT** | http://sift.bii.a-star.edu.sg/ |
| **Mutpred** | http://mutpred.mutdb.org/about.html |

### SIFT

SIFT is an online program that uses sequence homology to forecast how amino acid alterations may affect protein function. It aligns certain sequences, looks for proteins that are relevant to the query, and scores each nsSNP to show if it has a harmful or tolerable effect [18,19].

### MutPred

Mutpred predicts protein structure and function changes due to non-synonymous sequence variants (nsSNPs), categorizing them as neutral or disease-associated. It also predicts the molecular etiology of AAS linked to illness, providing likelihood scores for each variation. (Li et al., 2009a).

## RESULTS

SNPs analysis of a cell signaling pathway is summarized and shown in Table 2. It helps to identify whether the mutated position is harmful/disease causing or neutral.

***Table 2*** *(List of nsSNP analysis by SIFT in this study.)*

| Gene | Nucleotide change | Amino acid change | SIFT prediction | SIFT score |
|---|---|---|---|---|
| 1 | C>T | P4S | T | 0.09 |
| 2 | G>T | G9V | APF | 0.00 |
| 3 | G>C | R15P | APF | 0.00 |
| 4 | G>C | G21A | T | 0.78 |
| 5 | G>A | G22S | APF | 0.00 |
| 6 | G>A | G26E | T | 0.40 |
| 7 | C>T | R36W | APF | 0.00 |
| 8 | G>A | R36Q | APF | 0.00 |
| 9 | G>T | A37S | APF | 0.00 |
| 10 | T>C | L38P | APF | 0.00 |
| 11 | C>A | L39M | APF | 0.00 |
| 12 | C>G | S44C | APF | 0.01 |
| 13 | C>T | P46S | T | 0.64 |
| 14 | C>G | P46R | APF | 0.02 |
| 15 | G>C | R49P | T | 0.16 |
| 16 | G>A | V50M | T | 0.43 |
| 17 | G>A | G53R | APF | 0.00 |
| 18 | G>A | S56N | APF | 0.00 |
| 19 | G>C | W58S | T | 0.79 |
| 20 | G>T | W58C | T | 0.08 |
| 21 | C>T | R60W | APF | 0.01 |
| 22 | G>A | V63I | APF | 0.03 |
| 23 | A>G | E67G | T | 0.07 |
| 24 | C>T | P68S | T | 0.35 |
| 25 | T>C | L72P | APF | 0.03 |
| 26 | C>A | P79T | T | 0.10 |
| 27 | G>T | R80L | APF | 0.02 |
| 28 | G>A | A81T | APF | 0.00 |

| | | | | |
|---|---|---|---|---|
| 29 | C>T | T84I | APF | 0.02 |
| 30 | C>T | A85V | T | 0.11 |
| 31 | G>A | G100R | T | 0.86 |

**T= Tolerated **APF= Affect Protein Function*

P4S, G21A, R49P, and other "Tolerated" changes have higher SIFT scores (more than 0.05), suggesting that they are probably not harmful to the protein. G9V, R15P, G22S, and other changes categorized as "Affect Protein Function" have lower SIFT scores (0.00-0.05), suggesting that they are likely to negatively impact protein function. Acceptable (non-effectual) Higher SIFT scores suggest that an amino acid change may be tolerable without substantially impacting protein function, while other mutations are categorized as non-affective to protein function. SIFT ratings for these changes are usually higher (more than 0.05). Impact the Function of Proteins The majority of gene alterations are categorized as "affecting protein function" and frequently have SIFT scores that are extremely low (around 0.00). G9V, R15P, G22S, R36W, and other mutations that are categorized as "Affect Protein Function" and have a SIFT score of 0.00 may suggest that they are alterations that substantially impact the protein's function and may be linked to genetic illnesses or health issues.

Even while some mutations could seem dangerous based on SIFT, they might not have a big impact on the biological system as a whole. This is because the effects of these mutations may be influenced by additional elements in the protein, such as genetic background or gene interactions. Maximum SIFT score: The G100R mutation (with a SIFT score of 0.86), which is categorized as having "no effect" on the protein's function, is unlikely to be detrimental. SIFT score that is lowest Numerous mutations, including G9V, R15P, G22S, G36W, and others, have extremely low SIFT scores (0.00), indicating that they will have a major impact on the protein's functionality. G21A and G26E mutations yield moderate SIFT scores (0.40-0.78). Although they may not have as much of an impact on the protein as mutations with a very low SIFT score, these alterations may nevertheless have some impact on the protein's function. The protein may continue to function normally if tolerated mutations show that the amino acid change has no discernible effect on the protein's tertiary structure or interactions with other molecules. Mutations that affect a protein's structure or biological function have the potential to cause biological abnormalities or hereditary illnesses. If a mutation affects important regions of the protein (such as binding or active sites), it may be connected to a particular disease.

***Table 3*** *(nsSNP analysis list using the MutPred2 tool.)*

| No. | Nucleotide change | Amino acid change | MutPred2 score | MutPred2 prediction |
|---|---|---|---|---|
| 1 | C>T | P4S | 0.504 | Pathogenicity |
| 2 | G>T | G9V | 0.719 | Pathogenicity |
| 3 | G>C | R15P | 0.561 | Pathogenicity |
| 4 | G>C | G21A | 0.278 | neutral |
| 5 | G>A | G22S | 0.47 | neutral |
| 6 | G>A | G26E | 0.446 | neutral |
| 7 | C>T | R36W | 0.356 | neutral |
| 8 | G>A | R36Q | 0.243 | neutral |
| 9 | G>T | A37S | 0.094 | neutral |
| 10 | T>C | L38P | 0.857 | Pathogenicity |
| 11 | C>A | L39M | 0.16 | neutral |

| | | | | |
|---|---|---|---|---|
| 12 | C>G | S44C | 0.279 | neutral |
| 13 | C>T | P46S | 0.152 | neutral |
| 14 | C>G | P46R | 0.293 | neutral |
| 15 | G>C | R49P | 0.468 | neutral |
| 16 | G>A | V50M | 0.082 | neutral |
| 17 | G>A | G53R | 0.478 | neutral |
| 18 | G>A | S56N | 0.119 | neutral |
| 19 | G>C | W58S | 0.256 | neutral |
| 20 | G>T | W58C | 0.269 | neutral |
| 21 | C>T | R60W | 0.473 | neutral |
| 22 | G>A | V63I | 0.067 | neutral |
| 23 | A>G | E67G | 0.28 | neutral |
| 24 | C>T | P68S | 0.08 | neutral |
| 25 | T>C | L72P | 0.444 | neutral |
| 26 | C>A | P79T | 0.131 | neutral |
| 27 | G>T | R80L | 0.162 | neutral |
| 28 | G>A | A81T | 0.075 | neutral |
| 29 | C>T | T84I | 0.09 | neutral |
| 30 | C>T | A85V | 0.111 | neutral |
| 31 | G>A | G100R | 0.57 | Pathogenicity |

## DISCUSSION

The use of molecular methods to determine (SNP) can be expensive and difficult at the same time (Chen and Sullivan, 2003), so computational studies (Bioinformatics studies) will work to determine the number of mutations and their association in molecular studies, which helps in better understanding the structural and functional aspects of the protein. Preceding studies in the field of bioinformatics, specifically in the field of prediction, have facilitated many methods of splicing polymorphisms (nsSNPs) and their functional association in proteins such as G6PD (Rajith, 2011), ATM (Doss and Rajith, 2012), PTEN (Doss and Rajith, 2013), and BRAF (Hussain et al., 2012). In our current study, the results we obtained helped in using bioinformatics tools to determine the extent of the relationship between nsSNPs. Previous results of researchers who used bioinformatics tools showed the efficiency of the work and the accuracy of the results. Among these studies, two bioinformatics tools (SIFT and PolyPhen-2) were used to predict the future relationship between nsSNPs and the extent of their impact on the normal function of the protein. Therefore, we chose only two types of bioinformatics tools for the purpose of comparing them in predicting the extent of the effect of nsSNPs on the structure and function of the protein responsible for the disease condition. These tools are SIFT and Mutpred2 to examine nsSNPs in the HTRA2 gene with Parkinson's disease. Finally, we used bioinformatics tools (prediction) to find out the extent of the effect of harmful nsSNPs on the normal function of the protein. Studying harmful nsSNPs at the structural and sequence levels will give high value in the prediction results. The results showed a wide spectrum of variations in the study gene, and bioinformatics tools or algorithms were used, which showed that there is a high probability of Parkinson's disease occurring as a result of the presence of these harmful variations. However, it was noted that some variations showed tolerance in not changing the normal function of the protein due to the chemical and physical similarity of amino acids and even the closeness in structure, which gave normal functions to the

protein, but in general, our study showed that harmful polymorphisms are likely and greatly have a role in having a functional effect on the HTRA2 gene and causing Parkinson's disease.

## CONCLUSIONS

A combination of effector and non-effector mutations may exist, and their impact on the protein may differ. Additionally, it's critical to keep in mind that SIFT ratings are merely estimates based on available data. To fully comprehend the effects of these mutations, you might want additional testing, such as biological models or functional investigations. These results indicate that our approach successfully allowed us in selecting the deleterious SNPs that are likely to have functional impact on the HTRA2 gene and contribute to an individual's susceptibility to the disease.

## REFERENCES

1. Checkoway, H., Lundin, J. I. & Kelada, S. N. Neurodegenerative diseases. IARC Sci. Publ. 163, 407–419 (2011).
2. Ali, A. M. & Kunugi, H. Royal jelly as an intelligent anti-aging agent—A focus on cognitive aging and Alzheimer's disease: A review. Antioxidants 9(10), 1–46. https://doi.org/10.3390/antiox9100937 (2020).
3. Chekani, F., Bali, V. & Aparasu, R. R. Quality of life of patients with Parkinson's disease and neurodegenerative dementia: A nationally representative study. Res. Soc. Adm. Pharm. 12(4), 604–613. https://doi.org/10.1016/j.sapharm.2015.09.007 (2016).
4. Denell, R. Establishment of tribolium as a genetic model system and its early contributions to evo-devo. Genetics 180(4), 1779–1786. https://doi.org/10.1534/genetics.104.98673 (2008).
5. Bingsohn, L., Knorr, E. & Vilcinskas, A. Te model beetle Tribolium castaneum can be used as an early warning system for transgenerational epigenetic side efects caused by pharmaceuticals. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 185, 57–64 (2016).
6. Jia, M. et al. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 gene. PLoS ONE 9(8), e104311 (2014).
7. Mooney, S. D., Krishnan, V. G. & Evani, U. S. Bioinformatic tools for identifying disease gene and SNP candidates. Methods Mol. Biol. 628, 307–319. https://doi.org/10.1007/978-1-60327-367-1_17 (2010).
8. Bromberg, Y. Chapter 15: Disease gene prioritization. In PLoS Computational Biology. https://doi.org/10.1371/journal.pcbi.10029 02 (2013).
9. Tey, H. J. & Ng, C. H. Computational analysis of functional SNPs in Alzheimer's disease-associated endocytosis genes. PeerJ https:// doi.org/10.7717/peerj.7667 (2019).
10. Sim, N. L. et al. SIFT web server: Predicting efects of amino acid substitutions on proteins. Nucleic Acids Res. 40(W1), W452–W457 (2012).
11. Tomas, P. D. et al. PANTHER: A library of protein families and subfamilies indexed by function. Genome Res. 13(9), 2129–2141 (2003).
12. Bromberg, Y. & Rost, B. SNAP: Predict efect of non-synonymous polymorphisms on function. Nucleic Acids Res. 35(11), 3823–3835 (2007).
13. Capriotti, E. & Fariselli, P. PhD-SNPg: A webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Res. 45(W1), W247–W252 (2017).

14. Tahseen, T. H. (2019). The impact of the educational method using the training method in some physical variables of the muscles of the limbs and the strength of the transmissions in the game of tennis. University of Anbar Sport and Physical Education Science Journal, 4(18).28. Hoeppner, M. A. NCBI Bookshelf: Books and documents in life sciences and health care. Nucleic Acids Res. 41(D1), D1251–D1260 (2012).
15. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1), 27–30. https://doi.org/10. 1093/nar/28.1.27 (2000).
16. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. Publ. Protein Soc. 28(11), 1947– 1951. https://doi.org/10.1002/pro.3715 (2019).
17. Qahtan, O. (2023). Sorting Intolerant from Tolerant and PolyPhen-2 Algorithms: A Variation in Exon 14 of ATP7B Gene among 4 West Iraqi Families with Wilson's Disease. Al-Anbar Medical Journal, 19(2), 98-103.
18. Yaseen, O. Q., Al-Ani, M. Q., & Majeed, Y. H. (2020). In-Silico prediction of impact on protein function caused by non-synonymous single nucleotide polymorphism in human ATP7B gene associated with Wilson disease. Research Journal of Biotechnology Vol, 15, 3.
19. Constantinescu, A. E., Hughes, D. A., Bull, C. J., Fleming, K., Mitchell, R. E., Zheng, J., ... & Vincent, E. E. (2024). A genome-wide association study of neutrophil count in individuals associated to an African continental ancestry group facilitates studies of malaria pathogenesis. Human Genomics, 18(1), 26.
20. Aspatwar, A., Tolvanen, M. E., Barker, H., Syrjänen, L., Valanne, S., Purmonen, S., ... & Parkkila, S. (2022). Carbonic anhydrases in metazoan model organisms: molecules, mechanisms, and physiology. Physiological reviews, 102(3), 1327-1383.
21. Al-Ayari, E. A., Shehata, M. G., El-Hadidi, M., & Shaalan, M. G. (2023). In silico SNP prediction of selected protein orthologues in insect models for Alzheimer's, Parkinson's, and Huntington's diseases. Scientific Reports, 13(1), 18986.
22. Hussein, A. F., Khamees, H. H., Mohammed, A. A., Hussein, S. A. M., Ahmed, M. A., Saad, A., & Raoof, M. In-Silico Study of Destabilizing Alzheimer's Aβ42 Protofibrils With Curcumin.