Genetic analysis of Parkinson's disease using Next generation sequencing.

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von
Anamika Giri
aus Pune, Indien

6 Juli, 2018.

Tag der mündlichen Prüfung: 19.03.2019

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel

Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: Prof. Dr. Thomas Gasser
2. Berichterstatter: Prof. Dr. Peter Heutink

Prüfungskommission:

Prof. Dr. Thomas Gasser

Prof. Dr. Peter Heutink

Prof. Dr. Ludger Schöls

Dr. Julia Schulze-Hentrich

Erklärung / Declaration:

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel: "Genetic analysis of Parkinson's disease using Next generation sequencing" selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled "Genetic analysis of Parkinson's disease using Next generation sequencing" submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den 6 Juli / July 2018

Datum / Date Unterschrift /Signature

Table of Contents

Summary

Neurological diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), Epilepsy and Multiple Sclerosis are included in the Global burden of disease study as these disorders have a high impact on public health. Lack of effective treatment has motivated the researchers to perform early diagnostics, by identifying new gene mutations, which can improve the therapies. The aim of this thesis was a genetic analysis of PD using next-generation sequencing data. In this thesis, whole genome sequencing (WGS) and whole exome sequencing (WES) using DNA from familial PD patients and healthy individuals was performed in order to identify the PD causal genes. A large repository of sporadic PD WES data and a genotyping array was used to replicate our findings. The PD patients from Germany were stratified for clinical trials on the basis of mitochondrial endo-phenotype by performing risk profiling of associated Single Nucleotide Polymorphisms (SNPs) using exome genotyping array. The sporadic PD WES and genotyping array data from International Parkinson's disease Genomics Consortium was used to perform association tests, to determine the burden of rare variants in candidate genes of interest. Furthermore, mRNA sequencing of all the genes under the PD GWAS loci after knockdown with short hairpin RNAs was performed, to identify the actual genes contributing to PD risk and the novel pathways involved in PD. Finally, an epistatic interaction of a Mendelian PD gene and associated locus was performed to understand the joint contribution to PD risk.

Taking everything into account, we identified pathogenic variants in known and some novel genes causing PD in families. On the basis of risk profiling some of the German PD patients will undergo clinical trials with coenzyme Q10 and vitamin K2. The association tests using sporadic PD data helped to identify some novel genes significantly associated with PD risk. The knockdown experiments facilitated the identification of genes contributing to PD risk in some of the PD GWAS loci.

Abbreviations

AAO Age at onset

AD Alzheimer's disease

AD Autosomal Dominant

ALS Amyotrophic Lateral Sclerosis

AMPK AMP-activated protein kinase

AR Autosomal Recessive

BAM Binary Alignment Map

BP Biological Processes

BST1 Bone Marrow Stromal Cell Antigen 1

BWA Burrows Wheeler Algorithm

CADD Combined Annotation Dependent Depletion

CAGE Cap analysis gene expression

CAMKK Calcium/calmodulin-dependent protein kinase kinase

CBD Cortico basal degeneration

CNV Copy Number Variant

DGKQ Diacylglycerol Kinase Theta

DGV Database of Genomic Variants

DJ1 Parkinsonism associated deglycase

DLB Dementia with Lewy bodies

DNA Deoxyribonucleic acid

ECM Extra Cellular Matrix

eQTL Expression Quantitative trait loci

ER Endoplasmic Reticulum

ExAC Exome Aggregation Consortium

FTDP-17 Frontotemporal dementia with Parkinsonism 17

GAK Cyclin G Associated Kinase

GATK Genome Analysis Tool Kit

GBA Glucosylceramidase Beta

GERP Genomic Evolutionary Rate Profiling

GS Gene Significance

g.VCF Genomic Variant Call Format

GWAS Genome wide association studies

HBD Homozygous by Descent

HWE Hardy Weinberg Equilibrium

IBD Identity by Descent

IBS Identity by State

IGV Integrative Genomics Viewer

IMPI Integrated Mitochondrial Protein Index

Indels Insertions Deletions

IPDGC International Parkinson's disease Genomics Consortium

KD Knockdown

LBs Lewy bodies

LD Linkage Disequilibrium

Lns Lewy neurites

logFC log Fold Change

LRRK2 Leucine-rich repeat kinase 2

MAF Minor Allele Frequency

MAPT Microtubule-associated protein tau

MPTP 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine

mRNA messenger RNA

MS Multiple Sclerosis

MSA Multiple System Atrophy

NBIA Neurodegeneration with brain iron accumulation

ncRNA non-coding RNA

NGS Next Generation Sequencing

NHGRI National Human Genome Research Institute

NUCKS1 Nuclear Casein Kinase and Cyclin Dependent Kinase Substrate 1

OR Odds Ratio

PARK2 Parkin RBR E3 Ubiquitin Protein Ligase

PCR Polymerase Chain Reaction

PD Parkinson's disease

PDGSC Parkinson's disease Genetics Sequencing Consortium

PINK1 PTEN-induced putative kinase 1 gene

PLA2G6 Phospholipase A2 Group VI

PM20D1 Peptidase M20 Domain Containing 1

PSP Progressive supranuclear palsy

QC Quality Control

RAB29 Member RAS Oncogene Family

RLE Relative Log Expression

RNA-seq Ribonucleic acid sequencing

ROH Runs of Homozygosity

ROS Reactive Oxygen Species

RSX1 Rotterdam Study

SHP1 Src homology region 2 domain-containing phosphatases-1

shRNA short hairpin RNA

SKAT Sequence Kernel Association Test

SKAT-O Optimal Sequence Kernel Association Test

SLC41A1 Solute Carrier Family 41 Member 1

SLC45A3 Solute Carrier Family 45 Member 3

SNCA alpha-synuclein

SNpc Substantia nigra pars compacta

SNPs Single Nucleotide Polymorphism

SNV Single Nucleotide Variant

SR Single Read

STAT1 Signal Transducer and Activator of Transcription 1

SV Structural Variant

TMEM175 Transmembrane Protein 175

TREM2 Triggering Receptor Expressed on Myeloid Cells 2

UCHL1 Ubiquitin C-Terminal Hydrolase L1

UPS Ubiquitin Proteasome System

UTR3 3-prime Untranslated Region

UTR5 5-prime Untranslated Region

VCF Variant Call Format

WES Whole Exome Sequencing

WGCNA Weighted Gene Co-expression Network Analysis

WGS Whole Genome Sequencing

Statement of Contribution

Project 1: Identification of risk variants in Parkinson's disease using whole exome and whole genome sequencing.

DNA samples of Parkinson's disease patients were provided by following members:

1. Cornelia van Duijn. Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands.
2. Ebba Lohmann. Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany & Behavioral Neurology and Movement Disorders Unit, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey.
3. Eduardo Tolosa. Movement Disorders Unit, Neurology Service, Hospital Clínic de Barcelona, Barcelona, Spain; Institute of Biomedical Research August Pi i Sunyer (IDIBAPS), Barcelona, Spain; Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Barcelona, Spain.
4. Pau Pastor. Movement Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Barcelona, Spain.
5. Matthew Farrer. Department of Medical Genetics, University of British Columbia. Centre for Applied Neurogenetics, Vancouver, Canada.
6. Stefano Goldwurm. A.O. Istituti Clinici di Perfezionamento. Milan, Italy.
7. Joachim Ferreira. Instituto de Medicina Molecular, Faculty of Medicine, University of Lisbon, Lisbon, Portugal.
8. Bob van Hilten. Movement Disorder Neurology at the Leiden University Medical Center (LUMC).
9. Bart van de Barrenburg. Radboudumc and Donders Institute.
10. Thomas Gasser. Hertie Institute for Clinical Brain Research, University of Tübingen & German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.

My Contribution to Project 1:

In this project, I analyzed the Whole Genome Sequencing (WGS) (180 samples) and Whole Exome Sequencing (WES) (20 samples) data from 126 families with Parkinson's disease and performed association tests using large repository of sporadic PD WES and genotyping array data provided by our collaborators from International Parkinson's disease Genomics Consortium (IPDGC). I performed

weighted gene co-expression network analysis using publicly available RNA sequencing data.

In project 1 the following papers were published:

"PLA2G6 Mutations Related to Distinct Phenotypes: A New Case with Early-onset Parkinsonism."
Anamika Giri, Gamze Guven, Hasmet Hanagasi, Ann-Kathrin Hauser, Nihan Erginul-Unaltuna, Basar Bilgic, Hakan Gurvit, Peter Heutink, Thomas Gasser, Ebba Lohmann and Javier Simón- Sánchez

My contribution: I analyzed the WGS data and identified the PD causal variant in the affected person. I wrote the manuscript.

"A novel homozygous DJ1 mutation causes parkinsonism and ALS in a Turkish family."
Hasmet A. Hanagasi, **Anamika Giri**, Ece Kartal, Gamze Guven, Başar Bilgiç, Ann-Kathrin Hauser, Murat Emre, Peter Heutink, Nazlı Basak, Thomas Gasser, Javier Simón-Sánchez, Ebba Lohmann

My contribution: I performed the bioinformatics analysis of WGS data and identified the causal variant in the affected person. The manuscript was written by Hasmet A. Hanagasi and me.

Project 2: Mitochondrial endo-phenotype of Parkinson's disease.

My Contribution to Project 2:

I performed risk profiling of PD associated SNPs in the German population as well as association tests using the sporadic PD WES and NeuroX data from the IPDGC.

Project 3: Transcriptomic analysis of genes under the PD GWAS loci.

My Contribution to Project 3:

I performed the analysis of mRNA sequencing data to investigate the genes under the PD GWAS loci. The association tests using the resequenced data provided by IPDGC was done by me.

Project 4: Genetic interaction of LRRK2 and PARK16 locus in Parkinson's disease using next generation sequencing data.

My Contribution to Project 4:

I performed epistatic interaction analysis using the sporadic PD WES data from the IPDGC.

General Introduction

1.1 Parkinson's disease

Parkinson's disease (PD) was first described as 'shaking palsy' by James Parkinson in 1817 (Parkinson 2002). It is the second most common neurodegenerative disease affecting up to 1% of people above 65 years of age. Parkinson's disease, like other neurodegenerative diseases, is considered to be one of the major medical challenges in the society, as there are no treatments that can stop the degenerative process. The average age at onset of PD is approximately 70 years and is influenced by both genetic and environmental factors; hence it can be characterized as a complex polygenic disorder. Approximately, 10% of the cases are familial, and the remaining 90% belongs to a sporadic form of PD (Papapetropoulos, Adi et al. 2007).

The main clinical features of PD are the slowness of movements, muscle rigidity and tremor. Mild symptoms such as hyposmia, loss of dexterity, stiffness, dragging of a foot, slowness of movements and sleep disturbances are often misinterpreted or unnoticed by the patient (Jankovic 2008). In the later stages of PD, a speech of the patient is monotonous and slightly slurred, the face is expressionless, and the posture is with pill-rolling tremor of the hand (Kalia and Lang).

Diagnosis of PD requires autopsy which reveals PD's pathognomonic hallmark: Lewy bodies (LBs). LBs are abnormal protein aggregates inside the nerve cells. There is some heterogeneity depending on the distinct aetiology of PD. The postmortem PD brains typically reveal significant neuronal death in the

substantia nigra pars compacta (SNpc) with LB inclusions permeating surviving neurons (Dickson 2018). In addition to the formation of LBs, alpha-synuclein also accumulates in neuronal processes, called Lewy neurites (Lns). These are observed either in the brainstem (classical LBs) or in the cerebral cortex (which are more frequent in demented PD patients) and are routinely detected by α-synuclein antibodies (Dickson 2018). Pale bodies precede the formation of LBs (Spillantini, Schmidt et al. 1997). Interestingly, α-synuclein seems to follow a consistent six-stage spreading pattern starting from the olfactory bulb in early stages and resulting in widespread distribution affecting even the cerebral cortex in late stages of the disease (Braak, Del Tredici et al. 2003). α-synuclein accumulation is not specific to PD as it is observed in a number of other diseases including Multiple System Atrophy (MSA), certain forms of neurodegeneration with brain iron accumulation (NBIA) and Alzheimer's disease (AD), and even in aged healthy individuals (Dickson, Fujishiro et al. 2008, Lees, Hardy et al. 2009); however, accumulation patterns and subcellular distributions differ in each disease (Stefanis 2012). Identification of the reasons underlying these variable pathological features could contribute to the elucidation of the pathogenetic pathways of the disease.

1.2 Epidemiology of PD

Prevalence refers to the proportion of a population with a particular disease at a given time point (von Campenhausen, Bornschein et al. 2005). The prevalence of PD in Europe ranges between 100 and 300 cases per 100,000 population.

Previously published studies on the meta-analysis of prevalence show that it rose from 107 cases per 100,000 between ages 50 and 59 years to 1087 cases per 100,000 between 70 to 79 years (Pringsheim, Jette et al. 2014).

The term incidence refers to the number of new cases in a particular population over a period, usually one year (von Campenhausen, Bornschein et al. 2005). The incidence of PD has been estimated between 10 and 50 cases per 100,000 population in Europe (Elbaz, Carcaillon et al. 2016). However, the number of PD patients is most likely to double by 2030, due to increasing aging (Dorsey, Constantinescu et al. 2007). PD is more common in men than in women. Prevalence and incidence of PD are lower in Asian population than in European and American population (Bolger, Lohse et al. 2014).

1.3 Etiology of PD

In the 1980s, it was observed that a neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) damaged the neurons of the SNpc, resulting in PD symptoms (Langston, Ballard et al. 1983). Since that time, several studies have reported pesticides as a risk factor of PD (Betarbet, Sherer et al. 2000). Previous epidemiological studies examined twins based on cross-sectional information and suggested an absence of genetic factors (Wirdefeldt, Gatz et al. 2004). However, a few years later, variants were identified in many genes in families with an autosomal dominant and recessive mode of inheritance (Lesage and Brice 2009). Wirdefeldt et al. applied a longitudinal design and re-evaluated the previous cross-sectional data and reported that PD is heritable (Wirdefeldt, Gatz

et al. 2011). Hence, the etiology of PD involves both environmental and genetic factors.

1.3.1 Mendelian PD genes:

Although most of the PD cases occur in a sporadic form, in a subset of PD cases the disease is inherited in an autosomal dominant or autosomal recessive manner. Several genes including *SNCA*, *LRRK2*, *PARK2*, *PINK1, DJ1* and *VPS35* are linked to Mendelian PD and have provided an insight of molecular pathways underlying neurodegeneration.

***SNCA*:** In 1996, Polymeropoulos et al. identified that markers on chromosome 4q21-q23 were linked to PD phenotype in a large, multigenerational family termed "Contursi kindred" (Polymeropoulos, Lavedan et al. 1997) of Italian descent with an autosomal dominant mode of inheritance. One year later, the same group identified a missense *SNCA* mutation p. A53T in the same Italian family. The mutation p. A53T also segregated in three Greek kindreds with PD phenotype. Subsequently, two new *SNCA* mutations, p. A30P (Kruger, Kuhn et al. 1998) and p.E46K (Zarranz, Alegre et al. 2004) were reported in other PD families. Additionally, a p. H50Q *SNCA* mutation was reported in two sporadic PD cases sharing a common haplotype on the *SNCA* locus (Appel-Cresswell, Vilarino-Guell et al. 2013) (Kiely, Asi et al. 2013), and a p. G51D mutation in two familial cases (Kiely, Asi et al. 2013, Lesage, Anheim et al. 2013). Interestingly, whole gene duplications (Chartier-Harlin, Kachergus et al.) and triplications (Singleton, Farrer et al. 2003) also cause PD and severity and age at onset of disease seem to correlate with the number of copies of *SNCA*, indicating a gene-dosage effect

(Fuchs, Nilsson et al. 2007). *SNCA* mutation carriers tend to commonly present with dementia and hallucinations, which represents a similar situation to dementia with Lewy bodies (DLB) (Houlden and Singleton 2012). Shortly after the identification of *SNCA* mutations as a cause of PD, α-synuclein was identified as the major component of Lewy bodies, the pathological hallmark of PD (Spillantini, Schmidt et al. 1997) thus linking genetics and apparently sporadic forms of the disease (Houlden and Singleton 2012). In neuropathology, there is always widespread α-synuclein accumulation with occasional tau deposition in patients with *SNCA* mutations (Poulopoulos, Levy et al. 2012). In a proportion of cases, α-synuclein forms oligodendroglial inclusions that are similar to those seen in MSA (Markopoulou, Dickson et al. 2008, Obi, Nishioka et al. 2008, Kiely, Asi et al. 2013). *SNCA* plays a significant role in pre-synaptic signaling and membrane trafficking (Bendor, Logan et al. 2013).

***LRRK2* (*PARK8*):** In 2002, Funayama and colleagues performed two-point parametric linkage analysis in a Japanese family, which was already reported in 1997 with an autosomal dominant parkinsonism. They identified a new locus for PD, mapping to chromosome 12p11.2-q13.1 (Funayama, Hasegawa et al. 2002). Two years later, Paisán-Ruíz et al. identified a heterozygous mutation (p. R1441G) in *LRRK2* in PD patients of 4 families from the Basque region of Spain (Paisán-Ruíz, Jain et al.). Moreover, the same group also identified a heterozygous mutation (p. Y1699C) in PD patients from an English family (Paisán-Ruíz, Jain et al.). In the same year, Zimprich and colleagues identified heterozygous (p. R1441C) mutations in *LRRK2* in a family with an autosomal dominant PD (Zimprich, Biskup et al.). Some cases of discordant monozygotic

twins carrying the same *LRRK2* mutation have been reported in the literature (Xiromerisiou, Houlden et al. 2012). The most frequent mutation (p. G2019S) in *LRRK2* is identified in 20% of Ashkenazi Jewish patients, 40% of Arab ancestry and 1-7 % of PD patients of European ancestry (Lesage, Dürr et al. 2006, Ozelius, Senthil et al. 2006). Clinical features of *LRRK2* mutation carriers appear similar to sporadic PD with an onset of 50-60 years but with a predominance of tremor and dystonia (Healy, Falchi et al. 2008). In a majority of the cases, *LRRK2* mutations are associated with LB pathology. However, tau, TDP43 and ubiquitin positive inclusions are also occasionally reported (Rajput, Dickson et al. 2006).

LRRK2 also known as dardarin consists of 51 exons and encodes leucine-rich repeat kinase 2, a large protein consisting of an ankyrin repeat region, a leucine-rich repeat (LRR) domain, a kinase domain, a DFG-like motif, a RAS domain, a GTPase domain, a MLK-like domain and a WD40 domain (Bolger, Lohse et al. 2014). Dardarin has an enzyme function known as kinase activity and is involved in the process of phosphorylation. *LRRK2* regulates autophagy through a calcium-dependent activation of the CaMKK/AMPK signaling pathway and also plays a role in synaptic vesicle trafficking (Bolger, Lohse et al. 2014).

***PRKN* (*PARK2*):** In 1998, Kitada et al. identified a homozygous deletion in *PARK2* in a nuclear consanguineous Japanese family with an autosomal recessive juvenile parkinsonism (Kitada, Asakawa et al. 1998). *PARK2* mutations were identified by linkage analysis and consist of homozygous, compound heterozygous and heterozygous mutations, as well as exonic deletions and duplications. Most of these mutations have been identified in familial cases and some have also been identified in sporadic PD cases (Clark, Afridi et al. 2006). In

2007, Kay and colleagues reported that heterozygous mutations in *PARK2* do not contribute to PD (Kay, Moran et al. 2007). An early onset PD patient carries *PARK2* mutations often before 45 years of age. *PARK2* mutation carriers manifest a consistent response to levodopa treatment, and motor dysfunction progressively declines in patients at a young age. Clinical features of *PARK2*-linked PD are focal dystonia, psychosis, early instability or atypical L-dopa induced dyskinesia (Chan, Mok et al. 2008). Pathological alterations are usually restricted to the SNpc and LBs are uncommon (Poulopoulos, Levy et al. 2012, Doherty, Silveira-Moriyama et al. 2013).

PARK2 spans 1.38 Mb and encodes the protein, Parkin. It functions as an E3 ubiquitin ligase, which plays a major role in ubiquitin proteasome system (UPS) by tagging abnormal proteins for degradation (Chan and Chan 2011). *PARK2* also plays a significant role in autophagy of impaired mitochondria (Fiesel, Caulfield et al. 2015).

***PINK1* (*PARK6*):** Homozygosity screening in a *PARK2* negative large Marsala kindred from Sicily, identified the gene PTEN induced putative kinase 1 (*PINK1*) (Valente, Bentivoglio et al. 2001). Three years later, the same group identified two homozygous mutations (p. G309D and p. W437X) in a consanguineous Spanish family and two Italian families, respectively (Valente, Abou-Sleiman et al. 2004). *PINK1* is the second most frequent cause of early-onset PD (Hatano, Li et al. 2004, Valente, Abou-Sleiman et al. 2004, Piccoli, Ripoli et al. 2008). Moreover, the heterozygous mutations are speculated to be involved in the development of PD (Klein, Djarmati et al. 2005, Choi, Woo et al. 2008). Neuropathological features of *PINK1* mutation carriers are similar to those of

patients with *PARK2* mutations, and also the age at onset of the patients is less than 50 (Poulopoulos, Levy et al. 2012).

PINK1 encodes serine/threonine protein kinase and is localized in the mitochondria. It plays a major role in protecting cells from stress-induced mitochondrial dysfunction (Matsuda, Kitagishi et al. 2013).

***DJ1* (*PARK7*):** In 2003, Bonifati and colleagues performed homozygosity mapping and positional cloning in two consanguineous families from a genetically isolated population in the Netherlands and identified a homozygous deletion of exons 1 to 5 in *DJ1* (Bonifati, Rizzu et al. 2003). Moreover, a missense mutation (p. L166P) in *DJ1* was identified in an Italian family with an autosomal recessive PD (Bonifati, Rizzu et al. 2003). Mutations in *DJ1* are a rare cause of autosomal recessive PD (Abou-Sleiman, Healy et al. 2003, Bonifati, Rizzu et al. 2003, Hering, Strauss et al. 2004). Age at onset of PD patients is less than 30 years. They are responsive to levodopa and show atypical features such as dysarthria and myoclonic jerks. *DJ1* mutations are roughly characterized by 0.8% of familial and 0.4% of sporadic PD cases (Kilarski Laura, Pearson Justin et al. 2012). The neuropathology related to *DJ1* mutations has not been studied to date.

***VPS35*:** In 2011, Vilariño-Güell et al. performed exome sequencing in a multigenerational Swiss family with an autosomal dominant mode of inheritance and identified a heterozygous mutation p. D620N in *VPS35*. However, the *VPS35* p. D620N variant exists not only in familial PD cases but also in sporadic PD cases (Vilariño-Güell, Wider et al.). The clinical presentation of *VPS35* mutation

carriers resembles individuals with classic late-onset, levodopa-responsive PD. *VPS35* encodes a subunit of the retromer cargo recognition complex and serves as a key player in endosomal-lysosomal trafficking (Mukadam and Seaman 2015).

1.3.2 Other genetic risk factors

***GBA*:** Homozygous mutations in *GBA* causes a lysosomal storage disorder known as Gaucher's disease (Tsuji, Choudary et al. 1987, Koprivica, Stone et al. 2000). Major risk factors in PD are heterozygous *GBA* mutations (O'Regan, deSouza et al. 2017). Approximately 300 variants including insertions, deletions, frameshift and splice site mutations have been reported in this gene (Hruska, LaMarca et al. 2008). However, the two most common mutations in *GBA* are p. N370S and p. L444P (Tan, Tong et al. 2007, Gutti, Fung et al. 2008, Mata, Samii et al. 2008, Neumann, Bras et al. 2009). The frequency and prevalence of the mutations in *GBA* vary among different ethnicities (Mitsui, Mizuta et al. 2009). Age at onset for a patient with *GBA* mutation is before 50 years and is likely to develop dementia than sporadic PD patients. They show a good response to Levodopa. Recently, *GBA* mutations also increase the risk for DLB, thus showing a link between lysosomal biology and α-synuclein (Goker-Alpan, Giasson et al. 2006).

Published reports suggest that the loss of Glucocerebrosidase enzyme activity impairs the lysosomal function and results in endoplasmic reticulum stress, which contributes to the PD pathogenesis (Fernandes, Hartfield et al. 2016).

***MAPT*:** Non-synonymous mutations in *MAPT* cause fronto-temporal dementia with Parkinsonism (FTDP-17), which includes Pick's disease (Hutton, Lendon et al. 1998). The *MAPT* locus on chromosome 17 includes approximately 900 kilobase inversion that leads to 1.3 megabase linkage disequilibrium region named as H1 and H2 haplotypes (Allen, Kachadoorian et al. 2014). The most common *MAPT* H1 haplotype has been linked not only to Alzheimer's disease (AD), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) but also to PD (Pascale, Di Battista et al. 2016). Moreover, GWAS has confirmed H1 haplotype as a risk factor for PD (Kay, Moran et al. 2007). Neuropathology of PD demonstrates the presence of tau, localized within glial and neuronal inclusions (Cookson, Hardy et al. 2008). Although the absence of LB does not exclude PD in the differential diagnosis, it suggests pathogenic *MAPT* mutations are not responsible for typical PD and that it is likely that the pathological mechanism underlying FTPD-17 is distinct from that in typical PD (Coakeley and Strafella 2017). *MAPT* encodes the microtubule-associated protein tau, whose transcript undergoes complex alternative splicing and generates six different isoforms. It plays a major role in microtubule assembly and stability, axonal transport, cellular signaling and protein fibrilization (Zhang, Xing et al. 2016).

1.4 Genome wide association study (GWAS)

Risch and Merikangas formulated GWAS in 1996 (Risch and Merikangas 1996). In a GWAS study, statistical comparison of allele frequencies between cases and controls is performed, and it should include more than 1000 cases and controls and >300,000 markers (Balding 2006). All GWAS completed till date are

published in the GWAS catalogue (https://www.ebi.ac.uk/gwas/). GWAS has the power to study the genetic basis of disease by assaying the entire genome. However, several criticisms have been put forward for the usage of GWAS. Firstly, some of the risk loci identified through GWAS has small effect size and are not clinically useful as they fail to explain disease risk; nevertheless, these loci have shed light on the biological pathways involved in disease pathogenesis. Moreover, risk variants with odds ratio less than 1.5 could be false positives due to population stratification. Finally, the risk variants identified through GWAS are usually present in the non-coding region, which does not have a clear function, though they could act as proxies for the causative variant (Visscher, Wray et al. 2017).

GWAS in PD:

The first PD GWAS was performed in 2006 with an underpowered study of 537 cases and controls that failed to identify any risk loci (Maraganore, de Andrade et al. 2005, Fung, Scholz et al. 2006). In 2009, two independent GWAS were performed in the Caucasian and Asian population with a larger sample size (Satake, Nakabayashi et al. 2009, Simon-Sanchez, Schulte et al. 2009). *LRRK2*, *SNCA* and *PARK16* were significantly associated with PD in both studies. *BST1* and *MAPT* were linked to PD in the Asian and Caucasian study, respectively. In two independent studies, two new risk loci, *GAK*/*DGKQ* and HLA region were identified and the role of *SNCA* and *MAPT* in PD susceptibility were supported (Pankratz, Wilk et al. 2009, Hamza, Zabetian et al. 2010). Furthermore, a large-scale meta-analysis of 13,708 cases and 95,282 controls shows that 24 risk loci are significantly associated with PD risk (Nalls, Bras et al. 2015). The *SNCA* and *LRRK2* locus has been continuously detected in GWAS, confirming they are

related to both familial and sporadic PD. In 2016, Lesage and colleagues reported that a rare deleterious variant in *VPS13C* is a cause of autosomal recessive early-onset parkinsonism (Lesage, Drouet et al. 2016). This locus has also been associated with sporadic PD as per largest meta-analysis of 2014 (Nalls, Pankratz et al. 2014). The *PARK16* locus contains five protein-coding genes (*SLC45A3*, *NUCKS1*, *RAB29*, *SLC41A1* and *PM20D1*). Several lines of evidence show that a molecular interaction between *RAB29* and *LRRK2* may be associated with PD (MacLeod, Rhinn et al. 2013). The *GAK*/*DGKQ* locus contains protein-coding genes including *GAK*, *TMEM175* and *DGKQ*. *GAK* contributes to modify α-synuclein expression and toxicity in PD (Dumitriu, Pacheco et al. 2011). Moreover, it also plays a major role in regulating clathrin-mediated membrane trafficking (Zhang Claire, Engqvist-Goldstein Åsa et al. 2005).

Recently, Chang et al. carried out a GWAS of 6,476 PD cases and 302,042 controls of European ancestry and identified 17 new PD risk loci in a joint analysis of 26,035 PD cases and 403,190 controls (Chang, Nalls et al. 2017) (figure 1).

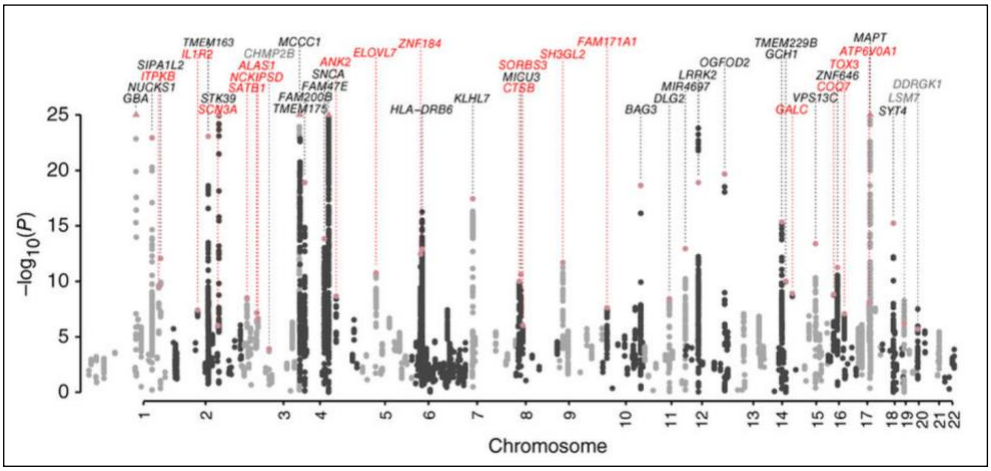


Figure 1: Manhattan plot from a GWAS meta-analysis for Parkinson's disease. Reproduced from Chang et al. 2017.

1.5 Pathways in Parkinson's disease

In both sporadic and hereditary form of PD, there is a loss of dopamine neurons in the SNpc. Moreover, impairment of endoplasmic reticulum (ER) function, protein degradation pathways, intracellular trafficking and calcium signaling enhances the degeneration of dopamine neurons (Michel, Hirsch et al. 2016). Down-regulation of autophagy leads to inclusion bodies, *i.e*. accumulation of aberrant proteins, contributing to the pathogenesis of neurodegenerative diseases. Two *SNCA* mutations p. A30P and p. A53P have been involved in the deterioration of Chaperone mediated pathway (Sala, Marinig et al. 2016). *LRRK2* plays a role in signaling pathways such as vesicle trafficking, mitochondrial function as well as endocytosis and autophagy. Two frequent *LRRK2* mutations p. G2019S and p. R1441C increases mitochondrial autophagy and accumulation of autophagic vacuoles, resulting in degeneration of dopaminergic neurons (Esteves and Cardoso 2016).

Endoplasmic reticulum (ER) plays a role in cellular homeostasis by maintaining proper protein folding and quality control. Aggregation of α-synuclein disrupts ER-Golgi vesicular trafficking and leads to ER stress. Furthermore, when Parkin is downregulated, it increases ER stress induced mitochondrial dysfunctions (Jiang, Gan et al. 2010). Hence, in order to prevent neurodegeneration and preserve normal physiology, it is crucial to prevent ER-mitochondria impairment, so that the calcium transfer between the two compartments is not reduced.

The major sources of oxidative stress such as mitochondrial dysfunction, impairment of protein degradation pathway and neuroinflammation are discussed below.

1.5.1 Mitochondrial dysfunction

MPTP is oxidized to MPP+ and inhibits the activity of complex I of the mitochondrial electron transport chain in the drug abusers (Langston, Ballard et al. 1983). Complex I inhibition increases the ROS production. In 1990, Schapira et al. have reported complex I deficiency in the SNpc of patients with sporadic PD (Schapira, Cooper et al. 1990).

Mutations in *PINK1* and *PARK2* are identified in patients with an autosomal recessive PD (Kitada, Asakawa et al. 1998) (Valente, Abou-Sleiman et al. 2004). Both these genes play a major role in mitophagy, *i.e.* degradation of impaired mitochondria. *PARK2* deficiency leads to oxidative stress, which in turn, impairs mitochondrial function (Muftuoglu, Elibol et al. 2004). Animal model's studies show that deficiency of *PINK1* leads to loss of dopamine neurons in SNpc, abnormal mitochondrial morphology, inhibition of complex I activity, and increase in oxidative stress (Clark, Dodson et al. 2006, Gautier, Kitada et al. 2008). *DJ1* binds to the mitochondrial complex I subunit and regulates its activity (Hayashi, Ishimori et al. 2009). It has been shown to be more protective against oxidative stress-induced cell death. A published study has reported a loss of nigrostriatal dopamine neuron in *DJ1* knockout mouse (Goldberg, Pisani et al. 2005). Moreover, this *DJ1* knockout mouse showed altered mitochondrial

morphology and respiration as well as reduced membrane potential and accumulation of impaired mitochondria (Goldberg, Pisani et al. 2005, Krebiehl, Ruckerbauer et al. 2010).

1.5.2 Impairment of protein degradation pathway

Protein misfolding and aggregation are common mechanisms in most of the neurodegenerative diseases including PD. A pathological hallmark of PD is an aggregation of α-synuclein. Ubiquitin Proteasome system (UPS) plays a major role in the removal of abnormal proteins. Mutations in PD-related genes increases oxidative stress, which impairs the function of UPS that degrades damaged and misfolded proteins (Blesa, Trigo-Damas et al. 2015). The first genetic link for involvement of UPS in neurodegeneration was provided by the discovery of variants in *PARK2* in a family with an autosomal recessive juvenile parkinsonism (Kitada, Asakawa et al. 1998). This gene encodes E3 ubiquitin ligase, which plays a significant role in UPS by modulating 26S proteasome activity (Chan and Chan 2011). Furthermore, mutations in *UCHL1* in a family with PD has provided additional support for the role of UPS in neurodegeneration (Leroy, Boyer et al. 1998). This gene encodes Ubiquitin C-Terminal Hydrolase L1 and is involved in the processing of ubiquitinated proteins.

1.5.3 Neuroinflammation

Neuroinflammatory mechanism, which is primarily controlled by activated microglia, contributes to the dopaminergic cell death in PD. Activated microglia have been found in the olfactory bulb and in the SNpc of both familial and sporadic PD patients (Le, Wu et al. 2016). Environmental toxins can shift microglia to an over-activated state and release ROS which can cause neurotoxicity (Blesa, Trigo-Damas et al. 2015). α-synuclein released from the neuronal cells activates the microglial inflammatory response. Mutations in *LRRK2* contributes to neurotoxicity by increasing the proinflammatory cytokine release from activated microglial cells (Gillardon, Schmid et al. 2012). Loss of function mutations in *PARK2* leads to inflammation-related degeneration of dopamine neurons (Frank-Cannon, Tran et al. 2008). *DJ1* negatively regulates inflammatory response of microglia and astrocytes by making easy the interaction between STAT1 and SHP1 (Kim, Choi et al. 2013).

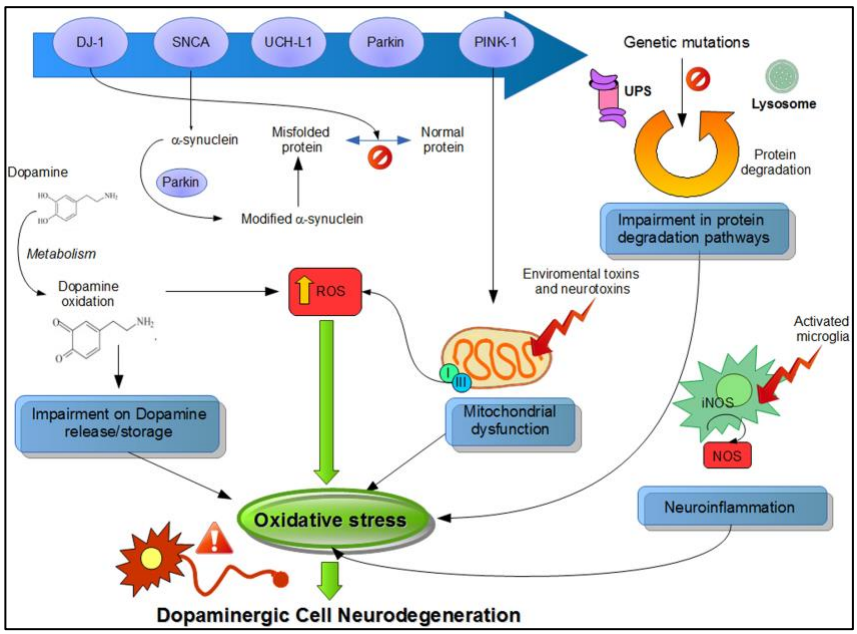


Figure 2: Pathways in Parkinson's disease. Reproduced from Blesa et al. 2015.

1.6 Next generation sequencing:

DNA sequencing was first established by Sanger in 1970s (Sanger and Coulson 1975). However, the technique was expensive and too laborious for the sequencing of the whole genome. Over the past decade, second generation sequencing has been developed and have reduced the time and cost required for sequencing. Moreover, Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) have enhanced identification of genetic defects in rare diseases (Saunders, Miller et al. 2012, Turro and BioResource 2016). Furthermore, genetically heterogeneous diseases can be diagnosed by simultaneously sequencing panel of genes associated with the disorder.

Whole Exome/Genome sequencing analysis:

The most important question for familial analysis is to use either WGS or WES. WGS is the most relevant option as it allows interrogating SNVs, indels, structural variants and copy number variants in both coding as well as a non-coding region of the genome (Gilissen, Hehir-Kwa et al. 2014). In the case of WES, reads are targeted to only protein-coding regions, so they represent less than 2% of the genome. WES can be useful in large population studies as its cost is less than WGS and it allows increasing the sample number. Moreover, the high coverage in WES helps to detect variants more accurately (Biesecker, Shianna et al. 2011). After sequencing, all the short sequence reads are mapped to the human reference genome. The accuracy of variant detection depends on the depth of coverage and DNA quality. Several tools have been developed to detect

SNVs, small INDELs as well as copy number variants like large deletions, insertions or structural variants such as inversions and chromosomal rearrangements. The identified variants can be annotated using several tools such as KGGSEQ (Cingolani, Platts et al. 2012), ANNOVAR (Wang, Li et al. 2010), and SNPEFF (Cingolani, Platts et al. 2012). Around 3 to 4 million SNVs and structural changes (deletions, insertions, inversions) exist in a genome of an individual. Moreover, approximately 20,000 variants are present in the coding regions, and an estimated 10,000 variants are non-synonymous, *i.e.* they change the protein sequence (Biesecker, Shianna et al. 2011). Hence, it is quite challenging to identify the variants relevant to the trait. However, 90% of the detected non-synonymous variants are common, *i.e.* their frequency is >1% in the population. These common variants are filtered out, as they are not often disease causal in rare Mendelian disorders. These rare genetic variants are functionally annotated by predicting their probable consequences at the protein level. Several prediction tools such as SIFT (Kumar, Henikoff et al. 2009), Polyphen (Chun and Fay 2009), GERP (Davydov, Goode et al. 2010), SNAP, CADD (Kircher, Witten et al. 2014), PhyloP (Pollard, Hubisz et al. 2010), and MutationTaster have been developed to distinguish deleterious variants from neutral ones and hence to determine the impact of non-synonymous SNVs on protein function. This in-silico prediction is based on previous knowledge from databases and it includes analysis such as conservation among species, splicing predictions and biochemical properties of the amino acids. Moreover, a genetic variant has a causal role in a patient only if it is present in all affected members and/or is carried by unaffected family members. Hence, family co-segregation using Sanger sequencing is often required, confirming variants of interest.

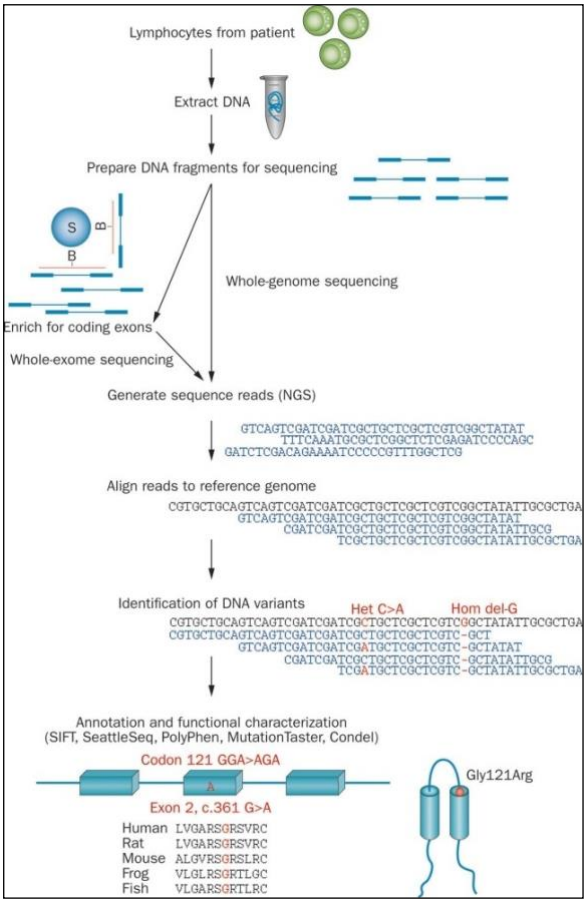


Figure 3: Whole genome/exome sequencing data analysis. Reproduced from Foo et al, 2012.

1.7 Aims of the thesis:

Advanced genomic technologies have illuminated the genetic underpinnings of devastating neurological disorders such as Parkinson's disease (PD), Alzheimer's disease (AD) and Amyotrophic Lateral Sclerosis (ALS). Such discoveries have been facilitated by genome-wide association studies (GWAS) and next-generation sequencing (NGS) investigations, which allow the identification of common and rare variants contributing to disease risk.

Project 1: The aim was to identify the causal gene by investigating WGS/WES data from familial PD samples (discovery cohort). Further, to screen the candidate genes from discovery cohort in a large repository of WES sporadic PD data from Parkinson's disease Genetics Sequencing Consortium (PDGSC). Moreover, to perform genetic replication by investigating the burden of rare variants in sporadic PD cases compared to controls using WES data and a genotyping array NeuroX. Finally, to perform weighted gene co-expression network analysis using publicly available RNA-sequencing data generated using prefrontal cortex tissues from postmortem PD patients.

Project 2: Several published studies have reported that impairment of mitochondrial function increases oxidative stress and promotes aging and neurodegeneration. Mitochondrial dysfunction plays a major role in PD pathogenesis. The goal of this project was to stratify PD patients for medical trials on the basis of mito-endophenotype. Additionally, to determine the burden of rare variants in mitochondrial-related genes in sporadic PD cases compared to controls.

Project 3: GWAS in PD has identified many loci implicated in disease risk. However, most of the genes responsible for this risk are still unknown. This project aimed to identify the gene(s) from each risk loci, which is responsible for the identified risk. In order to achieve this goal, genes under the GWAS loci were knockdown by shRNA and followed by mRNA sequencing to understand the behaviour of Mendelian PD genes. Further, to identify gene ontologies and pathways over-represented in differentially expressed genes.

Project 4: GWAS has provided evidence that *PARK16* locus and *LRRK2* are associated with PD. Previous published molecular biology studies have shown an

interaction between *PARK16* locus and *LRRK2*. In this study, the aim was to determine the genetic interaction between genes under the *PARK16* locus and *LRRK2* by interrogating a large repository of WES data.

Project 1: Identification of risk variants in Parkinson's disease using whole exome and whole genome sequencing.

1.1 Introduction

In the past two decades, Next Generation Sequencing (NGS) has been extensively used in genomic studies. As the cost of NGS has decreased, family-based sequencing analysis has been increasingly used to identify causal genes for Mendelian disorders. Family-based analysis has an advantage to utilize co-segregation of variants with the disease within families, which helps to distinguish causal from non-causal ones and also the joint variant calling across all family members increases the accuracy of variant calling. These advantages and continuing decrease in the cost of NGS technology, exome/genome sequencing have been widely used for research of Mendelian disorders (Stavropoulos, Merico et al. 2016).

Previously, several genes with Parkinson's disease (PD) causal mutations have been identified in families with Mendelian inheritance and suggest that increased oxidative stress, mitochondrial and lysosomal dysfunction and abnormal handling of misfolded proteins by ubiquitin-proteasome and autophagy-lysosomal systems, contribute to PD (Blesa, Trigo-Damas et al. 2015). In order to understand the genetic architecture of PD in the general population, it is essential to elucidate the strong effects of rare alleles on disease risk in families. However, causal variants have been facilitated using filtering strategies based on segregation, annotation and quality control.

We performed a two-stage study to face challenges that are ingrained in gene discovery in complex disorders. Firstly, we investigated exomes/genomes from 126 families with PD. Furthermore; exonic variants with minor allele frequency (MAF) less than 1% in all public databases and segregated in the family members were prioritized as interesting candidates. In the second stage, the most promising genes from the families were investigated in a replication cohort comprising the large repository of whole exome sequencing (WES) data of 2,859 sporadic PD patients and 690 controls from the Parkinson's disease Genetics Sequencing Consortium (PDGSC) as well as 1,732 controls from the Rotterdam study (RSX1) WES data. Furthermore, gene-based association tests were performed using sporadic PD WES data and NeuroX genotyping data (Nalls, Bras et al. 2015) to estimate the burden of rare variants in the sporadic PD cohort against controls. Finally, publicly available RNA-sequencing data from postmortem human brain samples of 29 PD cases and 41 controls were utilized to generate weighted gene co-expression network (Dumitriu, Golji et al. 2016).

1.2 Subjects and Methods:

1.2.1 Discovery phase

Subjects: A total of 180 DNA samples from the Netherlands, Turkey, Spain, Tunisia, Italy, Portugal and Germany were collected from 117 families with an autosomal dominant (AD) or autosomal recessive (AR) PD. Additionally, 20 DNA samples from 9 Turkish families with an AD and AR mode of inheritance were provided by Dr. Ebba Lohmann. For a detailed list of samples included in this study as well as the clinical characteristics for each cohort, see table 1. A total of

0.5-1ug of DNA from the 180 individuals was shipped to Macrogen (http://www.macrogen.com/) for Whole Genome Sequencing (WGS), and DNA from the additional 20 individuals from Turkey was sent to CeGAT (https://www.cegat.de/) for WES.

Sample preparation: Samples were prepared according to the Illumina TruSeq Nano DNA library preparation guide or TruSeq DNA PCR-free library preparation guide depending on quality and the total amount available. Libraries were sequenced using Illumina HiSeqX sequencer (www.illumina.com). 151 paired-end reads were produced.

Table 1: Sample details and clinical characteristics of WGS/WES.

Cohort	Number of families	Number of cases	Number of controls	AAO (mean ± SD)	Cases (Males)	Cases (Females)
Italy	23	28	2	56.03 ± 11.71	15	13
Nijmegen	10	10	0	47 ± 13.09	5	5
Spain	8	15	4	65.6 ± 10.96	5	10
Tübingen	10	10	0	54.8 ± 7.7	4	6
Tunisia	3	6	3	57.5 ± 15.35	4	2
Turkey	22	29	17	41.37 ± 16.09	17	15
Portugal	11	13	1	54.75 ± 14.85	7	6
SCOPA	3	3	0	51.79 ± 16.48	NA	NA
GRIP_NL	36	39	19	-	18	21

*AAO: Age at onset

1.2.1.1 SNPs and INDELs

Sequence reads were aligned to the human reference genome (hg19) using Burrows wheeler algorithm (BWA) (Chang, Nguyen et al. 2010). Sorting, indexing and PCR duplicate marking were performed using Picard tools (http://broadinstitute.github.io/picard). Variant calling and indel recalibration

was performed with GATK practices (McKenna, Hanna et al. 2010). A VCF file (Variant Call format) with variants identified by WGS was then generated for a total of 180 individuals as well as for variants identified by WES for 20 individuals from the respective 126 families. The VCF files comprising individuals from the same family were utilized for further analysis.

Quality control (QC) was performed using VCFtools (Danecek, Auton et al. 2011) and KGGSeq (Cingolani, Platts et al. 2012). In brief, genotypes with a Phred quality score below 20, a read depth of less than 10, the second smallest normalized Phred-scaled genotype likelihood below 20, those with the fraction of reads carrying alternative allele >= 5% at a reference-allele homozygous genotype, the fraction of reads carrying alternative allele <= 25% at a heterozygous genotype, or the fraction of the reads carrying alternative allele <= 50% at an alternative-allele homozygous genotype were set to missing (./.). Variants with the 'FILTER' field not matching the label 'PASS', a minimum overall sequencing Phred quality score below 50, an overall mapping quality Phred score below 20, an overall strand bias Phred-scaled p-value (using Fisher's exact test) above 6, and those within putative super-duplicate genomic regions as defined in genomic SuperDupsdataset (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/) were removed from further analysis. We also filtered out variants with a disease allele frequency above 0.01 in 1000 Genomes (www.1000genomes.org), dbSNP138, dbSNP141, gnomAD browser (http://gnomad.broadinstitute.org/) the Exome Variant Server (http://evs.gs.washington.edu/EVS/) or ExAC (http://exac.broadinstitute.org/); or our in-house whole-genome/exome database from more than 3,000 neurologically healthy individuals. Variants in

families in which an autosomal dominant mode of inheritance was suspected were further filtered for a disease allele frequency of 0.001.

In a last step of QC, variants with genotypes not consistent with the disease inheritance pattern (autosomal dominant or autosomal recessive) assumed for each family were removed from further analysis. BEAGLE 4.0 (http://faculty.washington.edu/browning/beagle/beagle.html) was used to estimate the phase in each family and, consequently, detect compound heterozygous variants.

Gene-based and deleteriousness prediction annotations were applied to each variant surpassing the QC filters mentioned above using UCSC RefGene (https://genome.ucsc.edu), GENCODE (https://www.gencodegenes.org), UCSC KnownGene (hg19) and dbNSFP database (https://sites.google.com/site/jpopgen/dbNSFP). These variants were also annotated for protein-protein interactions according to STRING (http://string-db.org/) and BioGRID (http://thebiogrid.org/) databases.

Segregating variants were also scored for their presence in Identity by Descent (IBD; autosomal dominant families) or Homozygous by Descent (HBD; recessive families) segments. These segments were detected using BEAGLE 4.0 and were only possible in some of the families included in this project. For the rest of families, Identity by State segments (IBS; dominant families) and Runs of Homozygosity (ROH; recessive families) were detected using PLINK 1.90 beta (Purcell, Neale et al. 2007). A final step of annotation included cross-reference with the PDGene database (http://pdgene.org/) as well as brain expression databases such as GTEx Portal (http://www.gtexportal.org) or BioGPS

(http://biogps.org/).

1.2.1.2 Structural variants/Copy number variants

Structural variants (SV) are large chromosomal rearrangements usually larger than 1kb that includes Copy Number Variation (CNV), inversions and interchromosomal translocations. CNV includes deletions and duplications with the region of DNA >1kb (Escaramís, Docampo et al. 2015).

SVs were called using Manta (Chen, Schulz-Trieglaff et al. 2016) from mapped paired-end sequencing reads. Manta identifies candidate SVs from the discordant pair and split-read alignments, then performs local assembly and realignment to refine candidates (Chen, Schulz-Trieglaff et al. 2016). The BAM files derived from the step mentioned in the previous section were used as input of Manta, and all the SVs were reported in VCF format. Manta has some limitations such as unable to detect dispersed duplications and fully-assembled large insertions.

CNV for all the genome sequenced samples were called using computationally efficient software ERDS (estimation by read depth with single-nucleotide variants) (Zhu, Need A et al. 2012). This tool combines paired-end information, read depth and polymorphism using Hidden Markov Model to predict CNVs. The BAM and VCF files were used as input of ERDS. The output includes copy numbers and scores calculated using the Poisson model (Zhu, Need A et al. 2012). Higher the score, more reliable is the CNV. The length of CNV is also an essential factor. Poisson scores are missing for small deletions (1kb).

Genomic regions including centromeres, telomeres, immunoglobulin, gap locations, repeat masked regions, GC percent greater than 90 and less than 10, common CNVs from Database of Genomic Variants (DGV) (dgv.tcag.ca/dgv/app/home) and low mappable regions overlapping with the CNVs and SVs were removed. All the small length CNVs with missing Poisson scores were removed. CNVs and SVs segregating in the families were retained and confirmed after visualization in Integrative Genomics Viewer (IGV) (Robinson, Thorvaldsdóttir et al. 2011).

1.2.2 Replication phase

1.2.2.1 Gene identification in sporadic WES dataset

The candidate genes identified in discovery cohort were screened in WES data comprising 2,859 sporadic PD cases and 690 controls from Parkinson's disease Genetics Sequencing Consortium (PDGSC) as well as 1,732 controls from the Rotterdam study (RSX1) WES data. This screening allowed not only to reduce the number of false positives but also to investigate the role of candidate genes in the general population.

The 100-bp paired-end reads from PDGSC were aligned to the human reference genome (hg19) using BWA-MEM (Chang, Nguyen et al. 2010). Variants were called and recalibrated using GATK (McKenna, Hanna et al. 2010). High-quality variants were selected by applying standard GATK filter steps with minimum genotype Phred score of 20, depth of 8 and all the variants with the 'FILTER' field matching the label 'PASS'. Colleagues from the PDGSC performed

quality control of this dataset. All the samples without 15X minimum coverage and not in 85% coverage of the broad canonical exome intervals were excluded. All the subjects with an excessive number of singletons (>500) were removed. Individuals with the ratio of transition to transversion (Ti/Tv) >3 SD from the mean and mean depth per alternative allele less than 25 were excluded. All the variants with minimal coverage of 15X per variant in cases or controls as well as a study-specific differential depth threshold (anywhere from P <1E-50 to P <1E-150) were removed. Moreover, variants were removed on the basis of maximal missingness of 15% per variant in cases or controls as well as a study-specific differential missingness threshold (anywhere from P <1E-6 to P <1E-25) based on distributions. Individual quality control was also performed which includes removal of divergent ancestry samples, duplicates, heterozygosity and missing genotype outliers and samples with discordant sex information. After quality control, we were left with 1,472,032 variants. We prioritized the rare (MAF<0.01) coding variants from the sporadic WES data and screened our candidate genes, which were identified using the discovery cohort.

1.2.2.2 Association analysis:

Gene-based optimal sequence kernel association test (SKAT-O) was performed using rare variants (MAF<0.01). SKAT-O is a combined test which selects significant result by comparing the one-sided model of burden test (all variants have the same direction of effect) and two-sided model of SKAT test (all variants have a distinct direction of effect) (Lee, Emond et al. 2012). It computes gene-based p-values by correcting population stratification with the inclusion of 20 multi-dimensional scaling components as covariates.

The following four criteria were used to see the enrichment of rare variants:

- all the rare variants.
- Functional variants (frameshift, non-frameshift, start-loss, stop-loss, stop-gain, splicing, missense, exonic, 5-prime untranslated region (UTR5), 3-prime untranslated region (UTR3), upstream (-100bp), downstream (+100bp) and non-coding RNA (ncRNA)).
- Coding variants (frameshift, non-frameshift, start-loss, stop-loss, stop-gain, splicing and missense).
- Loss of function variants (frameshift, start-loss, stop-gain and splicing). SKAT-O tests were performed using two independent datasets as in below sections 1.2.2.2.1 and 1.2.2.2.2.

1.2.2.2.1 IPDGC WES

The sporadic PD WES data comprises 1,450 PD cases and 535 controls from IPDGC consortium, and 1,732 controls from the RSX1 WES data. Paired-end sequencing using Illumina HiSeq 2000 was performed for IPDGC and RSX1 data. The sequencing reads were aligned to the human reference genome (hg19) using BWA-MEM algorithm (Chang, Nguyen et al. 2010). Binary Alignment Map (BAM) files were generated using Picard tools (http://broadinstitute.github.io/picard). Further, SNPs and small insertions/deletions were called using Genome Analysis Toolkit (GATK) (McKenna, Hanna et al. 2010). The IPDGC and RSX1 datasets were merged by the joint variant calling of the g.VCF files. Genotype and variant quality control was performed using KGGSeq (Cingolani, Platts et al. 2012) and VCFtools (Danecek, Auton et al. 2011) as already mentioned in section 1.2.1.1. It

was followed by individual quality controls by excluding samples with gender inconsistencies, heterozygosity outliers, duplicate samples, individuals of divergent ancestry and missing genotype. After QC, the IPDGC-RSX1 dataset consists of 545,752 SNPs and 70,548 indels. The sequence-kernel association test (SKAT-O) was performed using EPACTS tool (https://genome.sph.umich.edu/wiki/EPACTS) to estimate the burden of rare variants in PD cases compared to controls. Association tests helps to understand the cumulative impact of rare variants in the candidate genes on PD risk.

1.2.2.2.2 NeuroX

The genotyping array NeuroX consists of 6,801 PD cases and 5,970 controls. The QC of this dataset was performed by excluding variants for subsequent analyses when the minimum call rate was less than 85%, a HWE p-value less than 1E-06 in all the controls, or when the missingness rate was significantly different between cases and controls. Individuals with sex discordance, a heterozygosity rate of >3 standard deviations from the mean, more than 15% of missing genotypes, closely related to another individual in the dataset (>18.5%) or representing a population outlier after multi-dimensional scaling analysis, were removed from further analyses After QC, the data consists of 177,216 variants. Further, gene-based SKAT-O tests were performed to estimate the burden of rare variants in PD cases compared to controls.

1.2.3 Development of Neurochip

A genotyping array has been developed to verify our genetic findings in a diverse population. An Infinium iSelect Custom Genotyping Assay was designed in

collaboration with Illumina (www.illumina.org). This array has been designed as part of an international collaboration aiming to provide a new tool for the fine mapping and interrogation of various neurodegenerative diseases. This array was designed of the Illumina CoreExome 24+ v1.1 on the 24-sample format and has the following content: Illumina CoreExome 24+ v1.1 backbone, content from the original NeuroX array (Nalls, Bras et al. 2015), thorough review of known neurodegenerative disease genes for all coding/splice site variants reported in ExAC (http://exac.broadinstitute.org/), known pathogenic variants in each of these genes not found in ExAC, tagging SNPs to capture all common variation in these genes, additional novel content based on current and ongoing sequencing projects including WGS data of 180 samples and a full list of all GWAS hits in the NHGRI database.

The diseases covered by this array are Alzheimer's disease (AD), Parkinson's disease (PD), Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Progressive Supranuclear Palsy (PSP), Cortical Basal Degeneration (CBD), Multiple System Atrophy (MSA), Frontotemporal Dementia (FTD), and Dementia with Lewy Bodies (DLB).

1.2.4 Weighted Gene Co-expression Network analysis

Co-expression of all the candidate genes from the discovery phase with Mendelian PD genes was investigated using publicly available RNA-sequencing data from postmortem human brain of 29 PD and 44 controls (Dumitriu, Golji et al. 2016). The weighted gene co-expression network was constructed using WGCNA (Langfelder and Horvath 2008) package in R. Modules are clusters of highly interconnected genes. Modules were identified using an unsigned

network with the power of 7 to achieve scale-free topology. However, we used unsigned co-expression network so that the modules correspond to clusters of genes with high absolute correlations. Each module is assigned a particular color. Gene significance (GS) represents a correlation between gene expression and trait (i.e. case/control status). Higher the absolute value of GS_i, the more biologically significant is the i^{th} gene (Langfelder and Horvath 2008).

The over-represented gene ontologies and KEGG pathways for all the genes in the modules were determined using HTSAnalyzeR package in R (Wang, Terfve et al. 2011), to facilitate biological interpretation. Over-representation analysis is performed based on the hypergeometric test (Subramanian, Tamayo et al. 2005).

1.3 Results

1.3.1 Discovery phase

1.3.1.1 SNPs and INDELs

Filtering of the variants by genotype and variant quality control resulted in 17,072,717 SNPs and 4,322,295 indels across all samples. We investigated the families independently, and the variants were retained based on segregation among the family members. Moreover, only coding SNPs and indels with a minor allele frequency less than 1% were retrieved. Genes/variants were ranked on the basis of pathogenicity scores and their expression in brain.

We identified mutations in genes causing PD and other forms of parkinsonism in some of the families. A novel stop-gain mutation was identified

in *DJ1* (p. Q45*) in a Turkish family with an early-onset levodopa-responsive parkinsonism and signs of ALS (Hanagasi, Giri et al. 2016). In two Turkish families, a homozygous (p. R747W) (Giri, Guven et al. 2016) and compound heterozygous (p. A727T; p. C134) mutations in *PLA2G6* were identified. A novel homozygous mutation in *TREM2* (p. D86V) was identified in a Turkish family (figure 8 Turkey_PD240) with an autosomal recessive mode of inheritance (see table 2). The age at onset of the patient with the *TREM2* mutation is 37 years, and he has dementia with parkinsonism. A point mutation was identified in *LRRK2* (p. I723V) in one of the Spanish family with an autosomal dominant mode of inheritance.

Apart from the known PD causal genes, we also identified some novel genes in multiple families. In Turkish family Turkish_PD326 (see figure 8) with the X-linked recessive mode of inheritance, the pathogenic (CADD score: 28.5) mutation (p. R102W) in *SH3KBP1* is the only variant segregating in that family, and it was also identified in an early onset PD patient from an independent family in Rotterdam. The age at onset of the index-case from Turkish family (Turkish_PD326) is 40 years, and one year later he developed bradykinesia without resting or postural tremor. He has rapid eye movement (REM) sleep behavior disorder and levodopa treatment was effective for him.

In families with an autosomal dominant mode of inheritance rare heterozygous mutations in *DYSF* (p. Y1052C; p. T172M) (table 5) were identified in two Turkish families (figure 8 Turkey_PD291 and Turkey_PD314). Both these mutations are pathogenic with CADD scores as 27.2 and 27.9 respectively. In family Turkey_PD291, the age at onset of the PD patient is 44 years. He has right-arm tremor and REM sleep behavior disorder. Brain MRI showed enlargement of

the retrocerebellar region. He showed improvement with levodopa treatment. The age at onset of the two PD patients from family Turkey_PD314 is 59 and 60 years. In two families from the genetic isolate of the Netherlands and Spain, rare heterozygous mutations were identified in *APC2* (p. S1584G; p. P946T; p. T104N) (table 5). The clinical details of the PD patients from these families are not available.

We identified some of the candidate genes, which are not identified in multiple families but are identified in sporadic PD data from replication cohort and will be discussed again in a replication phase (see table 4). These candidate genes are *HTT*, *CCNB3*, *CCDC154*, *CXorf22*, *MYO15A*, *TMEM63C* and *ULK2*. A compound heterozygous mutation in *HTT* (p. P42T; p. P43Q) was identified in an Italian family (see figure 8 Italy_Fam034), and it is the only variant segregating in that family comprising 3 PD patients. In Turkish family (see figure 8 Turkey_PD313) rare homozygous mutations in *CCNB3* (p. M313I) and *CCDC154* (p. E133K), and in family Turkey_PD296 rare homozygous mutation in *CXorf22* (p. V547A) were identified. In a family from a genetic isolate of the Netherlands (see figure 8 GRIP252), a homozygous mutation in *TMEM63C* (p. R334H) was the most plausible one amongst the two variants segregating in that family. The variant p. R334H is pathogenic (CADD score: 24.5) and is present in the runs of homozygosity. A compound heterozygous mutation in *MYO15A* (p. L3160F; p. R201H) was identified in a Spanish family (figure 8 Spain_2933). Only two variants are segregating in this family comprising two PD patients. Furthermore, a rare homozygous stop-loss mutation in *ULK2* (p. *170G) was identified in a Tunisian family (figure 8 Tunisia_3086), and it is the only mutation segregating in that family with 2 PD patients.

Table 2: SNVs identified in genes causing parkinsonism and other neurological disorders.

Cohort	Family	Type	CHR	RefGene_Features	gnomAD
Turkey	PD262	AH	22:38508548:G/A	***PLA2G6***:NM_001199562:c.2239C>T:p.R747W:(16Exons):exon15:missense	0.000008994
Turkey	PD295	AH	1:8025426:C/T	***PARK7***:NM_001123377:c.133C>T:p.Q45*:(7Exons):exon3:stopgain	N
Turkey	PD240	AH	6:41129135:T/A	***TREM2***:NM_018965:c.257A>T:p.D86V:(5Exons):exon2:missense	N
Turkey	PD328	CH	22:38508248:C/T	***PLA2G6***:NM_001199562:c.2179G>A:p.A727T:(16Exons):exon16:missense	0.000222
Turkey	PD328	CH	22:38541467:AG/A-	***PLA2G6***:NM_001199562:c.?400del-C:p.C134:(16Exons):exon3:frameshift;	N
Turkey	PD270	AH	6:161969884:AC/A-	***PARK2***:NM_004562:c.1082+1GT>-T:(12Exons):exon9GTdonor	N
Spain	HCB3	Het	12:40671989:A/G	***LRRK2***: NM_198578:c.2167A>G:p.I723V:exon18:missense	0.04

* AH: Alternative Homozygous; CH: Compound Heterozygous; Het: Heterozygous; CHR: Chromosome.

1.3.1.2 Structural variants/Copy number variants

The SVs and CNVs were called for all the 180 WGS samples. The total number of losses and gains after removing regions such as centromeres, telomeres, immunoglobulin, gap locations, repeat masked regions, GC percent greater than 90 and less than 10, common CNVs from Database of Genomic Variants (DGV) and low mappable regions are illustrated in table 3. Further, all the SVs and CNVs, which are segregating among the family members and are confirmed after visualization in IGV were retrieved.

Copy number variants such as homozygous deletion of exon1 to exon5 in *DJ1* (Bonifati, Rizzu et al. 2003) and duplication in *PARK2* were identified in families from the genetic isolate of the Netherlands (table 10). All the CNVs, which are segregating in the families and were confirmed after IGV visualization are mentioned in table 10. After filtering the common SVs from DGV, we were not left with any SVs segregating in the families from all the cohorts.

Table 3: Copy Number Variants in all cohorts.

Cohort	Deletions after filtering	Duplications after filtering
Dutch	58	188
GRIP	631	4213
Italian	477	1397
Portugal	167	541
Spanish	257	777
Tunisian	144	432
Turkish	363	863

1.3.2 Replication

1.3.2.1 Gene identification in sporadic PD WES dataset

The prioritized variants/genes from discovery cohort were investigated in WES data from PDGSC. After screening the candidate genes from discovery cohort families with an autosomal recessive and compound heterozygous mode of inheritance, we identified 8 genes (with different variants), namely *HTT*, *CCNB3*, *CCDC154*, *CXorf22*, *MYO15A*, *SH3KBP1*, *TMEM63C* and *ULK2* in PDGSC data. The detailed description of these variants, their features and minor allele frequency is mentioned in table 4.

In the sporadic PD data from PDGSC, 9 PD patients carry hemizygous and 1 patient carries an alternative homozygous mutation in *SH3KBP1* (figure 4). All these variants in *SH3KBP1* have MAF less than 1% (table 4) in public databases such as ExAC, gnomAD and 1000 genomes. One PD patient carries a rare alternative homozygous mutation in *HTT* (p. Q38P). The frequency of this missense mutation in gnomAD browser is 0.01. Furthermore, 8 PD patients carry

rare hemizygous mutations in *CCNB3*, and 2 PD patients carry an alternative homozygous mutation in *CCDC154*. Alternative homozygous missense mutations in *TMEM63C* (p. R334H and p. P5S) were identified in two PD patients. A p. R334H variant in *TMEM63C* was also identified in a family (figure 8 GRIP252) from discovery cohort. An alternative homozygous mutation in *MYO15A* (p. R2942C) and *ULK2* (p. D500A) were identified in single PD patient, respectively (table 4). All these variants identified in PDGSC data are exclusively present in the PD cases.

DYSF and *APC2*, which replicated in multiple dominant families, were screened in sporadic PD data. A rare (MAF<0.01) heterozygous mutations in *DYSF* were identified exclusively in 23 PD patients from the sporadic data (figure 5). Furthermore, 12 sporadic PD patients carry rare heterozygous mutations in *APC2* (figure 6) (see table 5).

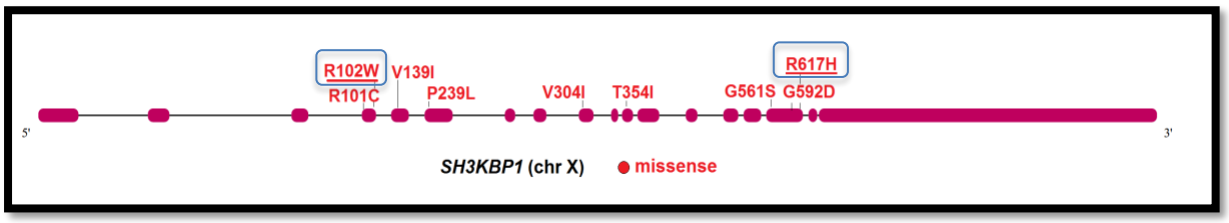


Figure 4: Mutations in *SH3KBP1*. Variants highlighted in blue rectangles are identified in two families and the rest other variants are identified in sporadic PD WES data.

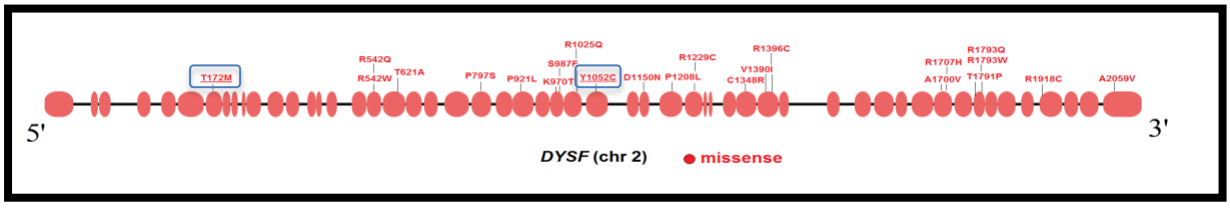


Figure 5: Mutations in *DYSF*. Variants highlighted in blue rectangles are identified in two families and the rest other variants are identified in sporadic PD WES data.

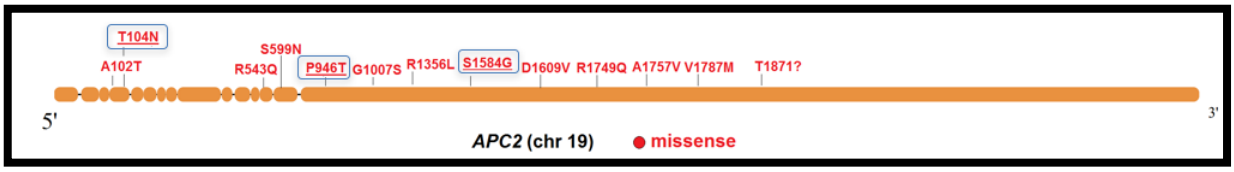


Figure 6: Mutations in *APC2*. Variants highlighted in blue rectangles are identified in three families and the rest other variants are identified in sporadic PD WES data.

1.3.2.2 Association analysis using two independent datasets

WES data from 1,450 sporadic PD cases and 2,267 controls from the IPDGC consortium and RSX1 as well as NeuroX genotyping data from 6,801 sporadic PD cases and 5,970 controls were used to estimate the genetic burden of rare variants in candidate genes identified in the discovery cohort. We performed gene-based SKAT-O tests using both these datasets. We used four criteria to see the enrichment of rare variants. The first criteria showed the burden of all rare variants in the respective genes, secondly to see an enrichment of functional variants and followed by coding and loss of function variants.

In the first replication cohort of sporadic PD WES data, the gene-based association tests showed that loss of function variants in *STON1-GTF2A1L* (Pvalue: 0.0001) and *IGFN1* (Pvalue: 0.01) are enriched to PD. Moreover, all the rare variants in *ULK2* (Pvalue: 0.01) show gene-based association to PD risk (table 6).

In the second independent dataset of a genotyping array, *i.e.* NeuroX, gene-based association shows that coding variants in *ULK2* (Pvalue: 0.001) and *IGFN1* (Pvalue: 0.002) are enriched to PD. Further, rare variants in *KNTC1* (Pvalue: 0.00001) and *MFSD8* (Pvalue: 0.007) are also associated with PD risk (table 7).

Table 4: Genes identified in AR families (discovery cohort) along with their replication.

Discovery cohort							Replication cohort
Cohort	Family	Type	Gene	chr:pos	Genefeature	gnomAD	Variants_in_PDGSC (No. of carriers)
Italy	Fam-034	CH	*HTT*	4:3076671:C/A ; 4:3076674:C/A	NM_002111:c.124C>A:p.P42T:missense ; NM_002111:c.128C>A:p.P43Q:missense	0.0006 ; 0.0005	HTT:NM_002111:c.119A>C:p.Q38P:missense (1)
Turkey	PD326	AH	*SH3KBP1*	X:19725085:G/A	NM_031892:c.304C>T:p.R102W:missense	N	SH3KBP1:NM_031892:c.1775G>A:p.G592D:missense (1); SH3KBP1:NM_031892:c.1681G>A:p.G561S:missense (4); SH3KBP1:NM_031892:c.1061C>T:p.T354I:missense (1); SH3KBP1:NM_031892:c.910G>A:p.V304I:missense (1); SH3KBP1:NM_031892:c.716C>T:p.P239L:missense (1); SH3KBP1:NM_031892:c.415G>A:p.V139I:missense (1); SH3KBP1:NM_031892:c.301C>T:p.R101C:missense (1)
Turkey	PD313	AH	*CCNB3*	X:50052108:G/A	CCNB3:NM_033031:c.939G>A:p.M313I:missense	N	CCNB3:NM_033031:c.290C>T:p.T97I:missense (1) ; CCNB3:NM_033031:c.1712G>A:p.R571K:missense (1); CCNB3:NM_033031:c.2939C>T:p.T980I:missense (1); CCNB3:NM_033031:c.3059C>T:p.T1020I:missense (1); CCNB3:NM_033031:c.3796C>T:p.R1266C:missense (1); CCNB3:NM_033031:c.2876A>G:p.E959G:missense (1); CCNB3:NM_033031:c.2419G>A:p.E807K:missense (1); CCNB3:NM_033670:c.216G>T:p.E72D:missense (1)
Turkey	PD313	AH	*CCDC154*	16:1493524:C/T	CCDC154:NM_001143980:c.397G>A:p.E133K:missense	0.00002	CCDC154:NM_001143980:c.1151T>G:p.L384R:missense (1); CCDC154:NM_001143980:c.1124G>A:p.R375Q:missense (1)
Turkey	PD296	AH	*CXorf22*	X:35985775:T/C	CXorf22:NM_152632:c.1640T>C:p.V547A:missense	N	CXorf22:NM_152632:c.1665G>T:p.K555N:missense (1); CXorf22:NM_152632:c.T2856del-C:p.F952:frameshift (1)
GRIP_NL	GRIP252	AH	*TMEM63C*	14:77706888:G/A	TMEM63C:NM_020431:c.1001G>A:p.R334H:missense	0.002	TMEM63C:NM_020431:c.1001G>A:p.R334H:missense (1); TMEM63C:NM_020431:c.13C>T:p.P5S:missense (1)
Spain	2933	CH	*MYO15A*	17:18064722:C/T ; 17:18075051:G/A	MYO15A:NM_016239:c.9478C>T:p.L3160F:missense ; ENST00000579848:c.602G>A:p.R201H:missense	0.006; 0.006	MYO15A:NM_016239:c.8824C>T:p.R2942C:missense (1)
Tunisia	3086	AH	*ULK2*	17:19685288:A/C	ENST00000575432:c.508T>G:p.*170G:stoploss	0.0003	ULK2:NM_001142610:c.1499A>C:p.D500A:missense (1)

* AH: Alternative Homozygous; CH: Compound Heterozygous

Table 5: Genes identified in multiple AD families (discovery cohort) along with their replication.

Discovery cohort							Replication cohort
Cohort	Family	Type	Gene	chr:pos	Genefeature	gnomAD	Variants_in_PDGSC (No. of carriers)
Turkey	PD291 ; PD314	Het	DYSF	2:71797798:A/G ; 2:71740427:C/T	DYSF:NM_001130987:c.3155A>G:p.Y1052C:missense ; DYSF:NM_001130980:c.515C>T:p.T172M:missense	N ; 0.00004	DYSF:NM_001130986:c.1624C>T:p.R542W:missense (1); DYSF:NM_001130986:c.1625G>A:p.R542Q:missense (1); DYSF:NM_001130986:c.1861A>G:p.T621A:missense (1); DYSF:NM_001130986:c.2389C>T:p.P797S:missense (1); DYSF:NM_001130986:c.2762C>T:p.P921L:missense (1); DYSF:NM_001130986:c.2909A>C:p.K970T:missense (1); DYSF:NM_001130986:c.2960C>T:p.S987F:missense (1); DYSF:NM_001130986:c.3074G>A:p.R1025Q:missense (1) ; DYSF:NM_001130986:c.3448G>A:p.D1150N:missense (1); DYSF:NM_001130986:c.3623C>T:p.P1208L:missense (1); DYSF:NM_001130986:c.3685C>T:p.R1229C:missense (1); DYSF:NM_001130986:c.4042T>C:p.C1348R:missense (1); DYSF:NM_001130986:c.4168G>A:p.V1390I:missense (1); DYSF:NM_001130986:c.4186C>T:p.R1396C:missense (1); DYSF:NM_001130986:c.5099C>T:p.A1700V:missense (1); DYSF:NM_001130986:c.5120G>A:p.R1707H:missense (1); DYSF:NM_001130986:c.5371A>C:p.T1791P:missense (3); DYSF:NM_001130986:c.5377C>T:p.R1793W:missense (1); DYSF:NM_001130986:c.5378G>A:p.R1793Q:missense (1); DYSF:NM_001130986:c.5752C>T:p.R1918C:missense (1); DYSF:NM_001130986:c.6176C>T:p.A2059V:missense (1)
GRIP_NL ; Spain	GRIP104 ; GRIP251 ; SpainHCB1	Het	APC2	19:1468050:A/G ; 19:1466136:C/A ; 19:1453508:C/A	APC2:NM_005883:c.4750A>G:p.S1584G:missense ; APC2:NM_005883:c.2836C>A:p.P946T:missense ; APC2:NM_005883:c.311C>A:p.T104N:missense	0.00002; 0.00006; N	APC2:NM_005883:c.304G>A:p.A102T:missense (1); APC2:NM_005883:c.1628G>A:p.R543Q:missense (1); APC2:NM_005883:c.1796G>A:p.S599N:missense (1); APC2:NM_005883:c.3019G>A:p.G1007S:(15Exons):exon15:missense (2); APC2:NM_005883:c.4067G>T:p.R1356L:missense (1); APC2:NM_005883:c.4826A>T:p.D1609V:missense (1) ; APC2:NM_005883:c.5246G>A:p.R1749Q:missense (1); APC2:NM_005883:c.5270C>T:p.A1757V:missense (1); APC2:NM_005883:c.5359G>A:p.V1787M:missense (1); APC2:NM_005883:c.5613C>*:p.T1871?:missense (2)

*Het: Heterozygous.

Table 6: Gene-based SKAT-O results of IPDGC exomes.

Gene	$N_{allrare}$	$P_{allrare}$	$P_{functional}$	P_{coding}	P_{lof}
APC2	22	0.04212	0.6078	0.6078	0.886927
BCHE	22	0.53183	0.51342	0.51342	8.0289e-18
CCDC129	81	0.33386	0.72253	0.72253	0.34602
CCDC154	23	1	0.66855	0.3398	0.2576
CCNB3	33	0.2299	0.45247	0.43821	0.75588
CRYZL1	15	0.86234	1	1	0.70161
CXorf22	31	0.81821	0.6444	0.6444	0.45899
DHX57	130	0.60213	0.70882	0.70882	0.83013
DYSF	220	0.28941	0.22977	0.27635	0.18173
DPP8	63	0.00040823	0.0011138	0.028692	0.7507
EPC1	22	0.016828	0.10013	0.10013	0.44955
FCF1	25	0.042565	0.031926	0.24796	0.78398
GART	81	0.011433	0.096183	0.08225	1
HTT	201	0.21039	0.054938	0.054938	1
IGFN1	225	0.46065	0.3122	0.35503	0.011291
KNTC1	90	0.23592	0.22598	0.22598	0.9097
MFSD8	26	1	0.89005	1	0.53128
MYO15A	179	0.5211	0.66788	0.66788	0.71526
PHKA2	55	0.029669	0.011683	0.011259	0.0052439
RPA2	18	0.27994	0.23606	0.23428	0.045132
SH3KBP1	34	0.7216	0.51781	0.51781	0.53622
SHANK2	119	0.0042875	0.044719	0.044719	0.51904
SLC22A10	65	0.20036	0.65099	0.65099	1
SLC45A3	20	0.89859	0.55154	0.55154	0.66004
STON1-GTF2A1L	65	0.027945	0.0019625	0.0024822	0.0001976

TBX4	21	0.43582	0.6483	0.6483	5.8597e-05
TMEM63C	62	1	1	1	0.23493
TTLL12	75	0.72034	0.81071	0.89576	1
ULK2	58	0.08077	0.018658	0.018658	0.11045
WWC1	89	0.75591	0.43159	0.54367	1
ZNF518A	53	0.27601	0.27601	0.27601	NA
ZNF521	55	0.60031	0.58536	0.56885	2.1233e-05

* $N_{allrare}$: Number of all rare variants; $P_{allrare}$: P-value for all rare variants; $P_{functional}$: P-value for functional variants; P_{coding}:P-value for coding variants; P_{lof}: P-value for loss of function variants.

Table 7: Gene-based SKAT-O results of NeuroX.

Gene	$N_{allrare}$	$P_{allrare}$	$P_{functional}$	P_{coding}	P_{lof}
APC2	7	0.16965	0.16965	0.16965	NA
BCHE	9	0.12471	0.12471	0.12471	NA
CCDC129	21	0.062061	0.037296	0.037296	NA
CCDC154	15	1	0.88392	0.88392	0.90874
CRYZL1	6	0.015284	0.015284	0.015284	NA
DHX57	24	0.01116	0.01116	0.01116	0.9148
DYSF	71	0.56924	0.56508	0.56508	NA
DPP8	10	0.62937	0.62937	0.62937	NA
EPC1	9	0.79752	0.78971	0.78971	NA
FCF1	2	0.17167	0.17167	0.17167	NA
GART	16	0.52262	0.52463	0.52463	NA
HTT	33	0.73869	0.73869	0.73869	NA
IGFN1	56	0.012481	0.0022045	0.0022045	0.20873
KNTC1	30	1.4171e-05	1.4171e-05	1.4171e-05	0.39821
MFSD8	14	0.0072169	0.0072169	0.0072169	NA
MYO15A	73	0.42765	0.44825	0.44825	NA
RPA2	3	0.76321	0.76321	0.76321	NA
SHANK2	31	0.69882	0.8667	0.8667	NA
SLC22A10	19	0.13099	0.13099	0.13099	0.049742

SLC45A3	13	0.00097305	0.00097305	0.00097305	NA
STON1-GTF2A1L	33	0.76908	0.56707	0.56707	0.3325
TBX4	8	0.84568	1	1	NA
TMEM63C	11	0.054963	0.066453	0.066453	NA
TTLL12	18	0.00066903	0.00066903	0.00066903	NA
ULK2	14	0.0011033	0.0016626	0.0016626	NA
WWC1	21	0.011888	0.011888	0.011888	NA
ZNF518A	24	0.0028038	0.0028038	0.0028038	NA
ZNF521	13	0.30372	0.30372	0.30372	NA

* $N_{allrare}$: Number of all rare variants; $P_{allrare}$: P-value for all rare variants; $P_{functional}$: P-value for functional variants; P_{coding}: P-value for coding variants; P_{lof}: P-value for loss of function variants.

1.3.3 Development of Neurochip

In the context of WGS performed in this study, we have submitted a total of 25,383 variants to Illumina (table 8). In brief, all the segregating variants that surpassed our QC filters have been selected for inclusion in the mentioned array. Additionally, variation in ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), as well as a rare variation (MAF < 0.01) in ExAC and an in-house database of 1,500 PD cases and 500 controls (IPDGC WES project) from the most interesting genes were included in this array. Variants of interest derived from WES efforts of other consortium partners, as well as variants needed for the Mendelian randomization and gene-environment interaction analyses, have also been selected for this array.

Table 8: List of variants selected from the WGS efforts for Neurochip.

Source	Number of variants
WGS segregating variants	1,180
Genes of interest: ClinVar	1,183
Genes of interest: ExAC	22,119
Genes of interest: IPDGC exomes	901
Total	25,383

1.3.4 WGCNA

In order to construct the weighted gene co-expression network, we need to choose the soft thresholding power β to which co-expression similarity is raised to calculate adjacency. The soft thresholding power is chosen based on the criterion of approximate scale-free topology. We chose the power 7, which is the lowest power for which the scale-free topology fit index curve flattens out upon reaching a high value (in this case 0.80) (figure 7).

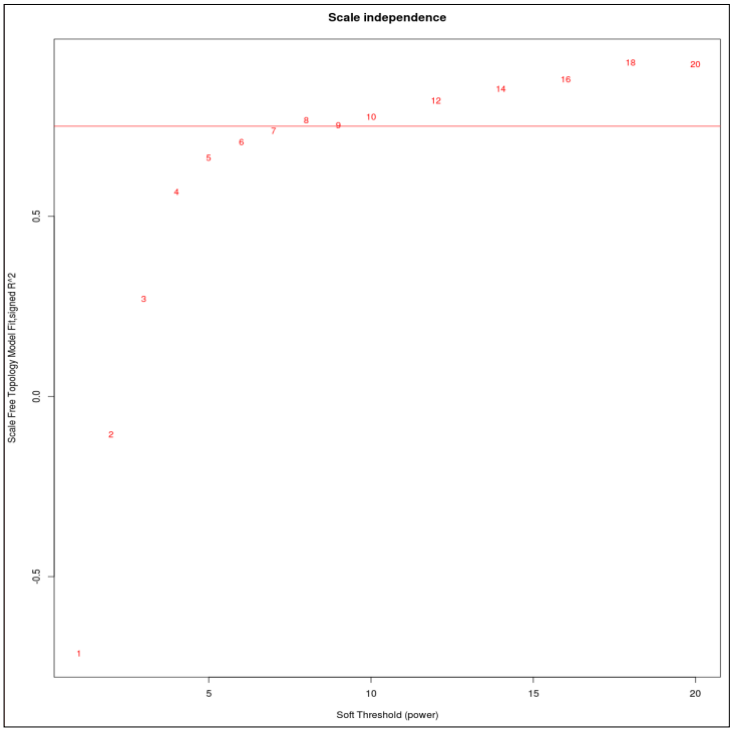


Figure 7: Analysis of network topology for various soft-thresholding powers. The figure shows the scale-free fit index (y-axis) as a function of the soft thresholding power (x-axis).

We would like to identify modules that are significantly associated with the measured clinical traits. A summary profile (eigengene) for each module was generated using WGCNA package in R, later we correlate eigengenes with external traits and look for the most significant associations.

WGCNA results for all the candidates from the discovery phase data are illustrated in table 11. Some of the candidate genes from the families (discovery cohort), which are co-expressed with PD causal genes are discussed as follows: we observed that the candidate gene *TMEM86B* is co-expressed with *LRRK2* in a blue module comprising 1503 genes. The blue module is positively correlated to the trait with a significant P-value of 0.001. Among the KEGG pathways, mitophagy and mTOR signaling pathway are over-represented in this module (table 9).

In a turquoise module, which consists of 2580 genes, *CCDC154*, *ZNF766* and *VPS35* are co-expressed with *PARK2.* The turquoise module is negatively correlated to the trait with a significant P-value of 1e-04. In this module regulation of postsynaptic membrane potential, synaptic plasticity and macroautophagy are over-represented (table 9).

APC2 is co-expressed with *PINK1* in a grey60 module, but this module is not significantly correlated to the trait. Furthermore, *DYSF* and *CCNB3* are co-expressed with *FBXO7* in a pink module. Although this module is not significantly correlated to the trait, clathrin-dependent endocytosis and regulation of apoptotic process are over-represented (table 9).

Table 9: Module-trait correlation and over-represented pathways and Gene ontology

Gene	Module	Cor.Module.trait	Pvalue	Over-represented Pathways & GO
TMEM86B	blue	0.37	0.001	mTOR signaling pathway ; mitophagy
LRRK2				
APC2	grey60	-0.061	0.6	GTPase activity
PINK1				
CCNB3	pink	0.07	0.5	Clathrin-dependent endocytosis ; regulation of apoptotic process
DYSF				
FBXO7				
CCDC154	turquoise	-0.44	1e-04	Chemical synaptic transmission ; regulation of postsynaptic membrane potential ; regulation of synaptic plasticity ; regulation of macroautophagy
VPS35				
PARK2				
ZNF766				

* Cor.Module.trait: Correlation of module and trait.

1.4 Discussion

In this study, analysis of WGS/WES data as discovery cohort of individuals with familial PD has led to the identification of mutations in already known PD causal genes such as *LRRK2*, *DJ1* and *PARK2*. A variant in *PLA2G6* was identified in a patient with early-onset parkinsonism, and the results confirm the clinical heterogeneity of *PLA2G6*-associated neurodegeneration. A novel alternative homozygous mutation in *TREM2* was identified in a patient with dementia and parkinsonism.

Exome sequencing of patients with PD helped to identify hemizygous mutations in *SH3KBP1* in two probands from unrelated families. Moreover, after screening the large repository of WES data from PDGSC consortium, we identified 7 missense mutations with MAF less than 1% in *SH3KBP1* in 10 sporadic PD cases. *SH3KBP1* is involved in many cellular processes such as down-regulation of receptor tyrosine kinase and apoptosis (Narita, Nishimura et

al. 2005, Bian, Yu et al. 2008). Previous studies have already reported inflammatory reaction and apoptosis as major causes of PD (Blesa, Trigo-Damas et al. 2015). Moreover, *SH3KBP1* is involved in pathways such as Clathrin-mediated endocytosis and vesicle-mediated transport or membrane trafficking. PD is defined by loss of dopaminergic neurons in the substantia nigra and dysfunction of the dopaminergic synapse heralds trafficking impairment. Hence, *SH3KBP1* looks like a promising candidate and might contribute to PD pathogenesis.

Apart from this gene, there are six more genes identified in discovery cohort families with AR mode of inheritance, which are not replicated in multiple families but are identified in a large WES data of sporadic PD. These genes are *HTT*, *CCDC154*, *CXorf22*, *TMEM63C*, *MYO15A* and *ULK2*. Non-synonymous and loss of functions variants with MAF less than 1% were identified in these 6 replicating genes.

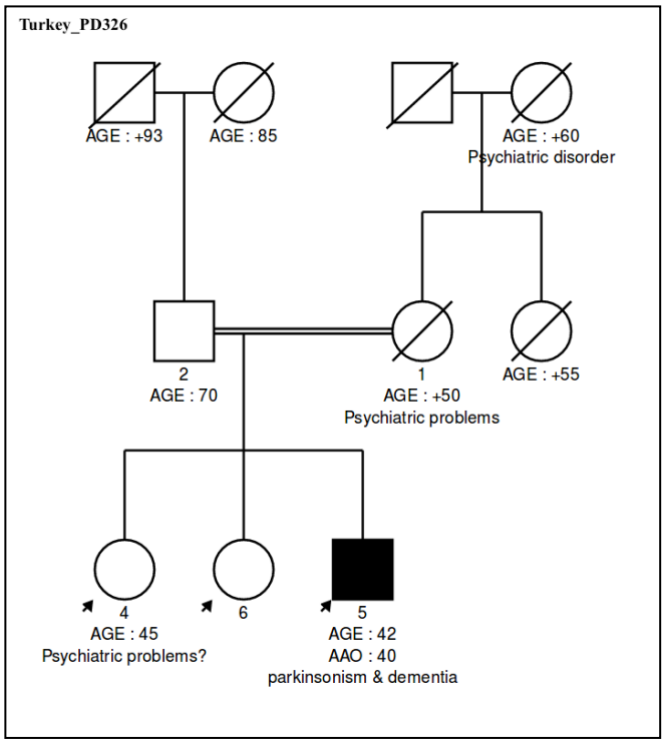
The first gene *HTT* (Huntingtin) is linked to neurodegenerative disease Huntington's disease and is involved in biological processes such as retrograde vesicle-mediated transport- Golgi to ER, apoptotic process and positive regulation of mitophagy (Schulte and Littleton 2011). Several lines of evidence have shown that *PARK2*/*PINK1* pathway plays a critical role in the autophagic removal of damaged mitochondria, *i.e.* mitophagy and defects in mitophagy cause impaired motor coordination, tremor and the accumulation of protein aggregates/inclusion bodies in residual neurons (Jin and Youle 2012). Hence, *HTT* seems like an interesting gene, but of course we need replication in more families to confirm it as a PD causal gene.

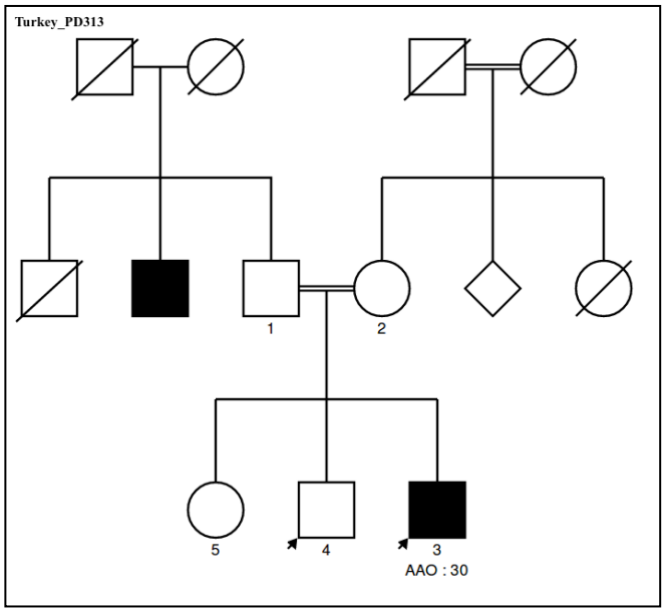
Molecular function and pathways for the second gene *CCDC154* are not known. As per the performed weighted gene co-expression network analysis, it is observed that *CCDC154* is co-expressed with PD Mendelian gene *PARK2*. The third gene *TMEM63C* is involved in calcium-activated cation channel activity (Hou, Tian et al. 2014). As per experimental evidence of Affinity Capture-Mass spectrometry *TMEM63C* interacts with *NDUFB8* (NADH: Ubiquinone Oxidoreductase Subunit B8), *i.e.* Complex I. Previous published studies have suggested mild defects of complex I in PD (Swerdlow, Parks et al. 1996, Schapira 1998). Furthermore, Myosins (*MYO15A*) are actin-based motor molecules with ATP activity. Mutations in this gene are associated with congenital, neurosensory, nonsyndromal deafness (Kalay, Uzumcu et al. 2007). Another promising candidate *ULK2* (Unc-51 Like Autophagy Activating Kinase 2) is involved in regulation of autophagy and mTOR signaling pathway (Jung, Seo et al. 2011, Lee and Tournier 2011). Autophagy is associated with the pathogenesis of Dementia with Lewy bodies and PD. Alteration of upstream autophagy-related proteins like *ULK2* has been observed in Lewy body disease (Miki, Tanji et al. 2016). In the gene-based sequence kernel association tests using both independent datasets from replication cohort, we observed a significant association of rare variants in *ULK2* to PD.

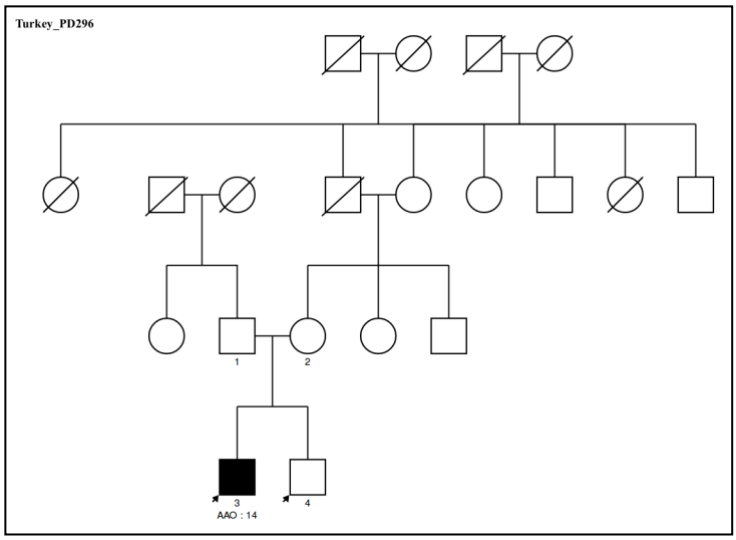
We identified rare heterozygous mutations in *DYSF* and *APC2* in multiple families with an autosomal dominant mode of inheritance. These genes were replicated further in sporadic PD data from PDGSC consortium. Dysferlin (*DYSF*) plays a significant role as calcium ion sensor and is involved in calcium ion triggered fusion of synaptic vesicle to plasma membrane (Fuson, Rice et al. 2014). The protein encoded by *APC2* helps to recruit, phosphorylates and targets

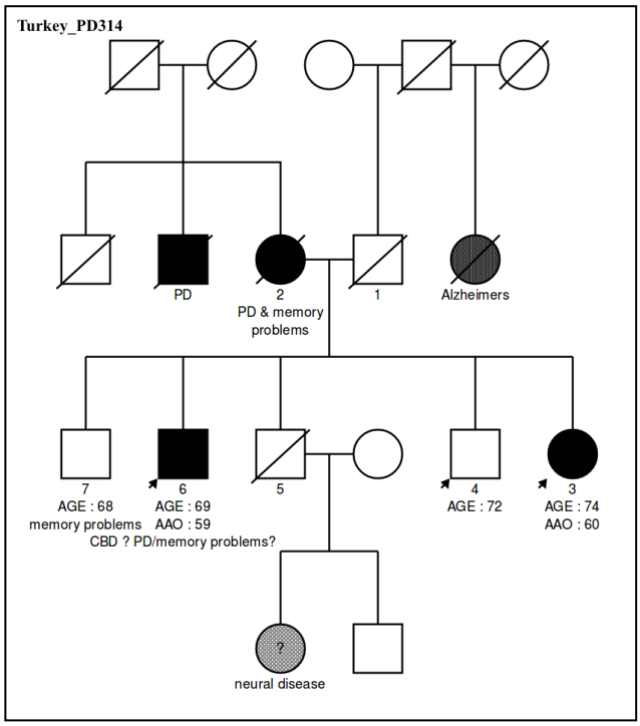
beta-catenin for ubiquitylation and proteasomal degradation (Nakagawa, Murata et al. 1998). Failure of ubiquitin-proteasome system has been identified in familial PD cases (McNaught, Olanow et al. 2001). *APC2* is involved in the PI3K-Akt signaling pathway.

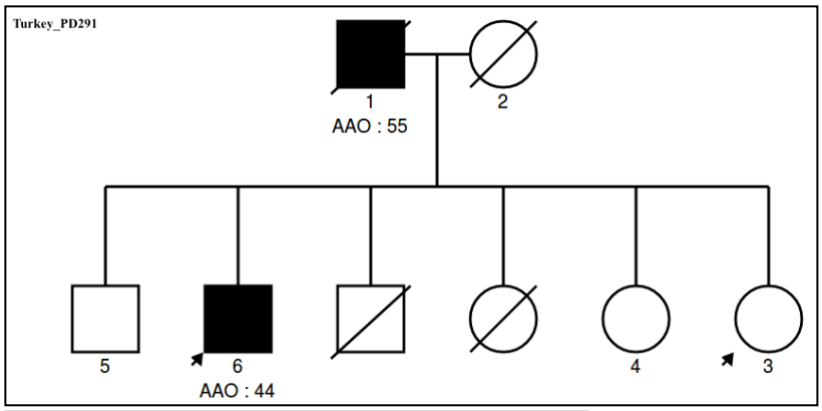
Approximately 25,383 variants from this study were submitted to Illumina for the development of genotyping array Neurochip. In this study, we identified promising candidate genes using NGS data, but in the future validation of these risk variants using the Neurochip, *i.e.* a genotyping array has to be performed. Furthermore, some functional validation using cellular or experimental animal model has to be done to prioritize genes for further investigation.

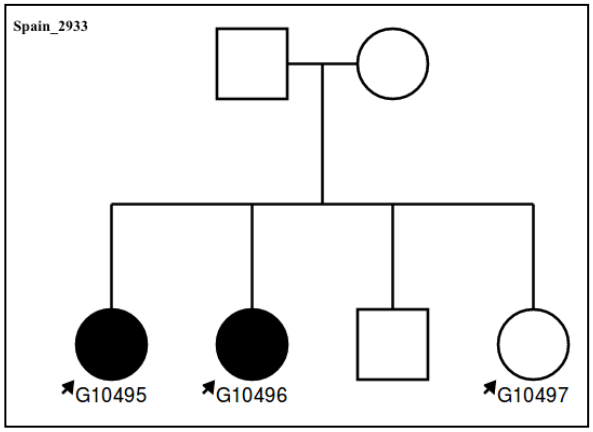

Turkey_PD326
AGE : +93
AGE : 85
AGE : +60
Psychiatric disorder
2
AGE : 70
1
AGE : +50
Psychiatric problems
AGE : +55
4
AGE : 45
Psychiatric problems?
6
5
AGE : 42
AAO : 40
parkinsonism & dementia

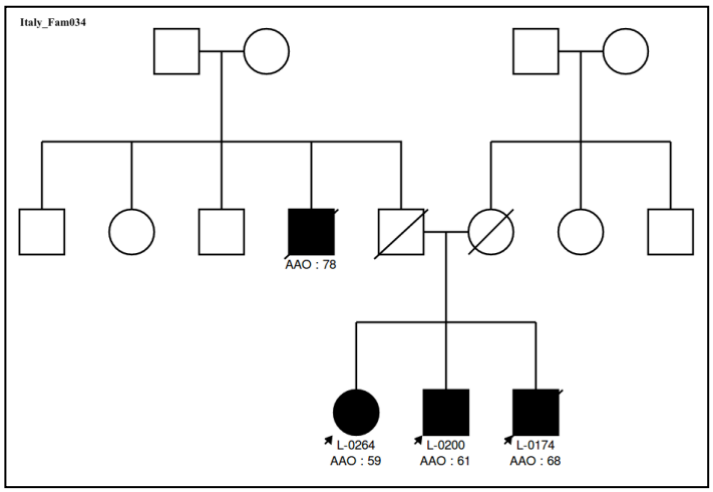

Turkey_PD313
1
2
5
4
3
AAO : 30

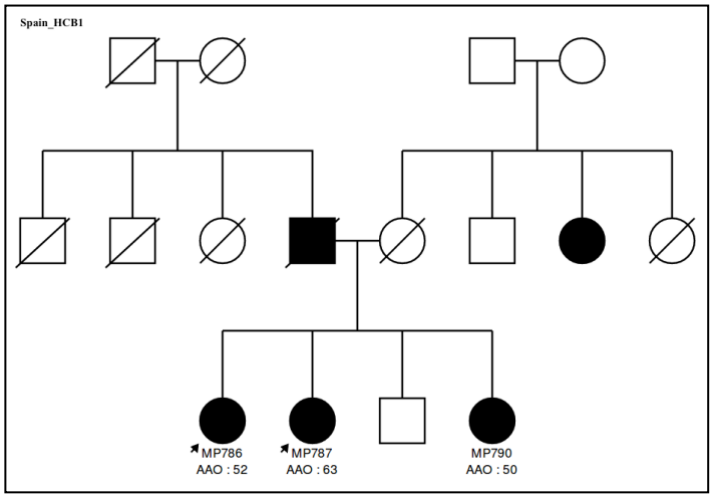

Turkey_PD296
1
2
3
AAO : 14
4


Turkey_PD314
PD
2
PD & memory
problems
1
Alzheimers
7
AGE : 68
memory problems
6
AGE : 69
AAO : 59
CBD ? PD/memory problems?
5
4
AGE : 72
3
AGE : 74
AAO : 60
?
neural disease


Turkey_PD291
1
AAO : 55
2
5
6
AAO : 44
4
3


Spain_2933
G10495
G10496
G10497


Italy_Fam034
AAO : 78
L-0264
AAO : 59
L-0200
AAO : 61
L-0174
AAO : 68


Spain_HCB1
MP786
AAO : 52
MP787
AAO : 63
MP790
AAO : 50

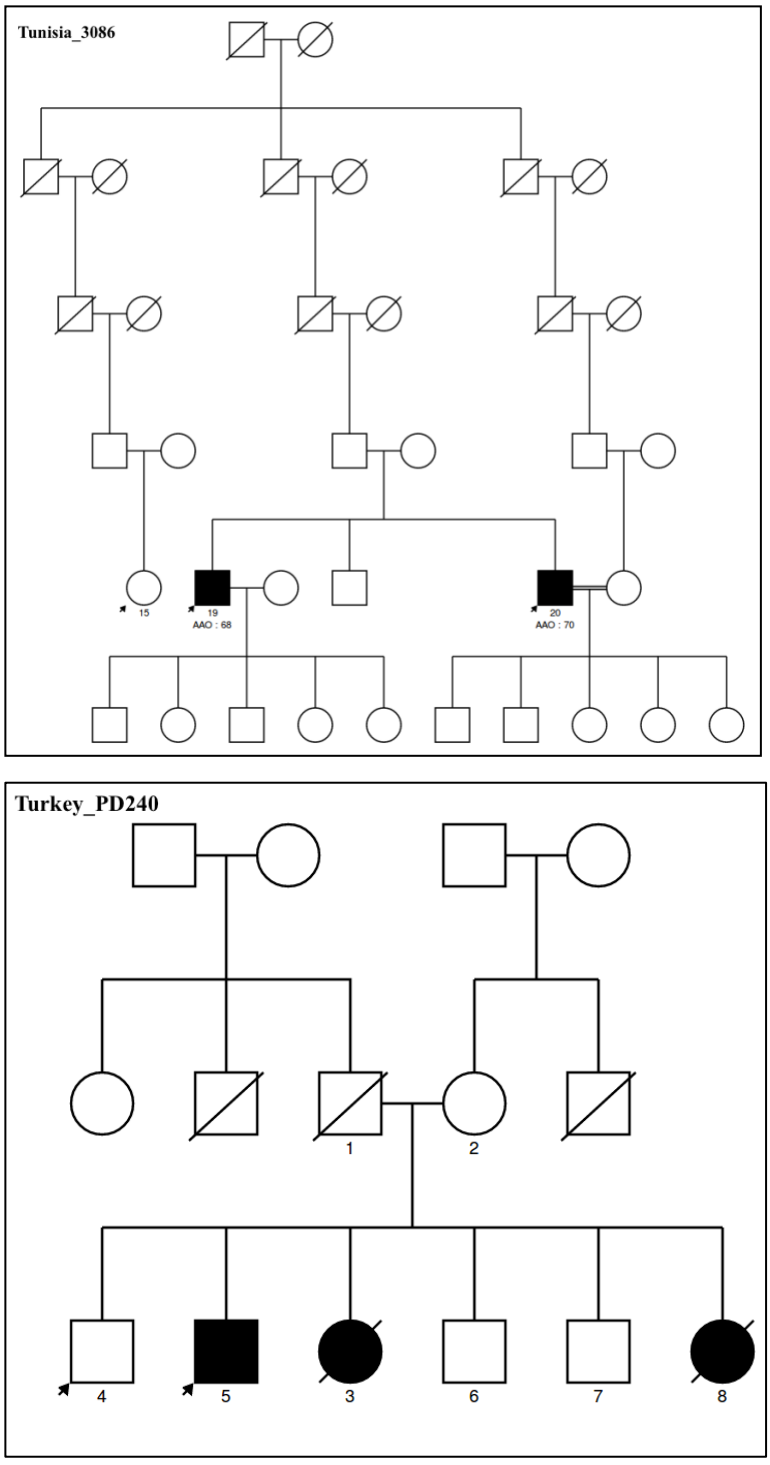


Figure 8: Pedigrees of families from discovery cohort.

Table 10: CNVs segregating in families from discovery cohort.

Cohort	sample	Poisson score	Copy Number	Chr	Start	End	Size	Genes	Feature
GRIP	D98-8553	1974.52	3	6	162204001	162439000	234999	PARK2	coding
GRIP	D98-8239, D98-8240	585	0	1	8018803	8032836	14033	PARK7	coding
GRIP	D98-8467	705.59	1	5	175791240	175817819	26579	HIGD2A; ARL10 ; NOP16 ; MIR1271	coding
GRIP	D98-8518	290.54	1	X	100213960	100227254	13294	ARL13A	5UTR
GRIP	D98-8541	521.5	1	7	66459649	66476339	16690	SBDS ; TYW1	coding
GRIP	D98-8623	185.06	1	2	56389768	56400376	10608	RP11-481J13.1	ncRNA
Italian	L-0174	141.88	1	11	76989993	76996482	6489	GDPD4	coding
Italian	L-0200	137.59	1	11	76989993	76996482	6489	GDPD4	coding
Italian	L-0264	140.99	1	11	76989993	76996482	6489	GDPD4	coding
Italian	F1326	130.62	1	16	67851933	67857174	5241	TSNAXIP1	coding
Italian	G-1261	284.41	1	1	245695557	245708504	12947	KIF26B	coding
Italian	P-0152	175.56	0	8	11708729	11714635	5906	CTSB	coding
Italian	R-0499	148.96	1	11	102209308	102215813	6505	BIRC3	coding
Italian	S-0074	269.75	1	1	242034087	242045637	11550	EXO1	coding
Italian	S-0074	217.12	1	4	152348381	152358593	10212	FAM160A1	coding
Portugal	1701195-71237	390.81	1	3	182179588	182196930	17342	RP11-139K4.1 ; FLJ46066	coding
Portugal	19120-75116	69.18	1	17	5273006	5275761	2755	RP11-420A6.2	coding
Portugal	19120-75116	397.18	0	19	48271007	48278181	7174	CTD-2571L23.6	coding
Portugal	19120-75116	520.17	1	22	32857793	32879303	21510	BPIFC ; FBXO7	coding
Portugal	367127102-20604	67.42	1	10	60995095	60997893	2798	PHYHIPL	coding
Portugal	367127102-20604	98.81	1	3	17741869	17746005	4136	TBC1D5	Intron
GRIP	D98-8225	214.11	3	4	146612001	146642000	29999	C4orf51	coding
GRIP	D98-8400	411.31	3	12	122999001	123077000	77999	KNTC1 ; RSRC2	coding
GRIP	D98-8463	527.5	3	7	107935001	107998000	62999	NRCAM	coding
GRIP	D98-8541	140.68	3	X	116081001	116099000	17999	SETP8	coding

GRIP	D98-8549	37.52	4	19	57776001	57782000	5999	ZNF805 ; CTC-444N24.8	downstream
GRIP	D98-8553	1974.52	3	6	162204001	162439000	234999	PARK2	coding
GRIP	D98-8558	45.29	3	19	4800001	4810000	9999	FEM1A ; AC005523.2	downstream
GRIP	D98-8558	61.05	3	2	55441001	55450000	8999	CLHC1	coding
GRIP	D98-8623	122.96	3	X	46315001	46337000	21999	KRBOX4	coding
GRIP	D98-8624	319.08	3	1	185095001	185132000	36999	TRMT1L ; SWT1	coding
GRIP	D98-8624	18.98	3	17	17122001	17126000	3999	FLCN	coding
GRIP	D98-8656	513.6	3	1	8077001	8152000	74999	ERRFI1 ; CTA-215D11.4 ; RP11-431K24.1	5UTR
GRIP	D98-8735	588.91	3	7	107935001	107999000	63999	NRCAM	coding
Turkish	PD296-3	1045.7	3	5	24437001	24544000	106999	CDH10	coding
Turkish	PD300-4	78.44	3	12	15063001	15074000	10999	ERP27 ; C12orf60	coding
Turkish	PD300-5	65.9	3	12	15063001	15075000	11999	ERP27 ; C12orf60	coding
Turkish	PD313-3	2142.05	3	16	60543001	60780000	236999	GNPATP ;RP11-354I13.1	coding
Turkish	PD317-11	240.78	3	9	14064001	14094000	29999	NFIB	coding
Turkish	PD317-4	280.78	3	9	14064001	14095000	30999	NFIB	coding
Turkish	PD317-8	247.92	3	9	14064001	14095000	30999	NFIB	coding
Italian	F1326	62.93	3	6	24373001	24385000	11999	DCDC2 ; LOC285819	5UTR
Italian	G-1261	645.05	3	X	33790001	33869000	78999	RP11-305F18.1 ; LOC105373153	Intron
Italian	M0160	17.75	4	6	41931001	41933000	1999	CCND3 ; BYSL	Intron
Italian	M-1249	90.24	3	8	90061001	90071000	9999	CALB1	coding
Italian	P-1806	85.34	3	4	125682001	125694000	11999	NUP58P1	coding
Portugal	1514090-93542	1327.91	3	22	29355001	29504000	148999	ZNRF3	coding
Portugal	1701195-71237	24.98	3	16	57050001	57054000	3999	NLRC5	coding

Table 11: WGCNA results for candidate genes from autosomal recessive and dominant families.

Gene	Module	GS.status	Pvalue.GS
OTOGL	black	-0.3390497667	0.0033431297
SLC22A10	black	-0.2586661437	0.0271292791
FRAS1	black	-0.2227132816	0.0582407541
TOX	black	-0.1429340601	0.2276800382
TMEM86B	blue	0.3757306037	0.0010538326
WBP4	blue	-0.3604966845	0.001730364
CCDC22	blue	0.3221938414	0.0054382815
DNAH9	blue	-0.2970224943	0.0107158678
SPEF2	blue	-0.2846376596	0.0146593845
SYTL5	blue	-0.2820478636	0.0156263661
TRIM68	blue	0.2431537282	0.0381843914
LRRK2	blue	-0.2269477618	0.0535015257
TRMT2A	blue	0.172252417	0.145050002
MAGIX	blue	0.0069715163	0.9533206673
SLC3A2	brown	0.6086642776	1.111321607394E-008
MFSD8	brown	-0.4607603161	4.0951370913064E-005
PRKDC	brown	-0.430762044	0.0001422426
KNTC1	brown	-0.3363285063	0.0036228854
ZBTB38	brown	-0.1203456047	0.3104978349
PARD3	cyan	0.5038537839	5.51294722264513E-006
WWC1	cyan	0.2713719203	0.0202142416
GPR143	cyan	0.2129128232	0.0705243791
PCSK5	cyan	0.1158287278	0.3291389848
ZNF837	darkgreen	-0.0398157252	0.7380424046
ARVCF	darkred	0.0428104806	0.7191278193
OBSCN	darkred	0.0192977455	0.8712670539
NEB	darkred	-0.0142479142	0.9047689923
SH3KBP1	darkturquoise	-0.1677937499	0.1559081526
RNF212	darkturquoise	0.036621944	0.7583862074
MAST3	darkturquoise	0.0210184253	0.8599000262
TCIRG1	green	0.5652909783	1.88282956042321E-007
TREM2	green	0.4500623614	6.4702821095638E-005
AKR1C3	green	0.3871920713	0.000714115
AHNAK	green	0.3867607486	0.0007248349
ADAMTS9	green	0.3638942997	0.0015523936
GART	greenyellow	-0.4336660898	0.0001267233
HEXA	greenyellow	-0.417645092	0.0002366419
ULK2	greenyellow	-0.3763425176	0.0010325286
FAM186A	greenyellow	0.3661514476	0.0014434631
MTMR2	greenyellow	-0.2147149102	0.0681216872
APC2	grey60	0.2210446139	0.0601996821
BCHE	grey60	-0.1568755363	0.1850271116
PINK1	grey60	-0.0937253383	0.4302835512

EPC1	lightcyan	0.265715492	0.0230803996
GBA	lightgreen	-0.3669296043	0.0014075385
PLA2G6	lightyellow	0.1217650028	0.304783699
AHI1	magenta	-0.4702892927	2.68946076723436E-005
SPHKAP	magenta	-0.4097714916	0.0003180653
DCHS2	magenta	-0.2484990311	0.0340139339
DOCK5	pink	0.2521478141	0.0313916937
CCNB3	pink	-0.2373083406	0.0432234204
DYSF	pink	0.214126626	0.0688986853
SLC45A3	pink	0.1743673621	0.1401022353
PLD1	pink	0.0995506435	0.4020504048
LRP2	pink	0.0835177903	0.4823743401
ZNF536	pink	-0.0752168871	0.5270875213
FBXO7	pink	-0.0738983522	0.5343755328
MAMLD1	pink	0.0168930267	0.8871961225
CRYZL1	purple	-0.485664752	1.32843176942749E-005
HCFC1	purple	0.3581667636	0.0018628058
ACACB	purple	0.3360473124	0.0036529486
RPA2	purple	-0.2114762178	0.0724881179
MAPKBP1	purple	0.1734501152	0.1422322327
DCAF10	purple	-0.1644919037	0.1643299994
CAAP1	purple	-0.0974746604	0.4119848622
PLEKHA6	red	-0.1193277341	0.3146378516
SBF1	red	0.0621563031	0.6013896255
ZNF521	royalblue	-0.1699201962	0.1506566768
LAMC3	salmon	0.3806669841	0.0008927501
ROBO4	salmon	0.3692898791	0.0013034533
TPM2	salmon	0.2905587814	0.012640462
IGFN1	salmon	0.075871542	0.523487547
SAMD9	salmon	0.0696433036	0.5582289863
PHC3	tan	-0.248536845	0.033985847
IGBP1	tan	0.1364529693	0.2496804072
PARK7	tan	0.0981667219	0.408657332
PLEKHH2	tan	-0.0422996931	0.7223423825
DHX57	turquoise	-0.5796680162	7.7117336889953E-008
DMXL2	turquoise	-0.5295977321	1.45162110059404E-006
DPP8	turquoise	-0.4862841059	1.2902939889743E-005
TENC1	turquoise	0.4583332905	4.54902471252079E-005
CCDC154	turquoise	0.4319540491	0.000135672
VPS35	turquoise	-0.423585785	0.0001884035
HS6ST2	turquoise	-0.4200393583	0.0002159787
ZNF223	turquoise	-0.4108469883	0.0003056042
SNCA	turquoise	-0.3956265127	0.0005314399
NLGN4X	turquoise	-0.3704344803	0.0012555251
PARK2	turquoise	-0.3566185123	0.0019557833
PHKA2	turquoise	0.3093730941	0.0077368411

DGKH	turquoise	-0.2865821522	0.0139676942
ZNF766	turquoise	0.2787451341	0.0169388229
SHANK2	turquoise	-0.2670018541	0.0223999276
ATP2B4	turquoise	-0.1596029675	0.1774089443
PITRM1	yellow	-0.3229444972	0.005324712
SPTBN2	yellow	-0.2450340863	0.0366712807
HTT	yellow	-0.2013999685	0.0875217282
RTN4R	yellow	-0.1450726385	0.2207266629
MSLNL	yellow	-0.1275507564	0.2822027431
GRB7	yellow	-0.1245424439	0.2938013021
MAP1A	yellow	-0.112379664	0.3438411479
TMEM63C	yellow	-0.0994998241	0.4022919047
FCF1	yellow	-0.0920893074	0.4384106881
TTLL12	yellow	0.0655148107	0.5818464394
PCNT	yellow	-0.0649449272	0.585142062
GCKR	yellow	-0.058020253	0.625847208
MYO15A	yellow	-0.0421332638	0.7233908201

*GS: Gene significance.

Project 2: Mitochondrial endo-phenotype of Parkinson's disease.

2.1 Introduction

Mitochondria play a prominent role in energy metabolism and are involved in various cellular processes such as cell death pathways, calcium homeostasis and stress response pathway (Osellame, Blacker et al. 2012). Mitochondria are a source of reactive oxygen species (ROS) and increase in ROS formation activates protective stress response pathway. Moreover, impairment of mitochondrial function increases oxidative stress, and it promotes aging and neurodegeneration (Fukae, Mizuno et al. 2007). In 1990s, Schapira et al. studied the structure and function of mitochondrial respiratory-chain enzyme proteins in substantia nigra pars compacta (SNpc) of postmortem samples from Parkinson's disease (PD) patients and controls and observed defects of complex I activity in the SNpc of these individuals (Schapira, Cooper et al. 1989). Hence, mitochondrial dysfunction plays a role in the pathogenesis of PD. However, mitochondrial dysfunction is a primary cause of disease in a small group of patients with *PARK2* and *PINK1* mutations in a family with an autosomal recessive mode of inheritance. Mitophagy, *i.e.* elimination of damaged mitochondria is disrupted due to loss of function mutation in these genes (Rakovic, Shurkewitsch et al. 2013). Parkin, which is encoded by *PARK2*, prevents mitochondrial swelling, cytochrome c release and caspase activation but Parkin mutations and proteasome inhibitor abrogate its protective effect

(Muftuoglu, Elibol et al. 2004, Palacino, Sagi et al. 2004). Moreover, the protective function of Parkin is also affected by S-nitrosylation, which impairs its ubiquitin-ligase activity (Clark, Dodson et al. 2006). *PINK1* deficiency results in loss of SNpc dopamine neurons, abnormal mitochondrial morphology, inhibition of complex I activity and ultimately increase in oxidative stress (Kitada, Pisani et al. 2007, Gautier, Kitada et al. 2008, Blesa, Trigo-Damas et al. 2015). Some studies have shown a decrease in complex I (i.e. NADH ubiquinone oxidoreductase) activity in patients with idiopathic PD (Parker, Boyson et al. 1989).

To understand whether a mitochondrial genetic subtype of PD could be identified from a large cohort of sporadic patients, all genes laying under the meta-analyses of PD Genome-Wide Association Studies (GWAS) (Nalls, Pankratz et al. 2014) were annotated with Gene Ontology, MitoCarta (Calvo, Clauser et al. 2016) and the Integrated Mitochondrial Protein Index databases. A total of 15 genes with a mitochondrial annotation were selected from the 24 PD risk loci (Nalls, Bras et al. 2015). These 15 genes were tagged with 8 Single Nucleotide Polymorphisms (SNPs) that were utilized to perform a series of risk profiling analyses in a large German cohort of 1,283 PD cases and 876 controls from exome genotyping array. On the basis of these risk scores, individuals were selected to undergo a series of clinical trials along with patients carrying mutations in Mendelian PD genes such as *PARK2* and *PINK1*. Results from these analyses will help to understand whether individuals with a high load of mitochondrial risk variants are comparable to those with causative mutations in nuclear mitochondrial genes.

The function of a gene is most likely affected by rare variants in protein coding regions, so we investigated the burden of rare variants in mitochondrial-related genes in sporadic PD cases compared to controls. The association tests were performed using a large repository of WES data from International Parkinson's disease Genomics Consortium (IPDGC). The results were further replicated using a genotyping array data (Nalls, Bras et al. 2015).

2.2 Methods

2.2.1 Selection of genes under PD GWAS peaks:

All significant 24 loci were selected from the PD GWAS published in 2014 (Nalls, Bras et al. 2015). PLINK (Chang, Chow et al. 2015) was used to identify all tagSNPs in strong linkage disequilibrium (LD) ($r2 \geq 0.8$) with the reported risk SNP, within a 1Mb window using data from the European population in 1000 genomes (http://www.internationalgenome.org/). Haplotype blocks were estimated by the default procedure implicated in PLINK by Gabriel et al., which defines it as regions with a low historical combination. Such haplotype blocks were determined for the most distant tagSNPs on both sides of the sentinel SNP (Gabriel, Schaffner et al. 2002). As low-frequency variants in coding regions are more likely to be functionally important, all coding regions within a 1Mb window surrounding each GWAS top-hit are included as targets, irrespective of estimated LD patterns.

2.2.2 Selection of Mitochondria-related genes:

We were interested only in genes with mitochondrial involvement, so we selected genes from MitoMiner (Smith, Blackshaw et al. 2012), which is a resource of mitochondrial localization evidence. It comprises of Gene Ontology, MitoCarta (Calvo, Clauser et al. 2016) and Integrated Mitochondrial Protein Index (IMPI) databases. MitoCarta comprises of 1,158 human and mouse genes encoding proteins with a strong support of mitochondrial localization (Calvo, Clauser et al. 2016). IMPI contains 1,550 human genes that encode mitochondrially localized proteins, out of which 1,130 are known to be mitochondrial, and 420 are predicted to be mitochondrial using the evidence in MitoMiner. Hence, we annotated the genes covered by the 24 risk loci in the PD GWAS (Nalls, Pankratz et al. 2014) using these three databases and selected the genes with mitochondrial annotation. The associated SNPs (Nalls, Pankratz et al. 2014) tagging these genes were further used for risk profiling.

2.2.3 Risk profiling study:

NeuroX consists of approximately 240,000 exonic variants standard to the Illumina Human Exome array, and 24,000 variants were focused on neurodegenerative diseases (Nalls, Bras et al. 2015). German population data from exome genotyping array NeuroX, comprising 1,283 PD cases and 876 controls were used for risk profiling. The associated SNPs from the meta-analysis of 2014 (Nalls, Pankratz et al. 2014) containing genes involved in mitochondrial function were extracted from this cohort. In order to estimate the cumulative

genetic risk per individual, allelic scoring was performed using PLINK1.9 (Chang, Chow et al. 2015). The allelic score is merely a sum of the number of effect alleles at each SNP multiplied by the natural logarithm (ln) of ODDs ratio (OR) for that SNP (Chang, Chow et al. 2015). The effect allele and the OR for the selected list of SNPs in the German population were retrieved from PDGENE (http://www.pdgene.org/). Moreover, by default, copies of unnamed allele contribute zero to score, while missing genotypes contribute an amount proportional to the imputed allele frequency (Chang, Chow et al. 2015). Age at onset (AAO) is also crucial along with the risk of the disease. AAO of PD has high heritability and has been reported to be associated with some of the common variants (Cai, Alp et al. 2002).To estimate a correlation between the age at onset (AAO) of PD and the determined risk scores, Pearson correlation coefficient and its significance were calculated using R.

2.2.4 Association analysis:

We performed association tests between a set of rare variants and dichotomous phenotypes. For the rare variants (MAF<0.01), gene-based optimal sequence kernel association test (SKAT-O) were performed. SKAT-O is a combined test, which selects significant result by comparing the one-sided model of burden test (all variants have the same direction of effect) and two-sided model of SKAT test (all variants have a distinct direction of effect) (Lee, Emond et al. 2012). It computes gene-level p-values by correcting population stratification with the inclusion of 20 multi-dimensional scaling components as covariates.

The following four criteria were used to see the enrichment of rare variants:

- all the rare variants.
- Functional variants (frameshift, non-frameshift, start-loss, stop-loss, stop-gain, splicing, missense, exonic, 5-prime untranslated regions (UTR5), 3-prime untranslated region (UTR3), upstream (-100bp), downstream (+100bp) and non-coding RNA (ncRNA)).
- Coding variants (frameshift, non-frameshift, start-loss, stop-loss, stop-gain, splicing and missense).
- Loss of function variants (frameshift, start-loss, stop-gain and splicing).

SKAT-O tests were performed using two independent datasets and is mentioned in sections 2.2.4.1 and 2.2.4.2.

2.2.4.1 IPDGC exomes:

The discovery cohort WES data consists of 1,450 PD cases and 535 controls from the IPDGC consortium, and 1,732 controls from the Rotterdam study (RSX1) WES data. Paired-end sequencing using Illumina HiSeq 2000 was performed for IPDGC and RSX1 data. The sequencing reads were aligned to the human reference genome (hg19) using BWA-MEM algorithm (Chang, Nguyen et al. 2010). Binary Alignment Map (BAM) files were generated using Picard tools (http://broadinstitute.github.io/picard). Further, variants and small insertions/deletions were called using Genome Analysis Toolkit (GATK) (McKenna, Hanna et al. 2010). The IPDGC and RSX1 datasets were merged by the joint variant calling of the g.VCF files. Genotype and variant quality control was

performed using KGGSeq (Cingolani, Platts et al. 2012) and is already mentioned in detail in the previous chapter. It was followed by individual quality controls by excluding samples with gender inconsistencies, heterozygosity outliers, duplicate samples, individuals of divergent ancestry and missing genotype. After QC, the IPDGC-RSX1 dataset comprises of 545,752 SNPs and 70,548 indels. The sequence-kernel association test (SKAT-O) was performed using EPACTS tool (https://genome.sph.umich.edu/wiki/EPACTS) to estimate the burden of rare variants in mitochondria-related genes.

2.2.4.2 NeuroX:

Data from 6,801 PD cases and 5,970 controls genotyped with the NeuroX array was used for replication. The QC of this data was performed by excluding variants with a minimum call rate less than 85%, a Hardy Weinberg Equilibrium (HWE) p-value less than 1E-06 in all the controls, or when the missingness rate was significantly different between cases and controls. Individuals with sex discordance, a heterozygosity rate of >3 standard deviations from the mean, more than 15% of missing genotypes, closely related to another individual in the dataset (>18.5%) or representing a population outlier after multi-dimensional scaling analysis were removed from further analyses. After QC, gene-based SKAT-O test was performed using EPACTS tool, same as in discovery cohort.

2.3 Results

2.3.1 Selection of genes under the PD GWAS peaks

The 24 PD risk loci from GWAS of 2014 consists of 697 genes, which were selected based on all the SNPs in LD to the associated SNP as well as all the genes under the haplotype block.

2.3.2 Selection of mitochondria-related genes

A total of 933 genes are annotated with a mitochondrial function, biological process or cellular component in all three databases utilized. Out of the 697 genes covered by the 24 risk loci, only 15 genes have the mitochondrial annotation. These 15 genes are tagged by 8 common SNPs associated with PD (table12).

2.3.3 Risk Profiling

The purpose of these analyses was to select all the associated variants with mitochondrial annotation from the sporadic PD data and perform risk profiling in the German population. The main aim of risk profiling was to stratify the patients in mito+ and mito- category. The mito+ group consists of all the patients with high risk and the mito- group with low risk to PD on the basis of mitochondrial phenotype. Results of the risk profile performed for the German population are shown in figure 9. Higher positive scores indicate a higher risk of developing PD due to mitochondrial phenotype. In the German population, high

allelic scores (score>0.35) are significantly more present in PD cases than in controls, while the low scores (score<-0.35) are over-represented in controls (table 13).

AAO of the disease might be genetically influenced and may have a similar effect to most of the risk genes. In our study, it was observed that AAO of all PD cases from Germany is significantly correlated to the risk scores (figure 10) (table 3). Several lines of evidence have shown mitochondrial impairment in PD patients due to *PARK2* and *PINK1* mutations (Clark, Dodson et al. 2006, Zanellati, Monti et al. 2015). Some of the German PD cases with allelic scores higher than 0.35 (mito+) or less than -0.35 (mito-) as well as PD patients carrying *PARK2/PINK1* mutations, will undergo clinical trials with coenzyme Q10 and vitamin K2. Coenzyme Q10 is an antioxidant and plays a vital role in mitochondrial electron transport pathway (Duberley, Heales et al. 2014).

Table 12: 8 associated SNPs whose locus has genes with mitochondrial annotation.

Associated SNP	Locus name	Effect allele	MAF (1000 Genomes, 2016)	OR (Germany)
rs329648	MIR4697	T	0.46	1.15
rs34311866	TMEM175-GAK-DGKQ	G	0.14	1.40
rs11868035	SREBF1-RAI1	A	0.49	0.97
rs14235	BCKDK-STX1B	A	0.36	1.19
rs11060180	CCDC62	G	0.25	0.90
rs71628662	GBA-SYT11	T	0.01	0.40
rs199347	GPNMB	C	0.48	0.97
rs12637471	MCCC1	A	0.34	0.67

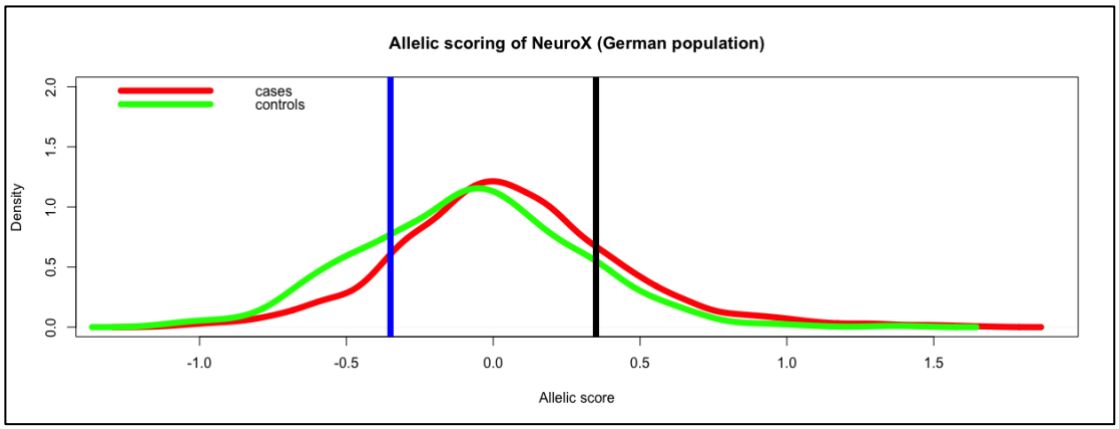


Figure 9: Density plot of the allelic scores calculated for German population: PD cases with score >0.35 represents the mito+ group, while cases with score <-0.35 represents the mito- group.

Table 13: Risk profiling of 8 GWAS associated SNPs in German population:

Allelic scores	German_Cases	German_Controls	Chi.sq test (P.value)
>0.35	240	104	0.0003303
<-0.35	152	197	3.747e-08

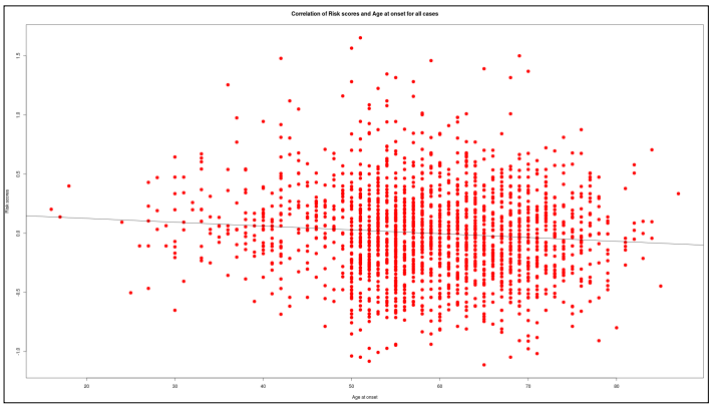


Figure 10: Correlation of risk scores and age at onset for all cases: AAO of all the German PD cases are significantly correlated to risk scores of the associated SNPs

.

Table 14: Correlation between allelic scores and age at onset of German PD cases

Correlation coefficient	Test statistic	P-value
-0.051	-2.18	0.02

2.3.4 Association analysis

2.3.4.1 IPDGC exomes:

The WES data from IPDGC consists of 840 mitochondria-related genes. The results of the gene-based SKAT-O analysis using the IPDGC discovery cohort for the mitochondria-related rare variants are presented in table 15. We used four different criteria to see the enrichment of rare variants. The first criteria show the burden of all rare variants (MAF<0.01) in the respective genes, secondly to see enrichment of functional variants and followed by coding and loss of function variants. In these gene-based analyses besides the expected association of *GBA* (P=4.2109E-006) to PD risk, we also identified an association of all rare and functional variants in *CS*, *MTHFD1L*, *FAM162A*, *POLRMT*, *AK4*, *MRPS11* and *ALDH1L1* to PD. Moreover, rare coding variants in *COX19*, and loss of function variants in *CISD2*, *HK2* and *ATPAF1* are highly enriched to PD.

Table 15: Gene-based SKAT-O results for IPDGC exomes.

Gene	$N_{allrare}$	$P_{allrare}$	$P_{functional}$	P_{coding}	P_{lof}
CS	31	1.706E-015	1.8029E-015	1.8029E-015	0.042088
FAM162A	1	0.000000045	0.000000045	0.000000045	0.000000045
MTHFD1L	67	7.9283E-008	2.6699E-006	0.000002693	0.000028237
POLRMT	75	1.0953E-007	8.9496E-008	8.9496E-008	0.73951
AK4	21	1.1047E-007	0.34662	0.33658	0.13605

GBA	67	4.2109E-006	0.000003508	1.7849E-006	0.058283
MRPS11	19	0.000015624	0.000045796	0.0016708	NA
ALDH1L1	159	0.000027259	0.071281	0.032693	0.88537
COX19	5	0.12036	0.062512	0.000012906	NA
CISD2	6	0.0724	0.13855	0.13855	2.2401E-012
HK2	68	0.20689	0.76275	0.75506	2.1593E-010
ATPAF1	32	1	0.91219	0.91219	0.000026057

* $N_{allrare}$: Number of all rare variants; $P_{allrare}$: P-value for all rare variants; $P_{functional}$: P-value for functional variants; P_{coding}: P-value for coding variants; P_{lof}: P-value for loss of function variants.

2.3.4.2 NeuroX:

The genotyping array NeuroX covers 814 mitochondria-related genes. This data was used to replicate the findings from discovery cohort SKAT-O analysis. This data consists of 6,801 PD cases and 5,970 controls of European ancestry. As already mentioned in the previous section, for this dataset as well, we used 4 criteria to see the enrichment of rare variants. The results of the gene-based SKAT-O results for NeuroX are mentioned in table 16. In the replication cohort, only *GBA* is replicated. In this cohort, Mendelian PD gene *LRRK2* and a novel gene *CHCHD6* shows genome-wide association to PD.

Table 16: Gene-based SKAT-O results for NeuroX.

Gene	$N_{allrare}$	$P_{allrare}$	$P_{functional}$	P_{coding}	P_{lof}
CHCHD6	9	5.1486E-041	5.1486E-041	5.1486E-041	0.33024
LRRK2	116	5.6104E-009	5.9088E-009	5.9088E-009	0.70593
GBA	13	3.1964E-006	2.1615E-006	2.1615E-006	NA
CS	1	0.42	0.42	0.42	NA
FAM162A	7	0.58	0.58	0.58	0.44
MTHFD1L	17	0.47	1	1	NA
POLRMT	6	0.14	0.14	0.14	NA

AK4	1	NA	NA	NA	NA
MRPS11	2	0.83	0.83	0.83	NA
ALDH1L1	27	1	1	1	0.53
COX19	2	0.41	0.41	0.41	NA
CISD2					
HK2	7	0.63	0.44	0.44	NA
ATPAF1	1	0.06	0.06	0.06	NA

* $N_{allrare}$: Number of all rare variants; $P_{allrare}$: P-value for all rare variants; $P_{functional}$: P-value for functional variants; P_{coding}: P-value for coding variants; P_{lof}: P-value for loss of function variants.

2.4 Discussion

Parkinson's disease is quite heterogeneous concerning genetic, environmental and clinical point of view, so there is a need for defining subtypes of PD that allows assigning therapeutic approaches. Genetic evidence and biochemical studies indicate that mitochondrial dysfunction, ensuing cellular energy failure and oxidative stress may be one of these crucial disease pathways in a subgroup of PD patients (Reeve, Grady et al. 2018). So, the important question in this study was to identify the sub-group, *i.e.* the mitochondrial endo-phenotype of PD using existing genomic data. We hypothesized that if an individual's allelic score represents an overall load of PD risk alleles, then an individual's disease liability should be somehow related to their allelic score. AAO of PD has high heritability, so we also investigated the relationship between age at onset and allelic scores. The cumulative genetic risk per individual in the German population was estimated using the 8 GWAS associated SNPs tagging genes with mitochondrial annotation. Our analysis shows that 240 PD cases and 104 controls have an allelic score greater than 0.35 and these cases were stratified as mito+ group. Moreover, 152 cases and 197 controls have allelic scores less than -0.35, and

these cases were stratified as mito- group. Hence, positive allelic scores were significantly present in more cases than in controls, while the negative scores were over-represented in controls. The age at onset of all German cases was significantly correlated to the risk scores. Some of the patients from these mito+ and mito- subgroups as well as PD patients with *PARK2* and *PINK1* mutations will undergo clinical trials with coenzyme Q10 and vitamin K2, in order to improve complex I activity. The application of drugs targeting complex I function has already been mentioned (Marella, Seo et al. 2009). Coenzyme Q10 is an anti-oxidant, and it influences mitochondrial activity including complex I (Zhu, Sun et al. 2017).

As rare variants in a coding region most likely affect the function of the gene, so we determined the burden of rare variants in nuclear mitochondrial genes. We investigated a large repository of WES data and identified that beside *GBA* some other genes such as *CS*, *MTHFD1L*, *FAM162A*, *POLRMT*, *AK4*, *CISD2*, and *HK2* also showed genome-wide association to PD. The protein encoded by the gene Citrate Synthase (*CS*) is a rate-limiting enzyme of Krebs tricarboxylic acid cycle and plays a role in regulating energy generation of mitochondrial respiration. Abnormal trafficking of *CS* isoform-a plays a role in mitochondrial dysfunction (Cheng, Liao et al. 2009). *MTHFD1L*, which encodes the methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like protein) influences homocysteine-related pathways as it plays a role in the generation of methionine from homocysteine and has been associated with late-onset Alzheimer's disease (AD) (Naj, Beecham et al. 2010). Moreover, high plasma homocysteine levels have been implicated in neurodegenerative diseases like AD and PD (Herrmann and Knapp 2002). As per experimental evidence of Affinity

Capture-Mass spectrometry *MTHFD1L* interacts with *NDUFS6* (NADH: Ubiquinone Oxidoreductase Subunit S6), *i.e.* Complex I. Diseases associated with NDUFS6 includes Mitochondrial Complex I deficiency and Leigh syndrome (Kirby, Salemi et al. 2004). Several lines of evidence have suggested that impairment of complex I contribute to PD pathogenesis (Schapira, Cooper et al. 1990). *FAM162A* is involved in positive regulation of apoptotic process and hypoxia-induced cell death of neuronal process (Lee, Kim et al. 2004). *POLRMT* encodes a mitochondrial RNA polymerase and is responsible for mitochondrial gene expression. The protein encoded by *CISD2* is a zinc finger protein and may play a role in calcium homeostasis. It also contributes to antagonizing BECN1-mediated cellular autophagy in the endoplasmic reticulum (Chang, Nguyen et al. 2010).

We further replicated our findings in genotyping array and observed that only *GBA* from the discovery cohort is significantly associated with PD. Moreover, rare variants in PD Mendelian gene *LRRK2* and a novel gene *CHCHD6* show genome-wide association to PD. The novel gene *CHCHD6* is involved in biological processes such as cristae formation and cellular response to DNA damage stimulus. It is involved in regulation of mitochondrial cristae morphology, cell growth and ATP production. The mitochondrial functions are dependent on the structural integrity of mitochondrial cristae (An, Shi et al. 2012). Our current study results implicate new PD associated genes apart from the known PD causal genes such as *LRRK2* and *GBA*. Two frequent *LRRK2* mutations p. G2019S and p. R1441C increases mitochondrial autophagy and accumulation of autophagic vacuoles, resulting in degeneration of dopaminergic neurons (Esteves and Cardoso 2016). The published report shows that in a cell

culture model inhibition of *GBA* causes impairment in mitochondrial function (Cleeter, Chau et al. 2013). Lack of replication could be due to differences in age of disease onset and the laboratory techniques applied to both WES and NeuroX data. Since the NeuroX data was restricted to variants included in the genotyping array, it could be possible that many pathogenic variants were missed. Further investigation on the functional level might improve our understanding of these candidate genes associated with PD.

Project 3: Transcriptomic analysis of genes under the PD GWAS loci.

3.1 Introduction

Genome-wide association studies (GWAS) helps to estimate an association of single nucleotide polymorphisms (SNPs) and trait. It helps to perform a statistical comparison of allele frequencies between cases and controls and has the power to study the genetic basis of disease by assaying the entire genome (Visscher, Wray et al. 2017). GWAS have identified many common genetic variants associated with PD. Early GWAS has produced inconsistent results due to small sample size (Evangelou, Maraganore et al. 2007) but recent GWAS meta-analysis has identified loci such as *LRRK2*, *SNCA*, *GAK*, *BST1* and *PARK16*, which were confirmed in the independent follow-up studies (Pankratz, Wilk et al. 2009, Satake, Nakabayashi et al. 2009, Simon-Sanchez, Schulte et al. 2009). In 2014, a large meta-analysis of GWAS identified 24 risk loci for PD (Nalls, Bras et al. 2015). Recently, a large meta GWAS identified 17 new loci for PD, so now 41 loci have been associated with Parkinson's disease risk in individuals of European ancestry (Chang, Nalls et al. 2017). Additionally, GWAS of PD in East Asians showed strong association signal at *SNCA*, *LRRK2* and *MCCC1*, which confirms a role of these loci in European and Asian PD (Foo, Tan et al. 2017).

mRNA sequencing is an accurate and highly sensitive means to quantify gene expression. It gives complete information of the coding transcriptome and can identify known and novel transcript isoforms and gene fusions. Assessment of

gene function by silencing gene expression using plasmids has been facilitated by RNA interference (RNAi) (Check 2007). Availability of short hairpin RNA (shRNA) libraries has made it possible to relate genes or cluster of genes to a normal or pathological biological process using gene expression analysis. Our interest in the study was to elucidate the genes responsible for PD risk at the loci implicated by the latest PD GWAS. Genes under the associated PD loci were silenced by shRNA and gene expression was quantified using mRNA sequencing. We assessed mRNA levels differences between the knockdown samples and scrambles (controls). Furthermore, we also investigated the over-represented gene ontology and pathways in differentially expressed genes.

GWAS approaches helps to identify common risk variants, while the missing heritability can be explained by rare variants. The function of a gene is most likely affected by rare variants in protein-coding regions. Hence, we investigated the genetic burden of rare variants in the genes under PD GWAS loci using exome-sequenced data generated for PD risk loci from a meta-analysis of 2011 (International Parkinson Disease Genomics, Nalls et al. 2011).

3.2 Methodology

3.2.1 Selection of genes under the GWAS loci:

All significantly associated loci from GWAS on PD were selected (Nalls, Pankratz et al. 2014, Chang, Nalls et al. 2017). PLINK was used to identify all tagSNPs at $r^2 \geq 0.8$ with the reported GWAS hit, within a 1Mb window using data from the European population in 1000 genomes. Haplotype blocks were estimated by the default procedure implicated in PLINK for the most distant tagSNPs on both

sides of the GWAS SNP (Gabriel, Schaffner et al. 2002). As low-frequency variants in coding regions are more likely to be functionally important, all coding regions within a 1Mb window surrounding each GWAS top-hit are included as targets, irrespective of estimated LD patterns (Purcell, Neale et al. 2007).

3.2.2 Knockdown of genes:

Genes such as *ATP13A2*, *RGS10*, *TIAL1*, *MCMBP*, *SEC23IP*, *INPP5F*, *BAG3*, *BST1*, *CCDC62*, *COASY*, *DLG2*, *FDFT1*, *ITGA8*, *MAP4K4*, *MCCC1*, *NUCKS1*, *RAB29*, *GCH1*, *WDHD1*, *DALRD3*, *ARIH2*, *IMPDH2*, *P4HTM*, *PRKAR2A*, *QARS*, *QRICH1*, *USP19*, *WDR6*, *PRSS8*, *VKORC1*, *KAT8*, *STX4*, *ZNF668*, *BIN3*, *PDLIM2*, *ELOVL7*, *ERCC8*, *SIPA1L2*, *SREBF1*, *SYT17*, *TMEM175*, *DGKQ*, *GAK* and *ZFYVE26* were successfully knockdown with at least 50% efficiency by introducing shRNA into human dopaminergic BE(2)-M17 neuroblastoma cell lines through infection with virus. The scrambled controls were replicates of the same treatment. The goal was to mimic the effects of the viral transduction without the shRNA having any effect. Thus, the scrambled control is a vector, which does not target any gene. The sequence of the shRNA has been scrambled. Since the viral transduction also affects gene expression, this is a much better control than cells which have not been transduced.

3.2.3 Transcriptomic analysis:

mRNA sequencing: RNA was isolated on EpMotion using a Promega 96 well kit. cDNA of normalized RNA has been synthesized using Mantis (https://formulatrix.com/liquid-handling-systems/mantis-liquid-handler/). Sequencing was performed at CeGat (https://www.cegat.de/). Samples were

sequenced on a SR Flowcell with 1*50bp and Dual-Index on the Illumina HiSeq4000.

Preliminary quality control or pre-processing of the reads was done using Trimmomatic (Bolger, Lohse et al. 2014). It performs sliding window trimming. Cutting is done once the average quality within the window of 4 bases falls below 20. Bases with a quality score less than 20 from the beginning and end of the read were removed. Reads were dropped if a length is less than 50bp. Quality control reports were generated using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were mapped against the human reference genome (build hg38) using HISAT2 (Kim, Langmead et al. 2015). This mapping information was used to quantify reads further. HISAT uses two types of indexes for alignment, whole-genome index and tens of thousands of small local indexes. Both types of the index are constructed using the Burrows-Wheeler transform (BWT)/Ferragina-Manzini(FM) index (Kim, Langmead et al. 2015). Gene expression quantification was performed using featureCounts (Liao, Smyth et al. 2014). This program assign reads to features by implementing chromosome hashing and feature blocking. Exons were defined to be features and genes as meta-features and quantification was done at meta-feature level. As this is mRNA-seq data, reads that overlapped multiple genes were excluded from the counts (Liao, Smyth et al. 2014). Gene annotation was done using hg38 Ensembl GTF (General Transfer Format) file (https://www.ensembl.org). Normalization between samples was performed using relative log expression (RLE) method. All the genes with counts per million greater than 1 in at least 50% of the KD cell lines and 20% of the control samples were retrieved. Furthermore, the estimation of the mRNA level difference

between the knockdown samples of each gene and scrambles (controls) was performed. All these steps of normalization, filtering and log fold change (logFC) calculation have been performed using EdgeR (Robinson, McCarthy et al. 2010).

Gene-set Enrichment:

Gene set over-representation has been performed using HTSAnalyzeR package in R (Wang, Terfve et al. 2011). Over-representation analysis is performed based on the hypergeometric test (Subramanian, Tamayo et al. 2005). It is followed by a superexact test to compute statistical distributions of multi-set intersections. This R package helps to intersect over-represented biological processes (BP) and pathways from all knockdown genes under the associated loci as well as Mendelian PD genes (*PARK7, PINK1* and *VPS35*) knockdown cell lines. This approach helps to know which over-represented biological processes and pathways are overlapped on silencing PD causal genes and genes under the associated loci of PD.

3.2.4 Association analysis:

We used NGS data generated for the 11 PD risk loci such as *MAPT, SNCA, HLA-DRB5, BST1, GAK, LRRK2, ACMSD, STK39, MCCC1/LAMP3, SYT11,* and *CCDC62/HIP1R* as per the meta-analysis of 2011 (International Parkinson Disease Genomics, Nalls et al. 2011). In the IPDGC resequencing project, there are total 4,514 samples. A total of 3,060 patients diagnosed with Parkinson's disease and 1,454 controls were included in the study. KGGSeq (Cingolani, Platts et al. 2012) was used to apply quality control (QC) filters on individual genotypes

and genomic variants. In brief, genotypes with a Phred quality score below 20, a read depth of less than 10, the second smallest normalized Phred-scaled genotype likelihood below 20 were removed. A minimum overall sequencing Phred quality score below 50, an overall mapping quality Phred score below 20, an overall strand bias Phred-scaled p-value (using Fisher's exact test) above 6 were considered. Variants within putative super-duplicate genomic regions were excluded. Only sites with Minor Allele Count greater than or equal to 1 were included. Sites with more than 5% missing data and markers with different genotype call rate between cases and controls (P<5.0E-8) were excluded. Related individuals, as well as individuals with genotype failure rate greater than 5% and heterozygosity rate >3 standard deviations from the mean, were excluded. Individuals with more than 5% missing genotype and divergent ancestry were removed.

After performing quality control 74,497 rare variants (MAF < 0.01) were retrieved. The sequence-kernel association test (SKAT-O) was performed using EPACTS tool to estimate the burden of rare variants in genes under PD GWAS loci between PD cases and controls.

3.3 Results

3.3.1 Transcriptomic analysis

The mRNA sequencing was performed after efficiently silencing 44 genes under the PD GWAS loci. About 78% of the reads were mapped to the human reference

genome. The mRNA levels differences between the knockdown samples and scrambles (controls) in terms of logFC are listed in table 17. α-synuclein is genetically and neuropathologically linked to PD and also contributes to its pathogenesis (Mezey, Dehejia et al. 1998), Moreover, several lines of evidence have shown that increase of *SNCA* dosage (duplications and triplications) cause familial PD (Chartier-Harlin, Kachergus et al. , Singleton, Farrer et al. 2003). Hence, we investigated *SNCA* mRNA levels in knockdown cell lines compared to controls. α-synuclein was upregulated in *COASY, ITGA8* and *P4HTM* KD cell lines and downregulated in *BAG3, GCH1, WDHD1, GAK* and *QRICH1* KD cell lines (figure 11). Over-represented biological processes and pathways for the differentially expressed genes between the knockdown samples and controls are listed in table 19. We prioritized the knockdown genes on the basis of differentially regulated α-synuclein mRNA levels. The enriched biological processes in all such prioritized knockdown genes are signaling pathways such as integrin-mediated signaling pathway, G-protein coupled receptor signaling pathway, cytokine-mediated signaling pathway, chemical synaptic transmission, inflammatory response, potassium ion transport, microtubule-based movement, positive regulation of synapse assembly and activation of protein kinase activity. Moreover, ECM-receptor interaction, Cardiomyopathy related pathways, Calcium Signaling pathway and Cytokine-cytokine receptor interaction are over-represented in the differentially regulated genes of the knockdown samples. Mendelian PD genes such as *PINK1*, *PARK7* and *VPS35* were also knockdown in the neuroblastoma cell lines. Targeted genes such as *ITGA8*, *DLG2*, *GCH1* and *PRSS8* were upregulated on silencing *PINK1* expression by shRNA. However, in *PARK7* knockdown samples compared to scrambles, it was observed that *ITGA8*,

DLG2, *SIPA1L2* and *PRSS8* were upregulated. Biological processes namely inflammatory response and G-protein coupled receptor-signaling pathway, and KEGG pathways such as Cytokine-cytokine receptor interaction and Systemic lupus erythematosus were over-represented in the differentially regulated genes of the knockdown samples *PARK7*, *VPS35* and *PLA2G6* compared to controls.

Table 17: logFC for PD causal genes in all knockdown samples.

Loci	KD_gene	KD_gene_logFC	*LRRK2*	*PRKN*	*SNCA*	*PARK7*	*UCHL1*	*VPS35*	*VPS13C*	*PINK1*
ATP13A2	*ATP13A2*	-3.4001489986	1.1213126192	-1.7332130356	-0.3762245145	-0.1885204575	-0.1801326195	0.0166611073	0.2400252982	-0.2386525382
BAG3	*RGS10*	-5.6839915855	0.1732731149	1.6048883787	-0.1446958878	-0.352208502	0.1051298671	0.0970311092	0.2924230163	1.2590587143
BAG3	*TIAL1*	-2.2725924059	-0.2004797731	1.1765423192	-0.0700936341	0.12323508	0.2011502427	-0.2010760343	0.1222476365	-0.5453874303
BAG3	*MCMBP*	-1.248804293	-0.1739839149	0.8686284941	-0.6633991485	-0.0305408472	-0.6352692575	-0.2086198415	0.582794857	-0.28659651
BAG3	*SEC23IP*	-1.749333671	-1.201417986	0.3629724111	-0.0370451061	-0.0981230816	-0.6095311137	0.0934845418	-0.4303787973	0.7767960256
BAG3	*INPP5F*	-1.562700554	1.2341259078	0.7697803294	0.2214197322	0.0067806853	0.0289317621	-0.320137424	0.2344492987	0.3477868803
BAG3	*BAG3*	-4.8520647604	-1.1389867618	2.7831876465	-2.0385274195	0.168619571	0.3508633254	-0.2154392144	0.7184181391	-0.5469920256
BST1	*BST1*	-5.4846570989	-0.0724520332	1.8726145056	-0.4571225779	-0.4201967678	0.1210338826	0.2251208694	0.1570401938	0.0207945147
CCDC62	*CCDC62*	-0.5529912921	-2.5743434995	-0.1927281824	0.026926012	0.0191856597	0.2441198597	-0.2118674274	0.0721776804	0.2063032723
COASY	*COASY*	-1.6016872511	-4.4579391869	0.4467583692	1.134467799	-0.3318350301	0.3694866527	-0.7385709039	0.9746855374	-0.1401101933
DLG2	*DLG2*	0.0839970286	0.9165742122	-0.147092634	1.8265965726	0.012151361	0.0503391647	0.2239031596	0.2427578617	0.6727340019
FDFT1	*FDFT1*	-1.8632102833	-0.5307521307	-0.4392129434	-0.3593633873	-0.0391818494	-0.0631387459	-0.2734277378	0.3317302549	0.1156346214
ITGA8	*ITGA8*	-5.154171029	-0.1707847161	0.3385726536	1.0242087623	-0.3432235444	-0.165475612	-0.0126993222	0.0613794474	-0.2049559398
MAP4K4	*MAP4K4*	-1.5620081396	0.7169362781	0.494039341	0.4572511753	-0.0653823416	0.0268390591	0.0203986169	0.0004982016	0.6228608735
MCCC1	*MCCC1*	-2.1109031623	-0.6398163754	0.4086431361	-0.5312585069	-0.4360169751	-0.4944473035	0.0963619414	-0.1169375276	0.0459323393
PARK16	*NUCKS1*	-2.8851325363	-4.4126919891	0.8657870583	-0.4586680954	0.0486245625	-0.6475400512	-0.2212506889	-0.2154837932	0.0438140816
PARK16	*RAB29*	-2.6066768641	0.7974053689	-0.3207021753	-0.8411159111	0.0006156777	-0.2571994658	0.2067113864	0.2222535032	-0.2823055881
rs11158026	*GCH1*	-2.0241468016	-2.1708627558	0.0122804265	-1.0386589238	-0.1717354183	-0.2024217653	-0.1286329549	-0.0788226612	-0.0534477707
rs11158026	*WDHD1*	-1.65012158	-0.7992780286	3.5390600304	-1.1374911725	-0.1484818341	-0.4407589932	-0.0279271638	0.6385843198	-0.5195465397
rs12497850	*DALRD3*	-4.2737216987	-0.4248411989	1.1976333617	0.392846637	-0.1461696674	-0.5893079144	-0.0720607021	-0.4146083087	-0.7180167912
rs12497850	*ARIH2*	-0.9472754935	0.1341090378	-0.7282346718	-0.1363021963	-0.2269595431	0.2541038207	-0.0777846402	0.3883354082	0.3872929447
rs12497850	*IMPDH2*	-0.0575059397	0.4566352755	1.7395651242	0.369795278	-	-	-	0.4118546761	0.2787305342

						0.2624250689	0.2185897592	0.4539276498		
rs12497850	*P4HTM*	-1.0533898178	0.5704631948	-1.5322867979	1.9421932857	-0.2344901076	0.4426121281	0.6866660488	0.1377922117	0.8930181453
rs12497850	*PRKAR2A*	-3.3264677175	1.8846506413	0.9821846677	-0.558949073	-0.3284393951	-0.6739589133	-0.2750235984	0.3332589614	0.3798352641
rs12497850	*QARS*	-0.6252627618	0.3948452451	0.7204573656	1.2825765935	0.2220618859	0.0639812742	0.2599526317	0.9696136776	0.8476093107
rs12497850	*QRICH1*	-1.7841169334	-0.0684554293	1.255073418	-1.222228491	-0.1913239121	-0.0043373183	-0.023238647	0.7600834048	0.4014726593
rs12497850	*USP19*	-2.0478829789	0.809059121	1.5518884182	-0.253263611	-0.0880613037	-0.136909176	0.1101560827	-0.2931406288	0.7424688873
rs12497850	*WDR6*	-3.0289406706	2.3059084452	-2.2526547227	0.6178067249	-0.3102528229	0.0831956881	-0.0755419741	0.4477298156	0.2216108138
rs14235	*PRSS8*	-0.0465932443	-1.3968959368	-1.7183627152	0.348897949	-0.1278541174	0.1474951938	-0.1096376857	0.459681461	0.0074062285
rs14235	*VKORC1*	-3.2401052651	1.0944056737	0.7998675142	0.8828306524	-0.2072590447	-0.4923051147	0.2757155894	0.4965117583	0.0981011805
rs14235	*KAT8*	-1.570074826	-1.8311318947	1.7085060134	-0.949627698	-0.0372948504	0.4483743776	-0.0073118299	0.2734308309	-0.8659019725
rs14235	*STX4*	-1.7441723824	0.0247153253	2.0258906846	0.0332929111	-0.0886878267	-0.8474356042	0.2896624418	-0.138462246	-0.0317020811
rs14235	*ZNF668*	-0.9382411373	0.0851992202	0.8977235156	0.4798709574	-0.2219393293	-0.3070905512	0.081111336	0.5235440906	0.2310726593
rs2280104	*BIN3*	-0.7463222907	-0.9468858039	0.8878247517	0.5337013916	-0.0223195623	-0.5385374473	-0.2364508321	-0.1698382679	0.2363334984
rs2280104	*PDLIM2*	-1.6263117035	0.3490086397	1.1872906181	0.4759124241	-0.2866509346	-0.3493364669	0.1316148729	0.322929798	0.353267631
rs2694528	*ELOVL7*	-4.6919684023	1.1748260007	-4.3221827298	0.0621673932	-0.1045713641	-0.4106390163	0.1459480366	0.0114206766	0.1052541572
rs2694528	*ERCC8*	-3.2783039739	-0.5329279711	-0.959679627	0.5240742184	0.0713227307	-0.1083426646	-0.5009776863	0.3066381483	0.3785529317
SIPA1L2	*SIPA1L2*	0.1386118429	0.0358088297	0.146917271	-0.4402458777	-0.8798593814	0.3442089194	0.4971851515	0.6629449464	0.346633125
SREBF1	*SREBF1*	-1.6542147224	-2.1911500569	1.1335302882	-0.2883126387	0.0545415309	0.3931279696	0.2680123783	0.287912313	0.7680107705
SYT17	*SYT17*	-2.308686327	0.4448659347	0.1003651719	0.2909696441	-0.1404947507	-0.514233593	0.0317583359	0.1515022683	0.3357771047
TMEM175	*TMEM175*	-1.0411630705	-4.5197032244	-4.3962580173	-0.5900674364	-0.0056393412	0.1536694835	0.0996739655	-0.8677081514	0.1767587798
TMEM175	*DGKQ*	-0.661887731	-0.0389941613	-0.6100671752	0.556540442	-0.3108220165	-0.1410464983	-0.0258887466	-0.060947155	-0.0027338225
TMEM175	*GAK*	-1.2555048424	-0.1254781513	1.1806117938	-1.75303873	0.1427104736	-0.8099527943	-0.4777903028	0.1014663699	-1.0858164517
ZFYVE26	*ZFYVE26*	-0.6312201894	1.413370213	0.5636465923	0.7781344678	-0.2411392395	0.1196892648	0.2967592783	0.3251596076	0.8503687596

*** KD_gene: knockdown gene; KD_genes_logFC: log Fold Change of knockdown gene.**

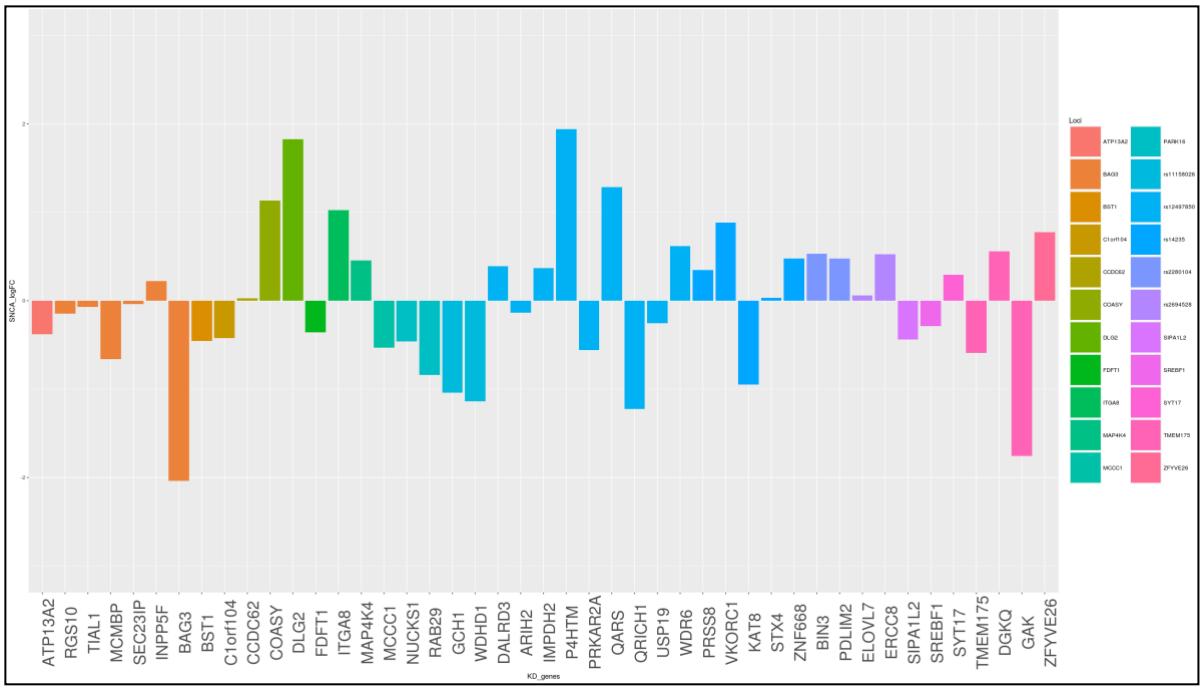


Figure 11: *SNCA* mRNA level differences (logFC) between knockdown cell lines and controls.

3.3.2 Association analysis:

The resequencing data from IPDGC comprises of only ~20 genes out of the ~50 genes, which were knockdown using shRNA. The results of the gene-based SKAT-O analysis for the rare variants in genes under the PD GWAS loci are presented in table 18. We used four criteria to see the enrichment of rare variants. The first criteria show the burden of all rare variants in the respective genes, secondly to see enrichment of functional variants and followed by coding and loss of function variants. In these gene-based analyses we identified an association of coding variants in *GAK* and *LRRK2* to PD (see table 18).

Table 18: Gene-based SKAT-O analysis for rare variants in knockdown genes.

Gene	$N_{allrare}$	$P_{allrare}$	$P_{functional}$	P_{coding}	P_{lof}
ATP13A2	56	0.36678	0.44989	0.53617	0.52576
BST1	608	0.51247	1	0.93091	0.49751
CCDC62	582	0.51461	0.54879	0.5641	0.46173
DGKQ	164	0.13441	0.26522	0.35556	0.22424
GAK	1711	0.34098	0.04545	0.022905	0.27866
KAT8	60	0.5162	0.60767	0.56946	0.78355
LRRK2	3309	0.17845	0.04513	0.029037	0.64283
MCCC1	1285	0.70264	0.62297	0.42299	0.44652
NUCKS1	687	0.28171	0.24955	0.068329	0.53388
PARK2	194	0.52878	0.43316	0.13434	0.084546
PARK7	114	0.35533	0.16721	0.10512	0.45157
PINK1	97	0.138	0.2203	0.2518	0.4011
PLA2G6	85	0.2725	0.30189	0.19178	0.34141
PRSS8	19	0.55601	0.54112	0.81371	NA
SNCA	2431	0.34682	0.33577	0.45705	0.1228
SREBF1	43	0.22589	0.44998	0.45611	0.63999
STX4	33	0.64142	0.42503	0.36977	0.27375
TMEM175	416	0.45443	0.63806	0.65661	0.35907
VKORC1	19	0.51004	0.43402	0.55915	NA
VPS35	72	1	0.45918	0.76235	NA
ZNF668	63	0.082066	0.073258	0.21373	0.55596

* $N_{allrare}$: Number of all rare variants; $P_{allrare}$: P-value for all rare variants; $P_{functional}$: P-value for functional variants; P_{coding}: P-value for coding variants; P_{lof}: P-value for loss of function variants.

3.4 Discussion

The meta-analysis of PD GWAS has identified 41 risk loci (Chang, Nalls et al. 2017). It is often difficult to identify specific genes responsible for risk at GWAS loci, as the regions showing significant association can span multiple genes and the nearby gene to SNP with the strongest P-value is not necessarily the one responsible for the association. We investigated the genes under the PD GWAS

loci to determine the most plausible gene contributing to PD risk and also to investigate their role in PD pathways. Our study offers an insight into the transcriptome analysis by silencing genes in neuroblastoma cell lines using short hairpin RNAs.

In this study, we discuss some of the loci in which the targeted genes were successfully silenced with at least 50% efficiency through shRNA. In the *BAG3* locus, amongst the six genes *BAG3, INPP5F, SEC23IP, MCMBP, RGS20* and *TIAL1*, which were knockdown, α-synuclein mRNA levels were downregulated by two folds in *BAG3* KD samples. Biological processes such as integrin-mediated signaling pathway, chemical synaptic transmission, immune response and cell adhesion related processes as well as ECM-receptor interaction and Cardiomyopathy related pathways were over-represented in *BAG3* KD samples. As per published studies, Extra Cellular Matrix (ECM) and integrin aberrations are likely to contribute to imbalanced synaptic functions (Wu and Reddy 2012). *BAG3* seems to be a plausible gene in this locus as it also plays a role in regulating *SNCA* clearance via macroautophagy (Cao, Yang et al. 2017). Experimental evidence of Affinity Capture-Luminescence in the BioGRID database (https://thebiogrid.org/) shows that *BAG3* interacts with *PLA2G6* (Chen, Yang et al. 2013). Published reports have suggested that a mutation in the *PLA2G6* gene causes early onset neurodegeneration with brain iron accumulation (Morgan, Westaway et al. 2006).

In the associated rs34311866 locus, amongst the tagged genes *TMEM175*, *GAK* and *DGKQ*, α-synuclein was downregulated in *GAK* KD cell lines. G-protein coupled receptor-signaling pathway, inflammatory response and Cytokine-cytokine receptor interaction pathways were enriched in *GAK* KD

samples. *GAK* plays a role in the vesicle-mediated signaling pathway and clathrin-mediated signaling pathway. Published studies reveal that *GAK* modifies α-synuclein expression levels and toxicity in Parkinson's disease (Dumitriu, Pacheco et al. 2011). According to the Ensembl genome browser (https://www.ensembl.org/index.html), *TMEM175* and *GAK* are encoded on the two different strands of the DNA in a head-to-head orientation containing only one promoter region between them. Moreover, deficiency of *TMEM175* increases α-synuclein aggregation by impairing lysosomal and mitochondrial function (Jinn, Drolet et al. 2017).

In *GCH1* KD cell lines, α-synuclein was downregulated and G-protein coupled receptor signaling pathway, cell surface receptor signaling pathway and chemokine-mediated signaling pathways were enriched. Published studies have reported *GCH1* variants segregating with both dopa-responsive dystonia and Parkinson's disease (Lewthwaite, Lambert et al. 2015). *GCH1* encodes the rate-limiting enzyme for tetrahydrobiopterin (BH_4) biosynthesis, an essential cofactor of tyrosine hydroxylase (TH) activity and dopamine production in nigrostriatal cells (Cai, Alp et al. 2002).

α-synuclein mRNA levels were upregulated in *COASY* and *ITGA8* knockdown samples with respect to controls. Signaling pathways, microtubule-based movement and activation of protein kinase activity were over-represented in *COASY* KD cell lines. A previous published report shows that a mutation in *COASY* causes neurodegeneration with brain iron accumulation (Annesi, Gagliardi et al. 2016). The protein encoded by *ITGA8* is a membrane protein that mediates cell-cell interactions and regulates neurite outgrowth of sensory and

motor neurons (Farias, Lu et al. 2005). It is involved in integrin-mediated signaling pathways and apoptotic pathways.

The associated SNP rs12497850 is among the 17 new loci identified in the latest meta-analysis of PD GWAS (Chang, Nalls et al. 2017). The genes such as *DALRD3*, *ARIH2*, *IMPDH2*, *P4HTM*, *PRKAR2A*, *QARS*, *QRICH1*, *USP19* and *WDR6* tagging this associated region (rs12497850) were knockdown. Amongst these knockdown genes α-synuclein was upregulated in *P4HTM* KD cell lines. The protein encoded by *P4HTM* plays a role in catalyzing the post-translational formation of 4-hydroxyproline in hypoxia-inducible factor (HIF) alpha proteins. It contributes to the hydroxylation of proteasomal degradation by targeting HIF von Hippel-Lindau ubiquitination complex (Koivunen, Tiainen et al. 2007). The inflammatory response, G-protein coupled receptor signaling pathway and Calcium signaling pathway were over-represented in the differentially expressed genes of *P4HTM* KD samples compared to controls. Henceforth, we detected that on silencing parkinsonism causal genes (*PARK7*, *VPS35* and *PLA2G6*) and targeted genes such as *GAK*, *ITGA8*, *GCH1* and *P4HTM*, biological processes such as inflammatory response, G-protein coupled receptor signaling pathway and Cytokine-cytokine receptor interaction were over-represented.

Both rare and common variants in multiple genes contribute to PD pathogenesis. Rare coding variants affect the function of the gene and are also involved in the genetic aetiology of PD. Hence, we investigated the genetic burden of rare variants in the genes tagging the PD risk loci using resequencing data from IPDGC. Gene-based association tests show a significant association of *LRRK2* and *GAK* to PD. These promising candidate genes such as *BAG3*, *GAK*, *GCH1*, *ITGA8*, *COASY* and *P4HTM* from the transcriptomic analysis will be

validated by functional assays such as Parkin translocation assay, an assay for α-synuclein aggregation, secretion and toxicity and by investigating mitochondrial function using Seahorse.

Table 19. Over-represented Biological processes and pathways in differentially regulated genes of knockdown samples.

Loci	KD_genes	Biological Process	Pathways
TMEM175	*GAK*	positive regulation of cell-substrate adhesion ; extracellular matrix organization ; response to lipopolysaccharide ; cellular response to mechanical stimulus ; immune response ; inflammatory response ; chondrocyte differentiation ; G-protein coupled receptor signaling pathway ; positive regulation of angiogenesis	Malaria ; Cytokine-cytokine receptor interaction
TMEM175	*TMEM175*	neuropeptide signaling pathway ; antibacterial humoral response ; defense response to Gram-negative bacterium	
PARK16	*NUCKS1*	cell surface receptor signaling pathway ; digestion ; response to lipopolysaccharide	Neuroactive ligand-receptor interaction
BAG3	*BAG3*	cell adhesion ; extracellular matrix organization ; cilium movement ; immune response ; positive regulation of cell-substrate adhesion ; cell-matrix adhesion ; cell adhesion mediated by integrin ; chemical synaptic transmission ; cell-substrate adhesion ; heterotypic cell-cell adhesion ; epidermis development ; integrin-mediated signaling pathway	ECM-receptor interaction ; Hypertrophic cardiomyopathy (HCM) ; Dilated cardiomyopathy ; Arrhythmogenic right ventricular cardiomyopathy (ARVC) ; Focal adhesion
BAG3	*INPP5F*	cell adhesion ; chemical synaptic transmission ; extracellular matrix organization ; keratan sulfate biosynthetic process ; positive regulation of transforming growth factor beta receptor signaling pathway ; response to wounding ; regulation of ion transmembrane transport ; startle response ; collagen fibril organization ; sensory perception of pain	ECM-receptor interaction
BAG3	*MCMBP*	immune response	Neuroactive ligand-receptor interaction
BAG3	*RGS10*	angiogenesis ; extracellular matrix organization ; positive regulation of angiogenesis ; endocrine pancreas development ; positive regulation of ERK1 and ERK2 cascade ; inflammatory response ; signal transduction ; chemical synaptic transmission ; positive regulation of Notch signaling pathway ; cell adhesion ; Notch receptor processing ; positive regulation of collagen biosynthetic process ; negative regulation of apoptotic process ; peripheral nervous system development ; positive regulation of cell proliferation ; positive regulation of transforming growth factor beta receptor signaling pathway ; positive regulation of fibroblast proliferation	Cytokine-cytokine receptor interaction ; Neuroactive ligand-receptor interaction
BAG3	*SEC23IP*	striated muscle contraction ; cell surface receptor signaling pathway ; G-protein coupled receptor signaling pathway ; G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger ; sarcomere organization ; cytokine-mediated signaling pathway ;	
BAG3	*TIAL1*	response to mechanical stimulus ; extracellular matrix organization ; response to estradiol ; cell surface receptor signaling pathway	Neuroactive ligand-receptor interaction
rs12497850	*DALRD3*		Neuroactive ligand-receptor interaction
rs12497850	*IMPDH2*		Systemic lupus erythematosus

rs12497850	*P4HTM*	cholesterol biosynthetic process ; cell adhesion ; cell surface receptor signaling pathway ; G-protein coupled receptor signaling pathway ; potassium ion transmembrane transport ; regulation of ion transmembrane transport ; ion transmembrane transport ; positive regulation of angiogenesis ; single organismal cell-cell adhesion ; regulation of membrane potential ; potassium ion transport ; positive regulation of synapse assembly ; regulation of cardiac conduction ; inflammatory response ; potassium ion import ; negative regulation of blood pressure ; extracellular matrix organization ; sensory perception of taste ; skeletal system development ; G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger ; sodium ion transmembrane transport ; nitric oxide mediated signal transduction ; regulation of heart contraction ; arachidonic acid metabolic process ; collagen catabolic process ; cholesterol homeostasis ; adherens junction organization ; cytokine-mediated signaling pathway ; bile acid and bile salt transport ; axonogenesis ; negative regulation of insulin secretion ; calcium ion-regulated exocytosis of neurotransmitter ; immune response ; keratinization ; linoleic acid metabolic process ; cholesterol transport ; cell development ; positive regulation of peptidyl-tyrosine phosphorylation ; oxidation-reduction process ; cellular response to follicle-stimulating hormone stimulus	Neuroactive ligand-receptor interaction ; Calcium signaling pathway ; Protein digestion and absorption ; Cytokine-cytokine receptor interaction ; Steroid biosynthesis ; Terpenoid backbone biosynthesis ; Hematopoietic cell lineage ; Glycosphingolipid biosynthesis - lacto and neolacto series ; Taste transduction
rs12497850	*PRKAR2A*	extracellular matrix organization ; G-protein coupled receptor signaling pathway ; negative regulation of adenylate cyclase activity ; neuropeptide signaling pathway ; cell adhesion ; phospholipase C-activating G-protein coupled receptor signaling pathway ; regulation of vasoconstriction ; mesodermal cell differentiation ; angiogenesis ; extracellular matrix disassembly ; cell surface receptor signaling pathway	Neuroactive ligand-receptor interaction ; Cytokine-cytokine receptor interaction ; ECM-receptor interaction
rs12497850	*QARS*	cell adhesion ; extracellular matrix organization ; homophilic cell adhesion via plasma membrane adhesion molecules ; wound healing ; negative regulation of angiogenesis ; positive regulation of cell proliferation ; G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger ; response to lipopolysaccharide ; keratan sulfate biosynthetic process ; cellular response to retinoic acid ; synapse assembly ; nervous system development ; lung development ; neural crest cell migration ; angiogenesis ; response to hypoxia ; chemokine-mediated signaling pathway ; inflammatory response ; cell surface receptor signaling pathway ; positive regulation of smooth muscle cell proliferation ; positive regulation of cell-substrate adhesion ; response to corticosterone ; signal transduction ; response to mechanical stimulus ; negative regulation of endothelial cell migration ; immune response ; heart development ; cell-cell signaling ; positive regulation of ERK1 and ERK2 cascade ; positive regulation of collagen biosynthetic process ; vasodilation ; skeletal system development ; response to activity ; positive regulation of cell migration ; cation transport ; response to wounding ; cellular response to interleukin-1 ; negative regulation of endothelial cell apoptotic process ; keratinocyte differentiation ; endothelial cell differentiation ; induction of positive chemotaxis ; regulation of postsynaptic membrane potential ; cell-matrix adhesion ; regulation of immune response	Neuroactive ligand-receptor interaction ; ECM-receptor interaction ; Cell adhesion molecules (CAMs) ; Axon guidance ; Arrhythmogenic right ventricular cardiomyopathy (ARVC) ; Dilated cardiomyopathy ; Hypertrophic cardiomyopathy (HCM) ; Malaria ; Calcium signaling pathway ; Focal adhesion
rs12497850	*QRICH1*	cytokine-mediated signaling pathway ; chemical synaptic transmission	
rs12497850	*WDR6*	striated muscle contraction	
rs14235	*KAT8*	cytokine-mediated signaling pathway ; skeletal muscle contraction	
rs14235	*STX4*	MAPK cascade ; chemical synaptic transmission ; regulation of phosphatidylinositol 3-kinase signaling ; positive regulation of Ras protein signal transduction ; positive regulation of ERK1 and ERK2 cascade ;	Cytokine-cytokine receptor interaction
rs14235	*VKORC1*	vasodilation	Neuroactive ligand-receptor interaction
rs14235	*ZNF668*	cell adhesion ; extracellular matrix organization	

SREBF1	*SREBF1*	G-protein coupled receptor signaling pathway,extracellular matrix organization,blood vessel remodeling	
ATP13A2	*ATP13A2*	chemical synaptic transmission ; cell surface receptor signaling pathway ; G-protein coupled receptor signaling pathway	-
BST1	*BST1*	extracellular matrix organization ; response to growth factor ; G-protein coupled receptor signaling pathway ; platelet degranulation ; inflammatory response ; gastrulation ; cytokine-mediated signaling pathway ; response to estradiol ; response to cAMP ; ERK1 and ERK2 cascade	Neuroactive ligand-receptor interaction ; Cytokine-cytokine receptor interaction ; Hematopoietic cell lineage
COASY	*COASY*	mitotic nuclear division ; sister chromatid cohesion ; cell division ; DNA replication ; regulation of transcription involved in G1/S transition of mitotic cell cycle ; chromosome segregation ; G1/S transition of mitotic cell cycle ; extracellular matrix organization ; cell adhesion ; mitotic sister chromatid segregation ; extracellular matrix disassembly ; DNA replication initiation ; metaphase plate congression ; mitotic cytokinesis ; potassium ion transport ; CENP-A containing nucleosome assembly ; cytokine-mediated signaling pathway ; mitotic spindle organization ; cell proliferation ; regulation of cyclin-dependent protein serine/threonine kinase activity ; mitotic metaphase plate congression ; glycosaminoglycan metabolic process ; retinoid metabolic process ; microtubule-based movement ; spindle organization ; positive regulation of cytokinesis ; activation of protein kinase activity ; retina homeostasis ; chromosome organization ; immune response ; cytokinesis ; spermatogenesis ; regulation of mitotic metaphase/anaphase transition ; DNA duplex unwinding ; strand displacement ; collagen fibril organization	Cell cycle ; ECM-receptor interaction ; Progesterone-mediated oocyte maturation ; Oocyte meiosis ; Hypertrophic cardiomyopathy (HCM) ; Dilated cardiomyopathy
FDFT1	*FDFT1*	immune response	-
rs11158026	*GCH1*	immune response ; G-protein coupled receptor signaling pathway ; cell surface receptor signaling pathway ; platelet degranulation ; defense response to Gram-negative bacterium ; chemokine-mediated signaling pathway ; response to pain	Cytokine-cytokine receptor interaction ; Intestinal immune network for IgA production

Project 4: Genetic interaction of *LRRK2* and *PARK16* locus in Parkinson's disease using next generation sequencing data.

4.1 Introduction

Genome-Wide Association Studies (GWAS) have provided unequivocal evidence of a genetic contribution of common variability to Parkinson's disease (PD) (Chang, Nalls et al. 2017). Although these studies have been successful in the identification of 41 risk loci for sporadic PD and supported the concept that sporadic and familial PD share multiple pathogenic pathways, the underlying mechanisms for the majority of them remain elusive. The main reason being that they only identify haplotypes of common variants, which in most cases are not functionally relevant but rather exist in linkage disequilibrium with the variants conferring the risk effect, which can only be identified by resequencing these loci in a large number of individuals. In our study, we used available NGS data and tried to understand the genetic mechanisms underlying the risk for PD in one of the first GWAS loci identified for this disorder, *i.e. PARK16*. This locus was identified for the first time in a Japanese GWAS (Satake, Nakabayashi et al. 2009) and was confirmed in European populations (Simon-Sanchez, Schulte et al. 2009). It spans 170kb region in chromosome 1 and contains five protein-coding genes, namely *SLC45A3*, *NUCKS1*, *RAB29*, *SLC41A1* and *PM20D1*.

Although several sequencing approaches aiming to understand the genetic variant responsible for PD in this locus, coding variants have been identified in *RAB29* (p. K157R) and *SLC41A1* (p. A350V) in a British PD cohort, but failed to replicate in a larger cohort (Tucci, Nalls et al. 2010). Lin et al.

reported that variant p. R244H in *SLC41A1* causes loss of Mg^{2+} efflux function (Evangelou, Maraganore et al.). These previous reports suggest that the identified variants are very rare and may or may not be a disease causal one, as they aren't replicated yet.

David MacLeod et al. performed human genetic, transcriptomic and cellular analysis and described that interaction between *LRRK2* and *PARK16* locus is associated with PD risk (MacLeod, Rhinn et al. 2013). However, amongst the five genes of *PARK16*, only *RAB29* overexpression suppressed the neurite length phenotype induced by *LRRK2* (MacLeod, Rhinn et al. 2013) pointing to this gene as the most interesting candidate within this locus. Defects in *RAB29* and *LRRK2* due to PD led to endolysosomal and Golgi apparatus defects and deficiency of the retromer complex component *VPS35* (MacLeod, Rhinn et al. 2013).

In this study, we interrogated next-generation sequencing (NGS) data of unrelated cases and controls to investigate the interaction between *LRRK2* and *PARK16* locus. Moreover, we presented the analysis of resequencing data in *LRRK2* and *PARK16* in a large series of PD cases and controls and tried to replicate the molecular interaction between these two loci.

4.2 Subjects

4.2.1 Discovery phase:

We performed a two-stage study. Firstly, we did a discovery analysis in which the Whole Exome Sequencing (WES) data was interrogated. To investigate for a

genetic interaction between *LRRK2* and *PARK16* we used our large repository of WES data of Caucasian population, which consists of 1,450 PD cases and 535 controls from the International Parkinson's Disease Consortium (IPDGC) and 1,732 controls from the Rotterdam Study Exome Sequencing Database.

4.2.2 Replication:

In order to replicate the results obtained, we used NGS data generated for the 11 PD risk loci such as *MAPT, SNCA, HLA-DRB5, BST1, GAK, LRRK2, ACMSD, STK39, MCCC1/LAMP3, SYT11,* and *CCDC62/HIP1R* as per the meta-analysis of 2011 (International Parkinson Disease Genomics, Nalls et al. 2011). Data from 3,060 PD cases and 1,454 controls from Caucasian series were available in this dataset including *LRRK2* and *PARK16* locus (International Parkinson Disease Genomics, Nalls et al. 2011).

4.3 Methodology:

Quality control of both the datasets are discussed in detail in the previous chapters. Genotype and variant quality control was performed for both datasets using KGGSeq (Cingolani, Platts et al. 2012) and VCFtools (Danecek, Auton et al. 2011, Cingolani, Platts et al. 2012). Individual quality control of the data consists of identification of gender inconsistencies, heterozygosity outliers, duplicate samples, individuals of divergent ancestry and missing genotype. All these individuals were removed from both datasets.

An epistatic interaction between *LRRK2* SNPs and *PARK16* locus SNPs along with marginal association was estimated using Bayesian Epistasis

Association Mapping (BEAM) tool (Zhang 2012), which is particularly useful for case-control studies. It uses Bayesian partitioning model and computes the posterior probability that each marker set is associated with the disease employing Markov chain Monte Carlo method (Zhang 2012). Also, the B-statistic of BEAM is more potent than the χ^2 statistic of exhaustive search. BEAM takes into account SNP linkage disequilibrium (LD), so the identified interactions are not due to LD effects but are associated with the disease. The interacting SNPs are jointly affecting the disease, and their joint contribution to disease risk is stronger than their marginal contribution individually (Zhang 2012).

4.4 Results

4.4.1 Discovery phase:

After performing epistatic analysis of *LRRK2* and *PARK16* SNPs, it was observed that there is a significant interaction between *LRRK2* intronic variant rs138343113 and *SLC41A1* intronic variant rs5780304. The joint posterior probability for marginal association (0.84) and interaction association (0.16) together for rs5780304 is 1.0 and for rs138343113 is 0.165 (table20).

4.4.2 Replication:

In order to replicate the findings, epistatic analysis of resequenced PD loci dataset was performed using Bayesian epistasis association mapping approach. It was observed that the SNP pairs from the discovery set did not replicate. However, both discovery and replication datasets are different, so we expect that

different SNPs pairs of *LRRK2* and *SLC41A1* might be involved in the interaction. Therefore, after inspecting the SNP pairs of *SLC41A1* and *LRRK2*, it was observed that *SLC41A1* intronic variant rs5780303 interacts with *LRRK2* intronic variant rs528054842. Posterior probability for marginal association and interaction association together for rs5780303 and all associated SNPs within 1Mb region is 2.100 and for rs528054842 along with all associated SNPs within 1Mb region of chr12 is 3.110. SNPs such as rs721711 and rs7307276 in *LRRK2* and rs5780304 variant in *SLC41A1* has interaction association probability greater than 50% (table21).

SLC41A1 SNPs such as rs5780304 (discovery set) and rs5780303 (replication set) are in LD (r^2 = 0.98) as well as *LRRK2* SNPs such as rs138343113 (discovery set) and rs721711, rs7307276 (replication set) are also in LD (r^2 = 0.61) [9]. Hence, we observed same effect in both discovery and replication dataset.

Table 20: Epistasis results of discovery phase.

Gene	rsID	CHR	POS	MARGINAL	INTERACTION	JOINT PROB
SLC41A1	rs5780304	1	205766159	0.84	0.16	1
LRRK2	rs138343113	12	40681142	0	0.155	0.155

* CHR: Chromosome; POS: position; MARGINAL: Posterior probability of marginal association; INTERACTION: Posterior probability of interaction association; JOINTPROB: Joint posterior probability.

Table 21: Epistasis results of replication phase.

Gene	rsID	CHR	POS	MARGINAL	INTERACTION	JOINT PROB
SLC41A1	rs5780303	1	205765590	0.005	0.995	1
LRRK2	rs528054842	12	40635005	0.22	0.78	1
LRRK2	rs721711	12	40709328	0	1	1
LRRK2	rs7307276	12	40715013	0	0.995	0.995
SLC41A1	rs5780304	1	205766159	0	0.645	0.645

* CHR: Chromosome; POS: position; MARGINAL: Posterior probability of marginal association; INTERACTION: Posterior probability of interaction association; JOINTPROB: Joint posterior probability.

4.5 Discussion

Previous studies draw our attention that variation in genes in *PARK16* locus and *LRRK2* interactively determine risk for PD (MacLeod, Rhinn et al. 2013). Furthermore, previous cellular studies implicate that *LRRK2* and *RAB29* (which is in *PARK16* locus) work together and mutation in *LRRK2* impairs lysosomal pathway. *RAB29* is localized in Golgi and retromer complex is involved in protein sorting between lysosomes and Golgi. Therefore, any defects caused due to *LRRK2* mutants could be due to impairment of retromer complex (MacLeod, Rhinn et al. 2013). Lisa Wang et al. reported that they did not identify an interaction between *LRRK2* and *PARK16* locus in a large multi-center series of PD cases and controls. Their analysis was limited to 5 SNPs from *PARK16* locus and 2 SNPs from *LRRK2*, which are shown to be associated with PD in the previous studies (Wang, Heckman et al.). However, NGS data has more coverage, so there is a possibility for other SNPs from *LRRK2* and *PARK16* locus to interact with

each other. We analyzed an interaction between all SNPs of *LRRK2* and *PARK16* locus using WES data of unrelated cases and controls. Additionally, resequencing data of PD risk loci was used to replicate our findings. To estimate genetic interaction, we used Bayesian partitioning model approach, which gives the posterior probability for both interaction and marginal association. In our analysis, we did not find an interaction between *RAB29* and *LRRK2*. In WES data (discovery dataset), an interaction between *SLC41A1* rs5780304 and *LRRK2* rs138343113 was observed, which is not only marginally associated to PD but also has a joint contribution to PD risk. Though the interacting SNP pairs from WES data (discovery dataset) didn't replicate in the resequencing data, there is still an interaction between *SLC41A1* rs5780303 and *LRRK2* rs52805484. *SLC41A1* variants such as rs5780303 and rs5780304 interacting with *LRRK2* SNPs are in strong LD and are present in the intronic region. Cornelis Blauwendraat et al. identified an eQTL for an antisense CAGE cluster to *SLC41A1.* Moreover, they mentioned eQTL association of two PD GWAS associated variants, *i.e.* rs947211, (which is located 8.5 kb upstream of *RAB29* and 5.6 kb downstream of *SLC41A1)* and rs823118 (4.2kb upstream of *NUCKS1)* with the expression of *SLC41A1* antisense transcript (Blauwendraat, Francescatto et al. 2016). *SLC41A1* acts as magnesium transporter and might play a role in magnesium homeostasis. It has been reported that magnesium interacts with α-synuclein to inhibit α-synuclein aggregation (Golts, Snyder et al. 2002). Moreover, an involvement of metal dyshomeostasis in neurodegeneration has been previously speculated (Myhre, Utkilen et al. 2013). Hence, *SLC41A1* of *PARK16* locus looks like an interesting candidate and may play a role in

developing a risk for PD. Moreover, functional study of this gene should be carried out to know the involvement of this gene in developing PD.

General Discussion

Parkinson's disease (PD) is a neurological disease that lacks therapies to modify the disease pathology. Medications can control the PD symptoms, but in an advanced stage of the disease, the medicines become less effective (Savitt, Dawson et al. 2006). The identification of causal rare genetic mutations facilitates the detection of deregulated biological processes in PD. Such findings are essential to improve our knowledge of PD pathogenesis. The main aim of this thesis was to identify genetic risk variants and pathogenic mutations contributing to PD pathogenesis.

The reduced cost of next-generation sequencing (NGS) has facilitated the family-based sequencing analysis to identify the causal genes for Mendelian disorders. In project 1 of this thesis, we sequenced and investigated the genomes (180 individuals) and exomes (20 individuals) from 126 families (discovery cohort). We identified pathogenic causal variants in already known parkinsonism causal genes such as *LRRK2*, *DJ1*, *PARK2* and *PLA2G6* in some families. A novel alternative homozygous mutation in *TREM2* was identified in a Turkish patient with parkinsonism and dementia. Lines of evidence have shown that mutations in *TREM2* not only cause Nasu-Hakola disease (Paloneva, Manninen et al. 2002) but were also identified in families with Frontotemporal dementia (FTD) (Guerreiro, Lohmann et al. 2013, Le Ber, De Septenville et al. 2014). Moreover, *TREM2* variants are also associated with Alzheimer's disease (AD) risk (Guerreiro, Wojtas et al. 2013). In our study, we identified a hemizygous mutation in a novel gene *SH3KBP1* in PD patients from two unrelated families.

On the basis of previous knowledge that *SH3KBP1* plays a role in apoptosis (Narita, Nishimura et al. 2005) Bian et al. silenced *SH3KBP1* expression in human dopaminergic SH-SY5Y cell lines using shRNA and observed that the cell death was decreased by 50% in comparison to controls (Bian, Yu et al. 2008). In families with an autosomal recessive mode of inheritance, alternative homozygous and compound heterozygous mutations were identified in *HTT*, *CCDC154*, *CXorf22*, *TMEM63C*, *MYO15A* and *ULK2.* Heterozygous mutations in *DYSF* and *APC2* were identified in multiple families with an autosomal dominant mode of inheritance. These 8 candidate genes were also identified in a large repository of sporadic PD WES data from Parkinson's disease Genetics Sequencing Consortium (PDGSC) with a minor allele frequency (MAF) less than 1% in public databases.

To test the association of rare variants/genes identified in the families to PD, we had a possibility of doing either burden test or variance-component test. In burden test, the phenotype is influenced by all the tested variants in the same direction (Auer and Lettre 2015). To overcome this limitation, we performed SKAT-O test, which is a combination of burden test (all variants have the same direction of effect) and SKAT test (all variants have a distinct direction of effect). These rare variant association studies are sensitive to confounding factors (Lee, Emond et al. 2012). To deal with these confounding factors, we used 20 multidimensional scaling components as covariates in the association analysis. By performing, gene-based SKAT-O tests using sporadic PD data, we observed that only rare variants in *ULK2* are enriched to PD in both replication cohorts. Among these 8 candidates, *HTT* and *ULK2* are highly promising as *HTT* is already linked to a neurological disorder and plays a role in vesicle-mediated transport,

apoptotic process and positive regulation of mitophagy (Schulte and Littleton 2011). Furthermore, *ULK2* is involved in autophagy and alteration of such autophagy-related proteins are observed in Lewy body disease (Jung, Seo et al. 2011, Miki, Tanji et al. 2016).

PD is heterogeneous in terms of genetic, pathological and clinical point of view, so targeting biological pathways will be helpful to treat a group of PD patients. Genetic evidence shows that mitochondrial dysfunction may be one of the important disease pathways in a subgroup of PD patients (Reeve, Grady et al. 2018). In project 2, we identified the sub-group, *i.e.* the mitochondrial endo-phenotype of PD using existing genomic data. The risk profiling of 8 GWAS associated SNPs showed that positive allelic scores were significantly present in more cases than in controls, while negative scores were present in more controls than in cases. The German PD cases were stratified as mito+ (allelic score > 0.35) and mito- (allelic score < -0.35) group on the basis of these allelic scores. Some patients from these two subgroups as well as PD patients with *PARK2* and *PINK1* mutations will undergo clinical trials with coenzyme Q10 and vitamin K2. Rare variant association analysis showed that only *GBA* is associated with PD in both datasets. The other genes, which showed association signals in WES data were not replicated in NeuroX, as the age of onset and the laboratory techniques applied to both the datasets are different.

Genome-wide association studies (GWAS) have identified many common genetic variants associated to PD (Pankratz, Wilk et al. 2009, Satake, Nakabayashi et al. 2009, Simon-Sanchez, Schulte et al. 2009, Nalls, Pankratz et al. 2014, Chang, Nalls et al. 2017). In project 3, the genes spanning the PD GWAS associated regions were investigated to identify the actual gene contributing to

PD risk. The genes under the associated PD loci were silenced by shRNA, and the gene expression was quantified using mRNA sequencing. α-synuclein, which is pathologically and genetically linked to PD, was differentially regulated in genes such as *BAG3*, *GAK*, *GCH1*, *ITGA8*, *COASY* and *P4HTM.* Functional validation of these genes is needed to detect their role in PD pathogenesis.

Despite the identified risk loci for PD through GWAS, a single SNP association explains the limited heritability of the disease, so multi-SNP interactions, *i.e. epistatic interactions* are a hope to discover more significant associations. The published report has shown that an interaction between Mendelian PD gene *LRRK2* and *PARK16* locus is associated with PD risk (MacLeod, Rhinn et al. 2013). In project 4, we identified an epistatic interaction between *LRRK2* and *SLC41A1* of *PARK16* locus using Bayesian partitioning model approach. An eQTL has already been identified for an antisense CAGE cluster to *SLC41A1* (Blauwendraat, Francescatto et al. 2016). *SLC41A1* plays a role in magnesium homeostasis and involvement of metal dyshomeostasis in neurodegeneration makes the gene more interesting for functional validation (Golts, Snyder et al. 2002, Myhre, Utkilen et al. 2013).

In this thesis, we identified pathogenic variants in some novel genes causing PD in families. Gene-based association tests using sporadic PD data also shows that rare variants in some novel genes are highly enriched. In the future, functional validation of these candidates has to be done using cellular or animal model to determine their role in PD pathogenesis.

Acknowledgements

My sincere gratitude to Prof. Dr. Thomas Gasser and Dr. Javier Simon-Sanchez, who accepted me at Hertie Institute for Clinical Brain Research as a PhD student and for all the interesting projects offered to me. It was a great pleasure to pursue PhD in a well-equipped and an extremely comfortable lab. Many thanks to my supervisor Dr. Javier Simon-Sanchez for all the guidance and mentorship throughout my PhD. I am also thankful to Prof. Dr. Peter Heutink and Dr. Patrizia Rizzu for giving me an opportunity to collaborate with their group at DZNE and for their advice and suggestions during our group meetings.

I am thankful to all the members of DZNE and Hertie Institute for Clinical Brain Research. Many thanks to all my colleagues: Tenzin, Burcu, Anupriya, Kevin, Anastasia, Natalia, Conny, Melissa, Iris, Cornelis, Margherita, Ashutosh, Manmeet, Noemia, Erika and Joachim for their kind help and a nice time at work. My special thanks to Dr. Ebba Lohmann for all the provided samples and clinical lessons and to Ann-kathrin Hauser for the Sanger sequencing performed on so many variants for confirming the segregation in the family. I am also thankful to all the collaborators of COURAGE-PD project for providing DNA samples and details of all the individuals. Special thanks to Mito-PD collaborators for all the productive and informative meetings. Thanks to all the funding bodies, who funded our research. I am grateful to all the members of International Parkinson's disease Genomics Consortium for providing such a large repository of sporadic PD data. I would like to thank all the patients and their relatives for providing their DNA samples and family history.

I am thankful to my family members for supporting me throughout my career so far. My special thanks to my husband Dinesh for encouraging and supporting me in my journey.

References

Abou-Sleiman, P. M., D. G. Healy, N. Quinn, A. J. Lees and N. W. Wood (2003). "The role of pathogenic DJ-1 mutations in Parkinson's disease." Ann Neurol **54**(3): 283-286.

Allen, M., M. Kachadoorian, Z. Quicksall, F. Zou, H. S. Chai, C. Younkin, J. E. Crook, V. S. Pankratz, M. M. Carrasquillo, S. Krishnan, T. Nguyen, L. Ma, K. Malphrus, S. Lincoln, G. Bisceglio, C. P. Kolbert, J. Jen, S. Mukherjee, J. K. Kauwe, P. K. Crane, J. L. Haines, R. Mayeux, M. A. Pericak-Vance, L. A. Farrer, G. D. Schellenberg, J. E. Parisi, R. C. Petersen, N. R. Graff-Radford, D. W. Dickson, S. G. Younkin and N. Ertekin-Taner (2014). "Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels." Alzheimers Res Ther **6**(4): 39.

An, J., J. Shi, Q. He, K. Lui, Y. Liu, Y. Huang and M. S. Sheikh (2012). "CHCM1/CHCHD6, Novel Mitochondrial Protein Linked to Regulation of Mitofilin and Mitochondrial Cristae Morphology." The Journal of Biological Chemistry **287**(10): 7411-7426.

Annesi, G., M. Gagliardi, G. Iannello, A. Quattrone, G. Iannello and A. Quattrone (2016). "Mutational analysis of COASY in an Italian patient with NBIA." Parkinsonism Relat Disord **28**: 150-151.

Appel-Cresswell, S., C. Vilarino-Guell, M. Encarnacion, H. Sherman, I. Yu, B. Shah, D. Weir, C. Thompson, C. Szu-Tu, J. Trinh, J. O. Aasly, A. Rajput, A. H. Rajput, A. Jon Stoessl and M. J. Farrer (2013). "Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease." Mov Disord **28**(6): 811-813.

Auer, P. L. and G. Lettre (2015). "Rare variant association studies: considerations, challenges and opportunities." Genome Med **7**(1): 16.

Balding, D. J. (2006). "A tutorial on statistical methods for population association studies." Nat Rev Genet **7**(10): 781-791.

Bendor, J. T., T. P. Logan and R. H. Edwards (2013). "The function of alpha-synuclein." Neuron **79**(6): 1044-1066.

Betarbet, R., T. B. Sherer, G. MacKenzie, M. Garcia-Osuna, A. V. Panov and J. T. Greenamyre (2000). "Chronic systemic pesticide exposure reproduces features of Parkinson's disease." Nature Neuroscience **3**: 1301.

Bian, M., M. Yu, S. Yang, H. Gao, Y. Huang, C. Deng, Y. Gao, F. Sun and F. Huang (2008). "Expression of Cbl-interacting protein of 85 kDa in MPTP mouse model of Parkinson's disease and 1-methyl-4-phenyl-pyridinium ion-treated dopaminergic SH-SY5Y cells." Acta Biochimica et Biophysica Sinica **40**(6): 505-512.

Biesecker, L. G., K. V. Shianna and J. C. Mullikin (2011). "Exome sequencing: the expert view." Genome Biology **12**(9): 128.

Blauwendraat, C., M. Francescatto, J. R. Gibbs, I. E. Jansen, J. Simon-Sanchez, D. G. Hernandez, A. A. Dillman, A. B. Singleton, M. R. Cookson, P. Rizzu and P. Heutink

(2016). "Comprehensive promoter level expression quantitative trait loci analysis of the human frontal lobe." Genome Med **8**(1): 65.
Blesa, J., I. Trigo-Damas, A. Quiroga-Varela and V. R. Jackson-Lewis (2015). "Oxidative stress and Parkinson's disease." Front Neuroanat **9**: 91.
Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.
Bonifati, V., P. Rizzu, M. J. van Baren, O. Schaap, G. J. Breedveld, E. Krieger, M. C. Dekker, F. Squitieri, P. Ibanez, M. Joosse, J. W. van Dongen, N. Vanacore, J. C. van Swieten, A. Brice, G. Meco, C. M. van Duijn, B. A. Oostra and P. Heutink (2003). "Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism." Science **299**(5604): 256-259.
Braak, H., K. Del Tredici, U. Rub, R. A. de Vos, E. N. Jansen Steur and E. Braak (2003). "Staging of brain pathology related to sporadic Parkinson's disease." Neurobiol Aging **24**(2): 197-211.
Cai, S., N. J. Alp, D. McDonald, I. Smith, J. Kay, L. Canevari, S. Heales and K. M. Channon (2002). "GTP cyclohydrolase I gene transfer augments intracellular tetrahydrobiopterin in human endothelial cells: effects on nitric oxide synthase activity, protein levels and dimerisation." Cardiovasc Res **55**(4): 838-849.
Calvo, S. E., K. R. Clauser and V. K. Mootha (2016). "MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins." Nucleic Acids Res **44**(D1): D1251-1257.
Cao, Y. L., Y. P. Yang, C. J. Mao, X. Q. Zhang, C. T. Wang, J. Yang, D. J. Lv, F. Wang, L. F. Hu and C. F. Liu (2017). "A role of BAG3 in regulating SNCA/alpha-synuclein clearance via selective macroautophagy." Neurobiol Aging **60**: 104-115.
Chan, D., V. Mok, P. Wing Ng, J. Yeung, J. Kwok, Z. Fang, R. Clarke, L. Wong, P. Schofield and N. Hattori (2008). PARK2 mutations and clinical features in a Chinese population with early-onset Parkinson's disease.
Chan, N. C. and D. C. Chan (2011). "Parkin uses the UPS to ship off dysfunctional mitochondria." Autophagy **7**(7): 771-772.
Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell and J. J. Lee (2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets." Gigascience **4**: 7.
Chang, D., M. A. Nalls, I. B. Hallgrimsdottir, J. Hunkapiller, M. van der Brug, F. Cai, C. International Parkinson's Disease Genomics, T. andMe Research, G. A. Kerchner, G. Ayalon, B. Bingol, M. Sheng, D. Hinds, T. W. Behrens, A. B. Singleton, T. R. Bhangale and R. R. Graham (2017). "A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci." Nat Genet.
Chang, N. C., M. Nguyen, M. Germain and G. C. Shore (2010). "Antagonism of Beclin 1-dependent autophagy by BCL-2 at the endoplasmic reticulum requires NAF-1." The EMBO Journal **29**(3): 606-618.
Chartier-Harlin, M.-C., J. Kachergus, C. Roumier, V. Mouroux, X. Douay, S. Lincoln, C. Levecque, L. Larvor, J. Andrieux, M. Hulihan, N. Waucquier, L. Defebvre, P.

Amouyel, M. Farrer and A. Destée "α-synuclein locus duplication as a cause of familial Parkinson's disease." The Lancet **364**(9440): 1167-1169.
Check, E. (2007). "RNA interference: hitting the on switch." Nature **448**(7156): 855-858.
Chen, X., O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Kallberg, A. J. Cox, S. Kruglyak and C. T. Saunders (2016). "Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications." Bioinformatics **32**(8): 1220-1222.
Chen, Y., L.-N. Yang, L. Cheng, S. Tu, S.-J. Guo, H.-Y. Le, Q. Xiong, R. Mo, Y. Li, J. Jeong, L. Jiang, S. Blackshaw, C. Li, H. Zhu, S.-C. Tao and F. Ge (2013). BAG3 Interactome Analysis Reveals a New Role in Modulating Proteasome Activity.
Cheng, T. L., C. C. Liao, W. H. Tsai, C. C. Lin, C. W. Yeh, C. F. Teng and W. T. Chang (2009). "Identification and characterization of the mitochondrial targeting sequence and mechanism in human citrate synthase." J Cell Biochem **107**(5): 1002-1015.
Choi, J. M., M. S. Woo, H. I. Ma, S. Y. Kang, Y. H. Sung, S. W. Yong, S. J. Chung, J. S. Kim, H. W. Shin, C. H. Lyoo, P. H. Lee, J. S. Baik, S. J. Kim, M. Y. Park, Y. H. Sohn, J. H. Kim, J. W. Kim, M. S. Lee, M. C. Lee, D. H. Kim and Y. J. Kim (2008). "Analysis of PARK genes in a Korean cohort of early-onset Parkinson disease." Neurogenetics **9**(4): 263-269.
Chun, S. and J. C. Fay (2009). "Identification of deleterious mutations within three human genomes." Genome Research **19**(9): 1553-1561.
Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu and D. M. Ruden (2012). "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3." Fly **6**(2): 80-92.
Clark, I. E., M. W. Dodson, C. Jiang, J. H. Cao, J. R. Huh, J. H. Seol, S. J. Yoo, B. A. Hay and M. Guo (2006). "Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin." Nature **441**: 1162.
Clark, L. N., S. Afridi, E. Karlins and et al. (2006). "Case-control study of the parkin gene in early-onset parkinson disease." Archives of Neurology **63**(4): 548-552.
Cleeter, M. W. J., K.-Y. Chau, C. Gluck, A. Mehta, D. A. Hughes, M. Duchen, N. W. Wood, J. Hardy, J. Mark Cooper and A. H. Schapira (2013). "Glucocerebrosidase inhibition causes mitochondrial dysfunction and free radical damage." Neurochemistry International **62**(1): 1-7.
Coakeley, S. and A. P. Strafella (2017). "Imaging tau pathology in Parkinsonisms." npj Parkinson's Disease **3**(1): 22.
Cookson, M. R., J. Hardy and P. A. Lewis (2008). "Genetic Neuropathology of Parkinson's Disease." International Journal of Clinical and Experimental Pathology **1**(3): 217-231.
Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin and G.

Genomes Project Analysis (2011). "The variant call format and VCFtools." Bioinformatics **27**(15): 2156-2158.
Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow and S. Batzoglou (2010). "Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++." PLoS Computational Biology **6**(12): e1001025.
Dickson, D. W. (2018). "Neuropathology of Parkinson disease." Parkinsonism Relat Disord **46 Suppl 1**: S30-S33.
Dickson, D. W., H. Fujishiro, A. DelleDonne, J. Menke, Z. Ahmed, K. J. Klos, K. A. Josephs, R. Frigerio, M. Burnett, J. E. Parisi and J. E. Ahlskog (2008). "Evidence that incidental Lewy body disease is pre-symptomatic Parkinson's disease." Acta Neuropathologica **115**(4): 437-444.
Doherty, K. M., L. Silveira-Moriyama, L. Parkkinen, D. G. Healy, M. Farrell, N. E. Mencacci, Z. Ahmed, F. M. Brett, J. Hardy, N. Quinn, T. J. Counihan, T. Lynch, Z. V. Fox, T. Revesz, A. J. Lees and J. L. Holton (2013). "Parkin Disease: A Clinicopathologic Entity?" JAMA neurology **70**(5): 571-579.
Dorsey, E. R., R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. Kieburtz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf and C. M. Tanner (2007). "Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030." Neurology **68**(5): 384.
Duberley, K. E., S. J. Heales, A. Y. Abramov, A. Chalasani, J. M. Land, S. Rahman and I. P. Hargreaves (2014). "Effect of Coenzyme Q10 supplementation on mitochondrial electron transport chain activity and mitochondrial oxidative stress in Coenzyme Q10 deficient human neuronal cells." Int J Biochem Cell Biol **50**: 60-63.
Dumitriu, A., J. Golji, A. T. Labadorf, B. Gao, T. G. Beach, R. H. Myers, K. A. Longo and J. C. Latourelle (2016). "Integrative analyses of proteomics and RNA transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci in Parkinson disease." BMC Med Genomics **9**: 5.
Dumitriu, A., C. D. Pacheco, J. B. Wilk, K. E. Strathearn, J. C. Latourelle, S. Goldwurm, G. Pezzoli, J. C. Rochet, S. Lindquist and R. H. Myers (2011). "Cyclin-G-associated kinase modifies α-synuclein expression levels and toxicity in Parkinson's disease: results from the GenePD Study." Hum Mol Genet **20**(8): 1478-1487.
Elbaz, A., L. Carcaillon, S. Kab and F. Moisan (2016). "Epidemiology of Parkinson's disease." Rev Neurol (Paris) **172**(1): 14-26.
Escaramís, G., E. Docampo and R. Rabionet (2015). "A decade of structural variants: description, history and methods to detect structural variation." Briefings in Functional Genomics **14**(5): 305-314.
Esteves, A. R. and S. M. Cardoso (2016). "LRRK2 at the Crossroad Between Autophagy and Microtubule Trafficking: Insights into Parkinson's Disease." The Neuroscientist **23**(1): 16-26.

Evangelou, E., D. M. Maraganore and J. P. Ioannidis (2007). "Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease." PLoS One **2**(2): e196.
Farias, E., M. Lu, X. Li and L. M. Schnapp (2005). "Integrin alpha8beta1-fibronectin interactions promote cell survival via PI3 kinase pathway." Biochem Biophys Res Commun **329**(1): 305-311.
Fernandes, H. J., E. M. Hartfield, H. C. Christian, E. Emmanoulidou, Y. Zheng, H. Booth, H. Bogetofte, C. Lang, B. J. Ryan, S. P. Sardi, J. Badger, J. Vowles, S. Evetts, G. K. Tofaris, K. Vekrellis, K. Talbot, M. T. Hu, W. James, S. A. Cowley and R. Wade-Martins (2016). "ER Stress and Autophagic Perturbations Lead to Elevated Extracellular alpha-Synuclein in GBA-N370S Parkinson's iPSC-Derived Dopamine Neurons." Stem Cell Reports **6**(3): 342-356.
Fiesel, F. C., T. R. Caulfield, E. L. Moussaud-Lamodiere, K. Ogaki, D. F. Dourado, S. C. Flores, O. A. Ross and W. Springer (2015). "Structural and Functional Impact of Parkinson Disease-Associated Mutations in the E3 Ubiquitin Ligase Parkin." Hum Mutat **36**(8): 774-786.
Foo, J. N., L. C. Tan, I. D. Irwan, W. L. Au, H. Q. Low, K. M. Prakash, A. Ahmad-Annuar, J. Bei, A. Y. Chan, C. M. Chen, Y. C. Chen, S. J. Chung, H. Deng, S. Y. Lim, V. Mok, H. Pang, Z. Pei, R. Peng, H. F. Shang, K. Song, A. H. Tan, Y. R. Wu, T. Aung, C. Y. Cheng, F. T. Chew, S. H. Chew, S. A. Chong, R. P. Ebstein, J. Lee, S. M. Saw, A. Seow, M. Subramaniam, E. S. Tai, E. N. Vithana, T. Y. Wong, K. K. Heng, W. Y. Meah, C. C. Khor, H. Liu, F. Zhang, J. Liu and E. K. Tan (2017). "Genome-wide association study of Parkinson's disease in East Asians." Hum Mol Genet **26**(1): 226-232.
Frank-Cannon, T. C., T. Tran, K. A. Ruhn, T. N. Martinez, J. Hong, M. Marvin, M. Hartley, I. Trevino, D. E. O'Brien, B. Casey, M. S. Goldberg and M. G. Tansey (2008). "Parkin deficiency increases vulnerability to inflammation-related nigral degeneration." J Neurosci **28**(43): 10825-10834.
Fuchs, J., C. Nilsson, J. Kachergus, M. Munz, E. M. Larsson, B. Schüle, J. W. Langston, F. A. Middleton, O. A. Ross, M. Hulihan, T. Gasser and M. J. Farrer (2007). "Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication." Neurology **68**(12): 916.
Fukae, J., Y. Mizuno and N. Hattori (2007). "Mitochondrial dysfunction in Parkinson's disease." Mitochondrion **7**(1-2): 58-62.
Funayama, M., K. Hasegawa, H. Kowa, M. Saito, S. Tsuji and F. Obata (2002). "A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1." Ann Neurol **51**(3): 296-301.
Fung, H. C., S. Scholz, M. Matarin, J. Simon-Sanchez, D. Hernandez, A. Britton, J. R. Gibbs, C. Langefeld, M. L. Stiegert, J. Schymick, M. S. Okun, R. J. Mandel, H. H. Fernandez, K. D. Foote, R. L. Rodriguez, E. Peckham, F. W. De Vrieze, K. Gwinn-Hardy, J. A. Hardy and A. Singleton (2006). "Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data." Lancet Neurol **5**(11): 911-916.

Fuson, K., A. Rice, R. Mahling, A. Snow, K. Nayak, P. Shanbhogue, A. G. Meyer, G. M. I. Redpath, A. Hinderliter, S. T. Cooper and R. B. Sutton (2014). "Alternate Splicing of Dysferlin C2A Confers Ca(2+)-Dependent and Ca(2+)-Independent Binding for Membrane Repair." Structure (London, England : 1993) **22**(1): 104-115.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly and D. Altshuler (2002). "The structure of haplotype blocks in the human genome." Science **296**(5576): 2225-2229.

Gautier, C. A., T. Kitada and J. Shen (2008). "Loss of PINK1 causes mitochondrial functional defects and increased sensitivity to oxidative stress." Proc Natl Acad Sci U S A **105**(32): 11364-11369.

Gilissen, C., J. Y. Hehir-Kwa, D. T. Thung, M. van de Vorst, B. W. M. van Bon, M. H. Willemsen, M. Kwint, I. M. Janssen, A. Hoischen, A. Schenck, R. Leach, R. Klein, R. Tearle, T. Bo, R. Pfundt, H. G. Yntema, B. B. A. de Vries, T. Kleefstra, H. G. Brunner, L. E. L. M. Vissers and J. A. Veltman (2014). "Genome sequencing identifies major causes of severe intellectual disability." Nature **511**: 344.

Gillardon, F., R. Schmid and H. Draheim (2012). "Parkinson's disease-linked leucine-rich repeat kinase 2(R1441G) mutation increases proinflammatory cytokine release from activated primary microglial cells and resultant neurotoxicity." Neuroscience **208**: 41-48.

Giri, A., G. Guven, H. Hanagasi, A. K. Hauser, N. Erginul-Unaltuna, B. Bilgic, H. Gurvit, P. Heutink, T. Gasser, E. Lohmann and J. Simon-Sanchez (2016). "PLA2G6 Mutations Related to Distinct Phenotypes: A New Case with Early-onset Parkinsonism." Tremor Other Hyperkinet Mov (N Y) **6**: 363.

Goker-Alpan, O., B. I. Giasson, M. J. Eblan, J. Nguyen, H. I. Hurtig, V. M. Y. Lee, J. Q. Trojanowski and E. Sidransky (2006). "Glucocerebrosidase mutations are an important risk factor for Lewy body disorders." Neurology **67**(5): 908.

Goldberg, M. S., A. Pisani, M. Haburcak, T. A. Vortherms, T. Kitada, C. Costa, Y. Tong, G. Martella, A. Tscherter, A. Martins, G. Bernardi, B. L. Roth, E. N. Pothos, P. Calabresi and J. Shen (2005). "Nigrostriatal Dopaminergic Deficits and Hypokinesia Caused by Inactivation of the Familial Parkinsonism-Linked Gene DJ-1." Neuron **45**(4): 489-496.

Golts, N., H. Snyder, M. Frasier, C. Theisler, P. Choi and B. Wolozin (2002). "Magnesium inhibits spontaneous and iron-induced aggregation of alpha-synuclein." J Biol Chem **277**(18): 16116-16123.

Guerreiro, R., A. Wojtas, J. Bras, M. Carrasquillo, E. Rogaeva, E. Majounie, C. Cruchaga, C. Sassi, J. S. Kauwe, S. Younkin, L. Hazrati, J. Collinge, J. Pocock, T. Lashley, J. Williams, J. C. Lambert, P. Amouyel, A. Goate, R. Rademakers, K. Morgan, J. Powell, P. St George-Hyslop, A. Singleton, J. Hardy and G. Alzheimer Genetic Analysis (2013). "TREM2 variants in Alzheimer's disease." N Engl J Med **368**(2): 117-127.

Guerreiro, R. J., E. Lohmann, J. M. Brás, J. R. Gibbs, J. D. Rohrer, N. Gurunlian, B. Dursun, B. Bilgic, H. Hanagasi, H. Gurvit, M. Emre, A. Singleton and J. Hardy (2013). "Using Exome Sequencing to Reveal Mutations in TREM2 Presenting as a Frontotemporal Dementia–like Syndrome Without Bone Involvement." JAMA neurology **70**(1): 78-84.
Gutti, U., H.-C. Fung, K. S. Hruska, M. E. LaMarca, C.-M. Chen, Y.-R. Wu and E. Sidransky (2008). "The Need for Appropriate Genotyping Strategies for Glucocerebrosidase Mutations in Cohorts With Parkinson Disease." Archives of neurology **65**(6): 850-851.
Hamza, T. H., C. P. Zabetian, A. Tenesa, A. Laederach, J. Montimurro, D. Yearout, D. M. Kay, K. F. Doheny, J. Paschall, E. Pugh, V. I. Kusel, R. Collura, J. Roberts, A. Griffith, A. Samii, W. K. Scott, J. Nutt, S. A. Factor and H. Payami (2010). "Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease." Nature genetics **42**(9): 781-785.
Hanagasi, H. A., A. Giri, E. Kartal, G. Guven, B. Bilgic, A. K. Hauser, M. Emre, P. Heutink, N. Basak, T. Gasser, J. Simon-Sanchez and E. Lohmann (2016). "A novel homozygous DJ1 mutation causes parkinsonism and ALS in a Turkish family." Parkinsonism Relat Disord **29**: 117-120.
Hatano, Y., Y. Li, K. Sato, S. Asakawa, Y. Yamamura, H. Tomiyama, H. Yoshino, M. Asahina, S. Kobayashi, S. Hassin-Baer, C. S. Lu, A. R. Ng, R. L. Rosales, N. Shimizu, T. Toda, Y. Mizuno and N. Hattori (2004). "Novel PINK1 mutations in early-onset parkinsonism." Ann Neurol **56**(3): 424-427.
Hayashi, T., C. Ishimori, K. Takahashi-Niki, T. Taira, Y.-c. Kim, H. Maita, C. Maita, H. Ariga and S. M.M. Iguchi-Ariga (2009). DJ-1 binds to mitochondrial Complex I and maintains its activity.
Healy, D. G., M. Falchi, S. S. O'Sullivan, V. Bonifati, A. Durr, S. Bressman, A. Brice, J. Aasly, C. P. Zabetian, S. Goldwurm, J. J. Ferreira, E. Tolosa, D. M. Kay, C. Klein, D. R. Williams, C. Marras, A. E. Lang, Z. K. Wszolek, J. Berciano, A. H. V. Schapira, T. Lynch, K. P. Bhatia, T. Gasser, A. J. Lees, N. W. Wood and L. C. on behalf of the International (2008). "Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study." Lancet Neurology **7**(7): 583-590.
Hering, R., K. M. Strauss, X. Tao, A. Bauer, D. Woitalla, E. M. Mietz, S. Petrovic, P. Bauer, W. Schaible, T. Muller, L. Schols, C. Klein, D. Berg, P. T. Meyer, J. B. Schulz, B. Wollnik, L. Tong, R. Kruger and O. Riess (2004). "Novel homozygous p.E64D mutation in DJ1 in early onset Parkinson disease (PARK7)." Hum Mutat **24**(4): 321-329.
Herrmann, W. and J. P. Knapp (2002). "Hyperhomocysteinemia: a new risk factor for degenerative diseases." Clin Lab **48**(9-10): 471-481.
Hou, C., W. Tian, T. Kleist, K. He, V. Garcia, F. Bai, Y. Hao, S. Luan and L. Li (2014). "DUF221 proteins are a family of osmosensitive calcium-permeable cation channels conserved across eukaryotes." Cell Research **24**(5): 632-635.

Houlden, H. and A. B. Singleton (2012). "The genetics and neuropathology of Parkinson's disease." Acta Neuropathol **124**(3): 325-338.
Hruska, K. S., M. E. LaMarca, C. R. Scott and E. Sidransky (2008). "Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA)." Hum Mutat **29**(5): 567-583.
Hutton, M., C. L. Lendon, P. Rizzu, M. Baker, S. Froelich, H. Houlden, S. Pickering-Brown, S. Chakraverty, A. Isaacs, A. Grover, J. Hackett, J. Adamson, S. Lincoln, D. Dickson, P. Davies, R. C. Petersen, M. Stevens, E. de Graaff, E. Wauters, J. van Baren, M. Hillebrand, M. Joosse, J. M. Kwon, P. Nowotny, L. K. Che, J. Norton, J. C. Morris, L. A. Reed, J. Trojanowski, H. Basun, L. Lannfelt, M. Neystat, S. Fahn, F. Dark, T. Tannenberg, P. R. Dodd, N. Hayward, J. B. J. Kwok, P. R. Schofield, A. Andreadis, J. Snowden, D. Craufurd, D. Neary, F. Owen, B. A. Oostra, J. Hardy, A. Goate, J. van Swieten, D. Mann, T. Lynch and P. Heutink (1998). "Association of missense and 5′ -splice-site mutations in tau with the inherited dementia FTDP-17." Nature **393**: 702.
International Parkinson Disease Genomics, C., M. A. Nalls, V. Plagnol, D. G. Hernandez, M. Sharma, U. M. Sheerin, M. Saad, J. Simon-Sanchez, C. Schulte, S. Lesage, S. Sveinbjornsdottir, K. Stefansson, M. Martinez, J. Hardy, P. Heutink, A. Brice, T. Gasser, A. B. Singleton and N. W. Wood (2011). "Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies." Lancet **377**(9766): 641-649.
Jankovic, J. (2008). "Parkinson's disease: clinical features and diagnosis." J Neurol Neurosurg Psychiatry **79**(4): 368-376.
Jiang, P., M. Gan, A. S. Ebrahim, W.-L. Lin, H. L. Melrose and S.-H. C. Yen (2010). "ER stress response plays an important role in aggregation of α-synuclein." Molecular Neurodegeneration **5**: 56-56.
Jin, S. M. and R. J. Youle (2012). "PINK1- and Parkin-mediated mitophagy at a glance." Journal of Cell Science **125**(4): 795.
Jinn, S., R. E. Drolet, P. E. Cramer, A. H.-K. Wong, D. M. Toolan, C. A. Gretzula, B. Voleti, G. Vassileva, J. Disa, M. Tadin-Strapps and D. J. Stone (2017). "TMEM175 deficiency impairs lysosomal and mitochondrial function and increases α-synuclein aggregation." Proceedings of the National Academy of Sciences of the United States of America **114**(9): 2389-2394.
Jung, C. H., M. Seo, N. M. Otto and D.-H. Kim (2011). "ULK1 inhibits the kinase activity of mTORC1 and cell proliferation." Autophagy **7**(10): 1212-1221.
Kalay, E., A. Uzumcu, E. Krieger, R. Çaylan, O. Uyguner, M. Ulubil-Emiroglu, H. Erdol, H. Kayserili, G. Hafiz, N. Başerer, J. G. M. Heister Angelien, C. Hennies Hans, P. Nürnberg, S. Başaran, G. Brunner Han, W. R. J. Cremers Cor, A. Karaguzel, B. Wollnik and H. Kremer (2007). "MYO15A (DFNB3) mutations in Turkish hearing loss families and functional modeling of a novel motor domain mutation." American Journal of Medical Genetics Part A **143A**(20): 2382-2389.
Kalia, L. V. and A. E. Lang "Parkinson's disease." The Lancet **386**(9996): 896-912.

Kay, D., D. Moran, L. Moses, P. Poorkaj, C. P Zabetian, J. Nutt, S. A Factor, C.-E. Yu, J. Montimurro, R. Keefe, G. Schellenberg and H. Payami (2007). Heterozygous parkin point mutations are as common in control subjects as in Parkinson's patients.
Kiely, A. P., Y. T. Asi, E. Kara, P. Limousin, H. Ling, P. Lewis, C. Proukakis, N. Quinn, A. J. Lees, J. Hardy, T. Revesz, H. Houlden and J. L. Holton (2013). "α-Synucleinopathy associated with G51D SNCA mutation: a link between Parkinson's disease and multiple system atrophy?" Acta Neuropathologica **125**(5): 753-769.
Kilarski Laura, L., P. Pearson Justin, V. Newsway, E. Majounie, M. D. W. Knipe, A. Misbahuddin, F. Chinnery Patrick, J. Burn David, E. Clarke Carl, M.-H. Marion, J. Lewthwaite Alistair, J. Nicholl David, W. Wood Nicholas, E. Morrison Karen, H. Williams-Gray Caroline, R. Evans Jonathan, J. Sawcer Stephen, A. Barker Roger, M. Wickremaratchi Mirdhu, Y. Ben-Shlomo, M. Williams Nigel and R. Morris Huw (2012). "Systematic Review and UK-Based Study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease." Movement Disorders **27**(12): 1522-1529.
Kim, D., B. Langmead and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements." Nat Methods **12**(4): 357-360.
Kim, J.-h., D.-j. Choi, H.-k. Jeong, J. Kim, D. W. Kim, S. Y. Choi, S.-M. Park, Y. H. Suh, I. Jou and E.-H. Joe (2013). "DJ-1 facilitates the interaction between STAT1 and its phosphatase, SHP-1, in brain microglia and astrocytes: A novel anti-inflammatory function of DJ-1." Neurobiology of Disease **60**: 1-10.
Kirby, D. M., R. Salemi, C. Sugiana, A. Ohtake, L. Parry, K. M. Bell, E. P. Kirk, A. Boneh, R. W. Taylor, H.-H. M. Dahl, M. T. Ryan and D. R. Thorburn (2004). "NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency." Journal of Clinical Investigation **114**(6): 837-845.
Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper and J. Shendure (2014). "A general framework for estimating the relative pathogenicity of human genetic variants." Nature Genetics **46**: 310.
Kitada, T., S. Asakawa, N. Hattori, H. Matsumine, Y. Yamamura, S. Minoshima, M. Yokochi, Y. Mizuno and N. Shimizu (1998). "Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism." Nature **392**(6676): 605-608.
Kitada, T., A. Pisani, D. R. Porter, H. Yamaguchi, A. Tscherter, G. Martella, P. Bonsi, C. Zhang, E. N. Pothos and J. Shen (2007). "Impaired dopamine release and synaptic plasticity in the striatum of PINK1-deficient mice." Proc Natl Acad Sci U S A **104**(27): 11441-11446.
Klein, C., A. Djarmati, K. Hedrich, N. Schafer, C. Scaglione, R. Marchese, N. Kock, B. Schule, A. Hiller, T. Lohnau, S. Winkler, K. Wiegers, R. Hering, P. Bauer, O. Riess, G. Abbruzzese, P. Martinelli and P. P. Pramstaller (2005). "PINK1, Parkin, and DJ-1 mutations in Italian patients with early-onset parkinsonism." Eur J Hum Genet **13**(9): 1086-1093.

Koivunen, P., P. Tiainen, J. Hyvärinen, K. E Williams, R. Sormunen, S. J Klaus, K. I Kivirikko and J. Myllyharju (2007). An Endoplasmic Reticulum Transmembrane Prolyl 4-Hydroxylase Is Induced by Hypoxia and Acts on Hypoxia-inducible Factor.

Koprivica, V., D. L. Stone, J. K. Park, M. Callahan, A. Frisch, I. J. Cohen, N. Tayebi and E. Sidransky (2000). "Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease." Am J Hum Genet **66**(6): 1777-1786.

Krebiehl, G., S. Ruckerbauer, L. F. Burbulla, N. Kieper, B. Maurer, J. Waak, H. Wolburg, Z. Gizatullina, F. N. Gellerich, D. Woitalla, O. Riess, P. J. Kahle, T. Proikas-Cezanne and R. Krüger (2010). "Reduced Basal Autophagy and Impaired Mitochondrial Dynamics Due to Loss of Parkinson's Disease-Associated Protein DJ-1." PLoS ONE **5**(2): e9367.

Kruger, R., W. Kuhn, T. Muller, D. Woitalla, M. Graeber, S. Kosel, H. Przuntek, J. T. Epplen, L. Schols and O. Riess (1998). "Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease." Nat Genet **18**(2): 106-108.

Kumar, P., S. Henikoff and P. C. Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nature Protocols **4**: 1073.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." BMC Bioinformatics **9**: 559.

Langston, J. W., P. Ballard, J. W. Tetrud and I. Irwin (1983). "Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis." Science **219**(4587): 979.

Le Ber, I., A. De Septenville, R. Guerreiro, J. Bras, A. Camuzat, P. Caroppo, S. Lattante, P. Couarch, E. Kabashi, K. Bouya-Ahmed, B. Dubois and A. Brice (2014). "Homozygous TREM2 mutation in a family with atypical frontotemporal dementia." Neurobiol Aging **35**(10): 2419 e2423-2419 e2425.

Le, W., J. Wu and Y. Tang (2016). "Protective Microglia and Their Regulation in Parkinson's Disease." Frontiers in Molecular Neuroscience **9**: 89.

Lee, E.-J. and C. Tournier (2011). "The requirement of uncoordinated 51-like kinase 1 (ULK1) and ULK2 in the regulation of autophagy." Autophagy **7**(7): 689-695.

Lee, M.-J., J.-Y. Kim, K. Suk and J.-H. Park (2004). "Identification of the Hypoxia-Inducible Factor 1α-Responsive HGTD-P Gene as a Mediator in the Mitochondrial Apoptotic Pathway." Molecular and Cellular Biology **24**(9): 3918-3927.

Lee, S., M. J. Emond, M. J. Bamshad, K. C. Barnes, M. J. Rieder, D. A. Nickerson, N. G. E. S. P.-E. L. P. Team, D. C. Christiani, M. M. Wurfel and X. Lin (2012). "Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies." Am J Hum Genet **91**(2): 224-237.

Lees, A. J., J. Hardy and T. Revesz (2009). "Parkinson's disease." Lancet **373**(9680): 2055-2066.

Leroy, E., R. Boyer, G. Auburger, B. Leube, G. Ulm, E. Mezey, G. Harta, M. J. Brownstein, S. Jonnalagada, T. Chernova, A. Dehejia, C. Lavedan, T. Gasser, P. J. Steinbach, K. D. Wilkinson and M. H. Polymeropoulos (1998). "The ubiquitin pathway in Parkinson's disease." Nature **395**: 451.

Lesage, S., M. Anheim, F. Letournel, L. Bousset, A. Honore, N. Rozas, L. Pieri, K. Madiona, A. Durr, R. Melki, C. Verny, A. Brice and G. French Parkinson's Disease Genetics Study (2013). "G51D alpha-synuclein mutation causes a novel parkinsonian-pyramidal syndrome." Ann Neurol **73**(4): 459-471.

Lesage, S. and A. Brice (2009). "Parkinson's disease: from monogenic forms to genetic susceptibility factors." Hum Mol Genet **18**(R1): R48-59.

Lesage, S., V. Drouet, E. Majounie, V. Deramecourt, M. Jacoupy, A. Nicolas, F. Cormier-Dequaire, S. M. Hassoun, C. Pujol, S. Ciura, Z. Erpapazoglou, T. Usenko, C. A. Maurage, M. Sahbatou, S. Liebau, J. Ding, B. Bilgic, M. Emre, N. Erginel-Unaltuna, G. Guven, F. Tison, C. Tranchant, M. Vidailhet, J. C. Corvol, P. Krack, A. L. Leutenegger, M. A. Nalls, D. G. Hernandez, P. Heutink, J. R. Gibbs, J. Hardy, N. W. Wood, T. Gasser, A. Durr, J. F. Deleuze, M. Tazir, A. Destee, E. Lohmann, E. Kabashi, A. Singleton, O. Corti, A. Brice, S. French Parkinson's Disease Genetics and C. International Parkinson's Disease Genomics (2016). "Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy." Am J Hum Genet **98**(3): 500-513.

Lesage, S., A. Dürr, M. Tazir, E. Lohmann, A.-L. Leutenegger, S. Janin, P. Pollak and A. Brice (2006). "LRRK2 G2019S as a Cause of Parkinson's Disease in North African Arabs." New England Journal of Medicine **354**(4): 422-423.

Lewthwaite, A. J., T. D. Lambert, E. B. Rolfe, S. Olgiati, M. Quadri, E. J. Simons, K. E. Morrison, V. Bonifati and D. J. Nicholl (2015). "Novel GCH1 variant in Dopa-responsive dystonia and Parkinson's disease." Parkinsonism Relat Disord **21**(4): 394-397.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics **30**(7): 923-930.

MacLeod, D. A., H. Rhinn, T. Kuwahara, A. Zolin, G. Di Paolo, B. D. McCabe, K. S. Marder, L. S. Honig, L. N. Clark, S. A. Small and A. Abeliovich (2013). "RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk." Neuron **77**(3): 425-439.

Maraganore, D. M., M. de Andrade, T. G. Lesnick, K. J. Strain, M. J. Farrer, W. A. Rocca, P. V. Pant, K. A. Frazer, D. R. Cox and D. G. Ballinger (2005). "High-resolution whole-genome association study of Parkinson disease." Am J Hum Genet **77**(5): 685-693.

Marella, M., B. B. Seo, T. Yagi and A. Matsuno-Yagi (2009). "Parkinson's disease and mitochondrial complex I: a perspective on the Ndi1 therapy." J Bioenerg Biomembr **41**(6): 493-497.

Markopoulou, K., D. W. Dickson, R. D. McComb, Z. K. Wszolek, L. Katechalidou, L. Avery, M. S. Stansbury and B. A. Chase (2008). "Clinical, neuropathological and genotypic variability in SNCA A53T familial Parkinson's disease. Variability in familial Parkinson's disease." Acta Neuropathol **116**(1): 25-35.
Mata, I. F., A. Samii, S. H. Schneer, J. W. Roberts, A. Griffith, B. C. Leis, G. D. Schellenberg, E. Sidransky, T. D. Bird, J. B. Leverenz, D. Tsuang and C. P. Zabetian (2008). "Glucocerebrosidase Gene Mutations: A Risk Factor for Lewy Body Disorders." Archives of neurology **65**(3): 379-382.
Matsuda, S., Y. Kitagishi and M. Kobayashi (2013). "Function and characteristics of PINK1 in mitochondria." Oxid Med Cell Longev **2013**: 601587.
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.
McNaught, K. S., C. W. Olanow, B. Halliwell, O. Isacson and P. Jenner (2001). "Failure of the ubiquitin-proteasome system in Parkinson's disease." Nat Rev Neurosci **2**(8): 589-594.
Mezey, E., A. M. Dehejia, G. Harta, N. Tresser, S. F. Suchy, R. L. Nussbaum, M. J. Brownstein and M. H. Polymeropoulos (1998). "Alpha synuclein is present in Lewy bodies in sporadic Parkinson's disease." Mol Psychiatry **3**(6): 493-499.
Michel, Patrick P., Etienne C. Hirsch and S. Hunot (2016). "Understanding Dopaminergic Cell Death Pathways in Parkinson Disease." Neuron **90**(4): 675-691.
Miki, Y., K. Tanji, F. Mori, J. Utsumi, H. Sasaki, A. Kakita, H. Takahashi and K. Wakabayashi (2016). "Alteration of Upstream Autophagy-Related Proteins (ULK1, ULK2, Beclin1, VPS34 and AMBRA1) in Lewy Body Disease." Brain Pathol **26**(3): 359-370.
Mitsui, J., I. Mizuta, A. Toyoda, R. Ashida, Y. Takahashi, J. Goto, Y. Fukuda, H. Date, A. Iwata, M. Yamamoto, N. Hattori, M. Murata, T. Toda and S. Tsuji (2009). "Mutations for Gaucher disease confer high susceptibility to Parkinson disease." Arch Neurol **66**(5): 571-576.
Morgan, N. V., S. K. Westaway, J. E. V. Morton, A. Gregory, P. Gissen, S. Sonek, H. Cangul, J. Coryell, N. Canham, N. Nardocci, G. Zorzi, S. Pasha, D. Rodriguez, I. Desguerre, A. Mubaidin, E. Bertini, R. C. Trembath, A. Simonati, C. Schanen, C. A. Johnson, B. Levinson, C. G. Woods, B. Wilmot, P. Kramer, J. Gitschier, E. R. Maher and S. J. Hayflick (2006). "PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron." Nature Genetics **38**: 752.
Muftuoglu, M., B. Elibol, O. Dalmizrak, A. Ercan, G. Kulaksiz, H. Ogus, T. Dalkara and N. Ozer (2004). "Mitochondrial complex I and IV activities in leukocytes from patients with parkin mutations." Mov Disord **19**(5): 544-548.
Mukadam, A. S. and M. N. J. Seaman (2015). "Retromer-mediated endosomal protein sorting: The role of unstructured domains." FEBS Letters **589**(19, Part A): 2620-2626.

Myhre, O., H. Utkilen, N. Duale, G. Brunborg and T. Hofer (2013). "Metal dyshomeostasis and inflammation in Alzheimer's and Parkinson's diseases: possible impact of environmental exposures." Oxid Med Cell Longev **2013**: 726954.

Naj, A. C., G. W. Beecham, E. R. Martin, P. J. Gallins, E. H. Powell, I. Konidari, P. L. Whitehead, G. Cai, V. Haroutunian, W. K. Scott, J. M. Vance, M. A. Slifer, H. E. Gwirtsman, J. R. Gilbert, J. L. Haines, J. D. Buxbaum and M. A. Pericak-Vance (2010). "Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities." PLoS Genet **6**(9): e1001130.

Nakagawa, H., Y. Murata, K. Koyama, A. Fujiyama, Y. Miyoshi, M. Monden, T. Akiyama and Y. Nakamura (1998). "Identification of a Brain-specific APC Homologue, APCL, and Its Interaction with β-Catenin." Cancer Research **58**(22): 5176.

Nalls, M. A., J. Bras, D. G. Hernandez, M. F. Keller, E. Majounie, A. E. Renton, M. Saad, I. Jansen, R. Guerreiro, S. Lubbe, V. Plagnol, J. R. Gibbs, C. Schulte, N. Pankratz, M. Sutherland, L. Bertram, C. M. Lill, A. L. DeStefano, T. Faroud, N. Eriksson, J. Y. Tung, C. Edsall, N. Nichols, J. Brooks, S. Arepalli, H. Pliner, C. Letson, P. Heutink, M. Martinez, T. Gasser, B. J. Traynor, N. Wood, J. Hardy, A. B. Singleton, C. International Parkinson's Disease Genomics and c. Parkinson's Disease meta-analysis (2015). "NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases." Neurobiol Aging **36**(3): 1605 e1607-1612.

Nalls, M. A., N. Pankratz, C. M. Lill, C. B. Do, D. G. Hernandez, M. Saad, A. L. DeStefano, E. Kara, J. Bras, M. Sharma, C. Schulte, M. F. Keller, S. Arepalli, C. Letson, C. Edsall, H. Stefansson, X. Liu, H. Pliner, J. H. Lee, R. Cheng, C. International Parkinson's Disease Genomics, G. I. Parkinson's Study Group Parkinson's Research: The Organized, andMe, GenePd, C. NeuroGenetics Research, G. Hussman Institute of Human, I. Ashkenazi Jewish Dataset, H. Cohorts for, E. Aging Research in Genetic, C. North American Brain Expression, C. United Kingdom Brain Expression, C. Greek Parkinson's Disease, G. Alzheimer Genetic Analysis, M. A. Ikram, J. P. Ioannidis, G. M. Hadjigeorgiou, J. C. Bis, M. Martinez, J. S. Perlmutter, A. Goate, K. Marder, B. Fiske, M. Sutherland, G. Xiromerisiou, R. H. Myers, L. N. Clark, K. Stefansson, J. A. Hardy, P. Heutink, H. Chen, N. W. Wood, H. Houlden, H. Payami, A. Brice, W. K. Scott, T. Gasser, L. Bertram, N. Eriksson, T. Foroud and A. B. Singleton (2014). "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease." Nat Genet **46**(9): 989-993.

Narita, T., T. Nishimura, K. Yoshizaki and T. Taniyama (2005). "CIN85 associates with TNF receptor 1 via Src and modulates TNF-alpha-induced apoptosis." Exp Cell Res **304**(1): 256-264.

Neumann, J., J. Bras, E. Deas, S. S. O'Sullivan, L. Parkkinen, R. H. Lachmann, A. Li, J. Holton, R. Guerreiro, R. Paudel, B. Segarane, A. Singleton, A. Lees, J. Hardy, H.

Houlden, T. Revesz and N. W. Wood (2009). "Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease." Brain **132**(7): 1783-1794.
O'Regan, G., R. M. deSouza, R. Balestrino and A. H. Schapira (2017). "Glucocerebrosidase Mutations in Parkinson Disease." J Parkinsons Dis **7**(3): 411-422.
Obi, T., K. Nishioka, O. A. Ross, T. Terada, K. Yamazaki, A. Sugiura, M. Takanashi, K. Mizoguchi, H. Mori, Y. Mizuno and N. Hattori (2008). "Clinicopathologic study of a SNCA gene duplication patient with Parkinson disease and dementia." Neurology **70**(3): 238-241.
Osellame, L. D., T. S. Blacker and M. R. Duchen (2012). "Cellular and molecular mechanisms of mitochondrial function." Best Pract Res Clin Endocrinol Metab **26**(6): 711-723.
Ozelius, L. J., G. Senthil, R. Saunders-Pullman, E. Ohmann, A. Deligtisch, M. Tagliati, A. L. Hunt, C. Klein, B. Henick, S. M. Hailpern, R. B. Lipton, J. Soto-Valencia, N. Risch and S. B. Bressman (2006). "LRRK2 G2019S as a Cause of Parkinson's Disease in Ashkenazi Jews." New England Journal of Medicine **354**(4): 424-425.
Paisán-Ruíz, C., S. Jain, E. W. Evans, W. P. Gilks, J. Simón, M. van der Brug, A. L. de Munain, S. Aparicio, A. M. n. Gil, N. Khan, J. Johnson, J. R. Martinez, D. Nicholl, I. M. Carrera, A. S. Peňa, R. de Silva, A. Lees, J. F. Martí-Massó, J. Pérez-Tur, N. W. Wood and A. B. Singleton "Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease." Neuron **44**(4): 595-600.
Palacino, J. J., D. Sagi, M. S. Goldberg, S. Krauss, C. Motz, M. Wacker, J. Klose and J. Shen (2004). "Mitochondrial dysfunction and oxidative damage in parkin-deficient mice." J Biol Chem **279**(18): 18614-18622.
Paloneva, J., T. Manninen, G. Christman, K. Hovanes, J. Mandelin, R. Adolfsson, M. Bianchin, T. Bird, R. Miranda, A. Salmaggi, L. Tranebjaerg, Y. Konttinen and L. Peltonen (2002). "Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype." Am J Hum Genet **71**(3): 656-662.
Pankratz, N., J. B. Wilk, J. C. Latourelle, A. L. DeStefano, C. Halter, E. W. Pugh, K. F. Doheny, J. F. Gusella, W. C. Nichols, T. Foroud, R. H. Myers, P. Psg, C. GenePd Investigators and L. Molecular Genetic (2009). "Genomewide association study for susceptibility genes contributing to familial Parkinson disease." Hum Genet **124**(6): 593-605.
Papapetropoulos, S., N. Adi, J. Ellul, A. A. Argyriou and E. Chroni (2007). "A prospective study of familial versus sporadic Parkinson's disease." Neurodegener Dis **4**(6): 424-427.
Parker, W. D., Jr., S. J. Boyson and J. K. Parks (1989). "Abnormalities of the electron transport chain in idiopathic Parkinson's disease." Ann Neurol **26**(6): 719-723.

Parkinson, J. (2002). "An Essay on the Shaking Palsy." The Journal of Neuropsychiatry and Clinical Neurosciences **14**(2): 223-236.
Pascale, E., M. E. Di Battista, A. Rubino, C. Purcaro, M. Valente, F. Fattapposta, G. Ferraguti and G. Meco (2016). "Genetic Architecture of MAPT Gene Region in Parkinson Disease Subtypes." Front Cell Neurosci **10**: 96.
Piccoli, C., M. Ripoli, G. Quarato, R. Scrima, A. D'Aprile, D. Boffoli, M. Margaglione, C. Criscuolo, G. De Michele, A. Sardanelli, S. Papa and N. Capitanio (2008). "Coexistence of mutations in PINK1 and mitochondrial DNA in early onset parkinsonism." J Med Genet **45**(9): 596-602.
Pollard, K. S., M. J. Hubisz, K. R. Rosenbloom and A. Siepel (2010). "Detection of nonneutral substitution rates on mammalian phylogenies." Genome Res **20**(1): 110-121.
Polymeropoulos, M. H., C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe and R. L. Nussbaum (1997). "Mutation in the alpha-synuclein gene identified in families with Parkinson's disease." Science **276**(5321): 2045-2047.
Poulopoulos, M., O. A. Levy and R. N. Alcalay (2012). "The neuropathology of genetic Parkinson's disease." Mov Disord **27**(7): 831-842.
Pringsheim, T., N. Jette, A. Frolkis and T. D. Steeves (2014). "The prevalence of Parkinson's disease: a systematic review and meta-analysis." Mov Disord **29**(13): 1583-1590.
Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly and P. C. Sham (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.
Rajput, A., D. W. Dickson, C. A. Robinson, O. A. Ross, J. C. Dächsel, S. J. Lincoln, S. A. Cobb, M. L. Rajput and M. J. Farrer (2006). "Parkinsonism, Lrrk2 G2019S, and tau neuropathology." Neurology **67**(8): 1506.
Rakovic, A., K. Shurkewitsch, P. Seibler, A. Grunewald, A. Zanon, J. Hagenah, D. Krainc and C. Klein (2013). "Phosphatase and tensin homolog (PTEN)-induced putative kinase 1 (PINK1)-dependent ubiquitination of endogenous Parkin attenuates mitophagy: study in human primary fibroblasts and induced pluripotent stem cell-derived neurons." J Biol Chem **288**(4): 2223-2237.
Reeve, A. K., J. P. Grady, E. M. Cosgrave, E. Bennison, C. Chen, P. D. Hepplewhite and C. M. Morris (2018). "Mitochondrial dysfunction within the synapses of substantia nigra neurons in Parkinson's disease." npj Parkinson's Disease **4**(1): 9.
Risch, N. and K. Merikangas (1996). "The Future of Genetic Studies of Complex Human Diseases." Science **273**(5281): 1516.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative Genomics Viewer." Nature biotechnology **29**(1): 24-26.
Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.
Sala, G., D. Marinig, A. Arosio and C. Ferrarese (2016). "Role of Chaperone-Mediated Autophagy Dysfunctions in the Pathogenesis of Parkinson's Disease." Frontiers in Molecular Neuroscience **9**: 157.
Sanger, F. and A. R. Coulson (1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase." Journal of Molecular Biology **94**(3): 441-448.
Satake, W., Y. Nakabayashi, I. Mizuta, Y. Hirota, C. Ito, M. Kubo, T. Kawaguchi, T. Tsunoda, M. Watanabe, A. Takeda, H. Tomiyama, K. Nakashima, K. Hasegawa, F. Obata, T. Yoshikawa, H. Kawakami, S. Sakoda, M. Yamamoto, N. Hattori, M. Murata, Y. Nakamura and T. Toda (2009). "Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease." Nat Genet **41**(12): 1303-1307.
Saunders, C. J., N. A. Miller, S. E. Soden, D. L. Dinwiddie, A. Noll, N. A. Alnadi, N. Andraws, M. L. Patterson, L. A. Krivohlavek, J. Fellis, S. Humphray, P. Saffrey, Z. Kingsbury, J. C. Weir, J. Betley, R. J. Grocock, E. H. Margulies, E. G. Farrow, M. Artman, N. P. Safina, J. E. Petrikin, K. P. Hall and S. F. Kingsmore (2012). "Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units." Sci Transl Med **4**(154): 154ra135.
Savitt, J. M., V. L. Dawson and T. M. Dawson (2006). "Diagnosis and treatment of Parkinson disease: molecules to medicine." J Clin Invest **116**(7): 1744-1754.
Schapira, A. H. (1998). "Human complex I defects in neurodegenerative diseases." Biochim Biophys Acta **1364**(2): 261-270.
Schapira, A. H., J. M. Cooper, D. Dexter, J. B. Clark, P. Jenner and C. D. Marsden (1990). "Mitochondrial complex I deficiency in Parkinson's disease." J Neurochem **54**(3): 823-827.
Schapira, A. H., J. M. Cooper, D. Dexter, P. Jenner, J. B. Clark and C. D. Marsden (1989). "Mitochondrial complex I deficiency in Parkinson's disease." Lancet **1**(8649): 1269.
Schulte, J. and J. T. Littleton (2011). "The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology." Current trends in neurology **5**: 65-78.
Simon-Sanchez, J., C. Schulte, J. M. Bras, M. Sharma, J. R. Gibbs, D. Berg, C. Paisan-Ruiz, P. Lichtner, S. W. Scholz, D. G. Hernandez, R. Kruger, M. Federoff, C. Klein, A. Goate, J. Perlmutter, M. Bonin, M. A. Nalls, T. Illig, C. Gieger, H. Houlden, M. Steffens, M. S. Okun, M. Cookson, K. D. Foote, H. H. Fernandez, B. J. Traynor, S. Schreiber, S. Arepalli, R. Zonozi, K. Gwinn, M. van der Brug, G. Lopez, S. J. Chanock, A. Schatzkin, Y. Park, A. Hollenbeck, J. Gao, X. Huang, N. W. Wood, D.

Lorenz, G. Deuschl, H. Chen, O. Riess, J. A. Hardy, A. B. Singleton and T. Gasser (2009). "Genome-Wide Association Study reveals genetic risk underlying Parkinson's disease." Nature genetics **41**(12): 1308-1312.
Singleton, A. B., M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum, S. Lincoln, A. Crawley, M. Hanson, D. Maraganore, C. Adler, M. R. Cookson, M. Muenter, M. Baptista, D. Miller, J. Blancato, J. Hardy and K. Gwinn-Hardy (2003). "alpha-Synuclein locus triplication causes Parkinson's disease." Science **302**(5646): 841.
Smith, A. C., J. A. Blackshaw and A. J. Robinson (2012). "MitoMiner: a data warehouse for mitochondrial proteomics data." Nucleic Acids Res **40**(Database issue): D1160-1167.
Spillantini, M. G., M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert (1997). "Alpha-synuclein in Lewy bodies." Nature **388**(6645): 839-840.
Stavropoulos, D. J., D. Merico, R. Jobling, S. Bowdin, N. Monfared, B. Thiruvahindrapuram, T. Nalpathamkalam, G. Pellecchia, R. K. C. Yuen, M. J. Szego, R. Z. Hayeems, R. Z. Shaul, M. Brudno, M. Girdea, B. Frey, B. Alipanahi, S. Ahmed, R. Babul-Hirji, R. B. Porras, M. T. Carter, L. Chad, A. Chaudhry, D. Chitayat, S. J. Doust, C. Cytrynbaum, L. Dupuis, R. Ejaz, L. Fishman, A. Guerin, B. Hashemi, M. Helal, S. Hewson, M. Inbar-Feigenberg, P. Kannu, N. Karp, R. H. Kim, J. Kronick, E. Liston, H. MacDonald, S. Mercimek-Mahmutoglu, R. Mendoza-Londono, E. Nasr, G. Nimmo, N. Parkinson, N. Quercia, J. Raiman, M. Roifman, A. Schulze, A. Shugar, C. Shuman, P. Sinajon, K. Siriwardena, R. Weksberg, G. Yoon, C. Carew, R. Erickson, R. A. Leach, R. Klein, P. N. Ray, M. S. Meyn, S. W. Scherer, R. D. Cohn and C. R. Marshall (2016). "Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine." Npj Genomic Medicine **1**: 15012.
Stefanis, L. (2012). "α-Synuclein in Parkinson's Disease." Cold Spring Harbor Perspectives in Medicine **2**(2): a009399.
Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A **102**(43): 15545-15550.
Swerdlow, R. H., J. K. Parks, S. W. Miller, J. B. Tuttle, P. A. Trimmer, J. P. Sheehan, J. P. Bennett, Jr., R. E. Davis and W. D. Parker, Jr. (1996). "Origin and functional consequences of the complex I defect in Parkinson's disease." Ann Neurol **40**(4): 663-671.
Tan, E., J. Tong, S. Fook-Chong and et al. (2007). "Glucocerebrosidase mutations and risk of parkinson disease in chinese patients." Archives of Neurology **64**(7): 1056-1058.
Tsuji, S., P. V. Choudary, B. M. Martin, B. K. Stubblefield, J. A. Mayor, J. A. Barranger and E. I. Ginns (1987). "A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease." N Engl J Med **316**(10): 570-575.

Tucci, A., M. A. Nalls, H. Houlden, T. Revesz, A. B. Singleton, N. W. Wood, J. Hardy and C. Paisán-Ruiz (2010). "Genetic variability at the PARK16 locus." European Journal of Human Genetics **18**(12): 1356-1359.
Turro, E. and N. BioResource (2016). "Whole Genome Sequencing of 10,000 Rare Disease Patients Reveals Digenic Inheritance of Blood Disorders." Blood **128**(22): 2540.
Valente, E. M., P. M. Abou-Sleiman, V. Caputo, M. M. Muqit, K. Harvey, S. Gispert, Z. Ali, D. Del Turco, A. R. Bentivoglio, D. G. Healy, A. Albanese, R. Nussbaum, R. Gonzalez-Maldonado, T. Deller, S. Salvi, P. Cortelli, W. P. Gilks, D. S. Latchman, R. J. Harvey, B. Dallapiccola, G. Auburger and N. W. Wood (2004). "Hereditary early-onset Parkinson's disease caused by mutations in PINK1." Science **304**(5674): 1158-1160.
Valente, E. M., A. R. Bentivoglio, P. H. Dixon, A. Ferraris, T. Ialongo, M. Frontali, A. Albanese and N. W. Wood (2001). "Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36." Am J Hum Genet **68**(4): 895-900.
Vilariño-Güell, C., C. Wider, Owen A. Ross, Justus C. Dachsel, Jennifer M. Kachergus, Sarah J. Lincoln, Alexandra I. Soto-Ortolaza, Stephanie A. Cobb, Greggory J. Wilhoite, Justin A. Bacon, B. Behrouz, Heather L. Melrose, E. Hentati, A. Puschmann, Daniel M. Evans, E. Conibear, Wyeth W. Wasserman, Jan O. Aasly, Pierre R. Burkhard, R. Djaldetti, J. Ghika, F. Hentati, A. Krygowska-Wajs, T. Lynch, E. Melamed, A. Rajput, Ali H. Rajput, A. Solida, R.-M. Wu, Ryan J. Uitti, Zbigniew K. Wszolek, F. Vingerhoets and Matthew J. Farrer (2011). "VPS35 Mutations in Parkinson Disease." American Journal of Human Genetics **89**(1): 162-167.
Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown and J. Yang (2017). "10 Years of GWAS Discovery: Biology, Function, and Translation." The American Journal of Human Genetics **101**(1): 5-22.
von Campenhausen, S., B. Bornschein, R. Wick, K. Bötzel, C. Sampaio, W. Poewe, W. Oertel, U. Siebert, K. Berger and R. Dodel (2005). "Prevalence and incidence of Parkinson's disease in Europe." European Neuropsychopharmacology **15**(4): 473-490.
Wang, K., M. Li and H. Hakonarson (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Research **38**(16): e164-e164.
Wang, L., M. G. Heckman, J. O. Aasly, G. Annesi, M. Bozi, S. J. Chung, C. Clarke, D. Crosiers, G. Eckstein, G. Garraux, G. M. Hadjigeorgiou, N. Hattori, B. Jeon, Y. J. Kim, M. Kubo, S. Lesage, J. J. Lin, T. Lynch, P. Lichtner, G. D. Mellick, V. Mok, K. E. Morrison, A. Quattrone, W. Satake, P. A. Silburn, L. Stefanis, J. D. Stockton, E. K. Tan, T. Toda, A. Brice, C. Van Broeckhoven, R. J. Uitti, K. Wirdefeldt, Z. Wszolek, G. Xiromerisiou, D. M. Maraganore, T. Gasser, R. Krüger, M. J. Farrer, O. A. Ross and M. Sharma "Evaluation of the interaction between LRRK2 and PARK16 loci in determining risk of Parkinson's disease: analysis of a large multicenter study." Neurobiology of Aging **49**: 217.e211-217.e214.

Wang, X., C. Terfve, J. C. Rose and F. Markowetz (2011). "HTSanalyzeR: an R/Bioconductor package for integrated network analysis of high-throughput screens." Bioinformatics **27**(6): 879-880.
Wirdefeldt, K., M. Gatz, C. A. Reynolds, C. A. Prescott and N. L. Pedersen (2011). "Heritability of Parkinson disease in Swedish twins: a longitudinal study." Neurobiol Aging **32**(10): 1923 e1921-1928.
Wirdefeldt, K., M. Gatz, M. Schalling and N. L. Pedersen (2004). "No evidence for heritability of Parkinson disease in Swedish twins." Neurology **63**(2): 305.
Wu, X. and D. S. Reddy (2012). "Integrins as Receptor Targets for Neurological Disorders." Pharmacol Ther **134**(1): 68-81.
Xiromerisiou, G., H. Houlden, A. Sailer, L. Silveira-Moriyama, J. Hardy and A. J. Lees (2012). "Identical Twins with Leucine Rich Repeat Kinase Type 2 Mutations Discordant for Parkinson's Disease." Movement Disorders **27**(10): 1323-1324.
Zanellati, M. C., V. Monti, C. Barzaghi, C. Reale, N. Nardocci, A. Albanese, E. M. Valente, D. Ghezzi and B. Garavaglia (2015). "Mitochondrial dysfunction in Parkinson disease: evidence in mutant PARK2 fibroblasts." Front Genet **6**: 78.
Zarranz, J. J., J. Alegre, J. C. Gomez-Esteban, E. Lezcano, R. Ros, I. Ampuero, L. Vidal, J. Hoenicka, O. Rodriguez, B. Atares, V. Llorens, E. Gomez Tortosa, T. del Ser, D. G. Munoz and J. G. de Yebenes (2004). "The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia." Ann Neurol **55**(2): 164-173.
Zhang, C. C., A. Xing, M. S. Tan, L. Tan and J. T. Yu (2016). "The Role of MAPT in Neurodegenerative Diseases: Genetics, Mechanisms and Therapy." Mol Neurobiol **53**(7): 4893-4904.
Zhang Claire, X., E. Y. Engqvist-Goldstein Åsa, S. Carreno, J. Owen David, E. Smythe and G. Drubin David (2005). "Multiple Roles for Cyclin G-Associated Kinase in Clathrin-Mediated Sorting Events." Traffic **6**(12): 1103-1113.
Zhang, Y. (2012). "A Novel Bayesian Graphical Model for Genome-Wide Multi-SNP Association Mapping." Genetic epidemiology **36**(1): 36-47.
Zhu, M., C. Need A , Y. Han, D. Ge, M. Maia J , Q. Zhu, L. Heinzen E , T. Cirulli E , K. Pelak, M. He, K. Ruzzo E , C. Gumbs, A. Singh, S. Feng, V. Shianna K and B. Goldstein D (2012). "Using ERDS to Infer Copy-Number Variants in High-Coverage Genomes." Am J Hum Genet **91**(3): 408-421.
Zhu, Z. G., M. X. Sun, W. L. Zhang, W. W. Wang, Y. M. Jin and C. L. Xie (2017). "The efficacy and safety of coenzyme Q10 in Parkinson's disease: a meta-analysis of randomized controlled trials." Neurol Sci **38**(2): 215-224.
Zimprich, A., S. Biskup, P. Leitner, P. Lichtner, M. Farrer, S. Lincoln, J. Kachergus, M. Hulihan, R. J. Uitti, D. B. Calne, A. J. Stoessl, R. F. Pfeiffer, N. Patenge, I. C. Carbajal, P. Vieregge, F. Asmus, B. Müller-Myhsok, D. W. Dickson, T. Meitinger, T. M. Strom, Z. K. Wszolek and T. Gasser "Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology." Neuron **44**(4): 601-607.