

GENOMIC DATA ANALYTICS IN PARKINSON'S DISEASE

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology Degree in Bioinformatics

by

Rakesh I (Register No: 37220014)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MARCH 2021



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119
www.sathyabama.ac.in



DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of Mr. **Rakesh I (Reg. No. 37220014)** who carried out the project entitled "**GENOMIC DATA ANALYTICS IN PARKINSON'S DISEASE**" under my supervision from September 2020 to March 2021.

Internal Guide
(Dr. Alex Anand)

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **Rakesh I (37220014)** hereby declare that the project report entitled “**GENOMIC DATA ANALYTICS IN PARKINSON’S DISEASE**” was done by me under the guidance of **Dr. Alex Anand, Associate Professor, Department of Bioinformatics, Sathyabama Institute of Science and Technology** is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Bioinformatics.

DATE :

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to Dr. **T.Sasikala M.E., Ph.D, Dean**, School of Computing , Dr.S.Vigneshwari M.E., Ph.D. and Dr.L.Lakshmanan M.E., Ph.D. , Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide Dr.Alex Anand, Associate professor, Department of Bioinformatics for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Computer Science and Engineering who were helpful in many ways for the completion of the project.

RAKESH. I

ABSTRACT

Parkinson's is a progressive, tremor-characterized, neurodegenerative disease. Parkinson's being second most affecting neuro-related disease after Alzheimer's and most adults develop the disease at around 60 years of age, but some have an early-onset of disease and forms of Parkinson's are inherited where certain forms have been related to unique gene mutations that can be genetic or environment associated but are not understood completely. This study concentrates on existing data from various repositories to find and predict SNP level severe damages in genes of Parkinson's disease especially in 3 of the important genes SNCA, LRRK2, PARK7. SNPs of the retrieved genes were analyzed by in-silico tools like SIFT, SNAP, iMutant suite, MetaSNP, Mutation Taster; Algorithms like CADD, DANN and various functions to determine the changes in structural and functional aspects. Sorting out only for harming and deleterious SNPs in all the 3 genes finally landed with 11 variants in PARK7, 4 variants in SNCA and 5 variants in LRRK2 which actually has an impact on functional level from initial set of 1000+ variant counts. Homology modeling using Swiss Model and validation using PROCHECK was done and further visualization for structure related studies were performed in HOPE server. Further, Stability related studies were done using all the data from various computational tools from which variants like R156, P73 in PARK7; E123K in SNCA and R1628, L1795 in LRRK2 were found to be damaging to great extent. Further extensive analysis required onto these variants and variants A107P in PARK7, K96I in SNCA and N2261L in LRRK2 had a positive impact on the protein stability which also requires separate study on how mutation increases the stability of diseased protein.

Keywords: Parkinson's Disease, SNPs in Parkinson, Mutations in Parkinson's, Early onset of Parkinson.

CONTENTS

CHAPTER NO	TITLE	PAGE NO
	LIST OF FIGURES	v
1	INTRODUCTION	
	1.1 BASIC INFORMATION	1
	1.2 SIGNS, SYMPTOMS & TREATMENT	2
	1.3 GENES INVOLVED	2
2	LITERATURE SURVEY	
	2.1 INTRODUCTION	3
	2.2 DJ-1/ PARK7	3
	2.3 SNCA	5
	2.4 LRRK2	6
	2.5 SNPs	8
	2.6 EOPD	8
3	AIM AND SCOPE	
	3.1 AIM AND OBJECTIVE	
	3.1.1 AIM OF THE PROJECT	11
	3.1.2 OBJECTIVE OF THE PROJECT	11
	3.2 SCOPE OF THE PROJECT	11
4	MATERIALS AND METHODS	
	4.1 DATA COLLECTION	12
	4.2 EVALUATING SNP EFFECTS	12
	4.2.1 CADD	12
	4.2.2 DANN	13
	4.2.3 FATHMM	13
	4.2.4 FUNSEQ2	14
	4.3 SECONDARY LEVEL STUDIES FOR AMINO	

	ACID VARIANTS	
	4.3.1 SNAP	14
	4.3.2 SIFT	15
	4.3.3 SNPdbe	16
	4.4 TOOLS FOR FILTERING AND FINDING	
	PRIME VARIANTS	
	4.4.1 MirSNPs	16
	4.4.2 MutationTaster2	17
	4.4.3 iMutant Suite	17
	4.4.4 PROVEAN	18
	4.4.5 SNPs & GO	18
	4.4.6 MetaSNP	19
	4.4.7 PANTHER	19
	4.4.8 PhD-SNP	20
	4.5 HOMOLOGY MODELING AND STRUCTURAL	20
	ANALYSIS	
	4.5.1 SWISS MODEL	20
	4.5.2 PROCHECK (UCLA SAVES V6.0)	20
	4.6 VISUALIZATION AND STABILITY CHANGES	
	4.6.1 DISCOVERY STUDIO	21
	4.6.2 HOPE SERVER	21
5	RESULTS AND DISCUSSIONS	
	5.1 ANALYSIS OF PARK7	23
	5.2 ANALYSIS OF SNCA	35
	5.3 ANALYSIS OF LRRK2	44
	5.4 DISCUSSION	
	5.4.1 PARK7	53
	5.4.2 SNCA	54
	5.4.3 LRRK2	55

6	SUMMARY AND CONCLUSION	57
	REFERENCES	58

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
4.1	Overall Workflow of Structured Protocol	22
5.1	Initial Variants from DbSNP- PARK7	24
5.2	PredictSNP2 summary for all algorithms	24
5.3	Results from PredictSNP2 to filter for exonic and non synonymous mutations	25
5.4	Top filtered variants in initial gene level variants of PARK7	25
5.5	Gene level mutations haven't been reported in 1000G.	25
5.6	Initial variants- amino acid level variants of PARK7	26
5.7	Variants filtered based on SNAP and SIFT scores	26
5.8	iMutant suite results for filtered variants of PARK7	27
5.9	PROVEAN results for filtered variants of PARK7	27
5.10	MetaSNP, SNPs & GO results of PARK7	28
5.11	HOPE SERVER VISUALZIATION RESULTS of PARK7	29
5.12	Initial gene variants of SNCA	35
5.13	Filtered variants in gene level of SNCA	36
5.14	Final gene level variants	36
5.15	SNCA amino acid level variants	37
5.16	SNCA filtered from aa level variants based on SIFT & SNAP	38
5.17	SNCA iMutant & Provean result	38
5.18	MetaSNP results of SNCA	39
5.19	Predictsnp1 results of SNCA	40
5.20	HOPE SERVER VISUALZIATION RESULTS of SNCA	40
5.21	Initial DbSNP variants of LRRK2	45
5.22	Filtered using PredictSNP2 for synonymous and exonic Mutationtaster2 result and sorting: Finally 9 filtered of	46
5.23	LRRK2	46
5.24	Initial aa level variants after sorting based on SIFT and	47

	SNAP	
5.25	iMutant Suite results of LRRK2	48
5.26	PROVEAN results for LRRK2	48
5.27	MetaSNP and final prime variants results for LRRK2	49
5.28	HOPE SERVER VISUALZIATION RESULTS of LRRK2	50

CHAPTER 1

INTRODUCTION

1.1 BASIC INFORMATION

Parkinson's disease (PD) is a neurodegenerative disease that primarily affects the substantia nigra's dopamine-producing neurons. In most cases, signs manifest progressively over time. Because of its diversity of the condition, the progression of symptoms varies a little from person to person. The cause is still unknown. Dopamine is a neurotransmitter that aids in movement. This loss of dopamine is the reason treatments aim to boost dopamine levels in the brain. Although there is no cure, there are a variety of treatment options available, including medications and surgery. While Parkinson's disease is not fatal in and of itself, it can cause serious complications. Parkinson's disease has been seen in some families for several generations, suggesting that some forms of the disease are hereditary. Many researchers believe that genetic factors play a role in Parkinson's disease. Many scientists believe that genetic factors play a part in the progression of Parkinson's disease. Parkinson's disease causes both motor and non-motor symptoms such as depression, constipation, and deterioration of taste and smell. The progression of idiopathic Parkinson's disease is slow and variable. The Hoehn and Yahr scale is frequently used by doctors to track the progression of a disease over time. The scale was first used in 1967, and it included stages zero through five, with zero indicating no signs of Parkinson's disease and five indicating advanced PD. Researchers believe that alphasynuclein build-up plays a role in the development of Parkinson's disease, and that new treatments could be developed based on this theory or on genetic level medications. Medications that increase or mimic the effects of dopamine in the brain help to reduce movement symptoms. Levodopa is a commonly prescribed drug. Many people find that levodopa is less effective at controlling their symptoms after few years. Medications as surgical type, such as Deep brain stimulation use a small device like pacemaker which when implanted below the skin of the chest to make some electrical signals to a part of the brain that controls movement via very thin wires.

1.2 SIGNS, SYMPTOMS & TREATMENT

Parkinson's disease (PD), characterized by four cardinal signs: postural instability, bradykinesia, rigidity and tremor. Other non motor symptoms include: trouble sleeping in comfortness and timings, bladder problems, change in sex drive, increased saliva production, fluctuation in weight, visual disturbances, fatigue, cognitive issues, such as trouble remembering things or frequent confusion with Motor symptoms being a part of diagnosis and are also a major symptom for detection for patients of all ages. These can include: resting tremor, or constant shaking movement even when relaxed, rigid muscles & balance problems all over. Certain genes are associated with early onset Parkinson's but entire genetics haven't been analyzed or known yet. Environmental causes of the condition may include exposure to chemical toxins such as certain insecticides, fungicides, and herbicides as reported by various researchers but there is still no proper analysis results yet on the causes and diagnosis. There is no single test to detect Parkinson's. A diagnosis may be difficult and take a while. Blood tests and other imaging tests, such as an MRI scan, don't diagnose Parkinson's. A DaTscan to visualize your brain's dopamine levels and it will help confirm diagnosis. However, they may be used to rule out other conditions. Researchers have reported that many changes happen in the brains of patients with Parkinson's disease, although it's still unclear why these changes happen in patients. These changes like higher amount of lewy bodies and Clumps of substances within brain cells are microscopic markers of PD. Researchers believe these Lewy bodies holds an important parameter in causing Parkinson's disease. Alpha-synuclein is found in enormous amount inside the Lewy bodies.

1.3 GENES INVOLVED

So concentrating on most reported genes LRRK2, SNCA & PARK7 might provide some deep insights into SNPs affecting stability & functions in population. Various Bioinformatics Methods / Tools and publicly available repositories & data are used to explore various nsSNPs / Mutations which might provide better insights into genetics based Early Onset Parkinson's disease and aid for future association studies.

CHAPTER 2

LITERATURE SURVEY:

2.1 INTRODUCTION

Sara Bandres-Ciga, Monica Diez-Fairen, Jonggeol Jeff Kim, Andrew B. Singleton (2020) reported that in last few decades, PD genetics has evolved a lot, from the reports and research questioning the Parkinson Disease genetics role to an enhanced integration and understanding on its impact on PD probability, initiation, response and growth to treatment, as well as presence of casual comorbidities and phenotypic etiologies make it difficult to detect the genetics of Parkinson disease. But only a fraction of the heritability is known and the relation between genetics and PD pathology is barely understood due to complications in detecting variants. The nascent initiative to expand this study at scale and to promote ethnic interest in PD genetics research would not only eradicate structural health gaps but also facilitate exploration of the missing pieces of PD genetics.

2.2 DJ-1/ PARK7:

Masanori Hijioka , Masatoshi Inden , Daijiro Yanagisawa , Yoshihisa Kitamura (2017) reported that three pathogenic variant mutations of DJ-1 (L166P, D149A and M26I) have been reported to disrupt stability & form improper structures and moreover lead to a loss of function. The administration of wild-type DJ-1 protein represses the neurological disruption associated with Parkinson's disease in rats. Moreover, in studies focusing DJ-1 as the causing stage in PD, components that have the capacity of binding to DJ-1 and research regarding has been made previously to find more information about the binding specifically at C 106 residue and it has been confirmed that it does exert some therapeutic effects and also has a link to disorders such as Parkinson's disease and Alzheimer's disease.

Ludmila P Dolgacheva , Alexey V Berezhnov, Evgeniya I Fedotova, Valery P Zinchenko, Andrey Y Abramov(2019) reported that DJ-1 protein has many numerous roles to protect dopaminergic neurons against neurodegeneration in Parkinson's

disease. Wild type PARK7 can act as an antioxidant and also as an oxidative stress sensor.

Ricardo Taipa, Conceição Pereira, Inês Reis, Isabel Alonso, António Bastos-Lima, Manuel Melo-Pires, Marina Magalhães(2016) reported that DJ-1 is a largely expressed protein that participates in the protection of cells over oxidative stress from various pathways. Recent work shows that DJ-1 can directly bind to α -synuclein mono- and oligomers, successfully reducing dimerization and reversing α -synuclein dependent cellular toxicity.

Eng-King Tan , Lisa M Skipper (2007) reported that the frequency of these gene mutations will vary according to origin, background and to the genetic variant and target population. A gene dose impact is seen in certain cases, and the phenotype of some of the mutation carriers closely fits standard PD. Penetrance of each of the repetitive mutations is insufficient and may vary with age.

Sapna Pandey(2019) reported that Mutation of two genes, DJ-1 and retromer complex factor gene are linked with disorders like Parkinson and Huntington's disease. DJ-1 gene located at 1p36 chromosomal position and has been implicated in Parkinson Disease pathogenesis by many pathways and also been reported by many researchers having its effect in including mitochondrial impairment and oxidative damage. DJ-1 D149A mutant complexes become more compact and shows the structural disruption and loss of protein function which in turn probably leading to a cause for PD affected with no context to onset type.

Darren J Moore , Li Zhang, Ted M Dawson, Valina L Dawson (2003) reported that the crystal structure of human PARK7 shows that it occurs in a homo-dimeric shape, but the L166P variant only exists as form of monomer. The L166P mutation confers highly reduced protein stability in PARK7 as been reported by various researchers, due to the rapid degradation of 20S/26S proteasome. The L166P mutant variant shows an harmful / deleterious motive to interaction studies and also to structure homo-oligomers. Compared to that, one of the most reported variants M26I mutation did not emerge to severely affect the protein stability, or even the capacity of DJ-1 to

form homo-oligomers. These properties of the L166P mutation which lead to the loss of DJ-1 function and are likely to be the underlying cause of early onset PD as a report says in patients of the Italian kindred.

Mariaelena Repici, Kornelis R. Straatman, Nadia Balduccio, Francisco J. Enguita, Tiago F. Outeiro, and Flaviano Giorgini (2018) reported that the E64D mutant variant of DJ-1 holds the potential to create homodimers. However, though wild-type DJ-1 dimers are balanced under oxidative stress conditions, we find that the E64D mutation inhibits this stabilization. The widely studied L166P variant mutated protein is briefly able to interrupt formation and also doesn't form wild-type homodimers.

2.3 SNCA:

Suzanne Lesage, Alexis Brice (2009) reported that Point mutations in the α -synuclein (SNCA) gene cause a rare dominant form of Parkinson Disease in irregular and inherited cases. Mutations in the leucine-rich repeat kinase 2 (LRRK2) gene contribute to more of a general cause of autosomal dominant PD. In addition, susceptibility variants that contribute to PD have been identified in several populations, such as polymorphisms in the SNCA, LRRK2 genes and heterozygous mutations in the β -glucocerebrosidase (GBA) gene.

Nevra Alkanli, Arzu Ay(2018) reported that there is a correlation between the likelihood of PD development and SNCA levels, and it is suggested that SNCA level is a significant predictor in PD diagnosis. Various polymorphisms have been established in the 5' and/or 3' UTR regions of the SNCA gene, and as a result of these polymorphisms, differences occur in the binding of transcription factors. The identification of the roles of SNCA gene polymorphisms in PD development may allow the development of new methods for the treatment of PD.

Todd L Edwards(2010) reported that SNPs in SNCA (rs2736990, p-value = 6.7×10^{-8} ; genome-wide adjusted p = 0.0109, odds ratio (OR) = 1.29. G vs. A allele, population attributable risk percent (PAR percent) = 12 percent) were genome-wide important. No other SNPs were genome-wide essential in this analysis. This study

confirms that SNCA is a significant gene whose usual variant are influencing risk of PD.

2.4 LRRK2:

Fernando Cardona , Marta Tormos-Pérez, Jordi Pérez-Tur (2014) reported that the LRRK2 gene (Leucine-Rich Repeat Kinase 2) is defective in a significant number of reports of autosomal dominant Parkinson's disease instances till now and in some rare cases of LOPD. LRRK2 is a huge, complicated protein that comprises multiple interaction domains: armadillo, ankyrin, leucine-rich repeats and WD40 domains; two catalytic domains: ROC-GTPase and serine/threonine kinase; and a COR domain (unknown function).

Currently, the evaluation of PD is based on professional criteria, aided neuroimaging as MRI or PET. Understanding the pathogenesis of the EOPD may be contributing to improving diagnostics and effectiveness of pharmacotherapy. Further studies on gene level variants may help in diagnosing primary research.

João Paulo Lima Daher (2017) reported that Mutations in the α -synuclein and LRRK2-encoding genes have been linked with traditional results in clinical and pathologic PD. LRRK2 tends to play a major role in the pathogenesis of PD as it is associated with α -synuclein pathology and other proteins involved in neurodegeneration. Thus, LRRK2 dysfunction can effect or have an impact on the accumulation of α -synuclein and its pathology through a lot of pathomechanisms altering signaling pathways and cellular functions including autophagy, immune system, retromer complex modulation and vesicle trafficking. A report suggests that abnormal α -synuclein accumulation can be tackled with novel LRRK2 inhibitors.

Suganya Selvaraj, Shanmughavel Piramanayagam (2019) reported that Genetic researches in PD have contributed to the recognition of several monogenic variants of the condition and of many genetic hazard factors raising the potential to experience the neuron degeneration. Under such circumstances, molecular examination is the most widely recommended method for individuals to detect the condition by molecular testing. Pedigree and population analyses reported several risk genes and loci were

linked to dopamine deficiency. During the past decade, few genes have been discovered that are significant in autosomal recessive and autosomal dominant type of PD. Whole genome linkage screening to identify chromosomal regions related to the likelihood of PD or the time of PD beginning. While SNCA mutation variants have been reported to start PD for lot of years, the process by which these variants contribute to disease remains poorly known. It is assumed that abnormal accumulation of the protein contributes to cell destruction and eventually neuronal death. However, more study is required to understand how mutations and multiplication in SNCA happens in Parkinsonism. In certain Parkinson's patients with LRRK2 mutated variant gene, have been detected with rare neuropathology in which it had involved diffuse Lewy bodies unlike most normal reports of over production, in the brain stem, front temporal lobar degeneration with intranuclear incorporations & pathological accumulation in neurofibrillary tangles. Moreover, LRRK2 gene was highly expressed in immune cells. Recent research shows that, in early stages of life, the extended LRRK2 action may ensure against opportunistic pathogenic infection however it can increase the chances of activating Parkinson's Disease. DJ-1 can be useful as a biomarker to neurodegenerative disorder. Comparatively DJ-1 induced PD tends to be very rare, there are not enough trials to say about the patients if they got diagnosed with PD owing to the genotype based mutation of the DJ-1 gene.

Justus C. Dächsel, Matthew J. Farrer(2010) reported that Pathogenic Lrrk2 mutations p.R1441C/G/H (Roc), p.Y1699C (COR), p.I2012T, p.G2019S, and p.I2020T (MAPK) and risk factors p.R1628P (COR) and p.G2385R (WD40) affect the formation of Lrrk2 in different ways. It is also reported that it has an effect on operational and functional impacts on LRRK2. The chances of disease conferred by Lrrk2 p.G2019S is almost 10-fold greater than conferred by p.G2385R. Nevertheless, most LRRK2 mutations eventually lead to the same phenotype and theoretically by the same pathogenic mechanism, whereby 1 mutant Lrrk2 molecule is all it takes to change the properties of the entire complex, leading to a dominant effect. The pathogenic LRRK2 mutations occurrences in patients with Parkinson show the importance of a genetic predisposition in this seemingly oddly neurodegenerative disorder. Nevertheless, Parkinson Disease is also considered a multi factorial disorder as it is caused by a

variety of environmental, genetic and stochastic factors, for patients which age remains a major factor. Disease incidence may have been more correctly measured for Lrrk2 p.G2019S but remains a hypothesis rather than a certainty. These findings affirm the association of LRRK2 with neurodegeneration, and identify a common mutation correlated with dominantly inherited Parkinson's disease.

2.5 SNPs:

A single nucleotide polymorphism usually short formed to SNP is a modification of any single nucleotide that happens at any given location in the genome. Each and every variant is present to some amount of degree within a population. (Miller RDet al., 2005). For eg, In the genome, A nucleotide may appear in lot of base position in majority of population but certain set of people have C in the place of A forming Minority, meaning presence of SNP in that location. C and A are two possible allele for this specific role in that location. SNPs result in a broad variety of diseases (Altshuler et al., 2000). A single base mutation in the APOE (apolipoprotein E) gene is correlated with a lower risk for alzheimer's disease. Single nucleotide polymorphisms might occur inside coding sequences of genes, non coding sections of genes. Often even seen in the regions of non coding parts. The SNPs inside a coding sequence do not actually alter the amino acid sequence of the protein owing to degeneracy of genetic code (Delport et al., 2008). The SNPs in the coding area are of two forms. They are synonymous and nonsynonymous SNPs. Synonymous SNPs are those which do not alter the protein sequence, whereas nonsynonymous SNPs modify the sequence of protein. The nonsynonymous SNPs again are of two forms : missense and nonsense.

2.6 EOPD:

Roy N Alcalay, Cheryl Waters (2010) reported that EOPD is known to be a complex disorder, which suggests that it is linked with the influence of several genes in association with lifestyle and environmental causes. Standard autosomal dominant and autosomal recessive genetic concepts may not completely occur in EOPD. Most autosomal dominant mutation carriers may not develop the disorder, indicating

inadequate penetrance (e.g., LRRK2 and potentially synuclein), though some autosomal recessive genes may significantly raise the risk of EOPD even though only one copy of the mutated gene is inherited. In 1997, α -synuclein (SNCA) was described as the first gene linked to EOPD. Synuclein is a significant component of Lewy bodies, which is a pathological characteristic in lewy body accumulation of PD and EOPD. Gain-of-function variants, duplications, and triplications have all been discovered in EOPD patients, with a long family background. However, there are just a few cases reported in the literature, and penetrance estimates differ. In one study, just 33% of duplication carriers in a single family developed PD. Carriers with gain-of-function mutations and copy number variance (CNV) in the genome, unfortunately, are at risk of developing dementia. According to one study, gene dosage can influence the intensity of the EOPD phenotype. Triplication carriers were more likely than replication carriers to develop EOPD and dementia, although duplication carriers were more likely to develop late-onset PD. Single-nucleotide polymorphisms (SNPs) in the promoter region of SNCA have been implicated in the pathogenesis of PD through genome-wide interaction studies (GWAS). SNPs associated with EOPD are still unknown to large extent. The correlation between mutations in DJ-1 and Early Onset of Parkinson's Disease was discovered in 2003 in two European families. Mutations in DJ-1 are less frequent and are very uncommon in Asian patients with PD. DJ-1 mutations has been reported in a study of genotyped 89 individuals of American region, a single heterozygote was found. DJ-1-associated EOPD is characterized as being close to PRKN65 and indistinguishable from idiopathic EOPD. One research found normal olfaction in DJ-1 carriers close to PRKN carriers, but the amount of mutation carriers was low. The relationship between LRRK2 mutations and PD was first discovered in 2004. In comparison to the other PARK genes, the occurrence of LRRK2 mutations is almost equivalent in EOPD and late-onset PD. The most widely recorded LRRK2 mutation, G2019S (nt. 2877510G>A), is present in up to 1 percent of spontaneous and 4 percent of hereditary PD cases worldwide. However, up to 39 percent of Northern African Arab and 18.3 percent of Ashkenazi Jewish individuals with PD carry the mutation. The I2020T mutation reported in Japanese patients and the R1441C, Y1699C in Caucasians, have been identified in that study.

The expression of LRRK2-related PD is heterogeneous, although it is normally consistent with positive response to treatment with dopamine agonists and levodopa and can be complicated further by dyskinesia. Thirty-four LRRK2 G2019S carriers and their analysis suggest that it might have an impact on onset of EOPD events. Disease characteristics between the groups were comparable, except that LRRK2 carriers were less prone to suffering from tremor and were more likely to exhibit the postural instability gait difficulty motor phenotype (PIGD). There were no other physiological variations between carriers and non-carriers, including other manifestations of PIGD (e.g. poorer motor function and cognitive impairment) (e.g. worse motor performance and cognitive impairment). Gan-Or and others have observed that carriers with G2019S mutations were more prone to emerge with gait deficiency compared with glucocerebrosidase (GBA) mutation carriers. Studies with LRRK2 carriers (not limited to EOPD cases) documented impaired olfaction. There were different results on anatomy, but most cases showed Lewy body pathology. In summary, though not only limited to early onset of PD, the LRRK2 mutation is among the most prominent genetic mutation that can be found in EOPD cases with and without a family history of PD. Fortunately, the same medications and other therapeutic strategies helpful in older patients may be particularly successful in young-onset PD. Those making use of drugs has to pay attention to the potential for long-term consequences such as dyskinesias. Due to this, many physicians suggest that levodopa usage should either be limited or that a dopaminergic agonist be used along with it. Young-onset PD has an increased chance to be a genetic condition.

Jack Kelly , Rana Moyeed , Camille Carroll , Diego Albani , Xinzhong Li (2019) reported that PD and AD share substantial common DEGs and pathways, and recognizes novel genes, upstream regulators and pathways which may be effective targets for therapy in both diseases.

CHAPTER 3

AIM & SCOPE:

3.1 AIM AND OBJECTIVE

3.1.1 AIM

Analyzing SNCA, LRRK2 & PARK7(All Parkinson associated) genes with various Bioinformatics Methods / Tools and publicly available repositories & data to explore various nsSNPs / Mutations which might provide better insights into genetics based Onset of Parkinson's disease and aid for future association studies.

3.1.2 OBJECTIVE

- To identify gene mutations associated with Parkinson's disease
- To find the mutations, structural and functional impact on the protein using various tools and databases.
- To perform Visualization and stability changes related studies

3.2 SCOPE

Parkinson's disease is second most affecting neurodegenerative disease after Alzheimers and its origin is tied with environmental & genetic factors in most cases. Many research have reported that many of the reported cases have genetic basis and especially in the early onset, genetics plays very important role. So analyzing the genes associated with parkinson's and its genetics to scrutinize for damage causing variants might help in studies related to analysis of disease even as early in diagnosis stage and also for other researchers. These data can also be helped in building for a disease prediction analysis and further can also be linked to mechanism level in various instances.

CHAPTER 4

METHODOLOGY

4.1 DATA COLLECTION:

Information regarding all the variants have been retrieved initially from Genome Viewer dbSNP. Structures from PDB has been used to refer with homology structured. Amino acid based variants are retrieved from dbSNP, SWISS-VAR, SWISS PROT, PMD etc...

4.2 EVALUATIONG SNP EFFECTS

4.2. PredictSNP2:

"A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions"

Methods have numeric ratings but not binary predictions of the deleteriousness of variations or belief scores that may be more readily interpreted. three databases covering various forms of disease-related variations, which were split into five categories: (i) regulatory, (ii) synonymous, (iii) missense, (iv) nonsense variants, and (v) splicing. These datasets were used to create category-optimal decision with a characteristic threshold and to test six methods for prioritizing the variants: CADD, GWAVA, FunSeq2, FATHMM, FitCons and DANN. The findings obtained for the five best-performing tools were then incorporated into a consensus ranking. Additional comparative analysis found that in the case of missense variants, protein-based predictors do better than DNA sequence-based predictors. The tool was used to find the effects on SNPs in various algorithms and a meta score to filter out if they are damaging or not. The web server is publicly open to the community at <http://loschmidt.chemi.muni.cz/predictsnp2>.

4.2.1 CADD:

CADD is a method for "scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome." Although several variant

annotation and rating methods are around, most annotations appear to manipulate a particular knowledge category (e.g. conservation) and/or are limited in nature (e.g. to missense changes) (e.g. to missense changes). Thus, a widely applicable metric that objectively weights and incorporates diverse knowledge is required. Combined Annotation Dependent Depletion (CADD) is a system that incorporates several annotations into one metric through comparing variants that escaped natural selection with simulated mutations. CADD can quantitatively prioritize deleterious, functional, and illness causal variations through a broad variety of functional types, impact genetic architectures and sizes and can be used prioritize causal heterogeneity across both scientific and clinical settings.

4.2.2 DANN:

Combined annotation-dependent depletion (CADD) is an “algorithm built to annotate both coding and non-coding versions, and has been shown to outperform other annotation algorithms. CADD trains a linear kernel support vector machine (SVM) to distinguish evolutionarily derived, likely benign, alleles from simulated, likely deleterious, variants. DANN uses the same feature set and training data as CADD to train a deep neural network (DNN)”. DNNs can catch non-linear interactions among features and are better suited than SVMs for with a large number of samples and features. DANN manages about a 19 percent relative reduction in the error rate and about a 14 percent relative boost in the region under the curve (AUC) metric over CADD’s SVM technique.

4.2.3 FATHMM:

Functional Analysis by Hidden Markov Models. A high-throughput web-server capable of forecasting the practical effects of all coding variants, i.e. non-synonymous single nucleotide variants (nsSNVs), and non-coding variants of the human genome. “A species-independent approach with optional species-specific weightings for the

estimation of the physiological effects of protein missense variants. Using a model weighted for human mutations”, It further obtained accuracy accuracies that outperformed standard prediction approaches. FATHMM may be easily extended to high-throughput/large-scale human and nonhuman genome sequencing programs with the additional advantage of phenotypic result associations. Unweighted / species-independent approach interrogates sequence conservation by the basic amino acid probabilities matched by the internal match states of many HMMs describing the arrangement of homologous sequences and conserved protein domains. <http://fathmm.biocompute.org.uk>.

4.2.4 FunSeq2:

Identification of noncoding drivers from thousands of somatic modifications in a normal tumor is a complex and unresolved task. FunSeq2, to annotate and prioritize these mutations. The architecture integrates an adjustable data background combining large-scale genomics and cancer tools with a simplified variant-prioritization pipeline. “The pipeline has a weighted scoring scheme combining: inter- and intra-species conservation; loss- and gain-of-function events for transcription-factor binding; enhancer-gene linkages and network centrality; and per-element recurrence through samples. Regulatory mutations may trigger transcriptional alterations through either loss-of or gain-of- function effects”. Loss-of-function events in the motifs of transcription factor binding are likely to cause effects which might cause deleterious mutants. Variants lowering the position weight matrix (PWM) scores may theoretically modify the binding power of transcription factors, or even abolish the binding. The structure consists of a module to detect motif-breaking events - identified as variants decreasing PWMs. It is mixture of FERP, 1000 Genomes, and several other data sources.

4.3 SECONDARY LEVEL STUDIES FOR AMINO ACID VARIANTS

4.3.1 SNAP:

The analysis of genome-wide association findings is startled by linkage disequilibrium between alleles around. “A scalable bioinformatics query platform for single-

nucleotide polymorphisms (SNPs) to classify and to annotate nearby SNPs in linkage disequilibrium (proxies) dependent on HapMap". By providing functionality to produce graphical plots for these statistics, the SNAP server can enable analysis and comparison of genome-wide interaction research findings, and the design of fine-mapping studies (by delineating genomic regions related variants and their proxy). SNAP is freely accessible at <http://www.broad.mit.edu/mpg/snap>, along with documents. For each question SNP, SNAP returns all proxy SNPs (after applying filters), annotated by physical and genetic location, recombination rate, r^2 , D' and neighboring genes. Tool was used to filter out SNPs in proxies and damaging SNPs. The server may also produce relationship plots and graphical plots of proxies for a specific SNP, or for a pair of SNPs.

4.3.2 SIFT:

"SIFT predicts whether an amino acid replacement influences protein function depending on sequence homology and the physical properties of amino acids". SIFT may be extended to spontaneously occurring nsSNP polymorphisms and insilico based induced missense mutations in laboratories. An nsSNP could theoretically influence the role of the protein, ultimately altering the carrier's phenotype. This protocol outlines the use of the 'Sorting Tolerant From Intolerant (SIFT) algorithm in predicting when an AAS affects protein work. To determine the impact of a substitution, SIFT assumes that essential position in a protein sequence have been conserved in evolution and so the substitutions at these positions on sequence can influence protein function. Thus, by utilizing sequence homology, SIFT estimates the results of all potential substitutions at - spot in the protein sequence. Reference SNP ids (rs ids) may be used as feedback. Each variant In the study was corporate to find the properties and changes for the replacement in our variants. <https://sift.bii.a-star.edu.sg>.

4.3.3 SNPdbe:

An nsSNP practical impacts archive. Each non-synonymous SNP (nsSNP) switches one amino acid in the gene product (single amino acid substitution). This transition will either impact protein function or be neutral in that regard. Most polymorphisms neglect experimental annotation of their functional effect. SNPdbe-SNP database of results, with predictions of computationally annotated practical impacts of SNPs. Database entries reflect nsSNPs in dbSNP and 1000 Genomes collection, as well as variants from UniProt and PMD. SAASs come from >2600 organisms; 'human' is the most prevalent. The effect of each SAAS on protein function is predicted using the SIFT and SNAP algorithms and supplemented with experimentally derived function/structure knowledge and disease associations from PMD, OMIM and UniProt. SNPdbe is regularly modified and conveniently supplemented with new sources of intelligence. SNPdbe accessible at <http://www.rostlab.org/services/snpdbe>.

4.4 TOOLS FOR FILTERING AND FINDING PRIME VARIANTS

4.4.1 MirSNPs:

“A database of polymorphisms altering miRNA target sites, detects miRNA-related SNPs in GWAS SNPs and eQTLs”. Numerous single nucleotide polymorphisms (SNPs) correlated with a lot of diseases have been reported through genome-wide interaction studies (GWAS) and expression quantitative feature loci (eQTLs) studies. However, few of these SNPs have explicit biochemical roles. Recent studies suggested that the SNPs within the 3'UTR regions of susceptibility genes could cause complex traits/diseases by influencing the role of miRNAs and these are the targets of mirSNPs. MirSNP (<http://cmbi.bjmu.edu.cn/mirsnp>), which is a list of human SNPs in predicted miRNA-mRNA binding sites of over 414,510+ SNPs that could impact miRNA-mRNA binding sites. Annotations were applied to these SNPs to predict whether an SNP inside the target site will decrease/break or enhance/create a miRNA-mRNA binding site. MirSNP may classify the putative miRNA-related SNPs from GWAS and eQTLs researches and provide the way for subsequent practical researches.

4.4.2 MutationTaster2:

Mutation prediction for the deep-sequencing age. evaluates the pathogenic potential of DNA sequence alterations. It is designed to predict the functional consequences of not only amino acid substitutions but also intronic and synonymous alterations, short insertion and/or deletion (indel) mutations and variants covering intron-exon boundaries. MutationTaster2 contains all freely accessible single-nucleotide polymorphisms (SNPs) and indels from the 1000 Genomes Project² (hereafter referred to as 1000G) as well as recognized disease variants from ClinVar³ and HGMD Public. MutationTaster2 dynamically maps the variant to all suitable genes and transcripts, analyzes the variant in all of them and shows a table summarizing the estimates for all transcripts and specific data for each transcript. , MutationTaster2 uses a Bayes classifier to make predictions. the software automatically identifies and categorizes confirmed polymorphisms and known disease mutations. In a real world example using exome info, the application yielded a false positive rate of 1 percent for homozygous alterations. This tool can be used to filter out disease causing and predict disease variants and also info if it is available on 1000G. <https://www.mutationtaster.org>

4.4.3 I-Mutant Suite:

I-Mutant is a suite of Support Vector Machine dependent predictors implemented in an unusual web server. It provides the potential to predict automatically protein stability changes upon single-site mutations beginning from protein sequence alone or protein structure when available. Moreover it offers the possibility to determine human deleterious Single Nucleotide Polymorphism beginning from the protein sequence alone. Tool can be used to find out about stability of the structure. <http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>

4.4.4 PROVEAN:

“A tool to predict the functional effect of amino acid substitutions and indels. A web server to predict the functional effect of single or multiple amino acid substitutions, insertions and deletions using the prediction tool PROVEAN”. The server provides rapid analysis of protein variants from any organisms, and also supports high-throughput analysis for human and mouse variants at both the genomic and protein levels. The web server is freely available and open to all users with no login requirements at <http://provean.jcvi.org>. The tool provides online access to the function of the stand-alone PROVEAN software package distribution. Its primary function is to provide a prediction for a protein sequence originated from any organisms. The tool accepts amino acid variations and in protein sequence as input, performs a BLAST search to identify homologous sequences (supporting sequences), and generates final PROVEAN scores. As a general binary cutoff predictions for normal analysis, the cutoff for PROVEAN scores was set to -2.5 for high balanced accuracy. However, users can reapply their own cutoffs for their analysis to achieve either higher sensitivity or higher specificity especially when analyzing novel organisms. Used to find if the structure is deleterious and to what extent it is. <http://provean.jcvi.org/index.php>

4.4.5 SNPs & GO:

Predicting disease associated variations using GO terms. The most investigated SNPs are missense mutations resulting in residue substitutions in the protein. “SNPs&GO collect in an unique framework information derived from protein sequence, 3D structure, protein sequence profile, and protein function”. The vast number of human SNPs available in the data bases poses the question of relating protein mutant variants in databases to diseases. “Server uses different sources of information, including that from the Gene Ontology annotation to predict if a given variation can be classified disease-related or neutral”. SNPs&GO was trained on a set of 38460 mutations and tested with a cross-validation procedure over sets in which

similar proteins were kept in the same dataset also for the calculation of the LGO score, as derived from the GO data base. The performance is also increasing, At increasing input degree of complexity, indicating that on top of sequence profile also LGO, derived from the protein GO annotation, is a key added benefit for distinguishing disease-related polymorphisms from neutral ones.

4.4.6 MetaSNPs:

Meta-predictor of disease causing variants. the number of human genetic variants stored into the publicly accessible databases has been growing exponentially. The new edition of dbSNP, for example, includes ~50 million validated Single Nucleotide Variants (SNVs) (SNVs). SNVs make up much of human diversity and are also the leading causes of disorder. The non-synonymous SNVs (nsSNVs) result in single nucleotide or amino acid substitutions and can influence protein structure, sometimes causing disease. While several methods for the identification of nsSNV effects have already been developed, the consistent growth in annotated data is offering the potential to improve prediction accuracy. A novel methodology for the identification of disease-associated nsSNVs (Meta-SNP) that incorporates four known methods: PANTHER, SIFT, PhD-SNP, and SNAP. Meta analysis of all algorithms to find out if it is disease causing and its prediction. <https://snps.biofold.org/meta-snp/>

4.4.7 PANTHER:

EVOLUTIONARY ANALYSIS OF CODING SNPS. “Estimates the likelihood of a particular nonsynonymous (amino-acid changing) coding SNP to cause a functional impact on the protein”. It calculates the length of time (in millions of years) a given amino acid has been preserved in the lineage leading to the protein of interest. The longer the preservation time, the greater the likelihood of functional impact.

4.4.8 PhD-SNP:

Predictor of human Deleterious Single Nucleotide Polymorphisms, PhD-SNP is based a SVM-based classifier. PhD-SNP is optimized to predict if a given single point protein mutation can be classified as disease-related or as neutral polymorphism and requires inputs as protein sequence, Position, Altered residue, Prediction method, Multi SVM model prediction.

4.5 HOMOLOGY MODELING AND STRUCTURAL ANALYSIS

4.5.1 SWISS MODEL:

Homology modeling of protein structures and complexes. Highly autonomous workflows and servers simplify and standardize the homology modelling process, also enabling users without a specialized computational background to produce accurate protein models and have convenient access to modelling data, their visualization and analysis. Functionality has been applied to the modeling of homo- and heteromeric structures. Starting from the amino acid residues of the protein complexes, both the stoichiometry and the general composition of the system are inferred by homology modeling. SWISS-MODEL is freely available at <https://swissmodel.expasy.org>.

4.6.2 PROCHECK in SAVES 6.0:

The PROCHECK suite of programs includes a quantitative check on the stereochemistry of a protein structure. Its outputs include a variety of plots in PostScript format and a detailed residue-by-residue listing. These provide an assessment of the overall consistency of the structure as contrasted with well defined frameworks of the same specification and also illustrate areas that may require further investigation. The PROCHECK programs are useful for evaluating the accuracy not only of protein complexes in the course of being resolved but also of existing systems and of those having modelled on known functions and properties. The most important parameter which is being tested in this work is for the homology model protein structure regions and whether they are correct based on the percentage of residues in all regions. <https://saves.mbi.ucla.edu>

4.6 VISUALIZATION AND STABILITY CHANGES

4.6.1 DISCOVERY STUDIO:

Molecular visualization is a key aspect of the analysis and communication of modeling studies. A graphics visualization tool for viewing, sharing, and analyzing protein and modeling data BIOVIA Discovery Studio Visualizer was used for interactive 3D visualization when HOPE server was unavailable due to no perfect model and in case of models generated with in silico tools for analysis.

4.6.2 HOPE SERVER:

“HOPE is an easy-to-use web service that analyses the structural effects of a point mutation in a protein sequence. The process starts by Input protein sequence and the mutation details and HOPE will collect and combine available information from a series of web services and databases and will produce a report, complete with results, figures and animations”. HOPE goes through a complex process to analyse the mutant of interest and has results from BLAST, UniProt, Reprof, Decision Tree and its report with homology modeling & information regarding 3D structure and stability and changes in amino acid state. Project HOPE has been tested by running a number of previously analyzed cases. HOPE's results has been compared with the results of our manual analysis and with the results of the well-know PolyPhen and SIFT methods.

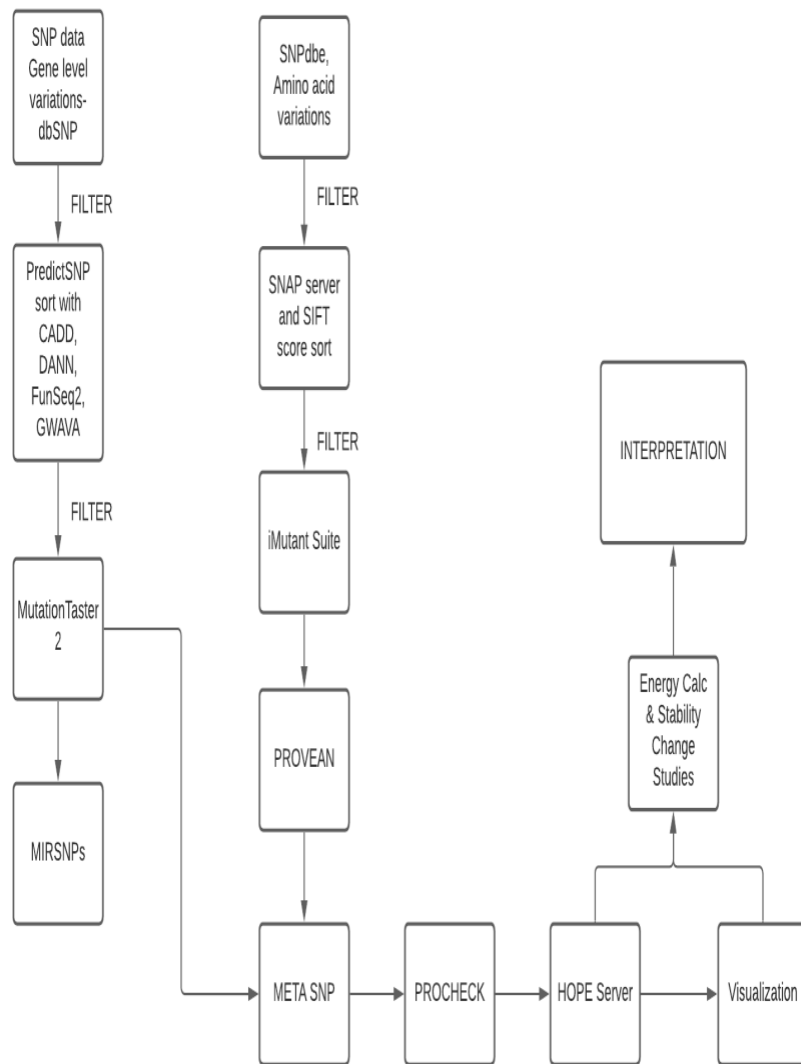


Fig. 4.1 Overall Workflow

5. RESULTS & DISCUSSION:

5.1 PARK7:

dna level INITIAL VARIANTS: 63 variants taken genome viewer & aa level: 41 variants from snpdb

63 variants from gene variants were loaded into predictsnp2. Tools used in predictsnp2 are CADD, DANN, FunSeq2, GWAVA. Filtered further to 13 gene variants with deleterious, harming, and been analyzed insilico further.

Each entry is posted into mutationtaster. 3 OF THEM(7962844,7969370, 7984932 NOT REPORTED IN 1000G) Noted. Then, MIRSNP analysis to find for miRNA-mRNA binding sites were analyzed. SNPs and GO Tool – Meta SNP tool used on 3 prime leads, found probability, then mixed with prime variants from aminoacid level(b) line.

From snpdb, AA initial 41 variants been filtered to 15 variants which are probably effecting, further filtered based on SNAP and SIFT scores. Imutant Suite analysis done for all 15 variants and then provean analysis for all 15 variants. 3 were neutral whereas all other were deleterious. 12 prime variants were analyzed in snps & go tool to find from all 3 tools PhD-SNP, PANTHER, SIFT, SNAP combined analysis. Based on MetaSNP results, 8 of them have been taken.

Combining 8 from (b) and 3 from (a) gene level variants, All have been modeled in SWISS model, studied for PROCHECK analysis. All 11 were in favored regions for about 90+ percent of residues. Then analyzed in HOPE server to study conservation of mutant and wild-type, domain region changes, amino acid changes, stability and amino acid characteristics as a comparison between wild-type and mutant with images of 3D structure visualizing the structural changes and its effects.

##fileformat=VCFv4.3									
##source=dbSNP									
##sequence_region=NC_000001.11 7958141 7989073"									
##INFO=<ID=RS.Number=1,Type=String,Description="dbSNP ID">									
##INFO=<ID=Source.Number=1,Type=String,Description="Variation source">									
#CHROM	POS	ID	REF	ALT	QUAL	FILTER		INFO	
NC_000001	7951680	rs175238	G	A	.	.		RS=rs17523802;Source=dbsnp	
NC_000001	7951690	rs557801	G	C	.	.		RS=rs557801400;Source=dbsnp	
NC_000001	7951718	rs226245	T	C,G	.	.		RS=rs226249;Source=dbsnp	
NC_000001	7951737	rs140230	T	A	.	.		RS=rs140230911;Source=dbsnp	
NC_000001	7951747	rs1112106	C	T	.	.		RS=rs11121064;Source=dbsnp	
NC_000001	7951787	rs886046	G	A,C	.	.		RS=rs886046544;Source=dbsnp	
NC_000001	7962764	rs115489	C	G,T	.	.		RS=rs11548933;Source=dbsnp	
NC_000001	7962813	rs130987	C	A,G	.	.		RS=rs1309873819;Source=dbsnp	
NC_000001	7962844	rs370430	T	C	.	.		RS=rs370430693;Source=dbsnp	
NC_000001	7962852	rs757247	G	A	.	.		RS=rs757247182;Source=dbsnp	
NC_000001	7962858	rs781346	G	A	.	.		RS=rs781346135;Source=dbsnp	
NC_000001	7962863	rs743153	G	A	.	.		RS=rs74315351;Source=dbsnp	
NC_000001	7965320	rs756040	G	A,C	.	.		RS=rs756040385;Source=dbsnp	
NC_000001	7965334	rs772272	C	T	.	.		RS=rs772272696;Source=dbsnp	
NC_000001	7965348	rs137853	G	A,T	.	.		RS=rs137853051;Source=dbsnp	
NC_000001	7965399	rs1146015	G	A	.	.		RS=rs114601558;Source=dbsnp	
NC_000001	7965425	rs743153	G	C	.	.		RS=rs74315353;Source=dbsnp	
NC_000001	7969370	rs367584	C	T	.	.		RS=rs367584305;Source=dbsnp	
NC_000001	7969375	rs886046	G	A	.	.		RS=rs886046545;Source=dbsnp	
NC_000001	7969386	rs115489	C	T	.	.		RS=rs11548937;Source=dbsnp	
NC_000001	7970934	rs716536	G	A	.	.		RS=rs71653619;Source=dbsnp	
NC_000001	7970951	rs774005	G	A,T	.	.		RS=rs774005786;Source=dbsnp	
NC_000001	7977638	rs728548	A	G	.	.		RS=rs72854882;Source=dbsnp	
NC_000001	7977728	rs398124	G	C	.	.		RS=rs398124657;Source=dbsnp	
NC_000001	7984885	rs143947	A	G	.	.		RS=rs143947415;Source=dbsnp	
NC_000001	7984930	rs743153	A	C	.	.		RS=rs74315352;Source=dbsnp	
NC_000001	7984932	rs368420	G	A	.	.		RS=rs368420490;Source=dbsnp	
NC_000001	7984971	rs743153	G	A	.	.		RS=rs74315354;Source=dbsnp	
NC_000001	7984981	rs289381	T	C	.	.		RS=rs28938172;Source=dbsnp	
NC_000001	7984984	rs886046	C	G,T	.	.		RS=rs886046546;Source=dbsnp	
NC_000001	7984985	rs716536	A	G	.	.		RS=rs71653621;Source=dbsnp	
NC_000001	7984995	rs777026	G	T	.	.		RS=rs777026628;Source=dbsnp	
NC_000001	7985179	rs779587	G	A,C	.	.		RS=rs779587926;Source=dbsnp	
NC_000001	7961680	rs175238	G	A	.	.		RS=rs17523802;Source=dbsnp	
NC_000001	7961913	rs356756	G	C,T	.	.		RS=rs35675666;Source=dbsnp	
NC_000001	7962137	rs376660	G	T	.	.		RS=rs3766606;Source=dbsnp	
NC_000001	7962863	rs743153	G	A	.	.		RS=rs74315351;Source=dbsnp	
NC_000001	7965215	rs751735	C	T	.	.		RS=rs7517357;Source=dbsnp	
NC_000001	7965348	rs137853	G	A,T	.	.		RS=rs137853051;Source=dbsnp	
NC_000001	7965425	rs743153	G	C	.	.		RS=rs74315353;Source=dbsnp	
NC_000001	7969434	rs2641116	T	A,C,G	.	.		RS=rs2641116;Source=dbsnp	
NC_000001	7969450	rs2641117	G	A,C,T	.	.		RS=rs2641117;Source=dbsnp	
NC_000001	7970796	rs670367	G	A	.	.		RS=rs6703670;Source=dbsnp	
NC_000001	7984930	rs743153	A	C	.	.		RS=rs74315352;Source=dbsnp	
NC_000001	7984971	rs743153	G	A	.	.		RS=rs74315354;Source=dbsnp	
NC_000001	7984981	rs289381	T	C	.	.		RS=rs28938172;Source=dbsnp	
NC_000001	7984985	rs716536	A	G	.	.		RS=rs71653621;Source=dbsnp	
NC_000001	7985019	rs716536	G	A	.	.		RS=rs71653622;Source=dbsnp	
NC_000001	7986612	rs127276	C	A	.	.		RS=rs12727642;Source=dbsnp	

Fig. 5.1 Initial Variants from DbSNP- PARK7



PredictSNP overview

Date: 2021/02/19

Genome version: GRCh38/hg38

PredictSNP version: 2.1

If you use PredictSNP and obtain scientific results that you publish, we would ask you to acknowledge the usage of PredictSNP by referencing the article below.

Bendl, J., Musil, M., Stourac, J., Zendulka, J., Damborsky, J., Brezovsky, J., 2016: PredictSNP2: A unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLOS Computational Biology* 12: e1004962

	PredictSNP2		CADD		DANN		FATHMM		FunSeq2		GWAVA	
	Neu	Del	Neu	Del	Neu	Del	Neu	Del	Neu	Del	Neu	Del
Regulatory	22	6	21	7	21	7	19	0	16	12	20	9
Splicing	1	0	1	0	1	0	0	1	11	6	5	12
Missense	10	7	8	9	2	15	1	9	0	1	1	0
Synonymous	1	0	1	0	1	0	0	0	0	0	0	0
Nonsense	0	0	6	10	10	6	8	8	6	10	6	10
Others	0	12	0	0	0	0	0	0	0	0	0	0

Fig. 5.2 PredictSNP2 summary for all algorithms

Fig. 5.3. Results from PredictSNP2 to filter for exonic and non synonymous mutations

Fig. 5.4. Top filtered variants in initial gene level variants of PARK7

Mutation Taster Result	
7962844	http://www.mutationtaster.org/cgi-bin/MutationTaster/MutationTaster69.cgi?sequence_snippet=GACGGT/CJATCC&transcript_stable_id=text=ENST00000338639&gene=PAK7&transcript_stable_id=radio=ENST00000338639&sequence_type=gDNA
7969370	http://www.mutationtaster.org/cgi-bin/MutationTaster/MutationTaster69.cgi?sequence_snippet=CTAC/CJAGGA&transcript_stable_id=text=ENST00000338639&gene=PAK7&transcript_stable_id=radio=ENST00000338639&sequence_type=CD
7984932	http://www.mutationtaster.org/cgi-bin/MutationTaster/MutationTaster69.cgi?sequence_snippet=AAAAAGAC/GJGCC&transcript_stable_id=text=ENST00000338639&gene=PAK7&sequence_type=gDNA

Fig. 5.5. Gene level mutations haven't been reported in 1000G.

Mutation	dbSNP	Gene	Organism	Sequency mds	SNAP pre	SNAP sco	SNAP RI	SIFT pred	SIFT score	Pssm wt	Perc wt	Psic wt	Pssm mt	Perc mt	Psic mt	Hydroph	Volume	Charge	Grantham score	Average heterozygosity
KAR	rs190738218	PARK7	human	MASKRAL/2f169eb0 no effect	-42	4	no effect	0.06	6	58	1.477	4	23	0.371	wt: -3.9 (pwt: 168.6	wt: positive, mt: positive	26	0 [0]		
M17L	rs147690459	PARK7	human	MASKRAL/2f169eb0 no effect	-42	4	no effect	0.26	3	6	1.691	3	27	0.453	wt: 1.9 (pwt: 162.9	wt: neutral, mt: neutral	15	0.000 [0.015]		
M17V	rs147690459	PARK7	human	MASKRAL/2f169eb0 no effect	-42	3	no effect	0.33	3	6	1.891	1	9	0.653	wt: 1.9 (pwt: 162.9	wt: neutral, mt: neutral	21	0.000 [0.015]		
M26I	rs14315351	PARK7	human	MASKRAL/2f169eb0 effect	14	1	effect	0.02	1	3	1.65	-1	3	0.364	wt: 1.9 (pwt: 162.9	wt: neutral, mt: neutral	10	0 [0]		
R28Q	rs142405016	PARK7	human	MASKRAL/2f169eb0 effect	25	2	effect	0.01	4	25	2.353	1	5	0.296	wt: -4.5 (pwt: 173.4	wt: positive, mt: neutral	43	0.000 [0.015]		
A39S	rs137853051	PARK7	human	MASKRAL/2f169eb0 no effect	-47	6	no effect	0.7	0	8	-0.107	1	8	0.234	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	99	0 [0]		
C46A	N/A	PARK7	human	MASKRAL/2f169eb0 effect	8	0	no effect	0.86	3	5	1.295	1	13	0.11	wt: 2.5 (pwt: 108.5	wt: neutral, mt: neutral	195	N/A		
C46S	N/A	PARK7	human	MASKRAL/2f169eb0 effect	23	2	no effect	0.72	3	5	1.295	3	20	-0.281	wt: 2.5 (pwt: 108.5	wt: neutral, mt: neutral	112	N/A		
R48H	rs145727915	PARK7	human	MASKRAL/2f169eb0 no effect	-49	4	no effect	0.6	2	10	1.454	5	12	1.989	wt: -4.5 (pwt: 173.4	wt: positive, mt: neutral	29	0.000 [0.015]		
V51A	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	0	0	no effect	0.46	0	6	0.082	0	7	-0.58	wt: 4.2 (pwt: 140.0	wt: neutral, mt: neutral	64	N/A		
C53A	N/A	PARK7	human	MASKRAL/2f169eb0 effect	4	0	no effect	0.52	0	2	0.595	0	8	-0.63	wt: 2.5 (pwt: 108.5	wt: neutral, mt: neutral	195	N/A		
C53S	N/A	PARK7	human	MASKRAL/2f169eb0 effect	3	0	no effect	0.52	0	2	0.595	-1	5	-0.493	wt: 2.5 (pwt: 108.5	wt: neutral, mt: neutral	112	N/A		
A56T	rs114601558	PARK7	human	MASKRAL/2f169eb0 no effect	-64	6	no effect	0.85	-1	5	0.264	0	5	0.529	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	58	0.009 [0.067]		
E64D	rs14315353	PARK7	human	MASKRAL/2f169eb0 no effect	-32	3	no effect	0.33	2	12	0.057	3	14	0.274	wt: -3.5 (pwt: 138.4	wt: negative, mt: negative	45	0 [0]		
P66Q	rs145093684	PARK7	human	MASKRAL/2f169eb0 no effect	-47	4	no effect	0.38	3	12	0.114	1	6	0.452	wt: -1.6 (pwt: 112.7	wt: neutral, mt: neutral	76	0.000 [0.015]		
A79T	rs45601537	PARK7	human	MASKRAL/2f169eb0 no effect	-47	4	no effect	0.35	2	14	0.824	1	10	0.783	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	58	0 [0]		
R98Q	rs17653619	PARK7	human	MASKRAL/2f169eb0 no effect	-13	1	no effect	0.57	3	14	0.338	2	9	0.762	wt: -4.5 (pwt: 173.4	wt: positive, mt: neutral	43	0.017 [0.092]		
R98W	rs182164240	PARK7	human	MASKRAL/2f169eb0 effect	30	3	effect	0.01	3	14	0.338	-1	1	-1.627	wt: -4.5 (pwt: 173.4	wt: positive, mt: neutral	101	0 [0]		
G150S	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-6	0	effect	0.02	-1	5	0.337	-1	4	-0.508	wt: -0.4 (pwt: 60.1	wt: neutral, mt: neutral	56	N/A		
L101V	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-24	2	no effect	0.19	1	12	0.502	0	7	-0.355	wt: 3.8 (pwt: 166.7	wt: neutral, mt: neutral	32	N/A		
A104T	N/A	PARK7	human	MASKRAL/2f169eb0 effect	58	5	effect	0	-1	6	1.483	0	6	-0.753	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	58	N/A		
C106A	N/A	PARK7	human	MASKRAL/2f169eb0 effect	73	7	effect	0.02	10	58	3.211	0	7	0.172	wt: 2.5 (pwt: 108.5	wt: neutral, mt: neutral	195	N/A		
C106D	N/A	PARK7	human	MASKRAL/2f169eb0 effect	84	8	effect	0	10	58	3.211	-1	4	-1.206	wt: 2.5 (pwt: 108.5	wt: neutral, mt: negative	154	N/A		
C106S	N/A	PARK7	human	MASKRAL/2f169eb0 effect	75	7	effect	0.01	10	58	3.211	-2	3	-0.123	wt: 2.5 (pwt: 108.5	wt: neutral, mt: neutral	112	N/A		
A107P	rs145196092	PARK7	human	MASKRAL/2f169eb0 effect	58	5	effect	0	2	18	1.593	-4	0	-0.904	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	27	0 [0]		
T110A	rs45577037	PARK7	human	MASKRAL/2f169eb0 no effect	-65	6	no effect	1	0	5	0.362	1	14	0.596	wt: -0.7 (pwt: 116.1	wt: neutral, mt: neutral	58	0 [0]		
F119C	N/A	PARK7	human	MASKRAL/2f169eb0 effect	25	2	effect	0.01	1	5	-0.466	2	5	-0.225	wt: 2.8 (pwt: 189.9	wt: neutral, mt: neutral	205	N/A		
H126A	N/A	PARK7	human	MASKRAL/2f169eb0 effect	63	6	effect	0.01	2	5	2.063	-1	5	-1.176	wt: -3.2 (pwt: 153.2	wt: neutral, mt: neutral	86	N/A		
K130R	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-7	0	no effect	0.22	1	10	0.505	2	9	0.206	wt: -3.9 (pwt: 168.6	wt: positive, mt: positive	26	N/A		
K132E	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-43	4	no effect	0.75	1	8	0.696	2	14	0.642	wt: -3.9 (pwt: 168.6	wt: positive, mt: negative	56	N/A		
M134V	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-43	4	no effect	0.35	0	2	0.188	0	6	-0.175	wt: 1.9 (pwt: 162.9	wt: neutral, mt: neutral	21	N/A		
N135D	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-59	5	no effect	0.72	3	14	0.329	1	6	0.304	wt: -3.5 (pwt: 114.1	wt: neutral, mt: negative	23	N/A		
T140A	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-34	3	no effect	0.74	0	5	0.134	-1	5	-0.386	wt: -0.7 (pwt: 116.1	wt: neutral, mt: neutral	58	N/A		
R145H	rs148682310	PARK7	human	MASKRAL/2f169eb0 effect	4	0	no effect	0.13	3	19	1.018	1	4	-0.005	wt: -4.5 (pwt: 173.4	wt: positive, mt: neutral	29	0.000 [0.015]		
D148A	rs14315352	PARK7	human	MASKRAL/2f169eb0 effect	65	6	effect	0.01	6	50	1.61	-2	3	-0.544	wt: -3.5 (pwt: 111.1	wt: negative, mt: neutral	126	0.002 [0.033]		
G150S	N/A	PARK7	human	MASKRAL/2f169eb0 effect	46	4	effect	0.04	5	42	0.967	-2	3	-0.001	wt: -0.4 (pwt: 60.1	wt: neutral, mt: neutral	56	N/A		
R156W	rs61772692	PARK7	human	MASKRAL/2f169eb0 effect	39	3	effect	0	4	19	1.442	-5	0	-1.412	wt: -4.5 (pwt: 173.4	wt: positive, mt: neutral	101	0.500 [0]		
E163K	rs14315354	PARK7	human	MASKRAL/2f169eb0 effect	36	3	no effect	0.14	3	15	0.856	-2	1	-0.496	wt: -3.5 (pwt: 138.4	wt: negative, mt: positive	56	0 [0]		
L166P	rs28938172	PARK7	human	MASKRAL/2f169eb0 effect	54	5	effect	0	4	43	1.218	-4	1	-1.436	wt: 3.8 (pwt: 166.7	wt: neutral, mt: neutral	98	0 [0]		
A171S	N/A	PARK7	human	MASKRAL/2f169eb0 no effect	-38	3	no effect	0.76	2	16	0.009	-2	2	-0.64	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	99	N/A		
A179T	rs17653622	PARK7	human	MASKRAL/2f169eb0 no effect	-70	7	no effect	0.56	1	14	-0.053	-1	3	-0.346	wt: 1.8 (pwt: 88.6	wt: neutral, mt: neutral	58	0.001 [0.026]		

Fig. 5.6. Initial variants - amino acid level variants of PARK7

SNAP prediction	SNAP score	SNAP RI	SIFT prediction	SIFT score	Pssm wt	Perc wt	Psic wt	Pssm mt	Perc mt	Psic mt	Hydrophobicity	Volume	Charge	Grantham score	Average heterozygosity	Mutant site Features
effect	58	5	effect	0	-1	6	1.483	0	6	-0.753	wt: 1.8 (nonpolar), mt: -0.7 (polar)	wt: 88.6 (tiny), mt: 116.1 (small)	wt: neutral, mt: neutral	58	N/A	N/A
effect	84	8	effect	0	10	58	3.211	-1	4	-1.206	wt: 2.5 (polar), mt: -3.5 (polar)	wt: 108.5 (small), mt: 111.1 (small)	wt: neutral, mt: negative	154	N/A	PTM: Cysteine sulfinic acid
effect	58	5	effect	0	2	18	1.593	-4	0	-0.904	wt: 1.8 (nonpolar), mt: -1.6 (nonpolar)	wt: 88.6 (tiny), mt: 112.7 (small)	wt: neutral, mt: neutral	27	0 [0]	N/A
effect	39	3	effect	0	4	19	1.442	-5	0	-1.412	wt: -4.5 (polar), mt: -0.9 (nonpolar)	wt: 173.4 (bulky), mt: 227.8 (bulky)	wt: positive, mt: neutral	101	0.500 [0]	N/A
effect	54	5	effect	0	4	43	1.218	-4	1	-1.436	wt: 3.8 (nonpolar), mt: -1.6 (nonpolar)	wt: 166.7 (bulky), mt: 112.7 (small)	wt: neutral, mt: neutral	98	0 [0]	N/A
effect	25	2	effect	0.01	4	25	2.353	1	5	0.296	wt: -4.5 (polar), mt: -3.5 (polar)	wt: 173.4 (bulky), mt: 143.8 (bulky)	wt: positive, mt: neutral	43	0.000 (0.015)	N/A
effect	30	3	effect	0.01	3	14	0.338	-1	1	-1.627	wt: -4.5 (polar), mt: -0.9 (nonpolar)	wt: 173.4 (bulky), mt: 227.8 (bulky)	wt: positive, mt: neutral	101	0 [0]	N/A
effect	75	7	effect	0.01	10	58	3.211	-2	3	-0.123	wt: 2.5 (polar), mt: -0.8 (polar)	wt: 108.5 (small), mt: 89.0 (tiny)	wt: neutral, mt: neutral	112	N/A	PTM: Cysteine sulfinic acid
effect	25	2	effect	0.01	1	5	-0.466	2	5	-0.225	wt: 2.8 (nonpolar), mt: 2.5 (polar)	wt: 189.9 (bulky), mt: 108.5 (small)	wt: neutral, mt: neutral	205	N/A	N/A
effect	63	6	effect	0.01	2	5	2.063	-1	5	-1.176	wt: -3.2 (polar), mt: 1.8 (nonpolar)	wt: 153.2 (bulky), mt: 88.6 (tiny)	wt: neutral, mt: neutral	86	N/A	ACTIVE SITE: Probable
effect	65	6	effect	0.01	6	50	1.61	-2	3	-0.544	wt: -3.5 (polar), mt: 1.8 (nonpolar)	wt: 111.1 (small), mt: 88.6 (tiny)	wt: negative, mt: neutral	126	0.002 (0.033)	N/A
effect	14	1	effect	0.02	1	3	1.65	-1	3	0.364	wt: 1.9 (nonpolar), mt: 4.5 (nonpolar)	wt: 162.9 (bulky), mt: 166.7 (bulky)	wt: neutral, mt: neutral	10	0 [0]	N/A
effect	73	7	effect	0.02	10	58	3.211	0	7	0.172	wt: 2.5 (polar), mt: 1.8 (nonpolar)	wt: 108.5 (small), mt: 88.6 (tiny)	wt: neutral, mt: neutral	195	N/A	PTM: Cysteine sulfinic acid
effect	46	4	effect	0.04	5	42	0.967	-2	3	-0.001	wt: -0.4 (nonpolar), mt: -0.8 (polar)	wt: 60.1 (tiny), mt: 89.0 (tiny)	wt: neutral, mt: neutral	56	N/A	N/A
Mutation	dbSNP	SwissProt	SwissVar	PMD	1000Genomes	Gene	Protein	Description	Organism	Sequency	md5	SNAP prediction	SNAP score	SNAP RI		
A104T	N/A	VARIANT	VAR_020495	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			58			
C106D	N/A	MUTAGEN	N/A	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			84			
A107P	rs145196092	N/A	N/A	N/A	N/A	PARK7	NP_009193	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			58			
R156W	rs61772692	N/A	N/A	N/A	N/A	PARK7	NP_009193	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			39			
L166P	rs28938172	VARIANT	VAR_020498	N/A	N/A	PARK7	NP_009193, C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			54			
R28Q	rs142405016	N/A	N/A	N/A	N/A	PARK7	NP_009193	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			25			
R98W	rs182164240	N/A	N/A	N/A	Yes	PARK7	NP_009193	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			30			
C106S	N/A	MUTAGEN	N/A	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			75			
F119C	N/A	CONFLICT	N/A	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			25			
H126A	N/A	MUTAGEN	N/A	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			63			
D148A	rs14315352	VARIANT	VAR_020496	N/A	Yes	PARK7	NP_009193, C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			65			
M26I	rs14315351	VARIANT	VAR_020492	N/A	N/A	PARK7	NP_009193, C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			14			
C106A	N/A	MUTAGEN	N/A	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			73			
G150S	N/A	VARIANT	VAR_020497	N/A	N/A	PARK7	C99497	Parkinson disease p[human]	MASKRALVILAKGAEEM/2f169eb09491Eae04956053657c3a	effect			46			

Fig. 5.7 Variants filtered based on SNAP and SIFT scores

Imutant result			
A104T	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut22905/output.html		
C106D	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut22748/output.html		
A107P	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23055/output.html		
R156W	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23199/output.html		
L166P	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23221/output.html		
R28Q	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23293/output.html		
R98W	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23376/output.html		
C106S	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23525/output.html		
F119C	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23663/output.html		
H126A	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut24962/output.html		
D149A	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut23881/output.html		
M26I	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut24031/output.html		
C106A	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut24059/output.html		
G150S	http://gpcr.biocomp.unibo.it/oldpredictors/I-Mutant/Mut24340/output.html		
	SVM2: INC/DEC, RI	DDG VALUE PREDICTION	
A104T	Decrease,7	-0.6	
C106D	Decrease,4	-0.55	
A107P	Increase,3	-0.04	
R156W	Decrease,4	-0.27	
L166P	Decrease,8	-1.75	
R28Q	Decrease,8	-0.85	
R98W	Decrease,5	-0.19	
C106S	Decrease,6	-0.53	
F119C	Decrease,4	-1.36	
H126A	DECREASE, 6	-0.58	
D149A	Decrease,8	-1.07	
M26I	Decrease, 7	-0.52	
C106A	Decrease, 4	-0.42	
G150S	DECREASE, 8	-1.5	
v20a	Decrease,9	-1.11	
p73L	Decrease,8	-0.82	

Fig. 5.8. iMUTANT suite results for filtered variants of PARK7

Variant	PROVEAN score	Prediction (cutoff= -2.5)	Variant	PROVEAN score	Prediction (cutoff= -2.5)
C106D	-11.719	Deleterious			
C106S	-9.704	Deleterious			
C106A	-8.646	Deleterious	V20A	-3.285	Deleterious
H126A	-6.812	Deleterious	P73L	-9.846	Deleterious
R156W	-6.438	Deleterious			
D149A	-6.328	Deleterious			
L166P	-6.225	Deleterious			
A107P	-4.863	Deleterious			
G150S	-4.628	Deleterious			
A104T	-3.923	Deleterious			
R28Q	-3.739	Deleterious			
R98W	-3.47	Deleterious			
F119C	-1.09	Neutral			
M26I	-0.767	Neutral			

Fig. 5.9. PROVEAN results for filtered variants of PARK7

Sequence File: fileseq.seq

Mutation	PANTHER	PhD-SNP	SIFT	SNAP	Meta-SNP	RI	Profile	
M26I	NA	Disease	Disease	Disease	Disease	4	F[M]=11% F[I]=1%	Nali=434
R28Q	NA	Disease	Disease	Disease	Disease	5	F[R]=95% F[Q]=1%	Nali=437
R98W	Disease	Neutral	Disease	Disease	Disease	3	F[R]=8% F[W]=0%	Nali=444
A104T	Disease	Disease	Disease	Disease	Disease	6	F[A]=90% F[T]=1%	Nali=445
C106A	Disease	Disease	Disease	Disease	Disease	6	F[C]=96% F[A]=1%	Nali=445
C106D	Disease	Disease	Disease	Disease	Disease	8	F[C]=96% F[D]=0%	Nali=445
C106S	Disease	Disease	Disease	Disease	Disease	7	F[C]=96% F[S]=1%	Nali=445
A107P	Disease	Disease	Disease	Disease	Disease	8	F[A]=84% F[P]=0%	Nali=446
F119C	Neutral	Disease	Disease	Disease	Disease	5	F[F]=9% F[C]=0%	Nali=438
H126A	Disease	Disease	Disease	Disease	Disease	5	F[H]=24% F[A]=0%	Nali=441
D149A	Neutral	Disease	Disease	Disease	Disease	6	F[D]=81% F[A]=1%	Nali=431
G150S	Neutral	Disease	Disease	Disease	Disease	3	F[G]=67% F[S]=2%	Nali=429
R156W	Disease	Disease	Disease	Disease	Disease	6	F[R]=55% F[W]=0%	Nali=442
L166P	Disease	Disease	Disease	Disease	Disease	8	F[L]=69% F[P]=0%	Nali=440
V20A	NA	Disease	Disease	Disease	Disease	3	F[V]=41% F[A]=1%	Nali=428
P73L	Disease	Disease	Disease	Disease	Disease	6	F[P]=97% F[L]=0%	Nali=444
G150S	Neutral	Disease	Disease	Disease	Disease	3	F[G]=67% F[S]=2%	Nali=429

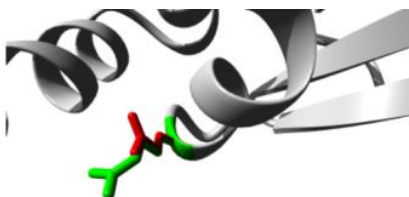
Sequence File: fileseq.seq

Mutation	Prediction	RI	Probability	Method	
R28Q	Disease	9	0.932	PhD-SNP:	F[R]=95% F[Q]=1% Nali=437
R28Q	Unclassified	NA	NA	PANTHER:	F[R]=NA F[Q]=NA
R28Q	Disease	4	0.710	SNPs&GO	
R98W	Neutral	2	0.410	PhD-SNP:	F[R]=8% F[W]=0% Nali=444
R98W	Disease	4	0.714	PANTHER:	F[R]=13% F[W]=0%
R98W	Neutral	6	0.179	SNPs&GO	
A104T	Disease	7	0.854	PhD-SNP:	F[A]=90% F[T]=1% Nali=445
A104T	Disease	5	0.762	PANTHER:	F[A]=80% F[T]=2%
A104T	Disease	5	0.729	SNPs&GO	
C106A	Disease	9	0.925	PhD-SNP:	F[C]=96% F[A]=1% Nali=445
C106A	Disease	5	0.748	PANTHER:	F[C]=85% F[A]=2%
C106A	Disease	7	0.827	SNPs&GO	
C106D	Disease	9	0.972	PhD-SNP:	F[C]=96% F[D]=0% Nali=445
C106D	Disease	8	0.922	PANTHER:	F[C]=85% F[D]=0%
C106D	Disease	8	0.917	SNPs&GO	
C106S	Disease	9	0.954	PhD-SNP:	F[C]=96% F[S]=1% Nali=445
C106S	Disease	6	0.789	PANTHER:	F[C]=85% F[S]=1%
C106S	Disease	7	0.868	SNPs&GO	
A107P	Disease	9	0.932	PhD-SNP:	F[A]=84% F[P]=0% Nali=446
A107P	Disease	6	0.816	PANTHER:	F[A]=75% F[P]=1%
A107P	Disease	6	0.789	SNPs&GO	
H126A	Disease	1	0.546	PhD-SNP:	F[H]=24% F[A]=0% Nali=441
H126A	Disease	3	0.659	PANTHER:	F[H]=45% F[A]=1%
H126A	Disease	2	0.604	SNPs&GO	
D149A	Disease	7	0.861	PhD-SNP:	F[D]=81% F[A]=1% Nali=431
D149A	Neutral	1	0.463	PANTHER:	F[D]=45% F[A]=3%
D149A	Disease	1	0.549	SNPs&GO	
G150S	Disease	5	0.766	PhD-SNP:	F[G]=67% F[S]=2% Nali=429
G150S	Neutral	3	0.357	PANTHER:	F[G]=29% F[S]=4%
G150S	Neutral	1	0.454	SNPs&GO	
R156W	Disease	6	0.791	PhD-SNP:	F[R]=55% F[W]=0% Nali=442
R156W	Disease	8	0.877	PANTHER:	F[R]=40% F[W]=0%
R156W	Disease	1	0.564	SNPs&GO	
L166P	Disease	9	0.933	PhD-SNP:	F[L]=69% F[P]=0% Nali=440
L166P	Disease	8	0.904	PANTHER:	F[L]=75% F[P]=0%
L166P	Disease	7	0.833	SNPs&GO	

Fig. 5.10. MetaSNP, SNPs & GO results of PARK7

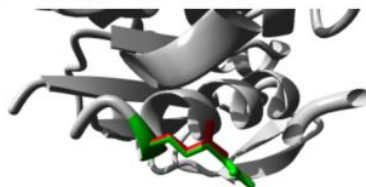
HOPE SERVER:

1. R28Q



[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

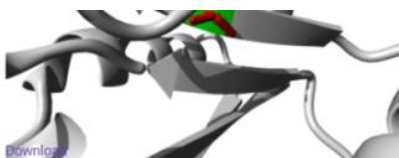


[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

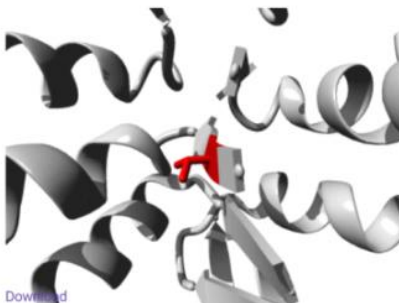
Fig. 5.11.1

2. A104T



[Download](#)

Close-up of the mutation. Both the wild-type and mutant side chain are shown in green and red respectively. The rest of the protein is shown in grey.

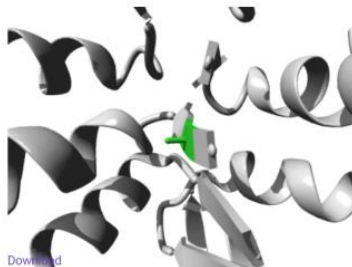


[Download](#)

Close-up of the mutation, same colours as animation 1. The animation shows alternating the wild-type side chain and the mutant side chain.



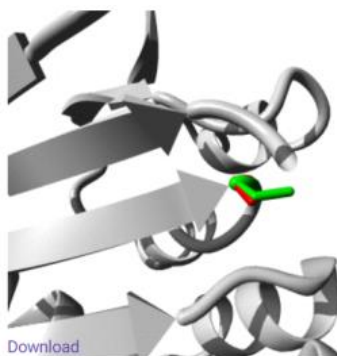
Close-up of the mutation. Both the wild-type and mutant side chain are shown in green and red respectively. The rest of the protein is shown in grey.



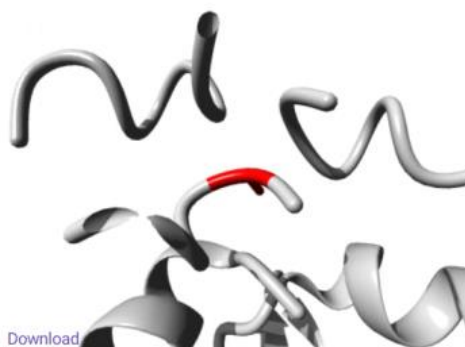
Close-up of the mutation, same colours as animation 1. The animation shows alternating the wild-type side chain and the mutant side chain.

Fig. 5.1.11.2

3. C106A:



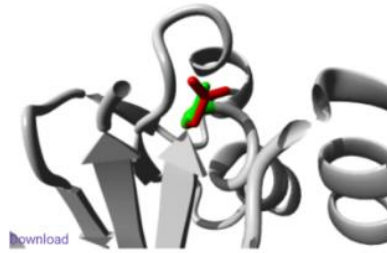
Close-up of the mutation. Both the wild-type and mutant side chain are shown in green and red respectively. The rest of the protein is shown in grey.



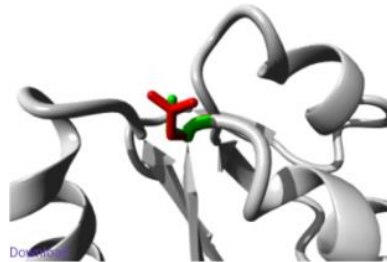
Close-up of the mutation, same colours as animation 1. The animation shows alternating the wild-type side chain and the mutant side chain

Fig. 5.1.11.3.

4. C106D:



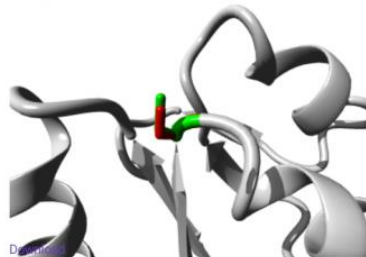
Close-up of the mutation. The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



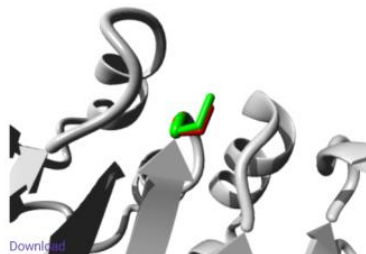
Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

Fig. 5.1.11.4

5. C106S:



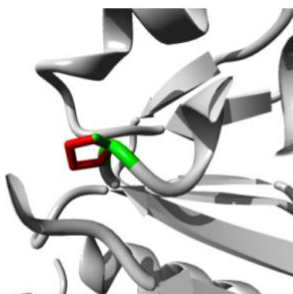
Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

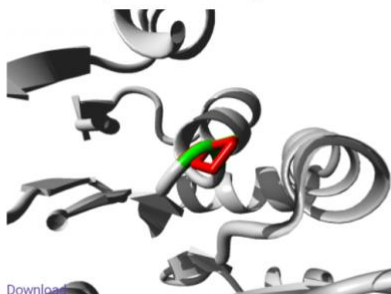
Fig. 5.1.11.5

6. A107P



[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

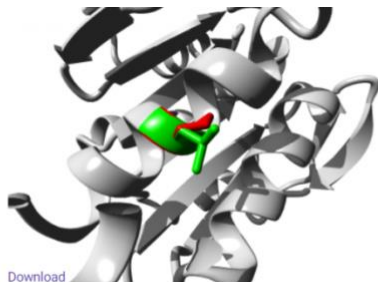


[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

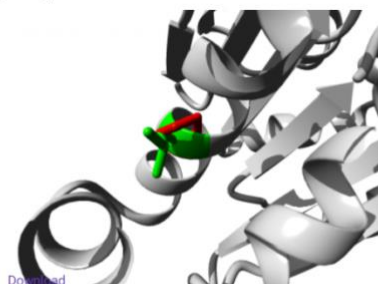
Fig. 5.1.11.6.

7. L166P:



[Download](#)

Close-up of the mutation. The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

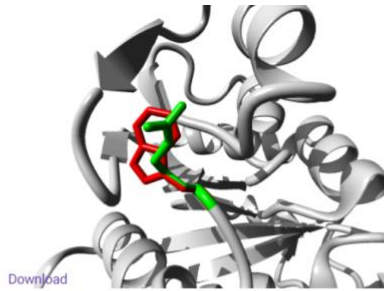


[Download](#)

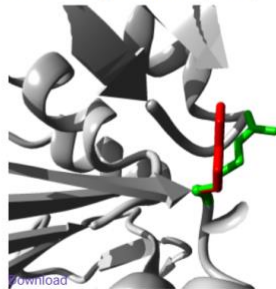
Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

Fig. 5.1.11.7.

8. R156W:



Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

Fig. 5.1.11.8.

9. G150S



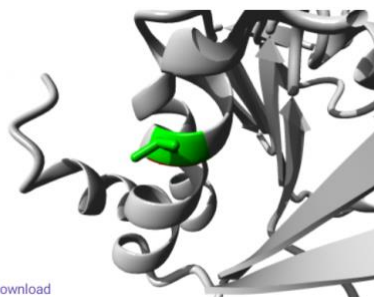
Close-up of the mutation. The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

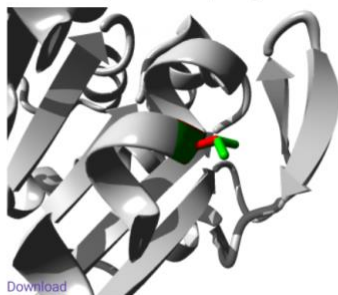
Fig. 5.1.11.9.

10. V20A



[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



[Download](#)

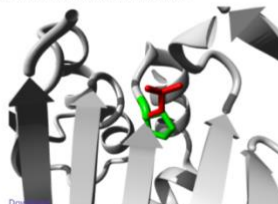
Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

Fig. 5.1.11.10

11. P73L:



Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

Fig. 5.1.11.11

5.2 SNCA:

In Genome browser, searched for gene, Genome viewer, Collected all clinical and cited variants for gene. Downloaded sSNPs. Sorted into 297 total SNV. All 297 loaded into predictsnp2. Filtered out for nssyn and exonic mutations. 26 had possible deleterious.. 3 exonic mutations were filtered harmful. Totally 8 were remaining.

TOTALLY 76 TAKEN FROM SNPDBE. 42 HAVE EFFECT ACCORDING TO SNAP, BASE ON SIFT 5 CENT EFFECT, filtered out for syn exonic mutataions.. 4 prime exonic mutations found. iMutant server used on 4 & provean analysis has been done for all of them. Further, MetaSNP analysis is done for finding disease causing probabilities. HOPE server couldn't be used for visualization as there is no stable available structure, So SWISS model homology modelling is performed for all variants & Discovery studio is used to show differences in structure and correlate with results from HOPE server.

4,89706976,A,G	4,89726434,A,G	4,89757647,G,C	4,89782082,A,T	4,89801720,C,A	4,89822180,C,G	4,89835399,C,T	4,89841046,A,C	4,89724520,T,G	4,89729235,G,T
4,89711617,G,T	4,89726489,A,C	4,89757647,G,A	4,89782082,A,G	4,89805385,C,T	4,89824305,A,C	4,89835399,C,G	4,89842109,A,G	4,89724520,T,A	4,89729235,G,A
4,89711617,G,C	4,89731983,C,T	4,89761176,G,T	4,89782082,A,C	4,89808451,A,T	4,89824556,G,A	4,89835399,C,A	4,89842552,A,T	4,89724523,C,T	4,89822254,A,G
4,89711617,G,A	4,89733513,G,C	4,89761176,G,C	4,89784213,A,T	4,89808451,A,G	4,89824982,T,C	4,89835580,C,G	4,89842552,A,C	4,89724591,A,G	4,89822254,A,C
4,89713401,A,T	4,89733513,G,A	4,89761176,G,A	4,89784213,A,G	4,89808596,T,G	4,89825315,G,T	4,89836121,G,A	4,89843548,C,T	4,89724732,G,A	4,89822265,T,C
4,89713401,A,G	4,89733852,T,C	4,89761353,C,T	4,89784213,A,C	4,89808596,T,C	4,89825315,G,A	4,89836158,A,T	4,89843548,C,A	4,89724926,C,T	4,89822265,T,A
4,89713401,A,C	4,89733852,T,A	4,89761420,A,G	4,89786796,T,G	4,89808596,T,A	4,89825685,A,G	4,89836158,A,G	4,89844062,C,T	4,89724951,G,A	4,89822301,C,A
4,89713869,T,C	4,89736035,C,T	4,89762017,T,G	4,89786796,T,C	4,89810558,A,G	4,89826032,G,A	4,89836243,G,C	4,89844062,C,G	4,89725303,T,C	4,89822309,T,C
4,89713869,T,A	4,89736340,C,T	4,89762017,T,C	4,89788590,G,C	4,89812518,A,G	4,89826600,C,T	4,89836243,G,A	4,89844129,A,T	4,89725318,T,A	4,89822336,C,T
4,89715410,G,A	4,89742391,T,C	4,89762017,T,A	4,89788590,G,A	4,89813384,T,G	4,89826600,C,G	4,89836243,G,T	4,89844129,A,G	4,89725339,A,G	4,89828136,A,C
4,89715859,A,T	4,89742391,T,A	4,89763127,A,T	4,89790619,T,C	4,89813577,A,G	4,89826600,C,A	4,89836354,T,C	4,89844129,A,C	4,89725341,C,T	4,89828149,C,T
4,89715859,A,G	4,89743643,C,T	4,89763127,A,G	4,89791360,G,A	4,89813577,A,C	4,89827044,T,G	4,89836354,T,A	4,89845835,T,C	4,89725341,C,A	4,89828150,C,A
4,89716450,G,A	4,89744890,C,T	4,89766356,C,T	4,89791478,T,C	4,89813810,G,C	4,89827044,T,C	4,89836584,C,A	4,89845944,A,G	4,89725341,C,G	4,89828154,C,T
4,89718364,G,T	4,89744890,C,A	4,89766356,C,G	4,89791913,G,T	4,89813810,G,A	4,89827044,T,A	4,89836689,T,C	4,89845944,A,C	4,89725550,T,A	4,89828156,A,C
4,89718364,G,C	4,89745140,G,A	4,89766776,G,T	4,89791913,G,A	4,89814855,A,T	4,89827337,T,G	4,89836689,T,A	4,89846282,C,T	4,89725581,A,C	4,89828159,C,T
4,89718364,G,A	4,89747463,T,C	4,89766776,G,C	4,89794080,C,T	4,89814855,A,C	4,89827337,T,C	4,89836694,C,T	4,89847220,C,A	4,89725595,A,G	4,89828170,C,T
4,89720189,T,C	4,89747463,T,A	4,89766776,G,A	4,89795026,C,T	4,89815366,T,C	4,89827337,T,A	4,89836694,C,A	4,89848271,A,G	4,89725626,A,C	4,89835560,A,G
4,89720189,T,A	4,89750519,A,T	4,89769981,T,G	4,89795026,C,G	4,89815366,T,A	4,89827495,C,T	4,89837238,G,T	4,89851700,A,C	4,89725735,G,A	4,89835580,C,G
4,89721313,A,T	4,89750519,A,C	4,89769981,T,C	4,89795865,A,T	4,89815434,C,T	4,89827495,C,G	4,89837238,G,C	4,89854061,A,T	4,89725768,A,T	4,89836947,G,A
4,89721972,C,T	4,89752619,A,G	4,89769981,T,A	4,89795865,A,G	4,89815434,C,A	4,89827981,A,T	4,89837238,G,A	4,89854061,A,G	4,89725998,A,C	4,89836961,C,T
4,89721993,C,G	4,89752619,A,C	4,89770386,T,C	4,89798041,A,T	4,89815772,G,A	4,89828149,C,T	4,89837794,A,T	4,89855014,G,T	4,89726035,G,T	4,89836976,C,T
4,89721993,C,A	4,89753280,G,A	4,89771317,T,G	4,89798041,A,C	4,89818104,A,T	4,89828154,C,T	4,89837794,A,G	4,89855014,G,C	4,89726035,G,A	4,89836978,A,G
4,89721993,C,T	4,89753300,T,G	4,89771317,T,C	4,89798041,A,G	4,89818104,A,G	4,89828156,A,C	4,89837896,G,T	4,89855014,G,A	4,89726127,G,A	4,89837065,C,T
4,89724520,T,G	4,89753300,T,C	4,89771317,T,A	4,89798045,A,T	4,89818388,C,T	4,89828170,C,T	4,89837896,G,C	4,89724127,G,T	4,89726163,T,A	4,89837155,C,T
4,89724520,T,A	4,89753300,T,A	4,89772325,C,G	4,89798045,A,G	4,89818388,C,A	4,89832326,G,A	4,89838736,A,G	4,89724298,G,A	4,89726351,C,T	4,89837155,C,G
4,89724523,C,T	4,89754087,A,T	4,89773346,A,G	4,89800409,A,G	4,89818841,T,A	4,89834579,G,A	4,89838736,A,C	4,89724337,G,A	4,89726351,C,A	4,89837187,T,G
4,89725735,G,A	4,89754087,A,G	4,89773436,T,G	4,89800486,C,T	4,89819727,G,A	4,89834788,A,G	4,89839070,A,T	4,89724351,T,G	4,89726417,C,T	4,89837210,C,T
4,89726127,G,A	4,89757390,G,T	4,89775184,A,G	4,89800486,C,G	4,89820622,A,T	4,89834926,T,G	4,89839070,A,C	4,89724359,G,A	4,89726455,C,T	
4,89726252,G,A	4,89757390,G,A	4,89776006,T,C	4,89801282,G,A	4,89820622,A,G	4,89835040,C,T	4,89839677,C,T	4,89724401,G,T	4,89726489,A,C	
4,89726425,C,T	4,89757647,G,T	4,89779010,C,G	4,89801720,C,T	4,89822180,C,T	4,89835270,G,C	4,89839677,C,A	4,89724401,G,A	4,89726643,G,A	

Fig. 5.12. Initial all gene variants of SNCA

[illegible]

	exonic and nssyn			
4	89828154	.	C	T
4	89828170	rs1048938	C	T
4	89835580	rs1048938	C	G
4	89822265	.	T	A
4	89822301	.	C	A

Mutation	dbSNP	Gene	Protein	Descriptio	Organism	Sequence md5	SNAP pre	SNAP sco	SNAP RI	SIFT pred	SIFT score	Pssm wt	Perc wt	Psic wt	Pssm mt	Perc mt	Psic mt	Hydrophi	Volume	Charge	Granthan	Average heterozygosity
L9V	rs28655310	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	0	0	effect	0	5	100	1.303	0	0	0.093	wt: 3.8 (n wt: 166.7 wt: neutri	32 0 (0)			
E30Q	rs18305655	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	5	0	effect	0	4	49	1.381	0	0	0.064	wt: -3.5 (wt: 138.4 wt: negati	29 0 (0)			
E40K	rs18779505	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	28	2	effect	0.02	4	31	1.06	-1	1	-0.372	wt: -3.5 (wt: 138.4 wt: negati	56 0 (0)			
D41G	N/A	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	21	2	effect	0.01	4	29	1.149	-2	1	-0.669	wt: -3.5 (wt: 111.1 wt: negati	94 N/A			
V44A	rs56285021	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-11	1	no effect	0.49	3	34	0.869	2	21	0.77	wt: 4.2 (n wt: 140.0 wt: neutri	64 0 (0)			
I57T	rs18384281	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	29	2	effect	0.04	4	39	1.109	1	9	-0.183	wt: 4.5 (n wt: 166.7 wt: neutri	89 0 (0)			
F73L	rs14217202	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	32	3	effect	0	7	71	1.869	-1	0	-0.132	wt: 2.8 (n wt: 189.9 wt: neutri	22 0.004 (0.044)			
R78W	rs14591951	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	34	3	effect	0	6	53	1.805	-4	0	-0.849	wt: -4.5 (wt: 173.4 wt: positi	101 0.000 (0.015)			
V79N	N/A	SNCAIP	Q9Y6H5	synuclein	human	MEAPEYLI0107d19c	effect	51	5	effect	0	5	52	1.378	-4	0	-0.807	wt: 4.2 (n wt: 140.0 wt: neutri	133 N/A			
P81N	N/A	SNCAIP	Q9Y6H5	synuclein	human	MEAPEYLI0107d19c	effect	32	3	effect	0	7	63	1.917	0	5	-0.582	wt: -1.6 (wt: 112.7 wt: neutri	91 N/A			
D112V	N/A	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	36	3	effect	0	6	43	1.508	-4	0	-0.802	wt: -3.5 (wt: 111.1 wt: negati	152 N/A			
G114D	rs14810009	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-6	0	no effect	0.06	3	21	0.375	-3	0	-0.65	wt: -0.4 (wt: 60.1 (wt: neutri	94 0.000 (0.015)			
L178W	rs14637430	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	22	2	effect	0.03	2	21	0.498	-3	0	-0.788	wt: 3.8 (n wt: 166.7 wt: neutri	61 0.000 (0.015)			
N185S	rs13934663	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-34	3	no effect	0.31	5	28	1.124	1	8	0.611	wt: -3.5 (wt: 114.1 wt: neutri	46 0.000 (0.015)			
S188F	N/A	SNCAIP	Q9Y6H5	synuclein	human	MEAPEYLI0107d19c	effect	40	4	effect	0	2	18	1.04	-1	2	-0.746	wt: -0.8 (wt: 89.0 (wt: neutri	155 N/A			
S200L	N/A	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-47	4	no effect	0.19	1	11	0.616	0	9	0.359	wt: -0.8 (wt: 89.0 (wt: neutri	145 N/A			
N203S	rs14438472	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-25	2	no effect	0.38	4	23	1.479	1	9	0.904	wt: -3.5 (wt: 114.1 wt: neutri	46 0.002 (0.033)			
P206A	N/A	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	32	3	effect	0.03	1	8	1.443	-2	4	-0.433	wt: -1.6 (wt: 112.7 wt: neutri	27 N/A			
S222L	rs14759464	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-9	0	effect	0.03	-1	5	0.523	-4	2	-0.812	wt: -0.8 (wt: 89.0 (wt: neutri	145 0.002 (0.033)			
D227N	rs14036649	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	4	0	effect	0	3	14	1.79	2	8	0.051	wt: -3.5 (wt: 111.1 wt: negati	23 0.000 (0.015)			
L231Q	rs14429177	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-64	6	no effect	0.2	2	19	0.286	-1	2	0.857	wt: 3.8 (n wt: 166.7 wt: neutri	113 0.000 (0.015)			
E235G	rs6867105	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	30	3	no effect	0.31	1	8	1.116	-1	5	0.697	wt: -3.5 (wt: 138.4 wt: negati	98 0.005 (0.049)			
F263C	rs14029529	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-54	5	no effect	0.08	0	5	1.158	2	5	1.484	wt: 2.8 (n wt: 189.9 wt: neutri	205 0.001 (0.021)			
S306P	rs11135614	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	37	3	effect	0.02	1	11	0.922	-4	1	-0.556	wt: -0.8 (wt: 89.0 (wt: neutri	74 0.500 (0)			
A363T	rs14503892	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-53	5	no effect	0.14	0	7	1.077	1	8	0.679	wt: 1.8 (n wt: 88.6 (wt: neutri	58 0.000 (0.015)			
T383N	rs14494983	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-62	6	no effect	0.86	-1	4	0.518	3	14	0.7	wt: -0.7 (wt: 116.1 wt: neutri	65 0.003 (0.039)			
R402H	rs15068702	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	2	0	effect	0	3	15	0.884	-1	2	-0.886	wt: -4.5 (wt: 173.4 wt: positi	29 0.000 (0.015)			
A427T	rs14192030	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0.01	7	83	0.235	-2	2	-0.098	wt: 1.8 (n wt: 88.6 (wt: neutri	58 0.000 (0.015)			
V461I	rs18551861	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-32	3	effect	0.04	-1	2	0.131	1	5	0.425	wt: 4.2 (n wt: 140.0 wt: neutri	29 0 (0)			
V483A	rs15069529	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	7	0	effect	0	4	30	0.273	-1	5	-0.461	wt: 4.2 (n wt: 140.0 wt: neutri	64 0.000 (0.015)			
M485I	N/A	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-9	0	no effect	0.34	2	4	0.887	2	13	-0.014	wt: 1.9 (n wt: 162.9 wt: neutri	10 N/A			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect	0.01	4	22	0.258	-2	1	0.376	wt: -4.5 (wt: 173.4 wt: positi	180 0.000 (0.015)			
T538M	rs14714722	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	effect	17	1	effect	0	1	9	0.553	0	2	-0.858	wt: -0.7 (wt: 116.1 wt: neutri	81 0.000 (0.015)			
R536C	rs14344033	SNCAIP	NP_0054	synuclein	human	MEAPEYLI0107d19c	no effect	-1	0	effect												

Descripti	Mutation	Gene	Protein	Descripti	Organism	Sequence md5	SNAP pre	SNAP sco	SNAP RI	SIFT pred	SIFT score	Pssm wt	Perc wt	Psic wt	Pssm mt	Perc mt	Psic mt	Hydrophic	Volume	Charge	Grantham	Average	Experime	Mutant site	Features
alpha-syn	S9A	SNCA	A004754	alpha-syn	human	MDVFMK(90cc5e82c	effect	30	3	effect	0.01	5	66	1.156	0	0	-0.463	wt: -0.8 (pwt: 89.0 (t	wt: neutre	99	N/A	Not valid	N/A		
alpha-syn	E123K	SNCA	NP_000333	alpha-syn	human	MDVFMK(90cc5e82c	effect	38	3	effect	0.01	5	66	1.529	1	10	-0.012	wt: -3.5 (pwt: 138.4 (wt: negati	56	N/A	Not valid	N/A			
alpha-syn	M127I	SNCA	NP_000333	alpha-syn	human	MDVFMK(90cc5e82c	effect	12	1	effect	0.02	3	16	1.296	-3	0	-0.626	wt: 1.9 (ncwt: 162.9 (wt: neutre	10	0 (0)	1000G	eno	N/A		
alpha-syn	A30P	SNCA	NP_000333	alpha-syn	human	MDVFMK(90cc5e82c	effect	27	2	effect	0.04	4	66	0.739	-2	0	-0.899	wt: 1.8 (ncwt: 88.6 (t	wt: neutre	27	0 (0)	by cluster	N/A		

Fig. 5.16. SNCA filtered from aa level variants

Imutant Suite				Gene Filtered			
aa filtered							
S9A	Decrease,6	-0.39	Neutral	G51D	Decrease,4	-0.91	Decrease stability
E123K	Decrease,6	-0.62	Decrease stability	E46K	Decrease,8	-0.69	Decrease stability
M127I	Decrease,3	-0.41	Neutral	G84V	Decrease,1	-0.21	Neutral
A30P	Decrease,0	-0.1	Neutral	K96I	Increase,2	0.49	Increase
Provean							
Variant	PROVEAN sc	Prediction (cutoff= -2.5)		Variant	PROVEAN sc	Prediction (cutoff= -2.5)	
S9A	-2.708	Deleterious		E46K	-3.48	Deleterious	
E123K	-1.592	Neutral		G51D	-2.3	Neutral	
M127I	-0.566	Neutral		G84V	-4.323	Deleterious	
A30P	-4.536	Deleterious		K96I	-5.573	Deleterious	

Fig. 5.17. SNCA iMutant & Provean result

Meta-SNP: Meta-predictor of disease causing variants
 Cite: Capriotti et al. BMC Genomics 2013. Suppl 3:S2.
 WEB: <http://snps.biofold.org/meta-snp>

Sequence File: fileseq.seq

Mutation	PANTHER	PhD-SNP	SIFT	SNAP	Meta-SNP	RI	Profile		
S9A	Neutral	Neutral	Disease	Disease	Neutral	3	F[S]=100% F[A]=0%	Nali=41	
	0.181	0.365	0.030	0.665	0.348				
A30P	Disease	Disease	Disease	Disease	Disease	3	F[A]=100% F[P]=0%	Nali=42	
	0.786	0.687	0.030	0.660	0.658				
E123K	Neutral	Neutral	Neutral	Disease	Neutral	7	F[E]=80% F[K]=0%	Nali=14	
	0.302	0.299	0.090	0.640	0.153				
M127I	Neutral	Neutral	Neutral	Disease	Neutral	8	F[M]=42% F[I]=0%	Nali=11	
	0.187	0.146	0.130	0.515	0.101				

Mutation: WT+POS+NEW

WT: Residue in wild-type protein

POS: Residue position

NEW: New residue after mutation

Prediction:

Neutral: Neutral variants

Disease: Disease causing variants

Outputs: Value reported under each prediction

PANTHER: Between 0 and 1. (If >0.5 mutation is predicted Disease)

PhD-SNP: Between 0 and 1. (If >0.5 mutation is predicted Disease)

SIFT: Positive Value (If >0.05 mutation is predicted Neutral)

SNAP: Output normalized between 0 and 1 (If >0.5 mutation is predicted Disease)

Meta-SNP: Between 0 and 1. (If >0.5 mutation is predicted Disease)

RI: Reliability Index between 0 and 10

F[X]: Frequency of residue X in the sequence profile

Nali: Number of aligned sequences in the mutated site

Meta-SNP: Meta-predictor of disease causing variants
 Cite: Capriotti et al. BMC Genomics 2013. Suppl 3:S2.
 WEB: <http://snps.biofold.org/meta-snp>

Sequence File: fileseq.seq

Mutation	PANTHER	PhD-SNP	SIFT	SNAP	Meta-SNP	RI	Profile		
E46K	Neutral	Disease	Neutral	Neutral	Disease	0	F[E]=68% F[K]=2%	Nali=43	
	0.363	0.682	0.100	0.500	0.505				
G51D	Disease	Neutral	Disease	Disease	Disease	1	F[G]=45% F[D]=0%	Nali=39	
	0.593	0.404	0.030	0.675	0.529				
G84V	Disease	Neutral	Disease	Disease	Disease	1	F[G]=77% F[V]=0%	Nali=38	
	0.639	0.385	0.040	0.635	0.530				
K96I	Disease	Neutral	Disease	Disease	Neutral	1	F[K]=71% F[I]=0%	Nali=37	
	0.734	0.309	0.040	0.660	0.435				

Mutation: WT+POS+NEW

WT: Residue in wild-type protein

POS: Residue position

NEW: New residue after mutation

Prediction:

Neutral: Neutral variants

Disease: Disease causing variants

Outputs: Value reported under each prediction

PANTHER: Between 0 and 1. (If >0.5 mutation is predicted Disease)

PhD-SNP: Between 0 and 1. (If >0.5 mutation is predicted Disease)

SIFT: Positive Value (If >0.05 mutation is predicted Neutral)

SNAP: Output normalized between 0 and 1 (If >0.5 mutation is predicted Disease)

Meta-SNP: Between 0 and 1. (If >0.5 mutation is predicted Disease)

RI: Reliability Index between 0 and 10

F[X]: Frequency of residue X in the sequence profile

Nali: Number of aligned sequences in the mutated site

Fig. 5.18. MetaSNP results of SNCA

RESULTS			neutral	deleterious	XX % expected accuracy				Expand all annotations
Annotation	Mutation	GRCh38/hg38	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
▶	A30P	4.89635580,C→G	87 %	88 %	82 %	74 %	81 %	79 %	81 %
▶	E46K	4.89828170,C→T	72 %	84 %	73 %	67 %	43 %	79 %	81 %
	G51D	4.89828154,C→T	61 %	88 %	68 %	67 %	41 %	79 %	62 %
	G84V	4.89822301,C→A	61 %	77 %	51 %	74 %	81 %	79 %	61 %
	K96I	4.89822265,T→A	87 %	82 %	73 %	74 %	45 %	79 %	81 %

Fig. 5.19 Predictsnp1 results of SNCA

HOPE SERVER:

1. A30P

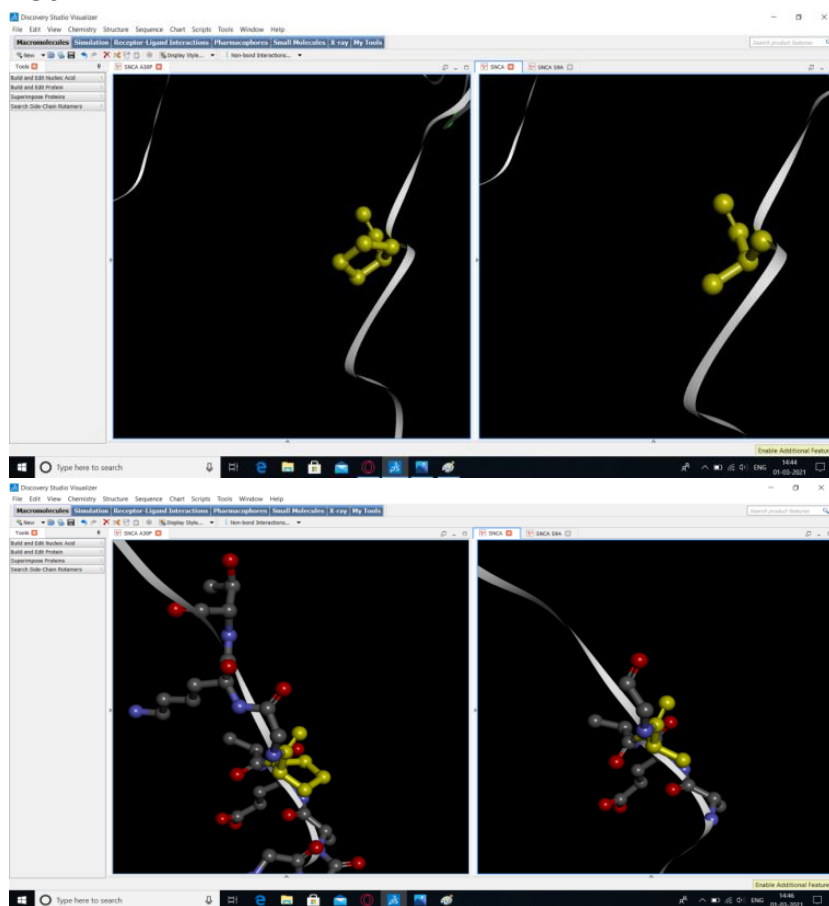


Fig. 5.20.1

2. E123K

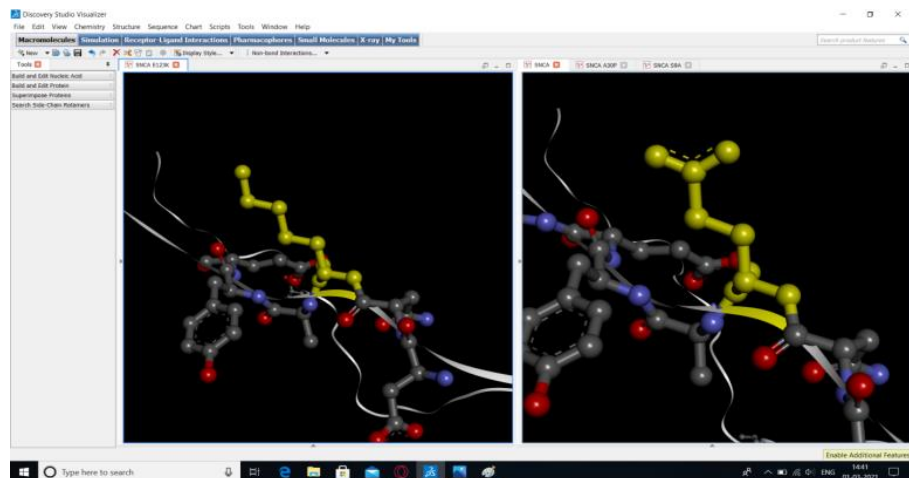
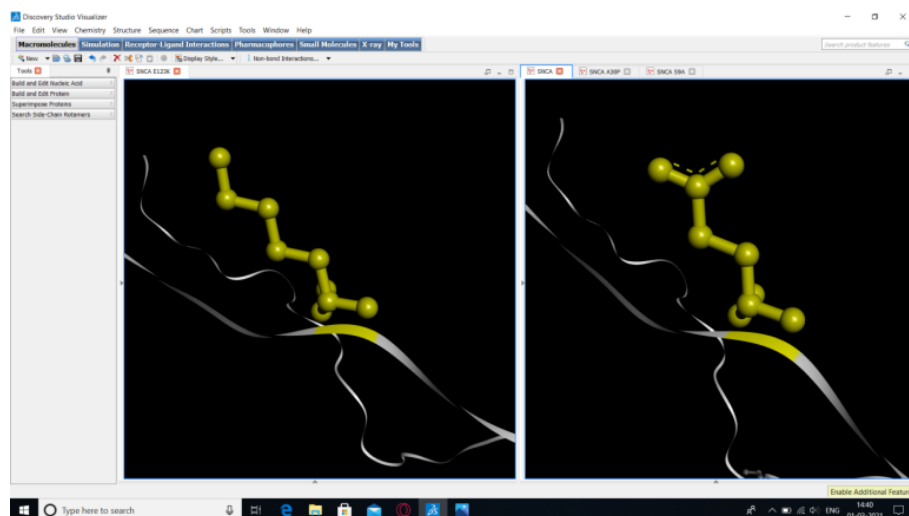
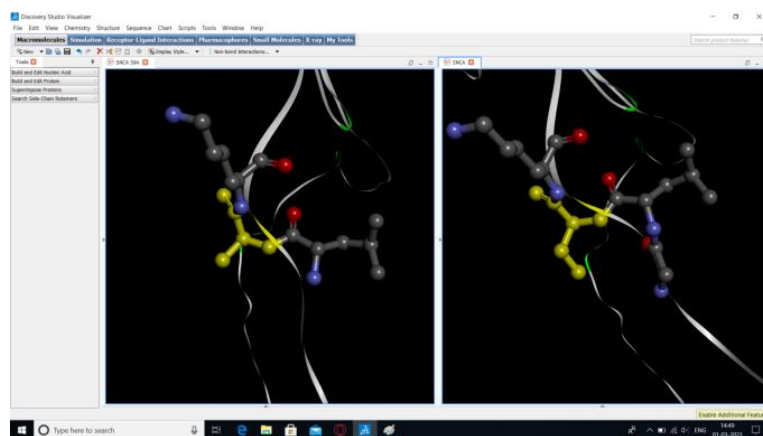


Fig. 5.20.2

3. S9A



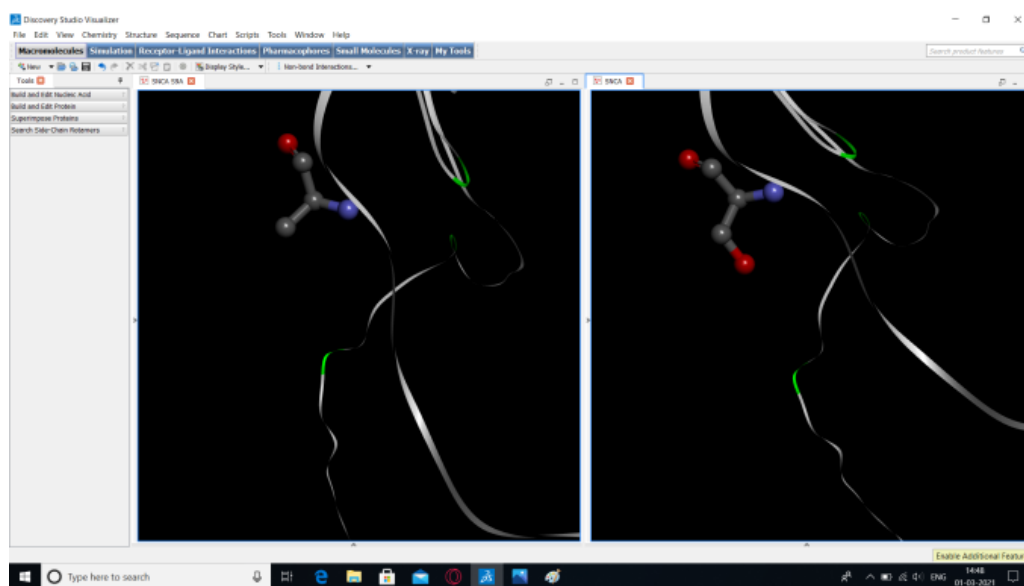


Fig. 5.20.3

4. E46K

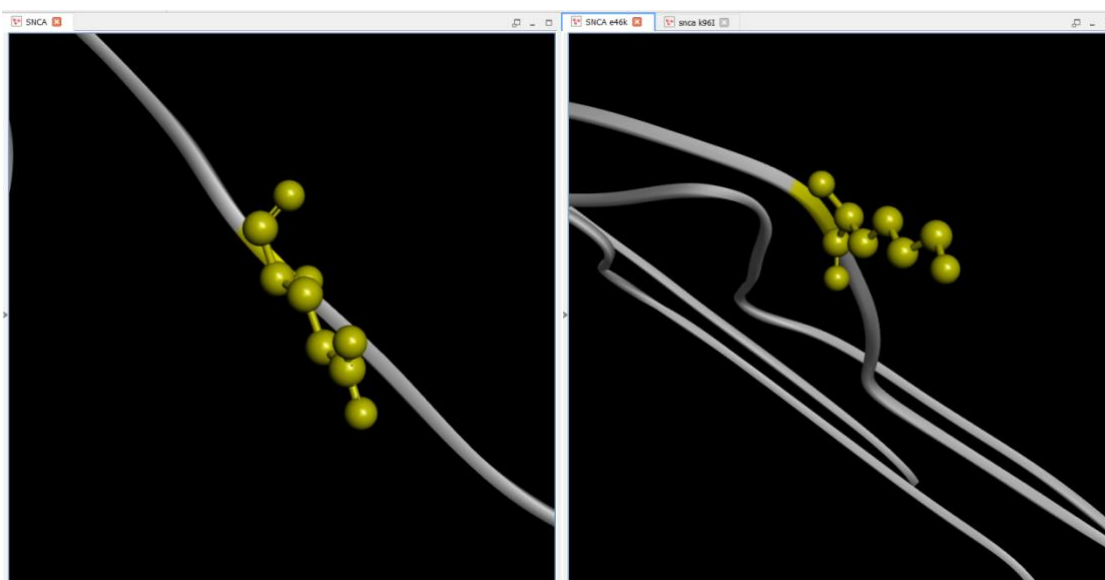


Fig. 5.20.4

5. G51D

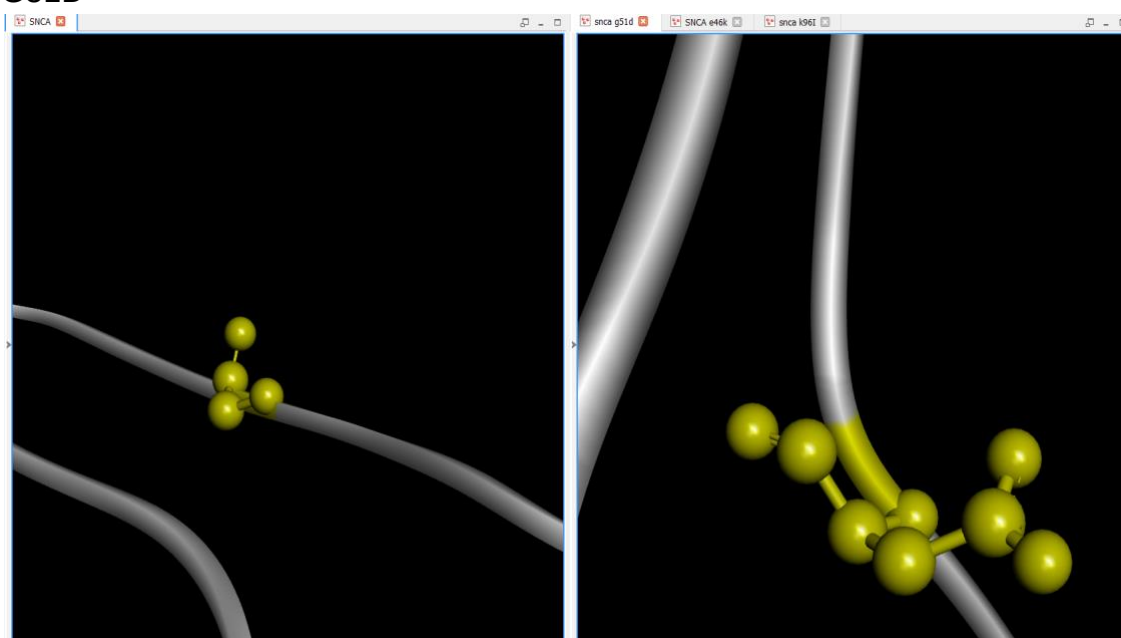


Fig. 5.20.5

6. G84V

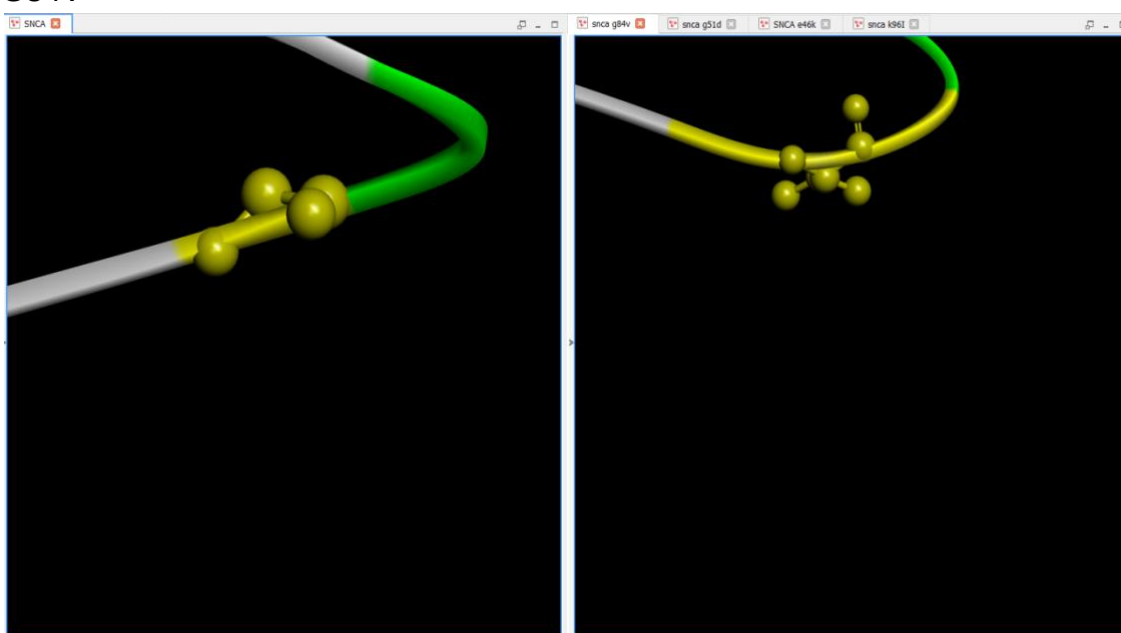


Fig. 5.20.6

7. K96I

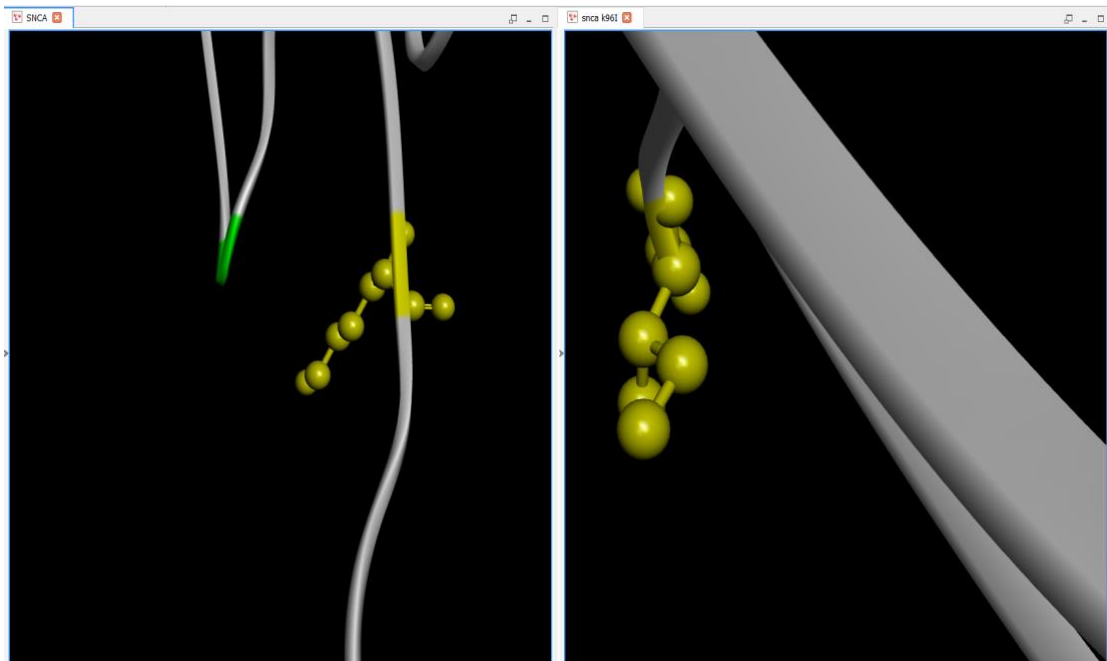


Fig. 5.20.7

5.3. LRRK2:

Initial data has been taken from dbSNP with total of 571 variants in SNPs. All of them have been loaded into predictsnp2. 308 non-synonymous exonic found. Rigorous filtering for finding deleterious resulted in 178 variants after sorting. After further filtration based on algorithm scores and sorting, 40 variants were taken further into Mutation Taster 2 tool. Of all the inputs, only 9 of them were both disease causing and haven't been reported in 1000G.

216 results from snpdb are further filtered into 56 with ones which are not validated. All of those entries into iMutant and only 38 of them had minimum threshold to decrease stability. Further from 38, only 28 of results were taken further based on RI and scores. 28 are incorporated into provean tool, 14 came out deleterious and probably del- neutral with mild chances of being deleterious

All of the results from MutationTaster and amino acid level filtered were put into metaSNP. Only 5 of them were harmful and further they were put into HOPE server, which gave results for all of them but only visualization was done for 2 of the variants in main LRRK2 domain. For other 3 leads, Homology modelling has been done in SWISS model and then Discovery studio was used to visualize and correlate the results with HOPE server.

12,40225008,G,C	12,40240610,G,A	12,40252984,C,T	12,40284011,G,T	12,40308618,A,G	12,40315306,T,A	12,40322458,T,C	12,40351680,G,A	12,40368111,A,G	12,40222364,G,C	12,40251369,G,A	12,40294893,T,C	12,40310450,C,T	12,40322468,C,A	12,40363541,G,A
12,40225008,G,A	12,40243556,A,T	12,40253016,G,A	12,40284011,G,A	12,40308632,C,A	12,40315309,C,A	12,40322468,C,A	12,40351680,G,A	12,40368129,A,T	12,40223890,G,A	12,40251828,G,A	12,40295466,G,A	12,40310461,G,A	12,40323255,A,G	12,40364489,G,T
12,40225088,G,T	12,40243579,A,C	12,40257311,A,G	12,40284012,G,A	12,40308663,A,G	12,40319979,A,T	12,40323151,A,G	12,40351723,A,G	12,40368129,A,G	12,40225159,G,A	12,40252884,C,T	12,40295566,A,G	12,40310515,A,G	12,40323256,T,C	12,40364489,G,A
12,40225088,G,A	12,40243593,T,C	12,40257324,T,C	12,40284061,A,T	12,40308663,A,C	12,40319979,A,G	12,40323244,C,G	12,40354304,T,C	12,40368129,A,C	12,40225280,G,A	12,40256700,G,A	12,40295569,T,C	12,40310561,G,A	12,40323260,G,T	12,40364843,G,A
12,40225159,G,A	12,40243668,T,C	12,40257342,T,C	12,40284061,A,G	12,40308675,C,T	12,40319988,T,C	12,40323255,A,G	12,40354410,G,A	12,40368149,C,T	12,40225558,C,T	12,40257342,T,C	12,40298346,G,A	12,40313976,G,C	12,40323270,G,T	12,40364850,T,C
12,40225206,T,C	12,40243668,T,A	12,40259507,A,G	12,40284114,T,C	12,40309109,G,T	12,40320032,C,T	12,40323256,T,C	12,40356126,A,T	12,40368170,C,T	12,40225628,G,T	12,40258718,T,G	12,40298433,C,T	12,40313976,G,A	12,40323270,G,A	12,40364884,G,A
12,40225245,C,A	12,40243843,C,T	12,40259507,A,T	12,40287367,A,G	12,40309109,G,A	12,40320032,C,G	12,40323260,G,T	12,40356145,C,T	12,40368170,C,A	12,40225628,G,A	12,40258718,T,C	12,40298433,C,G	12,40314059,C,T	12,40335031,G,A	12,40367012,T,A
12,40225280,G,A	12,40243843,C,G	12,40259537,A,G	12,40287438,A,G	12,40309145,C,T	12,40320032,C,A	12,40323270,G,T	12,40359253,G,T	12,40368235,G,T	12,40227006,C,G	12,40258718,T,A	12,40298479,G,T	12,40314101,C,T	12,40335205,C,T	12,40367050,A,G
12,40225558,C,T	12,40243854,C,T	12,40259578,G,A	12,40287444,C,T	12,40309174,G,C	12,40320041,G,T	12,40323270,G,A	12,40359316,G,A	12,40368235,G,A	12,40227013,G,T	12,40259525,A,T	12,40298488,A,G	12,40314101,C,A	12,40338592,T,C	12,40368149,C,T
12,40225563,T,C	12,40243854,C,A	12,40263806,A,G	12,40287461,A,G	12,40309174,G,A	12,40320041,G,A	12,40323297,T,C	12,40359345,C,T	12,40368380,T,A	12,40227013,G,C	12,40259578,G,A	12,40299125,A,G	12,40315266,T,A	12,40338592,T,A	12,40372127,T,A
12,40225580,T,A	12,40243875,A,G	12,40263875,A,G	12,40293607,C,T	12,40309185,G,A	12,40320043,G,T	12,40323297,T,C	12,40359422,A,G	12,40368486,G,T	12,40227013,G,A	12,40259708,G,A	12,40299212,G,A	12,40315611,T,G	12,40339098,C,A	12,40383514,T,G
12,40225628,G,T	12,40243881,T,C	12,40263898,C,G	12,40293616,G,A	12,40309206,C,T	12,40320043,G,C	12,40333503,G,A	12,40363425,C,T	12,40368501,G,T	12,40229925,G,A	12,40263806,A,G	12,40299255,T,C	12,40315611,T,C	12,40340361,T,C	12,40298488,A,G
12,40225628,G,A	12,40243923,G,T	12,40263898,C,A	12,40293624,G,C	12,40309225,A,C	12,40320043,G,A	12,40335079,G,T	12,40363440,C,T	12,40368501,G,C	12,40231279,G,A	12,40263875,A,G	12,40302866,A,G	12,40315611,T,A	12,40340380,T,C	12,40309352,C,T
12,40232303,A,G	12,40243923,G,A	12,40274600,G,A	12,40293624,G,A	12,40309230,A,G	12,40320054,G,C	12,40335079,G,A	12,40363465,T,C	12,40368501,G,A	12,40235634,T,C	12,40263898,C,G	12,40304004,G,A	12,40319957,T,C	12,40340400,G,A	12,40304141,C,T
12,40232303,A,C	12,40251273,G,A	12,40274681,T,C	12,40293626,G,T	12,40310425,C,T	12,40320071,A,G	12,40335140,C,T	12,40363526,G,A	12,40368736,C,A	12,40235642,C,T	12,40263898,C,A	12,40304040,G,T	12,40319957,T,A	12,40340404,T,C	12,40304841,G,A
12,40232380,A,C	12,40251346,C,T	12,40274899,A,G	12,40293626,G,A	12,40310434,C,T	12,40320090,A,G	12,40340361,T,C	12,40363528,A,G	12,40368744,T,C	12,40267601,A,T	12,40304040,G,G	12,40319998,T,C	12,40340436,T,C	12,40315266,T,A	
12,40232385,T,C	12,40251361,A,G	12,40274917,G,C	12,40293633,A,T	12,40310434,C,G	12,40320094,A,G	12,40340380,T,C	12,40363541,G,A	12,40368744,T,A	12,40237989,T,A	12,40267601,A,C	12,40304884,A,G	12,40320032,C,T	12,40340781,T,C	12,40322413,A,G
12,40235644,C,T	12,40251366,C,A	12,40277933,T,C	12,40293633,A,G	12,40310434,C,A	12,40320097,T,C	12,40340400,G,A	12,40364843,G,A	12,40368783,G,T	12,40238078,A,T	12,40269794,G,T	12,40305791,C,G	12,40320032,C,G	12,40346563,G,T	12,40351680,G,C
12,40235642,C,T	12,40251369,G,T	12,40277963,T,C	12,40293644,A,G	12,40310435,G,A	12,40320099,T,A	12,40340404,T,C	12,40364850,T,C	12,40368800,T,C	12,40238078,A,G	12,40274600,G,A	12,40308467,G,T	12,40320032,C,A	12,40346884,A,T	12,40367988,C,T
12,40235646,C,T	12,40251369,G,A	12,40277968,A,C	12,40294854,T,G	12,40310436,C,T	12,40320119,A,G	12,40340411,G,A	12,40364884,G,A	12,40368834,A,G	12,40238297,T,G	12,40274899,A,G	12,40308467,G,C	12,40320043,G,A	12,40346884,A,T	12,40223890,G,A
12,40235646,C,G	12,40251455,C,A	12,40278025,T,G	12,40294854,T,C	12,40310450,C,T	12,40321114,A,G	12,40340436,A,T	12,40364927,A,G	12,40368905,A,T	12,40239512,T,G	12,40277196,C,A	12,40308481,G,A	12,40320043,G,C	12,40348452,G,C	12,40251369,G,T
12,40237989,T,C	12,40251495,C,T	12,40278025,T,A	12,40294866,G,T	12,40310461,G,A	12,40321123,T,C	12,40340684,A,T	12,40364927,A,G	12,40368946,C,T	12,40239512,T,C	12,40277933,T,C	12,40308618,A,G	12,40320043,G,T	12,40348452,G,A	12,40294866,G,T
12,40237989,T,A	12,40252905,G,A	12,40278154,A,G	12,40294893,T,C	12,40310471,A,G	12,40321172,C,T	12,40340684,A,G	12,40364956,C,T	12,40368947,C,T	12,40239512,T,A	12,40277968,A,C	12,40308632,C,A	12,40320071,A,G	12,40348484,C,T	12,40310437,G,T
12,40238078,A,T	12,40252906,A,G	12,40278167,C,T	12,40295448,A,G	12,40310515,A,G	12,40321181,A,G	12,40340845,G,C	12,40364956,C,G	12,40368951,C,T	12,40240543,C,T	12,40278154,A,G	12,40309109,G,T	12,40320097,T,C	12,40348561,G,C	12,40322458,T,C
12,40238078,G,A	12,40252929,G,A	12,40278187,A,G	12,40295463,A,G	12,40310530,C,T	12,40321181,A,C	12,40340845,G,A	12,40364956,C,A	12,40369044,T,G	12,40240543,C,A	12,40278167,C,T	12,40309109,G,A	12,40320097,T,C	12,40348561,G,A	12,40363528,A,G
12,40240543,C,T	12,40252958,A,T	12,40283897,C,T	12,40295466,G,A	12,40310546,G,A	12,40322026,A,T	12,40340845,A,G	12,40364960,A,T	12,40369044,T,C	12,40240594,G,C	12,40278187,A,G	12,40309145,C,T	12,40320119,A,G	12,40350973,A,C	
12,40240543,C,A	12,40252958,A,C	12,40283947,C,T	12,40295535,A,T	12,40310554,G,A	12,40322026,A,C	12,40340846,C,T	12,40364960,A,G	12,40369153,T,C	12,40243556,A,T	12,40282398,G,T	12,40309174,G,C	12,40320207,A,G	12,40351572,T,A	
12,40240594,G,C	12,40252968,G,A	12,40283955,G,A	12,40295566,A,G	12,40310561,G,A	12,40322026,A,G	12,40351534,C,G	12,40365044,C,T	12,40369211,T,A	12,40243668,T,C	12,40283227,A,C	12,40309174,G,A	12,40321114,A,G	12,40351579,C,T	
12,40240602,T,C	12,40252974,C,T	12,40283985,C,A	12,40295569,T,C	12,40310652,A,G	12,40322037,C,T	12,40351534,C,A	12,40367012,T,A	12,40369229,T,G	12,40243668,T,A	12,40283897,C,T	12,40309185,G,A	12,40321181,A,G	12,40351585,G,A	
12,40240604,C,T	12,40252978,C,T	12,40283990,T,C	12,40295574,A,C	12,40313966,T,C	12,40322037,C,G	12,40351537,A,G	12,40367032,T,C	12,40369247,A,G	12,40244579,A,T	12,40284011,G,T	12,40309206,C,T	12,40321181,C,A	12,40351667,C,A	
12,40295581,G,A	12,40309352,A,G	12,40305577,C,T	12,40313976,G,C	12,40322037,C,A	12,40351541,C,G	12,40367044,C,T	12,40208138,C,T	12,40245759,A,G	12,40284011,G,A	12,40309225,A,C	12,40321500,G,A	12,40351680,G,T		
12,40295581,G,T	12,40299190,C,T	12,40304000,G,A	12,40305781,T,A	12,40313976,G,A	12,40322038,G,T	12,40351572,T,A	12,40367044,C,T	12,40208422,T,G	12,40247890,A,G	12,40284061,A,T	12,40309350,C,A	12,40322037,C,T	12,40351680,G,C	
12,40298330,A,G	12,40299193,A,G	12,40304004,A,G	12,40305791,C,G	12,40314043,G,T	12,40322047,G,T	12,40351579,C,T	12,40367050,A,G	12,40208422,T,C	12,40249854,C,T	12,40284061,A,G	12,40310434,C,T	12,40322037,C,G	12,40351680,G,A	
12,40298330,A,C	12,40299196,A,G	12,40304035,A,G	12,40305871,G,A	12,40314043,G,A	12,40322047,G,A	12,40351580,G,T	12,40367664,G,A	12,40214089,A,T	12,40249854,C,A	12,40284114,T,C	12,40310434,C,G	12,40322037,C,A	12,40351723,A,G	
12,40298333,C,T	12,40299196,A,C	12,40304035,A,C	12,40305946,T,A	12,40314046,G,A	12,40322134,A,G	12,40351580,G,A	12,40367849,G,A	12,40214089,A,C	12,40249881,T,C	12,40287461,A,G	12,40310434,C,A	12,40322038,G,A	12,40356126,A,T	
12,40298346,G,A	12,40299212,G,A	12,40304040,G,T	12,40308462,A,G	12,40314059,C,T	12,40322137,A,C	12,40351585,G,A	12,40367861,T,C	12,40215078,C,T	12,40249923,G,T	12,40293624,G,C	12,40310435,G,A	12,40322047,G,T	12,40362670,C,T	
12,40298433,C,T	12,40299255,T,C	12,40304040,G,C	12,40308467,G,T	12,40314062,A,G	12,40322364,T,C	12,40351623,G,T	12,40367888,A,C	12,40216081,G,A	12,40249923,G,A	12,40293624,G,A	12,40310436,C,T	12,40322047,G,A	12,40362670,C,A	
12,40298433,C,G	12,40299266,T,G	12,40304078,G,C	12,40308467,G,C	12,40314101,C,T	12,40322386,G,T	12,40351667,C,A	12,40367913,T,C	12,40220632,C,T	12,40251273,G,A	12,40293644,A,G	12,40310437,G,C	12,40322386,G,T	12,40363440,C,T	
12,40298479,G,T	12,40302866,A,G	12,40304078,G,A	12,40308476,A,G	12,40314101,C,A	12,40322386,G,A	12,40351680,G,T	12,40367913,T,A	12,40222364,G,C	12,40251361,A,G	12,40294279,A,G	12,40310437,G,A	12,40322386,G,A	12,40363526,G,A	

Fig. 5.21 Initial DbSNP variants of LRRK2

provean		
Mutation		
L212S	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut24940/oupt.D.9	-2.38
L437S	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut12632/oupt.DECREASE.9	-2.25
I554T	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut12636/oupt.DECREASE.9	-2.12
L805R	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut12381/oupt.DECREASE.9	-1.76
I201Q	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut15325/oupt.DECREASE.8	-1.72
V229F	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut19714/oupt.DECREASE.9	-1.7
L582P	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut25203/oupt.D.6	-1.68
L1273V	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut24353/oupt.D.6	-1.46
L119P	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut19904/oupt.DECREASE.6	-1.45
K39M	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut25271/oupt.D.6	-1.44
R1633G	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut14633/oupt.DECREASE.8	-1.43
N1437H	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut13010/oupt.DECREASE.8	-0.7
P1262A	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut19982/oupt.DECREASE.8	-1.24
D478N	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut12878/oupt.DECREASE.7	-1.18
R194H	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut25286/oupt.D.8	-1.17
L1870F	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut21735/oupt.D.6	-1.16
R1728H	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut15071/oupt.DECREASE.9	-1.15
N226I	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut21951/oupt.DECREASE.5	1.01
R1725Q	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut14943/oupt.DECREASE.8	-1.09
L1795F	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut15186/oupt.DECREASE.6	-0.99
M237T	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut25251/oupt.D.6	-0.98
T1612A	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut13046/oupt.DECREASE.9	-0.95
R1628P	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut19935/oupt.DECREASE.7	-0.89
I1652V	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut15306/oupt.DECREASE.7	-0.85
C2139S	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut15340/oupt.DECREASE.6	-0.85
C1266S	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut24338/oupt.D.6	-0.83
G606D	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut22020/oupt.D.5	-0.82
Q847H	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut22122/oupt.D.8	-0.81
I810V	http://igor.biocomp.unibo.it/oldpredictor/submit/Mutant/Mut22026/oupt.D.7	-0.74
http://provean.jcvi.org/provean_seq_report.php?jobid=62240936887245		

Fig. 5.25. iMutant Suite results of LRRK2

Variant	PROVEAN score	Prediction (cutoff = -2.5)
L805R	-3.188	Deleterious
L1273V	-2.544	Deleterious
L119P	-2.666	Deleterious
N1437H	-4.138	Deleterious
P1262A	-3.607	Deleterious
L1870F	-2.855	Deleterious
R1728H	-2.57	Deleterious
L1795F	-3.422	Deleterious
T1612A	-2.811	Deleterious
R1628P	-2.736	Deleterious
C2139S	-4.948	Deleterious
G606D	-3.204	Deleterious
L212S	-2.148	Neutral
L437S	-2.379	Neutral

Fig. 5.26. PROVEAN results for LRRK2

META SNP																
Mutation	R	Pd-SNP	SIFT	SNAP	SNP	RI		Mutation	R	Pd-SNP	SIFT	SNAP	SNP	RI		
L119P	NA	Disease	Disease	Disease	Disease	4		R1628P	NA	Disease	Disease	Disease	Disease	3		
	-	0.822	0	0.72	0.68				-	0.543	0.03	0.73	0.636			
L212S	NA	Disease	Neutral	Disease	Neutral	3		R1728H	NA	Neutral	Disease	Disease	Disease	2		
	-	0.585	0.11	0.535	0.363				-	0.43	0	0.635	0.573			
L437S	NA	Disease	NA	Disease	Disease	3		L1735F	NA	Disease	Disease	Disease	Disease	3		
	-	0.773	-	0.585	0.644				-	0.581	0.03	0.64	0.644			
G606D	NA	Disease	Neutral	Disease	Disease	5		L1870F	NA	Neutral	Neutral	Disease	Neutral	3		
	-	0.856	0.13	0.705	0.728				-	0.431	0.08	0.63	0.331			
L805B	NA	Disease	Neutral	Disease	Disease	5		C2133S	NA	Neutral	NA	Disease	Neutral	2		
	-	0.311	0.16	0.73	0.731				-	0.483	-	0.51	0.373			
P1262A	NA	Neutral	Neutral	Disease	Neutral	7		R506Q	NA	Neutral	Neutral	Neutral	Neutral	7		
	-	0.075	0.5	0.58	0.128				-	0.433	0.1	0.23	0.153			
L1273V	NA	Disease	Disease	Disease	Disease	2		M1288I	NA	Neutral	Disease	Disease	Neutral	6		
	-	0.512	0.03	0.63	0.583				-	0.123	0.01	0.665	0.138			
N1437H	NA	Neutral	Disease	Disease	Neutral	4		F1632Q	NA	Neutral	Disease	Neutral	Neutral	6		
	-	0.413	0.05	0.68	0.311				-	0.098	0	0.405	0.173			
T1612A	NA	Neutral	Neutral	Neutral	Neutral	6		Y1645C	NA	Neutral	Neutral	Disease	Neutral	5		
	-	0.31	0.1	0.47	0.217				-	0.418	0.03	0.62	0.275			
K2116R	NA	Neutral	NA	Disease	Neutral	6		F1883L	NA	Neutral	Disease	Disease	Neutral	5		
	-	0.078	-	0.6	0.215				-	0.202	0.01	0.6	0.254			

META	GENE LEVEL	FINAL FILTERED
L119P	R506Q	L119P
L437S	A716V	L437S
L1273V	R1067Q	L1273V
R1628P	L1870F	R1628P
L1795F	M1288I	L1795F
	E1632Q	
	Y1645C	
	F1883L	
	K2116R	

Fig. 5.27. MetaSNP and final prime variants results for LRRK2

HOPE SERVER RESULTS:

1. L119P

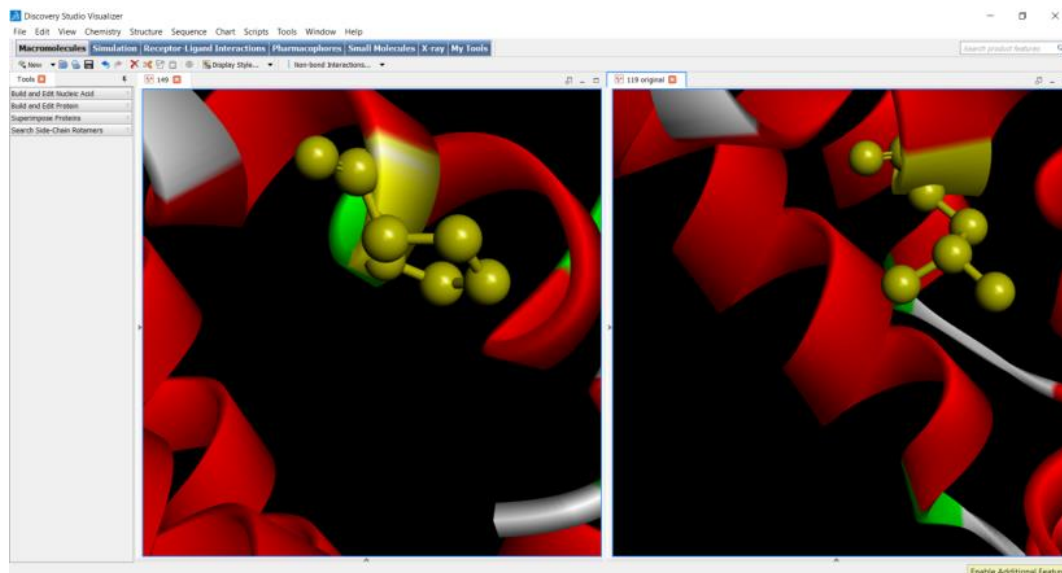


Fig. 5.28.1.

2. L437S

Conservation

Only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein.

The mutant and wild-type residue are not very similar. Based on this conservation information this mutation is probably damaging to the protein.

Your mutant residue is located near a highly conserved position.

Domains

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Armadiillo-Type Fold IPR016024	None	None
Armadiillo-Like Helical IPR011989	None	None

Amino Acid Properties

The wild-type and mutant amino acids differ in size.

The mutant residue is smaller, this might lead to loss of interactions.

The hydrophobicity of the wild-type and mutant residue differs.

Hydrophobic interactions, either in the core of the protein or on the surface, will be lost.

Images

Due to a lack of structural information, HOPE cannot generate pictures.

Fig. 5.28.2.

3. L1273V

Conservation

Only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein.

The mutant residue, however, has some properties in common with the wild-type residue. This mutation might occur in some rare cases, but it's more likely that the mutation is damaging to the protein.

Your mutant residue is located near a highly conserved position.

Domains

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Leucine Rich Repeat 4 IPR025875	None	None
Leucine-Rich Repeat, Typical Subtype IPR003591	None	None
Leucine-Rich Repeat Domain Superfamily IPR032675	None	None
Leucine-Rich Repeat IPR001611	Protein Binding GO:0005515	Molecular_Function GO:0003674 Binding GO:0005488

The mutated residue is located in a domain that is important for binding of other molecules. Mutation of the residue might disturb this function.

Amino Acid Properties

The wild-type and mutant amino acids differ in size.

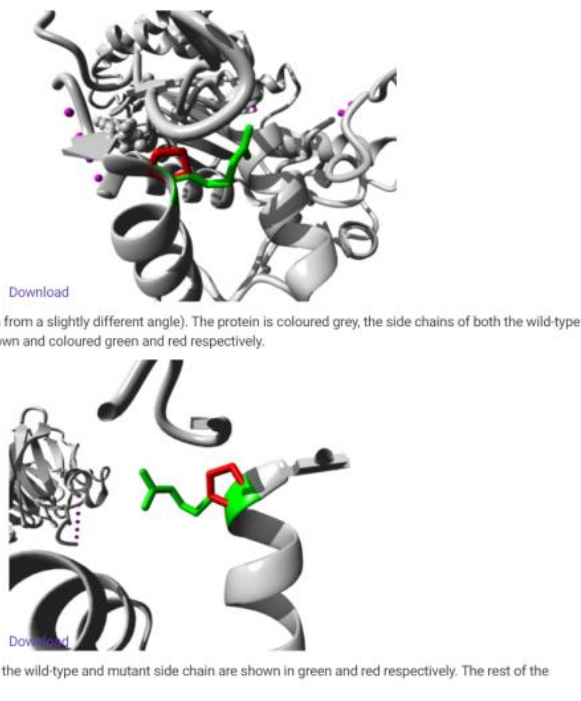
The mutant residue is smaller, this might lead to loss of interactions.

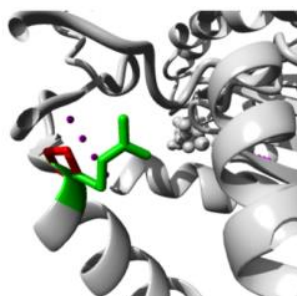
Images

Due to a lack of structural information, HOPE cannot generate pictures.

Fig. 5.28.3

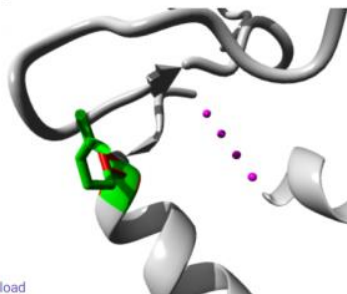
4. R1628P





[Download](#)

Close-up of the mutation. The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

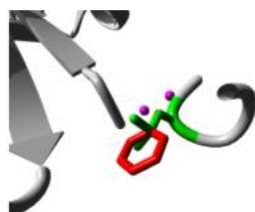


[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

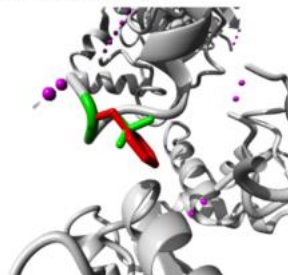
Fig. 5.28.4.

5. L1795F



[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.



[Download](#)

Close-up of the mutation (seen from a slightly different angle). The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively.

Fig. 5.28.5.

5.4 DISCUSSION:

5.4.1 PARK7:

Compared to other genes involved in Parkinson's disease, PARK7, also known as DJ-1 is less explored. DbSNP has around 7365 results for PARK7, with 11 being pathogenic and 2 being likely pathogenic and 140 of them are missense. As the input is taken for initial analysis, using CADD, DANN, FATHMM, FunSeq2, GWAVA; All the entries are individually sorted out for predictions on all the algorithms. MutationTaster2 is used to find if the change predicts any disease and found out 3 of them which haven't been reported in 1000G yet.

Starting with aa level variants from various sources, Rigorous filtering based on SIFT scores and then SNAP scores have been done to filter out for prime factor variants. Rigorous sorting is being done having the success rate of insilico application's success rate in mind. Both iMutant Suite gave Stability Decrease for 14 variants and PROVEAN results were deleterious for 12 variants. One Odd thing is the A107P mutant, which predicted a disease and is deleterious but is also having the protein's stability increased, by iMutant Suite. It should probably be an effect of the variants on C106 region, which is one of the most studied variant having multiple mutations into various residues as it is highly conservative, it causes severe damage to protein function & interaction. There should be a link between the C106 region's importance and Stability increase in A107 region which can be studied further. Later, the results of MetaSNP stated that 11 of them were harmful overall and they were subject to Homology Modelling. All the models were also inspected in PROCHECK algorithm in UCLA's SAVES v6.0 web server which resulted in about 90-95% of residues falling in core and allowed regions. HOPE server also gave extensive results on amino acid changes, stability changes, info about domain, region and more.

Out of all 11, the most affected by stability, function and other factors were mutations in C106 region(C106 ->A, D, S), L166P, R156W, P73L. C106 and L166P have been reported earlier in many studies, but insilico studies suggests even the R156 and P73

can also have dangerous impact on the protein and further studies employed with another set of tools are required to confirm them.

5.4.2 SNCA:

“SNCA” in DbSNP results over 26870 mutations and 133 of them are missense and only 5 of them were pathogenic. SNCA is reported to be a prime region for parkinson’s disease onset & progression as reported by various studies, but analyzing SNCA is a tough task as it is a disoriented protein and doesn’t have a stable structure. As the data been collected from DbSNP, Using PredictSNP2, all the algorithms have been acted on them to find out the status. Further resulted in 26 possible deleterious on rigorous filtering out of which only 3 of them were in exonic region & are non synonymous. Due to SNCA being dis oriented, MutationTaster2 couldn’t be used on to analyze if they are disease causing and have they been reported. Further, alternatives have been tried to use but in vain. This gives a scope to work further in future on specific tools targeting proteins like SNCA which are probably not properly understood and requires extensive studying.

From amino acid variations from various sources, SNAP server filtered 42 and based on SIFT with 5 percent effect threshold, 4 of them have been filtered. Those variants have been analyzed in iMutant and Provean, which both of them kind of gave mixed results which didn’t have any proper correlation as how it works generally for all other analysis in genes. Based on that, All the variants from gene level filtered and amino acid level filtered have been analyzed in metaSNP. Another tool by name PredictSNP1 was used, which incorporated MAPP, PhD-SNP, Polyphen scores to analyze how harmful the mutant is. HOPE server couldn’t be utilized to its maximum as it can only be helpful in visualization if there is a proper structure. All the variants have been homology modelled but PROCHECK in SAVES v6.0 had only 80-90% in allowed region, which is considered a borderline for analysing modelled proteins. Information regarding amino acid changes, interaction in domains and regions were studies in HOPE server.

Variant A30P has various effects on stability but it has been reported earlier in different studies, but E123K variant also has effect on protein due to bond formation and this should play important role in interactions with other parts and requires further analysis. Variant S9A didn't have major effects but had minimum effect predictions to disturb the region. Surprisingly, A variant K96I increases the stability by larger extent which is not the case in all other variants. This requires separate study on why it increases the stability at that region even being deleterious.

Other variants like E46K which had a bulky effect of change in structure and stability which also caused loss/ disruption of function in main region. The G51D structure gets bigger and changes the way the region functions, disturbances in interaction, stability degrade the potential but not very harmful as compared with other variants. G84V variant predicted as deleterious and harmful by various tools but didn't have a much effect while studying for changes in stability and didn't have negative effects due to variation, as predicted by various algorithms. Whereas the K96I mutant doesn't have major change in structure and its variation affects the least to structure compared to all the mutant structures

5.4.3 LRRK2:

LRRK2 had a total of 571 variants from dbSNP is gene level and after loading all into predictSNP2, filtered down to 178 and then rigorously filtered to 40 variants. When all of them were loaded into MutationTaster, 9 of them were never reported in 1000G and have never been studied further too.

Initially starting off with 216 amino acid level mutations, and only taking ones that are not validated, 56 of them were filtered and analyzed using iMutant and PROVEAN tools. Since there were lot of inputs, Reliability Index and the scores were taken into account to filter out only to important residues. 28 of them decreased stability. Further in provean, 14 came out as deleterious. Incorporating amino acid and gene level prime variants into MetaSNP, finally 5 of the variants which have been predicted

harmful and causing serious damage have been filtered and studied in HOPE server. PROCHECK also had above 86% in core region and more in allowed regions.

Out of these 5, R1628P variant seems to be the most harmful and serious change of all stability change analysis it has on its domain and region. Also the L1795F on the main region of LRRK2 will have serious effects on functionality based on predictions from these tools. LRRK2 has lot of domains to analyze for and requires more in depth studies on its effects and function. Unexpectedly, the variant N2261I shows massive increase in stability(increase as +1.01) which generally for a mutation would be negative. This is something which requires further analysis on what is causing this effect. For L119P location, the mutation disrupts the α -helix structure and so the whole region stability changes happen.

L437S mutant is present in armadillo domain and is highly conserved. Mutant and wild type have lot of structural differences and also lead to loss of interaction and proper functionality, stability. Mutant is smaller than wild-type and the variant is highly damaging. L1273V is present in repeats of LRRK, mutation is usually damaging as it is in conserved region but this unexplored region of this vast protein has lot of properties in common with the wild type and mutant, which indicates that this shouldn't affect the protein to much extent but it can also affect and lead to loss of interaction.

CHAPTER 6

CONCLUSION & FUTURE WORK:

Three of the most important genes have been analyzed using various databases and bioinformatics tools. The idea behind using numerous tools is that can be further be used to form a more structured protocol as most of the planned tools were under maintenance or non workable in project work period. So finally this structured workflow starting from gene level and amino acid level variants into joining the prime variants from different tools to further classify and then model and analyze them can be incorporated into a single click pipeline for mass analysis in other studies.

Analysis using this protocol resulted in 11 variants of PARK7, 7 in SNCA and 5 in LRRK2 of which few of them haven't ever been reported. In silico analysis of these genes have resulted in data that might effect the functioning in vivo which just cannot be confirmed by computational studies. Further correlation with other wet lab studies are required to match with all the computational results.

Some distinctive results in the study are increase in stability as detected by iMutant Suite in N226I in LRRK2, K96I in SNCA & A107P in PARK7 surprisingly increases with positive effect even being deleterious & harmful. The act at A107 can be explainable as C106 region being one of the most studied and very highly conservative and damaging position affecting stability and function and also been reported in various patients. The other 2 positions in 2 genes results need to be worked further onto why the increase in stability happens. By constructing a Structured Protocol from all the mentioned tools, This process can also be automated and possibly made simple by integrating into a pipeline. All the data obtained from the study can be possibly made into part of disease prediction analysis at gene level studies.

REFERENCES

1. Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiol Dis.* 2020 Apr;137:104782. doi: 10.1016/j.nbd.2020.104782. Epub 2020 Jan 25. PMID: 31991247; PMCID: PMC7064061.
2. Samii A, Nutt JG, Ransom BR. Parkinson's disease. *Lancet.* 2004 May 29;363(9423):1783-93. doi: 10.1016/S0140-6736(04)16305-8. PMID: 15172778.
3. Scholz SW, Mhyre T, Ransom H, Shah S, Federoff HJ. Genomics and bioinformatics of Parkinson's disease. *Cold Spring Harb Perspect Med.* 2012 Jul;2(7):a009449.
4. Sara Bandres-Ciga, Monica Diez-Fairen, Jonggeol Jeff Kim, Andrew B. Singleton, Genetics of Parkinson's disease: An introspection of its journey towards precision medicine, *Neurobiology of Disease*, Volume 137, 2020, 104782, ISSN 0969-9961,
5. Kitamura Y, Inden M, Kimoto Y, Takata K, Yanagisawa D, Hijioka M, Ashihara E, Tooyama I, Shimohama S, Ariga H. Effects of a DJ-1-Binding Compound on Spatial Learning and Memory Impairment in a Mouse Model of Alzheimer's Disease. *J Alzheimers Dis.* 2017;55(1):67-72. doi: 10.3233/JAD-160574. PMID: 27662308.
6. Dolgacheva LP, Berezhnov AV, Fedotova EI, Zinchenko VP, Abramov AY. Role of DJ-1 in the mechanism of pathogenesis of Parkinson's disease. *J Bioenerg Biomembr.* 2019 Jun;51(3):175-188. doi: 10.1007/s10863-019-09798-4. Epub 2019 May 3. PMID: 31054074; PMCID: PMC6531411.
7. Nevra Alkanli and Arzu Ay (January 8th 2019). The Relationship between Alpha-Synuclein (SNCA) Gene Polymorphisms and Development Risk of Parkinson's Disease, *Synucleins - Biochemistry and Role in Diseases*, Andrei Surguchov, IntechOpen, DOI: 10.5772/intechopen.82808.
8. Lesage S, Brice A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet.* 2009 Apr 15;18(R1):R48-59. doi: 10.1093/hmg/ddp012. PMID: 19297401.
9. Cardona F, Tormos-Pérez M, Pérez-Tur J. Structural and functional in silico analysis of LRRK2 missense substitutions. *Mol Biol Rep.* 2014;41(4):2529-42. doi: 10.1007/s11033-014-3111-z. Epub 2014 Feb 2. PMID: 24488318.
10. Daher JP. Interaction of LRRK2 and α -Synuclein in Parkinson's Disease. *Adv Neurobiol.* 2017;14:209-226. doi: 10.1007/978-3-319-49969-7_11. PMID: 28353286.

11. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001 Jan 1;29(1):308-11.
12. Bendl J, Musil M, Štourač J, Zendulka J, Damborský J, Brezovský J (2016) PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS Comput Biol* 12(5): e1004962. <https://doi.org/10.1371/journal.pcbi.1004962>
13. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008;24(24):2938-2939. doi:10.1093/bioinformatics/btn564
14. Liu, C., Zhang, F., Li, T. et al. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics* 13, 661 (2012). <https://doi.org/10.1186/1471-2164-13-661>
15. Schwarz, J.M., Cooper, D., Schuelke, M., & Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods*, 11, 361-362.
16. Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics (Oxford, England)*, 31(16), 2745–2747. <https://doi.org/10.1093/bioinformatics/btv195>
17. Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids research*, 31(13), 3381–3385. <https://doi.org/10.1093/nar/gkg520>
18. Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.*, 26, 283-291.
19. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 2010 Nov 8;11:548. doi: 10.1186/1471-2105-11-548. PMID: 21059217; PMCID: PMC2992548.
20. Accelrys Software, Discovery Studio Modeling Environment, Release 4.0, Accelrys Software, San Diego, Calif, USA, 2013, <http://accelrys.com/products/discovery-studio>.
21. Triarhou LC. Dopamine and Parkinson's Disease. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK6271/>

22. Inamdar, N. N., Arulmozhi, D. K., Tandon, A., & Bodhankar, S. L. (2007). Parkinson's disease: genetics and beyond. *Current neuropharmacology*, 5(2), 99–113. <https://doi.org/10.2174/157015907780866893>
23. Stefanis L. (2012). α -Synuclein in Parkinson's disease. *Cold Spring Harbor perspectives in medicine*, 2(2), a009399. <https://doi.org/10.1101/cshperspect.a009399>
24. Bereczki D. The description of all four cardinal signs of Parkinson's disease in a Hungarian medical text published in 1690. *Parkinsonism Relat Disord*. 2010 May;16(4):290-3. doi: 10.1016/j.parkreldis.2009.11.006. Epub 2009 Nov 30. PMID: 19948422.
25. Meade, R.M., Fairlie, D.P. & Mason, J.M. Alpha-synuclein structure and Parkinson's disease – lessons and emerging principles. *Mol Neurodegeneration* **14**, 29 (2019). <https://doi.org/10.1186/s13024-019-0329-1>
26. <https://www.parkinson.org/understanding-parkinsons/what-is-parkinsons>
27. <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Deep-Brain-Stimulation>
28. <https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/diagnosis-treatment/drc-20376062>
29. <https://www.genengnews.com/news/parkinsons-disease-treatment-may-be-unlocked-by-protein-structures/>