Exploring the genetic contribution to idiopathic Parkinson disease

Johannes Jernqvist Gaare

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

Exploring the genetic contribution to idiopathic Parkinson disease

Johannes Jernqvist Gaare

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 05.03.2021

Year: 2021

Title: Exploring the genetic contribution to idiopathic Parkinson disease

Name: Johannes Jernqvist Gaare

Print: Skipnes Kommunikasjon / University of Bergen

Table of contents

Acknowledgements

I would like to thank the patients and control subjects who participated in these studies. Without them, none of this research would have been possible.

My main supervisor, Prof. Charalampos “Haris” Tzoulis, has been an invaluable source of inspiration and support these last few years, and I’ve long since lost count of the number of meetings, e-mail threads, and coffee fueled discussions we’ve had. With his open-door policy and infectious enthusiasm for science, he constantly inspires all those who work alongside him. I consider myself extremely lucky to have been given the opportunity to be part of the scientific community he has created.

My co-supervisor, Dr. Kristoffer Haugarvoll, has been a constant source of grounded advice and direction, for which I am forever grateful. It is only through the combined efforts of him and Haris that I have been able to carry out the research presented in this thesis.

My co-supervisor and former head of the Department of Neurology at Haukeland University Hospital, Prof. Ole-Bjørn Tysnes, has been an inspiration ever since I attended his lectures in clinical neurology as a medical student 10 years ago. I am thankful for his support and guidance.

I want to thank all of my co-authors for their contributions to the work presented in this thesis. They are, in alphabetical order: Prof. Guido Alves, Prof. Olav Dahl, Dr. Kristoffer Haugarvoll, Prof. Stefan Johansson, Prof. Per Knappskog, Prof. Jan Petter Larsen, Prof. Morten Lund-Johansen, Dr. Jodi Maple-Grødem, Dr. Gonzalo Nido, Prof. Geir Olve Skeie, Dr. Paweł Stromwasser, Prof. Ole-Bjørn Tysnes and Prof. Charalampos Tzoulis. I would also like to thank all of those involved with the ParkWest study for their invaluable contributions.

A special thanks to my co-author, Neuromics group member and, sometimes, fellow runner Gonzalo Nido. He has been an inexhaustible source of help and guidance, and his knowledge of bioinformatics and statistics has been invaluable. I would also like to thank all of the other members of the Neuromics research group: Birgitte, Brage, Chris, Fiona, Gia, Gon, Irene, Janani, Lilah, Nelson, Romain and Thomas. I am grateful to be part of such a wonderful group of people.

I would also like to thank all of my colleagues at the Department of Neurology, who for years have cultivated an environment where science and research is both encouraged and supported.

A special thanks to Irene, who for close to 15 years has been a friend and fellow neurology enthusiast. I am grateful for all of our conversations, both work-related and not. I look forward to our continued exploration of the subject of garden gnomes.

My colleagues in the Rotunda tower office have provided me with a wonderful work-environment, where lunch discussions of both scientific and less scientific topics have been equally stimulating. I am grateful to have been able to share an office with Irene, Kjersti, Kristin, Margrethe, Elisabeth, Ingrid Anne, Hilde Marie, Vojtech, Annette, Aliona, Brage and Sander (though thankfully not all at once).

My parents, Eldar and Elisabeth, have been a constant source of inspiration for me throughout my education and scientific undertakings, and provided invaluable logistic support during the final phases of writing this thesis. Likewise, my mother-in-law, Elisabeth, and her husband, Edvin, have been very supportive, for which I am grateful.

Kjersti, thank you for always being there for me, and for your constant support in all of my endeavors, both the sensible ones and the slightly less so. Our daughter, Vilde, is the joy of my life, and inspires me to be the best I can be every day.

Scientific environment

These studies were carried out in the Neuromics research group at the Department of Clinical Medicine, University of Bergen, Norway, and the Department of Neurology, Haukeland University Hospital, Bergen, Norway.

The ParkWest study data used in all papers (I-IV) were gathered through a collaborative effort between the neurological departments of Haukeland University Hospital (Bergen, Norway), Stavanger University Hospital (Stavanger, Norway), Førde Central Hospital (Førde, Norway), Haugesund Hospital (Haugesund, Norway) and Sørlandet Hospital Arendal (Arendal, Norway).

Whole-exome sequencing (paper II, III, IV) was performed at HudsonAlpha Institute for Biotechnology (Huntsville, Alabama).

List of publications

I. Gaare JJ, Skeie GO, Tzoulis C, Larsen JP, Tysnes OB. *Familial aggregation of Parkinson's disease may affect progression of motor symptoms and dementia.* Movement Disorders, 2017. 32(2):241-5.

II. Gaare JJ, Nido GS, Stromwasser P, Knappskog PM, Dahl O, Lund-Johansen M, Alves G, Tysnes OB, Johansson S, Haugarvoll K, Tzoulis C. *No evidence for rare TRAP1 mutations influencing the risk of idiopathic Parkinson's disease.* Brain, 2018. 141(3):e16

III. Gaare JJ, Nido GS, Stromwasser P, Knappskog PM, Dahl O, Lund-Johansen M, Alves G, Tysnes OB, Johansson S, Haugarvoll K, Tzoulis C. *Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease.* Movement Disorders, 2018. 33(10):1591-600

IV. Gaare JJ, Nido GS, Stromwasser P, Knappskog PM, Dahl O, Lund-Johansen M, Alves G, Tysnes OB, Johansson S, Haugarvoll K, Tzoulis C. *Meta-analysis of whole-exome sequencing data from two independent cohorts finds no evidence for rare variant enrichment in Parkinson disease associated loci.* PLoS One, 2020. 15(10):e0239824.

List of abbreviations

AD	Autosomal dominant
ADHD	Attention deficit hyperactivity disorder
ALS	Amyotrophic lateral sclerosis
AMD	Age related macular degeneration
ANOVA	Analysis of variance
AR	Autosomal recessive
aSum	Data-adaptive sum test
CADD	Combined annotation dependent depletion
CAST	Cohort allelic sums test
CMC	Combined multivariate and collapsing
DLB	Dementia with Lewy bodies
DZ	Dizygotic
EOPD	Early onset Parkinson disease
eQTL	Expression quantitative trait locus
EREC	Estimated regression coefficient
ExAC	Exome Aggregation Consortium
FDR	False discovery rate
GATK	Genome Analysis Toolkit
GEE	Generalized estimating equations
gnomAD	Genome Aggregation Database
GO	Gene ontology
GWAS	Genome wide association study
JOPD	Juvenile onset Parkinson disease
KEGG	Kyoto encyclopedia of genes and genomes
LD	Linkage disequilibrium
LOPD	Late onset Parkinson disease
LoF	Loss of function
MAF	Minor allele frequency
MAP	Minimum achievable p-value
MMSE	Mini mental state examination

MPP+	1-methyl-4-phenylpyridinium
MPPP	1-methyl-4-phenyl-4-propionoxypiperidine
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
mQTL	Methylation quantitative trait locus
mRNA	Messenger RNA
MSA	Multiple system atrophy
mtDNA	Mitochondrial DNA
mtSSB	Mitochondrial single stranded DNA binding protein
MULVR	Multiple traits based on variable reduction
MZ	Monozygotic
MiST	Mixed effects score test
NGS	Next generation sequencing
OR	Odds ratio
PD	Parkinson disease
PET	Positron emission tomography
PIGD	Postural instability gait difficulties
PLAN	Phospholipase-associated neurodegeneration
POLG	DNA polymerase gamma
POLRMT	Mitochondrial DNA polymerase
PPMI	Parkinson Progression Markers Initiative
PtbD	Predicted-to-be-damaging
ROS	Reactive oxygen species
RR	Relative risk
SCA	Spinocerebellar ataxia
SGA	Single gene association
SKAT	Sequence kernel association test
SKAT-O	Optimal sequence kernel association test
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SNc	Subtantia nigra pars compacta
sQTL	Splicing quantitative trait locus

SSU	Sum of squares
SVA	Single variant association
TD	Tremor dominant
TFAM	Mitochondrial transcription factor A
UPDRS	Unified Parkinson's Disease Rating Scale
UTR	Untranslated region
VCF	Variant call format
VT	Variable threshold
WES	Whole exome sequencing
WGS	Whole genome sequencing
WST	Weighted sum test

Abstract

Background: Parkinson disease (PD) is a major cause of death and disability and has a devastating global socioeconomic impact. It affects 1-2% of the population above the age of 65 and its prevalence increases as the population ages. Several biological processes have been implicated in Parkinson disease, including mitochondrial dysfunction, aberrant protein clearance, and neuroinflammation. To which degree these processes are cause, effect or bystander to disease initiation and progression, remains however largely unknown. Having limited understanding of the mechanisms underlying the pathogenesis and pathophysiology of Parkinson disease, we are unable to develop disease-modifying therapies and patients face a future of progressive disability and premature death.

There is a clear hereditary component to idiopathic PD, established through both twin studies and genome-wide association studies. However, only a minor fraction of the total estimated heritability can be explained by known associated genetic variability. It has been hypothesized that the cumulative effects of rare, low-impact mutations spread across genes and biological pathways could explain some of this "missing heritability".

Aims: The aim of this work was to explore the genetic contribution to idiopathic PD, focusing on the cumulative effects of rare mutations.

Materials and methods: The main study population utilized in all four papers was the ParkWest cohort, a Norwegian population-based cohort of incident PD. In paper I, ParkWest provided both cases and controls, including clinical longitudinal data up to and including 7 years after baseline. All ParkWest cases were whole-exome sequenced and combined with previously sequenced control samples to form the genetic cohort utilized in papers II-IV. Additionally, a whole-exome sequencing cohort from the Parkinson Progression Markers Initiative was used in papers II-IV. Finally, a publicly available chip-genotyped dataset (NeuroX) from the International Parkinson's Disease Genomics Consortium was used as a replication cohort in paper IV. In paper I, we characterized the familial aggregation of Parkinson disease in the ParkWest cohort and

explored the effect of family history on disease progression. Subsequently, we used genetic data from multiple cohorts to assess the impact of rare, protein-altering mutations in mitochondrial biological pathways (paper III) and in genes previously linked to PD (paper II and IV).

Results and conclusions: We show that, while familial aggregation is present in our Norwegian cohort, this has a slightly lower effect size compared to previous studies. Through regression analysis we also show that having a family history of PD among first degree relatives is associated with a slightly milder phenotype, which may be due to genetic variability.

In paper II, we attempted to replicate the results of a recently published study reporting an association between genetic variation in the *TRAP1* gene and Parkinson disease. Our analyses did not replicate this association in our Norwegian cohort. Moreover, using stricter quality control parameters abolished the association in the same dataset used in the original study. Our results do not support the proposed role of *TRAP1* in idiopathic PD.

In paper III, we sought to investigate the role of rare, amino acid changing variation in molecular pathways related to mitochondrial function. Using the sequence kernel association (SKAT) test, we detected a statistically significant enrichment in the pathway of mitochondrial DNA maintenance. Impaired mitochondrial DNA homeostasis has previously been shown to be present in PD neurons, and our results indicate that this dysfunction could be partly mediated by inherited genetic mutations.

In paper IV, we performed a targeted single gene and gene-set association study on genes that had previously been implicated in PD through genome-wide association studies. We identified 303 genes of interest, but did not find statistically significant associations, either in the single gene or gene-set analyses. Our results do not therefore support a major role for rare variant enrichment in genes tagged by GWAS, but cannot rule out effects with small effect sizes.

1. Introduction

1.1 Parkinson disease

In 1817, James Parkinson published "An essay on the shaking palsy", describing the main clinical features of the disease that would later bear his name[1]. Today, Parkinson disease (PD) is recognized as the second most common neurodegenerative disorder after Alzheimer disease[2]. The clinical spectrum of PD comprises both motor and non-motor features. Motor features comprise bradykinesia, resting tremor, rigidity and postural instability. Non-motor symptoms are diverse, and include olfactory loss, neuropsychiatric dysfunction, autonomic dysregulation, gastrointestinal dysmotility, sleep disorders, cognitive impairment and dementia[3]. Available treatments for PD are purely symptomatic and can achieve partial control of primarily motor symptoms for a period of time. In the absence of neuroprotective therapies, however, neuronal loss progresses inexorably, leading to increasing disability and premature death[2].

The most important risk factor for PD is increasing age, with the prevalence ranging from 1.8% above the age of 65 years to approximately 3.0% above the age of 80 years in Western populations[4,5]. Additionally, epidemiological studies have suggested several environmental factors as possible modulators of PD risk, including dairy consumption, exposure to pesticides, methamphetamine use and traumatic brain injury. Conversely, a negative association has been shown with tobacco use, caffeine intake, high serum urate and physical activity, suggesting that these may be protective against PD[6]. The sexes are unequally affected, with males having an increased lifetime risk compared to females; 2.0% and 1.3% respectively[7].

Due to high levels of disability, which affect both patients and caregivers, and the need for medical treatment for both PD and PD-related complications, there are high societal costs linked to PD. In the United States alone, the total economic burden of PD has been estimated at > $50 billion per year[8], and the cost is expected to rise due to an ever increasing number of patients[9]. Given its ever-increasing prevalence, severe disability and high socioeconomic cost, PD is now one of the greatest challenges facing science, healthcare and society in the 21st century.

1.2 A primer on genetics and genetic association analyses

The human genome consists of approximately 3.2 billion base pairs, spread across 23 chromosomes[10]. It can broadly be divided into genes and intergenic regions, each making up roughly half the total size of the genome[11]. Genes consist of exons, introns and untranslated regions (UTRs). Exons are the protein-coding parts of the genome, while introns are spliced away during the synthesis of mRNA. Exons make up a relatively small part of the total size of the genome, <1%, and the complete protein-coding part of the genome is referred to as the exome[12]. A typical human genome differs from the reference sequence at, on average, roughly 4.1 to 5 million sites, depending on the population. The vast majority (>85%) of this variation is in the form of single nucleotide variation/polymorphisms (SNV/SNP), but less than 1% of these variants are situated in coding regions[13]. The majority are located in intergenic or intronic regions.

The sequencing of the human genome was first completed during the first half of the 2000s, at the end of a more than a decade long effort by the Human Genome Project[14,15]. The technique used was Sanger sequencing, also referred to as first generation sequencing, which relies on capillary electrophoresis to read the sequence. As exemplified by the number of years needed to complete the human genome sequence, this technique would be too laborious for sequencing the entire genome of a large number of individuals. In the years after the completion of the Human Genome Project, next generation sequencing (NGS) techniques, which employ mass parallelization of sequencing reactions and allow the sequence to be read in real time, were developed[16]. However, these methods were initially prohibitively expensive, and there was another type of genotyping technique, building on the groundwork from the Human Genome project as well as the concept of linkage disequilibrium (LD), that revolutionized the field of complex disease genetics.

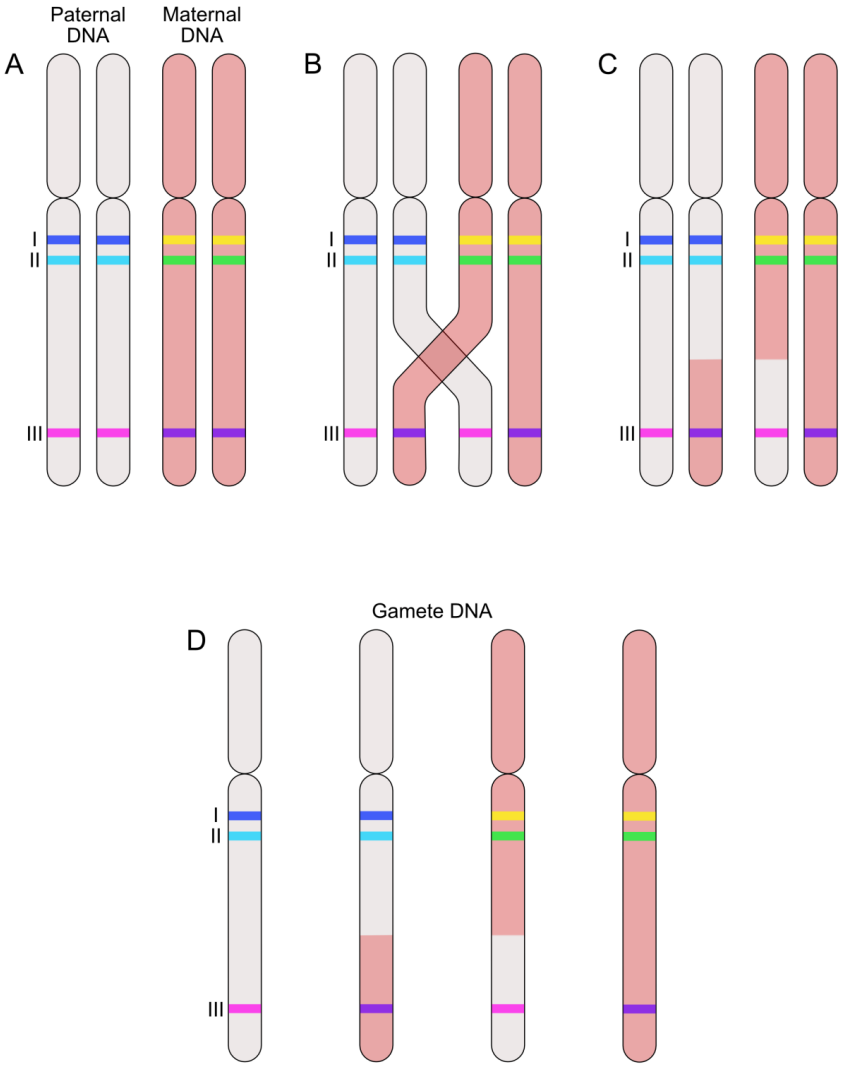

Figure 1. Recombination. *During meiosis, paternally and maternally derived DNA (A) undergo recombination (B and C), or chromosomal crossover, to form novel chromatids in the gamete cells (D). Genetic regions in close proximity (I and II) are more likely to be passed on together than regions far apart (III), because increasing genetic distance increases the likelihood of a recombination event taking place between the two loci.*

During meiosis, homologous chromosomes undergo recombination where the DNA breaks and crossover to form two new, unique mixtures of maternally and paternally derived DNA (see Figure 1). This is an important process in all sexually reproducing eukaryotes, ensuring genetic diversity in subsequent generations. Across the genome, some sites are more likely to undergo recombination events, so-called hotspots, while other areas are more conserved[17]. This results in the phenomenon of LD, where genetic variation on loci in close proximity are not randomly distributed in the population[18]. This non-random distribution of SNPs can be quantified given a large enough sample, and the sequence of whole regions of the DNA can therefore be inferred by the genotyping of only a few, selected SNPs. In the early 2000s, The International HapMap Consortium developed an ever-growing public database with detailed information on LD structures in the human genome[19]. Using relatively cheap and fast SNP-arrays, researchers could use that information to preselect SNPs and genotype a large number of individuals in a genome wide association study (GWAS)[20]. Here, one takes advantage of the LD-structures in the genome and look for genotype-phenotype associations between common SNPs, typically situated in non-coding regions, and disease. The actual causal variation is then assumed to be other variants in high LD with the associated SNPs[20]. Collectively, GWAS studies have revolutionized the field of complex trait genetics, and novel discoveries are continuously being made. As of December of 2019, a total of 166,103 SNP-phenotype associations have been described[21]. Despite their success, GWAS studies are not without their disadvantages. An important limitation is that the associated SNPs in most cases only act as proxies for the actual causal genetic variation. Additional assumptions and analyses are needed to connect a GWAS hit to an assumed causal gene. Furthermore, GWAS studies are not well suited to investigate rare mutations[22].

GWAS studies are largely based on the hypothesis that common diseases are caused by common variants, but for complex traits and disorders, GWAS have generally failed to account for the majority of the observed heritability. This discrepancy has been described as the “missing heritability” problem[23]. For PD, even the most recent GWAS of more than 37,000 cases and 1.4 million controls explain, by their own estimation, only 16 – 36% of the total estimated genetic heritability[24]. One

hypothesis that has been put forth is that much of the missing heritability can be explained by rare variants with small effect sizes that are not well covered by the SNP-arrays used in GWAS studies[23]. Rare variants (minor allele frequency [MAF] < 1%) make up more than 95% of all missense, nonsense and splicing mutations, making them an attractive target for genetic association studies[25]. Genetic sequencing is needed to capture and study these variants.

NGS techniques became commercially available around 2007-2008, and the price per whole genome quickly dropped from approx. $10 million to just $10,000 in just a few years, and the cost have continued to drop ever since[26]. A less expensive type of sequencing is exome sequencing, where you, as opposed to whole genome sequencing, capture only the protein-coding parts of the genome, i.e. the exome[27]. Sequencing data allows for novel approaches to genetic association testing. Since each rare mutation is expected to have only a minor effect by itself, prohibitively large sample sizes are needed to detect single variant associations (see Figure 1). Therefore, statistical methods have been developed that allow for testing the effects of an aggregate of multiple variants across a whole gene or region[28]. Using these methods, studies have documented region-based enrichment of rare variants for a multitude of complex disorders, including schizophrenia[29], type 2 diabetes[30], amyotrophic lateral sclerosis (ALS)[31], age related macular degeneration (AMD)[32] and hypertension[33].

1.2.1 Rare variant association testing

At lower MAFs, the number of individuals needed to detect a single variant association with acceptable power (80%) at a genome-wide significance level (5×10^{-8}) increases exponentially (see Figure 2). For example, assuming a population prevalence for PD of 1% and an equal number of cases and controls, the number of individuals needed to detect a variant with a MAF of 1% and OR of 1.4 would be approx. 60,000 (calculated using Quanto v1.2.4[34]). A similar variant with a MAF of 0.1% would require approx. 600,000 individuals. In addition, each individual also carries a number of unique mutations (singletons), with estimates ranging from 10,000-20,000 singletons per person depending on the population[13].

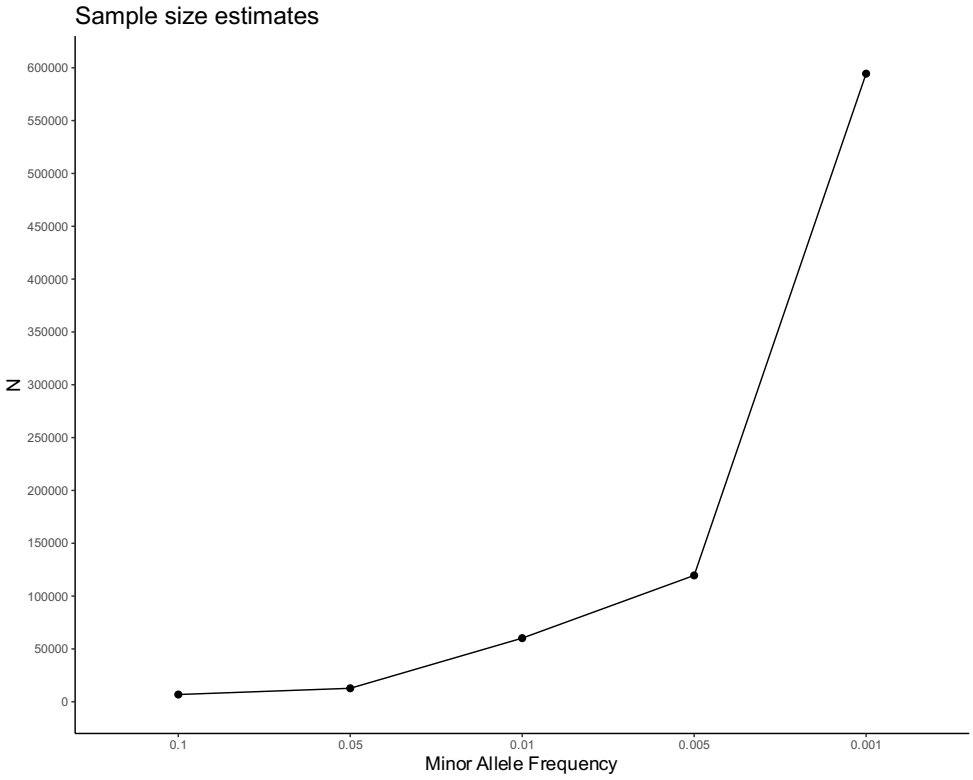

Figure 2. **Sample size estimates**. *Sample sizes needed to achieve 80% power at different MAFs for a single variant with OR 1.4, assuming an equal number of cases and controls and at a genome-wide significance level (5e-08).*

Being present in only one individual, they cannot be used in a traditional single variant association analysis. Given these limitations, it is clear that alternative approaches are needed to elucidate the contribution of rare variants to complex diseases such as PD.

A general strategy in rare variant association analyses is to group multiple variants together and perform the analysis on different aggregated measures. An *a priori* hypothesis is necessary in order to define meaningful groupings and interpret the results of the analysis and must specify both variant- and region-based parameters. Variant-based parameters define which variants are to be included in the analysis, and categories to consider are localization (exonic, intronic, intergenic), type (synonymous, nonsynonymous, splicing, stopgain, stoploss), function (loss-of-

function, gain-of-function, likelihood of altered protein function) and MAF. Region-based parameters define how to group the subsequently selected variants, and can for example be all genes, a subset of genes, promotor regions, or pathways. This strategy accomplishes two things: 1) an increase in power as multiple low-impact variants are grouped together, and 2) reducing the need for multiple testing correction compared to single-variant analyses. The standard genome-wide significance threshold for single variant analyses (5e-08)[35] translates to Bonferroni-correcting for 1,000,000 tests[36]. Reducing the number of tests, even to all genes (~20,000), drastically lowers the significance threshold.

A wide variety of statistical methods of aggregating variants have been developed, and can broadly be divided into four main categories: 1) burden tests, 2) variance component tests, 3) combination tests and 4) other tests[37]. In addition, several methods for meta-analysis have been developed based on tests from these four categories.

Burden tests

Burden tests are based on the principle of summarizing genetic information across a region into one score statistic per individual, which can then be used for different methods of association testing[37]. One of the simplest versions of this is the cohort allelic sums test (CAST)[38]. Here, a binary score is generated by checking for the presence of at least one variant in the specified region. For a binary phenotype the results can then be collapsed into a 2x2 table, and a χ^2 test or Fisher's exact test then be used to test for an association. The obvious limitation of this approach is that there is no differentiation between having one and multiple mutations in the region. The combined multivariate and collapsing (CMC) test is similar to CAST, but allows for subgrouping of variants, for example based on allele frequencies[39].

Other burden tests incorporate individual variant weights into the calculation of the genetic score statistic. A common approach is to upweight rare variants, with the assumption that rarer alleles have a larger impact on the phenotype in question than common alleles. The Weighted Sum Test (WST) uses MAF to calculate individual variant weights (w_j)[28,40]:

$$w_j = 1/[MAF_j(1 - MAF_j)]^{1/2}$$

Another weighting method upweights rare variants using beta densities[41]:

$$\sqrt{w_j} = Beta(MAF_j, \alpha_1, \alpha_2)$$

Once the score statistic is generated, different statistical methods of association testing can be applied. For example, Madsen and Browning proposes the nonparametric Wilcoxon rank-sum test for the WST[40], while others use the score statistic in a regression framework[41]. A major drawback with the burden tests is that all variants included are assumed to be causal and have the same effect size[28]. A subclass of burden tests, the so-called adaptive burden tests, attempt to address this issue. For example, the variable threshold (VT)[42] test supposes that there is a value for MAF where variants below are much more likely to be causal, but that the threshold can vary between different genomic regions. The method calculates genetic scores for each region at different MAF thresholds and estimates the optimum threshold by permutation of phenotypes for each genomic region. VT also allows for weighing variants according to their predicted potential for disrupting protein function by using different algorithms, for example PolyPhen2[43] or CADD[44]. Other examples of adaptive burden tests include the data-adaptive sum test (aSum)[45], the estimated regression coefficient test (EREC)[46] and the step-up test[47].

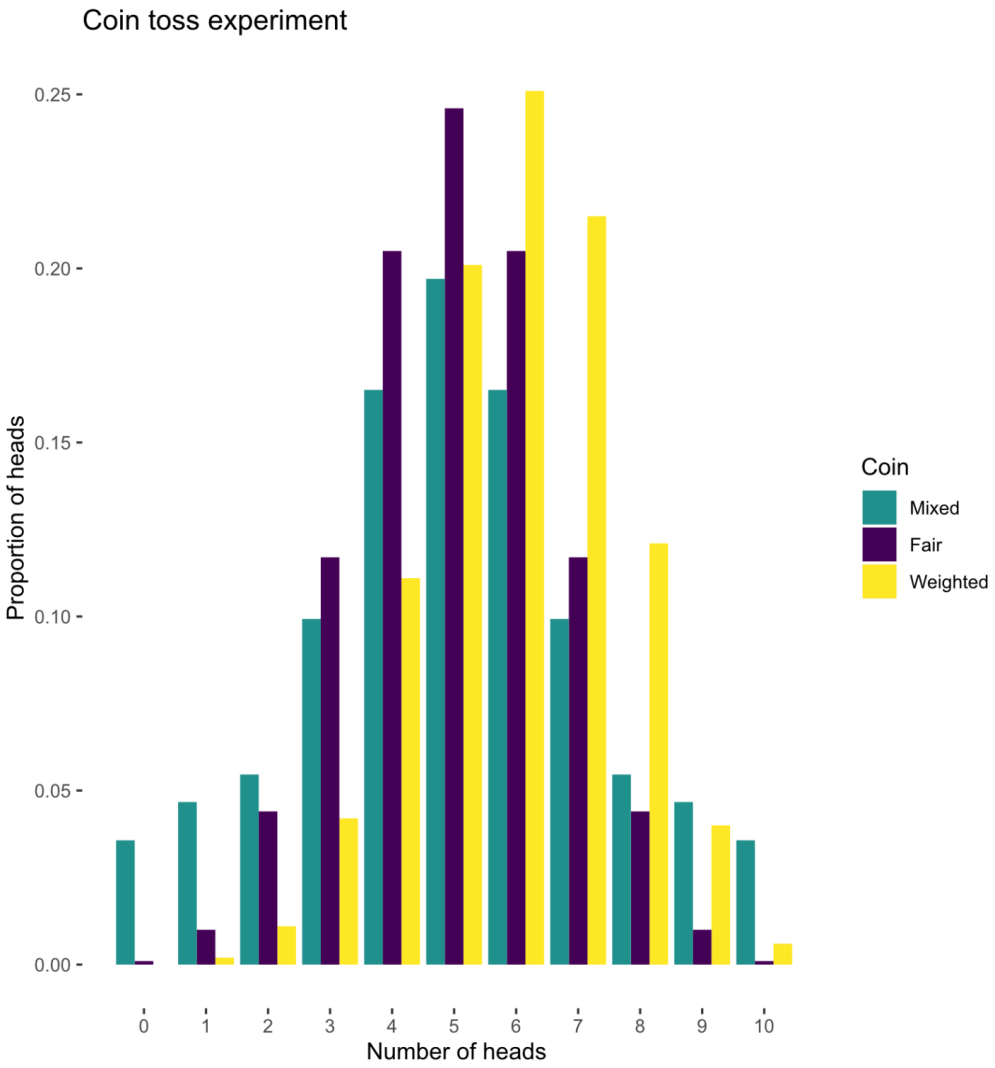

Figure 3. Coin toss experiment. *Using a coin toss experiment to visualize the difference between burden and variance component tests, originally described by Neale et al*[48]*. Using a set of 10 coins, the plot shows the probability (y-axis) of obtaining a given number of heads (x-axis) using different types of coin sets (colors). The "Fair" coin set contains only coins with a probability (P) of P=0.5 of coming up heads, the "Weighted" set contains coins with a P=0.6 of coming up heads, and the "Mixed" set contains a 10:80:10 mixture of P=0.9:P=0.5:P=0.1 of coming up heads. Imagine then that each coin represents one rare variant in a given gene, which is*

either present (heads) or not (tails), and that the number of heads obtained after tossing all 10 coins once represents the number of rare variants for one individual in that gene. When, in a case-control setting, the variants have no correlation with a given phenotype, the number of variants will follow the binomial distribution with P=0.5, as for the "Fair" coin set. Burden tests are designed to detect instances where mean number of variants (heads) is either increased or decreased across the entire gene, like for the "Weighted" coin set (P=0.6). However, variance component tests excel in situations where both protective and risk variants are present, exemplified by the "Mixed" set. Here, the overall probability of coming up heads is the same as for the "Fair" set (P=0.5), but the increased variance of the outcomes (larger probabilities in the tail ends of the distribution) shows that there is likely a few unfair coins (variants) with high probabilities of coming up either heads or tails (risk or protective).

Variance component tests

Variance component tests were developed to address some of the limitations of the burden tests, and are able to accommodate variants with opposing directions of effect and different effect sizes within the same region[37]. While burden tests are able to detect differences in mean across a region, variance component tests look for differences in individual variant variances[48]. For the C-alpha test, the test statistic is calculated by comparing the observed individual variant counts with the expected variance, which within a null hypothesis of no effect will follow the binomial distribution[48]. Another example of a variance component test is the sum of squares (SSU) test[49]. One of the most flexible tests is the sequence-kernel association test (SKAT)[41]. SKAT uses a mixed model approach and can accommodate covariates and variable variant weights. These tests are all statistically related, and under certain conditions (flat weights, binary phenotype and no covariates), the C-alpha test, SSU test and SKAT are statistically equivalent[28,41].

Figure 3 visualizes the main difference between burden and variance component tests through a hypothetical coin toss experiment.

Combination tests

Combination tests attempt to combine burden and variance-component test statistics into one. The rationale behind is that burden tests are generally more powerful when all or most variants are causal, whereas variance-component tests are more powerful when a region contains a mixture of causal, neutral and protective variants[50]. One suggested approach is to use Fisher's method of combining p-values from burden and variance-component tests[51]:

$$Fisher = -2\log(p_{burden}) - 2\log\left(p_{variance-component}\right)$$

Instead of combining p-values, SKAT-O combines tests statistics (Q) from burden and SKAT[52]:

$$Q_p = (1-p)Q_{SKAT} + pQ_{burden}$$

where $0 \leq p \leq 1$, and an optimal value for p is estimated by calculating the minimum p-value across a range of values for p. The mixed effects score test (MiST) is another combination test that uses hierarchical (or mixed effects) modeling, and uses known variant characteristics (for example insertion, deletion, nonsense etc.) in an attempt to increase power[53]. The Q-tests also include a combination test approach, with a special focus on gene sets[54]. A weakness of the combination tests is that if the assumptions for either the burden or variance component tests are largely true, the combination test will be less powerful than either of them[28].

Other tests

This category includes the exponential combination (EC) procedure, where the test statistic is the sum of exponential variant statistics, which improves power when only a few variants in the tested set is associated with the phenotype[55]. A replication-based strategy calculates test statistics for variants enriched in cases and controls separately, and uses permutation to compute p-values from a combined maximum statistic[56]. A

number of tests have also been developed to detect pleiotropic effects, including Multi-SKAT[57] and MULVR/MULVR-O[58]. The IGOF tests are based on Pearson's goodness-of-fit, and were developed specifically to detect gene-gene interaction effects[59].

Meta analyses

As the number of sequencing studies continues to grow, using multiple datasets in a joint analysis has become an important tool in the detection of novel associations. The gold standard is considered to be a joint calling of all available datasets, a so-called mega-analysis, but this is both labor intensive and not always possible due to ethical concerns and privacy restrictions[60]. Meta-analyses are therefore an attractive alternative and has been shown to have comparable power when compared to mega-analyses[61]. General statistical methods of combining p-values across different studies can also be used in genetic association studies, for example Fisher's method[62], Stouffer's method[63] or the weighted Z-test[64]. Instead of combining p-values, several specific tools have been developed that combine test scores, derived from the methods outlined above, across datasets. Examples of tools with available tests in parenthesis include RAREMETAL[65] (CMC, WST using Wilcoxon rank-sum test, VT and SKAT), MetaSKAT[61] (WST, SKAT and SKAT-O) and Meta-Qtest[66] (Q-tests).

1.3 Parkinson disease genetics

Familial clustering of PD has been observed as early as the late 19th century[67], but the first attempt at a systematic study was published by the Swedish psychiatrist Henry Mjönes in 1949[68]. He concluded, erroneously, that PD was a strictly autosomal dominant disease with low penetrance, but even at the time there were doubts regarding the validity of his results[69]. A more thorough exploration of the heritability of PD began with twin studies in the 1980s.

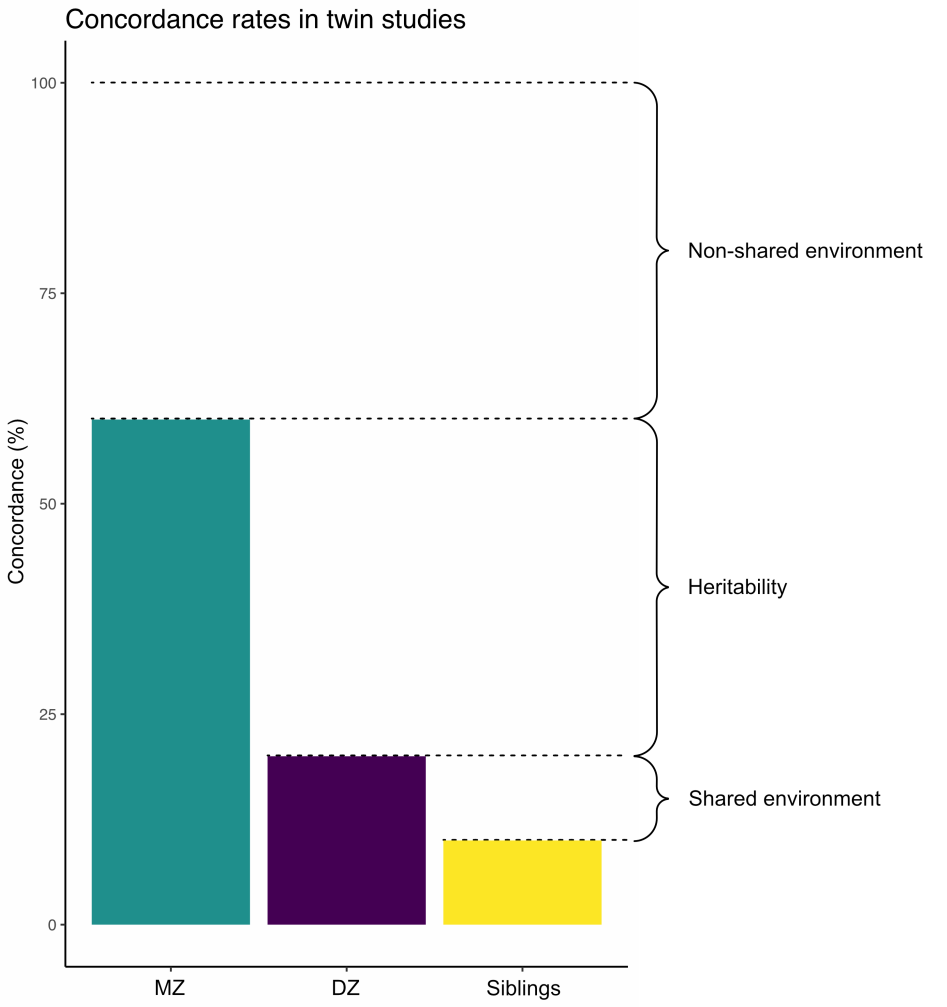

Figure 4. Concordance rates in twin studies. *The figure shows concordance rates for a hypothetical disease. In general, the difference in concordance between MZ and DZ twins is attributed to heritable factors, for example genetic variation. By comparing DZ twins to siblings, the contribution of early environmental factors (shared) can also be differentiated.*

1.3.1 Twin studies

Early twin studies found little evidence for a genetic component of PD in the general population[70-73], but the studies were largely based on cross-sectional data with little or no follow-up. The most recent study found concordance rates of 11% for monozygotic

(MZ) and 4% for dizygotic (DZ) twins, with an estimated heritability of 34%[74]. The considerable difference between the concordance rates of MZ and DZ twins strongly indicates the presence of a heritable factor (see Figure 4). As PD is a strongly age-correlated disease with a long subclinical phase[75], studying clinical concordance rates can lead to underestimates of the true genetic contribution. When using positron emission tomography (PET) to identify and include dopaminergic dysfunction in heritability estimates, concordance rates have been as high as 75% for MZ and 22% in DZ twins[76]. Evidence from twin studies therefore strongly indicate a genetic component for PD.

1.3.2 Monogenic PD

The first time a genetic change was conclusively linked to PD was in 1997 when a mutation in the *SNCA* gene was found to cause an autosomal dominant form of PD[77]. Since then, many more genes, of both autosomal dominant and recessive inheritance, have been identified. Genes robustly associated with PD and/or parkinsonism include *SNCA*, *LRRK2* and *VPS35* causing autosomal dominant disease, and *PRKN*, *PARK7*, *PINK1*, *ATP13A2*, *FBXO7*, *PLA2G6*, *DNAJC6*, *SYNJ1* and *VPS13C* as causes of autosomal recessive disease[78]. The phenotypical presentation varies among the different monogenic forms of PD. Broadly, the autosomal dominant forms (*SNCA*, *LRRK2*, *VPS35*) causes a phenotype fairly similar to that of classical, sporadic PD, with a later age of onset (around 50 years of age) and similar clinical characteristics[79]. Of the recessive genes, *PRKN*, *PARK7* and *PINK1* cause early-onset disease (around 30 years of age), with a clinical phenotype similar to that of classical PD, but where dystonia is typically more prevalent and cognitive decline less frequent[80]. The remaining recessive genes generally cause a variety of atypical forms of PD. *FBXO7* causes a juvenile-onset (<20 years of age) syndrome characterized by equinovarus foot deformity, pyramidal signs and parkinsonism[81,82]. Mutations in *ATP13A2* causes Kufor-Rakeb syndrome with juvenile-onset, parkinsonism, dementia, pyramidal signs and supranuclear upgaze paresis[83,84]. Various phenotypes have been reported for different mutations in *DNAJC6*. The c.801-2A>G splice-cite mutation causes juvenile-onset parkinsonism[85], while the c.2371C>T (p.Gln791*) nonsense mutation have been

reported in a broader syndrome of juvenile-onset parkinsonism, mental retardation and pyramidal signs[86]. Mutations in *SYNJ1* causes juvenile-onset parkinsonism, typically accompanied by seizures and generalized dystonia[87-90], while mutations in *VPS13C* cause early-onset parkinsonism similar to that of classical PD, but with rapid and severe disease progression and cognitive decline[91]. Finally, *PLA2G6* mutations were first associated with infantile neuroaxonal dystrophy and neurodegeneration with brain iron accumulation[92], terminology which were later consolidated into phospholipase-associated neurodegeneration (PLAN). In 2009, mutations in *PLA2G6* were described in patients with a clinical syndrome of early-onset parkinsonism and dystonia, or PLAN-DP, and later studies have confirmed the association[93-95].

Several other genes have been nominated as possible causes of mendelian PD, including *UCHL1*, *GIGYF2*, *HTRA2*, *EIF4G1*, *DNAJC13*, *TMEM230*, *LRP10* and *CHCHD2*[78,96,97]. However, at present, they are either lacking replication or supportive evidence, or have been contradicted by later studies.

In addition, to the genes mentioned above, mutations in a number of other genes are known to cause degeneration of the dopaminergic neurons of the *substantia nigra*, with or without clinical parkinsonism. Mutations in *POLG* or *TWNK* encoding the mtDNA polymerase and helicase, respectively, cause mitochondrial disease with severe loss of the dopaminergic neurons in the *substantia nigra* similar to that seen in PD[98]. Intriguingly, this is not always accompanied by clinical parkinsonism, even in cases with severe nigrostriatal degeneration[99,100]. Perry syndrome is caused by mutations in *DCTN1*, and neuropathological studies shows gliosis and neuronal loss in the *substantia nigra*[101]. Clinically, parkinsonism and psychiatric symptoms are prominent. The spinocerebellar ataxias (SCAs) are a heterogeneous group of dominantly inherited ataxias caused by a mutations in a wide range of genes, with the number of distinct SCAs being close to 50 at present[102]. Degeneration of the *substantia nigra* has been documented for both SCA2 and SCA3, which are caused by CAG-repeat expansions in *ATXN2* and *ATXN3* respectively[103,104]. *Substantia nigra* degeneration is also seen in ataxia-teleangiectasia, an autosomal recessive disease causes by mutations in the *ATM* gene[105]. Finally, nigrostriatal dopaminergic

dysfunction has also been documented in patients with missense mutations in *OPA1*, a gene typically associated with optic atrophy[106].

It's estimated that approx. 5-10% of PD cases worldwide have monogenic causes[107], but this varies widely between different population groups. For example, *LRRK2*-mutations are seen in as many as ~40% of cases among North African Arabs and ~25% of Ashkenazi Jews[108], likely due to founder effects. It is worth noting the high degree of complexity associated with the link between PD and genetic mutations. Many of the genes that cause monogenic PD are in reality causing syndromes where parkinsonism is accompanied by a varying degree of other clinical features. Even the autosomal dominant genes, where the phenotype closely resembles classical PD, are complex in terms of genetic inheritance. Following the classical laws of Mendelian inheritance, genetically inherited diseases are either dominant or recessive, with 50% and 25% disease risk in offspring respectively. However, monogenic forms of PD display highly variable and age-dependent penetrance, suggesting that other factors, either genetic or environmental, act as important modulators[78]. This is particularly noticeable for *LRRK2*-mutations, where the penetrance estimates ranges from 26%-42.5% and the incidence is highly age-dependent, similar to that of sporadic PD[109,110].

Table 1. Overview of relevant PD-associated genes

Category	Inheritance	Phenotype	Genes
Confirmed monogenic PD genes	AD	Classical LOPD	*SNCA, LRRK2, VPS35*
	AR	Classical EOPD	*PRKN, PARK7, PINK1*
		Atypical JOPD/EOPD	*FBXO7, ATP13A2, DNAJC6, SYNJ1, VPS13C, PLA2G6*
Unconfirmed monogenic PD genes	AD	-	*UHCL1, HTRA2, GIGYF2, EIF4G1, DNAJC13, TMEM230, LRP10, CHCHD2*
Genes associated with degeneration of the *substantia nigra*	-		*POLG, DCTN1, ATXN2, ATXN3, ATM, OPA1*

AD: autosomal dominant. AR: autosomal recessive. LOPD: late-onset Parkinson disease, EOPD: early-onset Parkinson disease, JOPD: juvenile-onset Parkinson disease, -: not relevant

1.3.3 Familial aggregation of PD

Familial aggregation of PD has been documented by numerous studies, with a 2008 meta-analysis estimating the relative risk (RR) for PD patients to have a first degree relative with PD to be 2.9[111]. This estimate likely includes cases of monogenic PD, so the true estimate for idiopathic PD could be lower. A more recent analysis of death certificates in Utah found that first degree relatives of individuals who had PD as a cause of death had themselves a RR of 1.82 of death with PD compared to the non-relatives[112]. As mentioned, a problem with these studies is that cases of monogenic PD are likely mixed in with sporadic cases, making it difficult to estimate the genetic contribution to non-mendelian PD from these results. In addition, many of the studies

are also based on data from specialized movement disorder clinics, where PD patients typically have a younger age of onset compared to the general population. Both familial aggregation[113] and monogenic PD[114] has been shown to be more prevalent in these patient groups, which would make the RR estimate less representative for sporadic PD.

1.3.4 Risk variants

The first GWAS of PD was published in 2006, but it and subsequent studies for the next few years yielded no genome-wide statistically significant associations[115,116]. The first GWAS that was able to find significantly associated SNPs was published in 2009, and identified three loci in close proximity to *SNCA*, *MAPT* and *LRRK2*[117]. To date, the strongest genetic risk factor for PD is mutations of the *GBA* gene[118], which were, interestingly, not discovered by GWAS. Suspicion originally arose due to the high prevalence of PD among relatives of patients with Gaucher disease, an autosomal recessive disorder caused by *GBA* mutations[119]. A subsequent analysis found a >5-fold increase in PD risk among mutation carriers[120], and an association between PD and non-coding variants around the *GBA* gene have later been established by GWAS[121]. The most recent and largest GWAS, comprising approx. 37,000 cases, 18,000 proxy cases (1st degree relatives of individuals with PD) and 1.4 million controls, identified 90 independent variants across 78 loci[24]. Some of the variants are tagging genes known to cause monogenic PD, mainly *SNCA* and *LRRK2*, while the majority are near genes not otherwise known to be implicated in PD pathology.

1.3.5 Parkinson disease genetics in the era of next generation sequencing

NGS technologies have been employed in a variety of ways in the study of PD, ranging from diagnostic case-reports sequencing a few individuals, to large association studies with hundreds to thousands of cases. Table 2 contains a chronological overview of genetic association studies in PD where NGS was employed. Studies that focused on discovering mutations underlying mendelian PD are not listed, as in the case of the discovery of causal *VPS35*-mutations[122,123].

The main message from this table is that the efforts to utilize NGS in the discovery of novel genetic contributions to the etiology of PD have, to some extent, been hamstrung by small sample sizes and the lack of a consistent methodological approach. Over the years, the methodologies of GWAS have gradually matured and consolidated, offering researchers a fairly clear-cut path[124]. In contrast, the variant selection process and statistical approaches utilized in the NGS studies are much more complex, and there is generally no clear consensus on how to best perform these studies. Even so, some guidelines and recommendations have been suggested. In 2012, Do et al.[125] recommended including, at minimum, the following analyses as a baseline for exome sequencing studies in complex diseases:

1) Whole-exome single variant association (SVA) analysis
2) Two types of burden analyses, where rare variants are grouped together (for example within each gene) to increase statistical power:
    a. A traditional burden analysis where variants are assumed to have the same effect size and direction of effect
    b. A burden test that allows for opposite directions of effect within the same group
3) Optionally, perform a restricted analysis of a subset of rare variants predicted to have a large impact on protein function (for example nonsense mutations, or variants predicted to be damaging by prediction algorithms)

Very few of the studies listed in Table 2 follow these guidelines. This is likely due to the fact that larger sample sizes than what is currently available are probably needed to detect rare variant gene-enrichment on an exome wide scale. Studies have therefore focused on specific parts of the genome, either implicated by previous studies or *a priori* hypotheses. Examples include analyzing only genes implicated by GWAS[126] or specific genes with suspected links to PD[127]. This reduces the need for multiple testing correction, thus increasing power, but also limits the overall scope of the study.

The most promising results from NGS-based genetic association studies in PD have thus far come from pathway-based analyses. Pathways are groups of genes that

encode proteins that share some predefined trait or function. This can for example be proteins with subcellular co-localization (e.g. endoplasmic reticulum or mitochondria), proteins with similar mechanisms of signal transmission (e.g. G-protein coupled receptors), or proteins that together perform a specific function (e.g. DNA repair). There are several different databases of pathways available, each with different curating strategies. The most extensive is the Gene Ontology (GO) database, which currently contain >20,000 human pathways divided into three categories: biological process, cellular component and molecular function[128,129]. The pathways are generated both through manual curation and computational algorithms. Other databases include the Kyoto Encyclopedia of Genes and Genomes (KEGG)[130], Reactome[131] and Biocarta[132], all of which rely more heavily on manual curation and annotation than GO.

Pathway analyses can take many forms. One of the most straight forward methods is to take the *n* number of genes with the lowest p-values from a single gene analysis (SGA), and, through statistical analysis, identify pathways that contain a higher number of these genes than expected[133]. Sandor et al[134] used this method to tentatively identify pathways related to extracellular matrix proteins as being enriched with genes from their list of top 300 genes from their single gene associations (SGA). Another method of pathway analysis is to expand the collapsing methods of single gene analyses and consider the pathway as a "mega-gene". Using this method, Robak et al[135] showed an enrichment of rare mutations in genes linked to lysosomal storage disorders , and we have shown a similar enrichment in mitochondrial pathways[136].

Table 2. Chronological overview of genetic association studies in PD using next generation sequencing

Study	Sequencing	Sample type	Sample size* (case/control)	Analyses	Variant focus	Brief summary of analyses and results
Nuytemans et al, 2013[127]	WES	Sporadic PD**	213/272	Targeted SVA/SGA	All	Targeted analysis of *VPS35* and *EIF4G1*. Supportive evidence for one *EIF4G1*-mutation. No statistically significant SGA
Foo et al, 2014[137]	WES	EOPD	375/399	SVA/SGA GWA-targets	Missense	39 genes implicated by previous GWAS were examined. Statistically significant enrichment of missense variants in *LRRK2*
Mencacci et al, 2014[138]	WES	Sporadic PD/EOPD/Familial PD	1318/5935	Targeted SVA	Disruptive	Targeted analysis of *GCH1* variants. Statistically significant difference in number of disruptive mutations in *GCH1* in PD compared to control populations
Quadri et al, 2015[139]	WES	Sporadic PD	100/-***	SVA	All	No statistically significant associations
Petersen et al, 2015[140]	WES	Sporadic PD	91/96	SVA	All	No statistically significant associations
Chen et al, 2015[141]	WES	Familial PD/EOPD	350/350	Targeted SVA	All	Targeted analysis of *SYNJ1*. No statistically significant results
Simon-Sanchez et al, 2015[142]	WES	Sporadic PD/EOPD	1189/469	Targeted SVA/SGA	All/Disruptive	Targeted analysis of *PARK10*. No statistically significant results
Farlow et al, 2016[143]	WES	Familial PD	93/- (discovery) 49/- (replication)	SVA	PtbD	Two genes, *TNK2* and *TNR*, were found to contain rare, likely deleterious variants in both the discovery and replication dataset
Nuytemans et al, 2016[144]	WES	Sporadic PD	396/222 444/153****	Targeted SVA	PtbD/LoF	Targeted analysis of *ABCA7*. No statistically significant results
Lubbe et al, 2016[145]	WES	Sporadic PD/EOPD	1255/473	Targeted SVA/SGA	All rare	Targeted analysis of genes involved in cutaneous malignant melanoma. The primary dataset was a chip-genotyped dataset, WES was used as replication. No statistically significant results
Sandor et al, 2017[134]	WES	Sporadic PD	228/884	SVA/SGA GWA-targets	Missense	Tentative evidence for the extracellular matrix pathway. No statistically significant SGA. Restricting analysis to variants within GWA-

				Top-hit pathway enrichment		nominated regions yielded tentative evidence for a variant in *RAD51B*
Jansen et al, 2017[146]	WES	Sporadic PD/EOPD	1148/503	SVA	LoF	The study identified 27 genes with homozygous or compound heterozygous LoF mutations in the PD group. Five genes, *GPATCH2L*, *UHRF1BP1L*, *PTPRH*, *ARSB* and *VPS13C*, were supported by functional and/or genetic replication
Siitonen et al, 2017[147]	WES	EOPD	185/440	SVA/SGA	All	Genome-wide statistically significant enrichment of variants in *MPHOSPH10*, *TAS2R19* and *SERPINA1*. Tentative evidence for single variants in *CEL*
Jansen et al, 2017[126]	WES	Sporadic PD/EOPD	1167/1685	GWA-targets	All/Disruptive/PtbD	Statistically significant rare-variant enrichment in *STBD1* for disruptive variants
Giri et al, 2017[148]	WES	Sporadic PD Monogenic PD	1450/2267 86/10	Targeted SVA/SGA	All	Targeted analysis of *TMEM230*. No statistically significant results
Robak et al, 2017[135]	WES	Sporadic PD/EOPD	1156/1679 (main) 436/139 (replication)	Pathway	Missense/PtbD/LoF	Targeted analysis of genes linked to lysosomal storage diseases as a single pathway. Statistically significant enrichment of missense/PtbD variants in the main dataset, not replicated in the WES replication cohort. The results did replicate in a third, chip-genotyped dataset
Ylönen et al, 2017[149]	WES	EOPD	225/-	Targeted SVA	All	Targeted analysis of PD risk loci. The study identified 2 variants in SMPD1 and 3 in LRRK2 that were present in the cases and not in the publicly available Finnish ExAC[150] population
Gaare et al, 2018[136]	WES	Sporadic PD	192/219 459/181****	SVA/SGA Pathway	Missense/PtbD	Statistically significant enrichment of missense mutations in the mtDNA homeostasis pathway
Guo et al, 2018[151]	WES	Familial PD	39/98	SVA	De novo	Tentative evidence for *NUS1* as a possible candidate gene

Blauwendraat et al, 2018	WES	Sporadic PD	2440/5774	Targeted SVA/SGA	LoF	Targeted analysis of LoF variants in *LRRK1* and *LRRK2*. No significant association between LoF variants and PD were detected.
Chew et al, 2019[152]	WES	EOPD	99/99	Targeted SVA/SGA	Missense/LoF	Targeted analysis of 33 candidate genes in Chinese individuals. No associations detected.
Siitonen et al, 2019[153]	WES	EOPD	439/855 (main) 60/8214 (replication)	Targeted SVA/PRS	All	Targeted analysis of 2305 genes derived from a protein-protein network analysis. PRS-analysis of candidate variants rs2230288 and rs2291312 from the main dataset suggest an association with PD in the replication dataset.
Rudakou et al, 2020[154]	Targeted sequencing	Sporadic PD	1567/1667	Targeted SVA/SGA	Rare/PtbD	Targeted sequencing of *VPS13C*. No association detected.
Gialluisi et al, 2020[155]	WES	Familial and sporadic PD	123/-	SVA/PRS	Common	Case-only analysis of the effect of common variants on endophenotypes. No statistically significant associations for single variants. Association between some PRS affecting subcortical brain volume and motor symptoms.
Oluwole et al, 2020[156]	Targeted sequencing	Sporadic PD	47/-	Identification of deleterious mutations	PtbD	The study explores possible deleterious mutations in a targeted sequencing of 751 genes with neurological relevance in sub-Saharan PD patients.
Bobbili et al, 2020[157]	WES	Sporadic PD	340/146	Exome-wide burden	Singleton LoF	Significant exome-wide burden of singleton LoF variants in PD
Gaare et al, 2020[158]	WES	Sporadic PD	192/219 459/181****	GWA-targets	Disruptive/LoF	No evidence for rare variant enrichment in GWAS-identified candidate genes

SVA: single variant association. SGA: single gene association. PRS: polygenic risk score. LoF: loss-of-function. PtbD: predicted-to-be-damaging variants. Top-hit enrichment: using the n genes with the lowest p-values in a PANTHER[133] enrichment analysis, where the gene list is screened against a database of gene pathways to look for overrepresentation. GWA-targets: focused enrichment analysis in regions/genes implicated by GWAS. Disruptive: all variants except synonymous, i.e. missense, nonsense and splicing mutations. Predicted to be damaging variants: Variants identified by a prediction algorithm as having a high likelihood of being damaging to the encoded protein.

**only WES datasets listed. Many studies use the same or heavily overlapping datasets. Some studies included chip-genotyped replication datasets, which are not listed in the table.*

***of the 213 cases, 25 were related, and could be considered cases of familial PD.*

****WES was performed on 100 cases. Functional variants that were not present in public databases and shared by ≥5 PD cases were subsequently genotyped in 500 control individuals.*

*****parallel analysis of two separate WES-datasets*

1.4 Pathophysiological mechanisms in PD

The motor symptoms that characterize PD are caused by the loss of dopaminergic neurons in the *substantia nigra pars compacta* (SNc)[2]. These neurons have projections to the basal ganglia, including the striatum, and exhibit an autonomous pacemaker function essential for the coordination of movement[159]. Parkinsonism becomes clinically apparent when at least 30% of the dopaminergic neurons in the SNc are lost[160]. Beyond the SNc, neuronal loss is also seen in the pedunculopontine nucleus, *locus coeruleus*, dorsal motor nucleus of the vagus, raphe nuclei, *nucleus basalis* of Meynert, ventral tegmental area, thalamus, hypothalamus, olfactory bulb, and the enteric nervous system[161,162]. The most consistent neuropathological finding in PD are Lewy bodies and Lewy neurites, collectively referred to as Lewy pathology. These neuronal inclusions were first described by Fritz Heinrich Lewy in 1912, and later named after him[163]. After the discovery of *SNCA*-mutations as a cause of monogenic PD in 1997, alpha-synuclein was identified as the main component Lewy pathology[164,165]. Under normal conditions, alpha-synuclein exists as both soluble monomers and multimers in the cytosol, predominantly in presynaptic terminals[166,167]. At some point, due to mechanisms that are currently largely unknown, alpha-synuclein becomes misfolded into a beta-pleated sheet structure, and goes on to form fibrils that later aggregate in Lewy bodies and Lewy neurites[168]. Although it is recognized as a neuropathological hallmark of PD, Lewy pathology is itself not exclusive to PD. It has been found in multiple disorders, including dementia with Lewy bodies (DLB), multiple system atrophy (MSA), Gaucher disease, Alzheimer's disease and several lysosomal storage disorders[169]. In addition to alpha-synuclein, aggregates of tau and beta-amyloid can also be found in the brains of PD patients[2]. The predominant protein deposition can be used to classify the parkinsonian disorders into alpha-synucleinopathies (PD, MSA) and tauopathies (progressive supranuclear palsy, corticobasal degeneration)[170].

In 2003, Braak et al described a spreading pattern of Lewy pathology throughout the brain that correlated with increasing severity of clinical symptoms[171]. The proposed staging system divides the distribution of Lewy pathology into six stages, where the initial affected areas are located in the brainstem (stages 1-2,

preclinical disease), and later affects the midbrain (stages 3-4, early clinical disease) and neocortex (stages 5-6, late clinical disease)[172]. Since early sites of Lewy pathology were the olfactory bulb and enteric plexus of the stomach, the Braak hypothesis postulates that some unknown pathogen enters the body through the nasal or gastric cavities and triggers alpha-synuclein aggregation which then spreads throughout the nervous system[173]. Both this hypothesis and the staging system have been subject to criticism. In as much as 20-50% of patients, Lewy pathology does not appear to follow the proposed spreading pattern[174,175]. Moreover, the distribution of Lewy pathology does not fully correlate with neuronal cell death[176,177]. Selective neuronal vulnerability of the affected neuronal populations in PD has been suggested as a possible explanation for the spreading pattern of Lewy pathology and neurodegeneration[178]. Irrespective of the nature of the spreading pattern or the initial cause of alpha-synuclein misfolding, there is evidence suggesting that misfolded alpha-synuclein can propagate in a prion-like fashion. For example, grafted nigral neurons in the striatum of PD patients has been found to be increasingly affected by Lewy pathology over time[179,180], and introducing misfolded alpha-synuclein to healthy neurons have triggered the formation of Lewy pathology in mouse models[181,182]. However, there is ongoing debate whether alpha-synuclein aggregation is the true driving force behind the neurodegeneration in PD, or rather an epiphenomenon[183]. Regardless, Lewy pathology, together with neuronal degeneration in the SNc, remains the defining neuropathological characteristic of idiopathic PD, and is present in some, but not all cases of monogenic PD, (Table 3).

Table 3. Neuropathological findings in genetic PD and genes associated with degeneration of the *substantia nigra*

Group	Gene	No. autopsies	Neuropathological findings
Autosomal dominant PD	*SNCA*	27	Lewy pathology present in all (27/27) and tau pathology present in many autopsied cases (17/25)
	LRRK2	55	Lewy (34/55) and tau (32/54) pathology present in most autopsied cases
	VPS35	-[a]	No data[a]
Autosomal recessive PD	*PRKN*	22	Lewy (6/19) and tau (7/16) pathology absent in most autopsied cases
	PARK7	1(2[b])	Lewy pathology present
	PINK1	1	Lewy pathology present, tau pathology not present
Autosomal recessive atypical PD	*FBXO7*	-	No data
	ATP13A2	-[c]	No data[c]
	DNAJC6	-	No data
	SYNJ1	-	No data
	VPS13C	1	Lewy pathology present
	PLA2G6	10	Lewy pathology present in all (10/10) and tau pathology present in most (7/10) cases. Brain iron accumulation in most (6/10) cases.
Degeneration of the *substantia nigra*	*POLG*	24	Lewy pathology absent in most (2/24) cases. Tau pathology absent in all cases (0/19).
	TWNK	2	No Lewy or tau pathology (0/2)
	DCTN1	8	Lewy and tau pathology absent in all cases
	ATXN2/ATXN3	23	Lewy pathology absent (2/23) in most autopsied cases
	ATM	33	Lewy and tau pathology reported in only a few, older cases
	OPA1	-	No data

The table is based on a comprehensive review by Schneider and Alcalay[184], and updated with additional publications for SNCA[185,186], LRRK2[187], PRKN[188], ATXN2/ATXN3[104], POLG[189-191], TWNK[191], DCTN1[101,192,193] and ATM[194].

[a]One autopsy of a patient who were later found to harbor a VPS35 mutation has been performed, but without crucial regions such as the substantia nigra, locus coeruleus and the brainstem. No Lewy pathology was found in the available samples of the cortex and basal ganglia[195].

[b]One case only underwent skin biopsy. Peripheral alpha-synuclein depositions were detected[196].

[c]One autopsy of a patient with a ATP13A2-mutation has been performed, but the patient presented with neuronal ceroid lipofuscinosis rather than Kufor Rakeb disease with parkinsonism. No Lewy pathology was found[197].

1.4.1 Pathophysiological clues from monogenic PD

Mutations in *SNCA*, encoding alpha-synuclein, was the first identified cause of monogenic PD[77]. Beyond single point mutations, multiplications of *SNCA* also cause PD, with duplications[198], triplications[199] and quadruplications[200] having been described thus far. Phenotypically, point mutations are varied in their presentation, while multiplications tend to cause earlier age of onset and a more rapid disease progression with increasing *SNCA* copy number[79]. This seemingly dose-dependent relationship, as well as the presence of alpha-synuclein containing Lewy pathology in several of the monogenic forms of PD (see Table 3), emphasizes the role of alpha-synuclein in PD pathology. In healthy neurons, alpha-synuclein is predominantly located in the presynaptic terminals, and is believed to regulate synaptic function[201]. Under normal conditions, there is an equilibrium between folded monomeric alpha-synuclein and multimeric, primarily tetrameric, alpha-synuclein[167]. The point mutations associated with monogenic PD have all been found to likely disrupt the folding of alpha-synuclein, shift the balance towards unfolded monomers and induce aggregation[167]. As to how this, in turn, leads to neurodegeneration is the subject of ongoing debate[202]. Mutations and increased expression of alpha-synuclein have been found to affect mitochondria in multiple ways, possibly causing mitochondrial fragmentation, mtDNA damage, increased reactive oxygen species (ROS) production, and impaired respiratory chain function[203]. Reversely, mitochondrial dysfunction can itself lead to aggregation of alpha-synuclein[204]. A recent discovery provides a possible mechanism, where mitochondria seem to play a key role in the degradation of aggregation-prone cytosolic proteins by import into the mitochondria and subsequent destruction by mitochondrial proteases[205].

The disease-causing mutations in *LRRK2* are all located in or near the two core enzymatic domains, and have been found to either increase kinase or decrease GTPase activity[206]. Given that these are relatively broad enzymatic functions, the normal function of LRRK2 has been linked to a wide range of cellular processes, including cytoskeletal dynamics, autophagy, neuroinflammation, vesicle dynamics and mitochondrial function[207]. Regarding its role in pathology, *LRRK2* mutations have

been associated with increased mitochondrial oxidative stress, as well as altered fission/fusion, mitophagy and mitochondrial trafficking[208].

VPS35 encodes a key component of the retromer, and its main function is believed to be the sorting and retrograde transport of proteins from the endosome to either the plasma membrane or the *trans*-Golgi network[209]. It remains unclear how the mutations causing monogenic PD impact the function of VPS35, as both overexpression and knock-out models have been shown to cause neurodegeneration[210,211]. The underlying mechanisms remain unclear, but defective autophagy, disrupted synaptic transmission (through reduced recycling of the AMPA receptor GluR1) and impaired mitochondrial fission/fusion have been suggested[209].

Multiple studies have described a causal relationship between *CHCHD2*-mutations and familial PD[96,97]. The *CHCHD*-genes encode proteins that are mainly located in mitochondria[212]. In addition to the link between *CHCHD2* and PD, *CHCHD10* has also been linked to neurodegenerative diseases, specifically ALS and frontotemporal dementia, possibly by causing mitochondrial dysfuction[213]. CHCHD2 acts as a transcription factor for proteins in the respiratory chain, especially during periods of hypoxia[214]. PD-associated mutations have been shown to cause both fragmentation of the mitochondrial reticulum and reduced oxidative phosphorylation in the respiratory chain[215], as well as aggregation of alpha-synuclein[216].

PINK1, parkin (*PRKN*) and FBXO7 are crucial modulators of mitophagy. In healthy mitochondria, PINK1 is continuously being imported into the mitochondria and degraded. Damaged mitochondria with depolarized membranes lose this ability, and PINK1 accumulates on the mitochondrial surface. Here, it phosphorylates and activates parkin, which in turn recruits ubiquitin and flags damaged mitochondria for mitophagy[217]. The PINK1/parkin-system has also been found to interact with other monogenic PD genes. Deficiency of DJ-1, encoded by *PARK7*, compromises mitochondrial function in a similar fashion to impairment of PINK1/parkin, and is believed to act as part of a parallel system of maintaining mitochondrial homeostasis where defects in one system can seemingly be rescued by up-regulation of the other[218,219]. Increased LRRK2 kinase activity, as caused by the most frequent PD-associated mutation (G2019S), has also been found to inhibit PINK1/parkin-dependent

mitophagy by interfering with mitochondrial fission[220]. Reversely, *VPS13C*-mutations reduces mitochondrial membrane potential, and increases PINK1/parkin-induced mitophagy[91]. Finally, alpha-synuclein has been found to disrupt mitochondrial protein import, possibly affecting clearance of PINK1 from the mitochondrial membrane[221].

ATP13A2 is a cation transporter, and contributes to zinc cation homeostasis in the cell[222]. Disruption of ATP13A2 has been shown to decrease mitochondrial membrane potential through accumulation of intra-mitochondrial Zn^{2+}, resulting in mitochondrial fragmentation and ATP depletion[223]. A link between ATP13A2 and alpha-synuclein has also been suggested, where downregulated ATP13A2 function causes lysosomal dysfunction and reduced alpha-synuclein clearance from the cell[224].

DNAJC6, encoding auxillin, and *SYNJ1*, encoding synaptojanin-1, both have functions important for synaptic vesicle endocytosis, i.e. recycling of vesicles in the synaptic terminal after neurotransmitter release. Auxillin facilitates recycling by stripping the vesicles of clathrin in cooperation with synaptojanin-1[225]. Finally, *PLA2G6* has been linked to multiple previously discussed PD-associated cellular processes. It may affect the function of the retromer by binding to, among other, VPS35, and loss of PLA2G6 has been shown to disrupt mitochondrial function by causing respiratory chain dysfunction, abnormal morphology and impaired mitophagy[226-228].

A common thread linking the various genes causing monogenic PD is that many encode proteins that are important for mitochondrial function. This is further emphasized when considering the genes where mutations are known to cause degeneration of the *substantia nigra* (Table 1). *POLG* and *TWNK* encode proteins in the mtDNA replisome, crucial for repairing and replicating mtDNA, while *OPA1* is important for mitochondrial fusion and cristae organization[229]. Table 4 gives an overview of the PD-associated genes, including protein localization within the cell and theorized function under normal conditions.

Table 4. Localization and function of proteins encoded by genes associated with monogenic PD and degeneration of the *substantia nigra*

Gene	Protein	Subcellular localization	Protein function
SNCA	Alpha-synuclein	Nucleus, cytoplasm, cell membrane, synaptic terminal	Vesicle trafficking and neurotransmitter release
LRRK2	Leucine-rich repeat serine/threonine-protein kinase 2	Mitochondrion, Golgi, endoplasmatic reticulum, lysosome, endosome, cytoplasmic vesicle, perikaryon, axon, dendrite, synaptic vesicle membrane	Involved in neuronal plasticity, vesicle trafficking and autophagy through protein phosphorylation
VPS35	Vacuolar protein sorting-associated protein 35	Endosome, cytoplasm, cell membrane	Component of the retromer cargo-selective complex involved in endosome-to-Golgi retrieval of membrane proteins
CHCHD2	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2	Nucleus, mitochondrion	Activates transcription of complex IV of the respiratory chain, especially during periods of hypoxia
PRKN	E3 ubiquitin-protein ligase parkin	Nucleus, mitochondrion, endoplasmatic recitulum, cytosol	Catalyzes ubiquitination of proteins, promotes mitophagy after activation and translocation to the mitochondrial membrane
PARK7	Protein/nucleic acid deglycase DJ-1	Mitochondrion, cell membrane, nucleus, cytoplasm	Repairs glycated proteins as well as glycated guanine, both in the free nucleotide pool and in RNA/DNA
PINK1	Serine/threonine-protein kinase PINK1	Mitochondrion, cytosol	Protection against mitochondrial dysfunction, promotes mitophagy through activation/translocation of *parkin* to damaged mitochondria
FBXO7	F-box only protein 7	Cytosol, mitochondrion, nucleus	Mediates ubiquitination of proteins, contributes to recruitment of parkin to damaged mitochondria
ATP13A2	Cation-transporting ATPase 13A2	Lysosome, cell membrane	Intracellular cation homeostasis, maintenance of mitochondria, lysosomes and neuronal integrity
DNAJC6	Putative tyrosine-protein phosphatase auxilin	Cytosol, cytoplasm	Promotes uncoating of clathrin-coated vesicles, involved in clathrin-mediated endocytosis in neurons
SYNJ1	Synaptojanin-1	Perinuclear region	Involved in clathrin-mediated enrocytosis, rearrangement of actin filaments
VPS13C	Vacuolar protein sorting-associated protein 13C	Mitochondrion	Maintenance of mitochondrial transmembrane potential, regulates PINK1/parkin-mediated mitophagy
PLA2G6	85/88kDa calcium-independent phospholipase A2	Cell membrane, cytoplasm	Catalyzes release of fatty acids from phosopholipids
POLG	DNA polymerase gamma	Mitochondrion	Replication of mtDNA
TWNK	Twinkle protein	Mitochondrion	Replication of mtDNA
DCTN1	Dynactin subunit 1	Nucleus, cytoskeleton, cytoplasm,	Promotes dynein-mediated retrograde transport of vesicles and organelles by recruitment of dynein to microtubuli

ATXN2	Ataxin-2	Cytoplasm	Involved in trafficking of endoplasmatic growth factor receptor
ATXN3	Ataxin-3	Nucleus	Deubiquitination enzyme, regulates autophagy, regulates transcription through histone binding
ATM	Serine-protein kinase ATM	Nucleus, cytoplasmic vesicle	DNA damage sensor by detecting double stranded DNA breaks
OPA1	Dynamin-like 12kDa protein	Mitochondrion	Regulator of mitochondrial fission/fusion

The table is based on information obtained from UniProtKB[230], accessed October 14, 2019.

1.4.2 Mitochondrial involvement in PD

In the summer of 1976, Barry Kidston, a 23-year old college student, successfully synthesized 4-propyloxy-4-phenyl-N-methylpiperidine, a recreational drug with a supposedly opiate-like "high". After self-administering the drug both intravenously and intramuscularly, and producing several additional batches of the drug, he eventually synthesized what he later referred to as a "sloppy batch" using reduced reaction times and higher temperatures in November of 1976. After injecting this batch, he quickly developed severe levodopa-responsive parkinsonism, with rigidity, tremor, a flat facial expression and muteness[231]. After succumbing to a cocaine overdose 18 months later, an autopsy revealed degeneration of the *substantia nigra*, and Lewy body-like pathology[231]. A few years later, in 1982, neurologist William Langston discovered a total of 7 similarly afflicted individuals, all having injected a type of "synthetic heroin"[232]. After obtaining samples from local dealers and police raids, a batch was eventually discovered that consisted almost entirely of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a byproduct in the synthesis of the pethidine-analogue 1-methyl-4-phenyl-4-propionoxypiperidine (MPPP)[232,233]. While MPTP is itself not toxic, it crosses the blood-brain-barrier and is converted by monoamine oxidase B to the toxic metabolite 1-methyl-4-phenylpyridinium (MPP+)[234]. MPP+ is taken up in mitochondria, and acts as a complex I inhibitor in the respiratory chain[235]. Deficiency of complex I was ultimately documented in the *substantia nigra* of idiopathic PD patients, linking mitochondrial function to the pathogenesis of PD[236]. Subsequent studies of other substances that inhibit complex I, e.g. rotenone, have revealed consistent nigrostriatal degeneration and Lewy pathology formation in animal models[237,238]. Complex I is an essential component of the mitochondrial respiratory chain, where it generates electrons by oxidizing NADH from the Krebs cycle and glycolysis to NAD+. Complex I is thus essential for ATP production and redox balance in the cell[239,240]. Dopaminergic neurons, with their long, branching, unmyelinated axons and pacemaker firing-pattern, have particularly high energy demands[241], which may explain their high vulnerability to impaired mitochondrial function and energy failure. Indeed, a long and highly branched axon

with a large number of synapses seem to be a common trait connecting most affected neuronal populations in PD[178].

Mitochondria contain their own genome, mtDNA, which encodes 13 peptide subunits of the respiratory chain, as well as the tRNAs and rRNAs required for their transcription and translation (Figure 5). In contrast with the diploid nucleus, each mitochondrion contains multiple copies of mtDNA[242]. The total number of copies in each cell ranges from a few hundred to over 100,000 and correlates with the number of mitochondria, which again is reflective of the energy demand of the cell[243-245].

Replication of mtDNA is controlled by several proteins encoded by nuclear DNA (see Figure 6). Mitochondrial transcription factor A (TFAM) is a key regulator of both mtDNA transcription and replication, as it attaches to mtDNA and distorts the structure in order to enable binding of additional enzymes[246]. The core proteins responsible for mtDNA replication are Twinkle, DNA polymerase gamma (POLG), mitochondrial RNA polymerase (POLRMT) and mitochondrial single stranded DNA binding protein (mtSSB)[247]. Twinkle is a DNA helicase that separate the strands of the mtDNA, initiating binding and mtDNA replication by POLG. POLRMT is responsible for synthesizing RNA primers necessary for replication of the lagging strand[247], and mtSSB, encoded by the gene *SSBP1*, binds to single stranded mtDNA and protects it from damage during the replication process[248].

Repair of damaged mtDNA encompasses many of the same processes utilized in the repair of nuclear DNA. Single-stranded mtDNA damage can be corrected by way of base excision repair and mismatch repair, while nucleotide excision repair, important for nuclear DNA repair, has not been conclusively shown in mitochondria[249,250]. The main mechanisms of double stranded DNA breaks are non-homologous end joining and homologous recombination, but there is conflicting evidence for whether these processes are occurring in mitochondria[249,250].

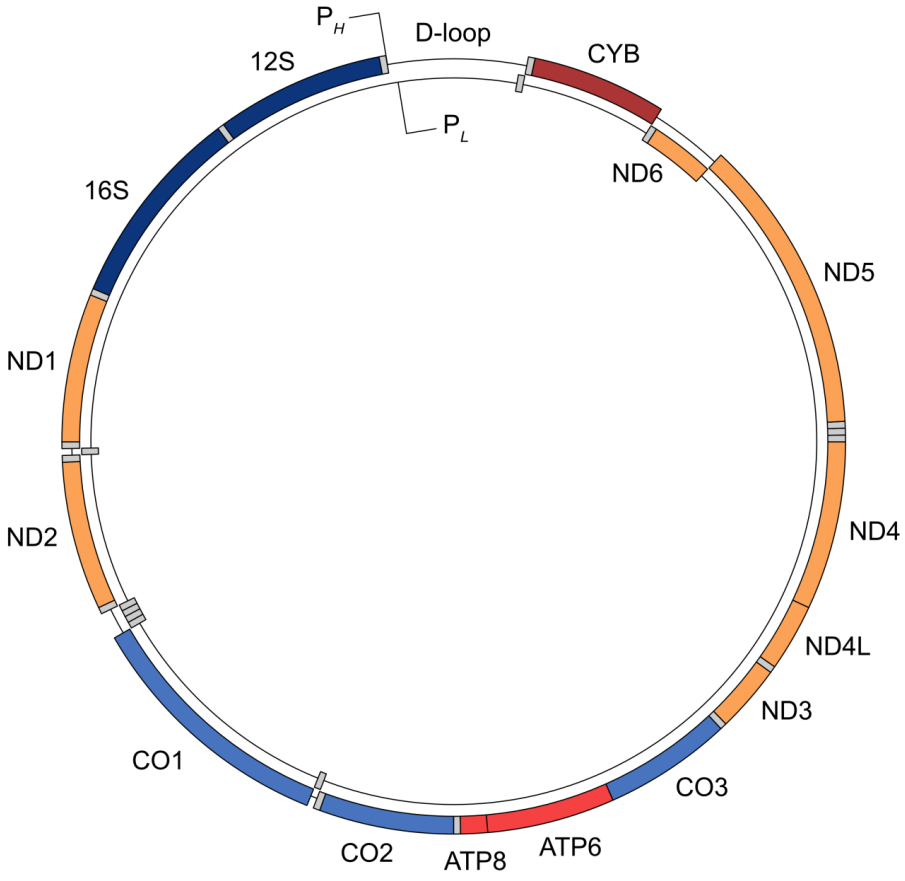

Figure 5. Mitochondrial DNA. *mtDNA encodes seven subunits of complex I (ND1-6 and ND4L), one subunit of complex III (CYB), three subunits of complex IV (CO1-3), and two subunits of complex V (ATP6 and ATP8). In addition, it also encodes transfer RNA (gray) and ribosomal RNA (dark blue). Most are encoded by the heavy strand (outer ring), but ND6 and some transfer RNAs are encoded by the light strand. P_H and P_L indicate the promotor regions of the heavy and light strand respectively. The figure is based on data from Chocron et al*[251] *and Mitomap*[252].

mtDNA is particularly susceptible to oxidative damage in the form of both point mutations and deletions, as it is physically located near the ROS-generating respiratory chain[253]. Both accumulation of deletions and point mutations can lead to

impaired respiratory chain function[254,255]. Deletions has been shown to accumulate with age in multiple tissues, including the *substantia nigra* of both normally aged individuals and subjects with PD[256]. Increasing mtDNA copy number has been suggested as a compensatory mechanism to maintain a healthy supply of wild type mtDNA, ensuring that the cells energy demands are met. This increase in copy number has been documented in healthy subjects but seems to be deficient in *substantia nigra* neurons from PD cases[257]. In addition, an increased load of point mutations has also been described in early stage PD neurons[258]. Maintaining a healthy pool of wild type mtDNA is reliant on a multitude of biological processes, including base excision repair, mitochondrial biogenesis, mitochondrial dynamics (fission/fusion) and mitophagy[259]. Many monogenic PD genes encode proteins that serve important functions for all of these processes.

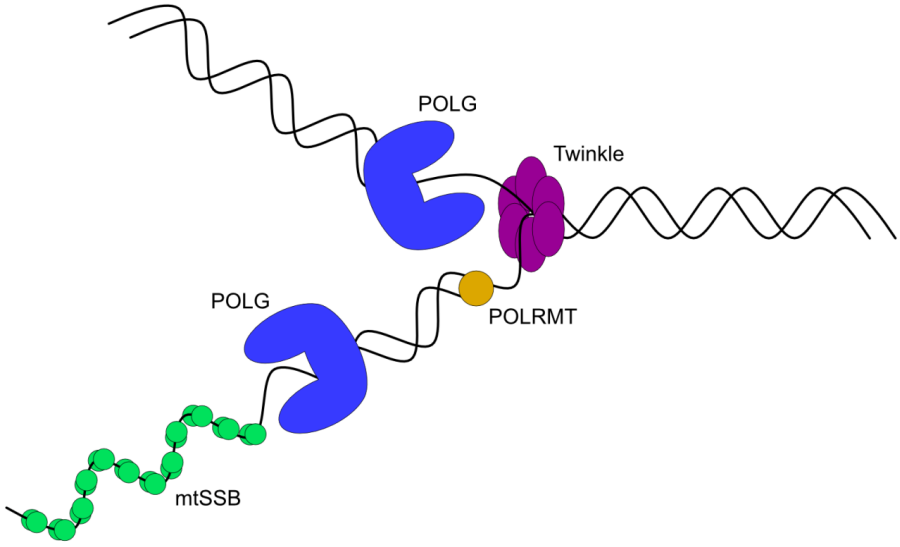

Figure 6. mtDNA replisome. *Twinkle separates the strands of the mtDNA, triggering mtDNA replication by POLG. On the lagging strand, POLRMT synthesizes primers to initiate replication. mtSSB attaches to and stabilizes single stranded mtDNA.*

2. Aims

Familial aggregation of PD has been observed in multiple cohorts, but generalizing risk estimates is complicated by the possibility undetected monogenic causes and the proportion of early onset cases. GWAS have identified a large number of risk variants for PD, but combined they only explain a small portion of the estimated heritability. Many of the genes responsible for monogenic PD encode proteins that are important for mitochondrial function, and mitochondrial dysfunction has been shown in idiopathic cases. Whether this dysfunction is influenced by genetic variation is not known. Finally, the variants identified by GWAS are typically located in non-coding regions. Identifying the functional impact of these variants is important in order to better understand the pathophysiology of PD.

Paper I: The aim of this study was to characterize the familial aggregation of idiopathic PD in a cohort where known monogenic causes were excluded, and correlate family history with phenotype and progression.

Paper II: The aim of this study was to independently replicate the finding that rare variation in *TRAP1* is modulating the risk of PD.

Paper III: The aim of this study was to investigate the contribution of rare genetic variation in genes important for mitochondrial function to the risk of PD.

Paper IV: The aim of this study was to evaluate rare genetic variation in genes implicated by GWAS in PD.

3. Summary of results

3.1 Paper I: Familial aggregation of Parkinson's disease may affect progression of motor symptoms and dementia.

In this study, we investigated the familial aggregation of PD in a Norwegian population-based cohort of incident PD, and explored the association between a positive family history and disease progression. The patients were diagnosed and included in the study between 2004-2006, and longitudinal data was available up until and including 7 years after baseline. All underwent whole exome sequencing to screen for known monogenic causes of PD. Controls were recruited during the same timeframe. A simplified family history was taken at baseline, and a detailed, validated questionnaire was introduced at year 3.

Whole exome sequencing revealed one patient with a G2019S *LRRK2*-mutation, who was subsequently removed from the analysis. Using the simplified questionnaire, there was no statistically significant difference between cases and controls regarding first degree relatives with PD. For the extended questionnaire, there was significant familial aggregation of first degree relatives with PD among the cases (OR = 1.99, p = 0.036). The effect size is comparable to estimates from previous studies, but at the conservative end of the spectrum. Possible reasons for this include our use of a validated questionnaire and exclusion of patients with known monogenic causes of PD.

There was no association between a positive family history and motor phenotype when comparing patients with tremor dominant PD to those with postural instability and gait disorder (PIGD). In a regression analysis using generalized estimating equations (GEEs), however, we detected a statistically significant association between having a first degree relative with PD and the progression of motor function (as measured by the Unified Parkinson Disease Rating Scale [UPDRS] II, p = 0.008) and cognitive decline (as measured by the mini mental state examination [MMSE, p = 0.042). Specifically, patients with a positive family history seemed to deteriorate more slowly both in terms of UPDRS II progression and MMSE decline over the 7-year long follow up. Our results indicate that heritable factors, either

genetic or environmental, could contribute not only to inducing disease, but also modulate disease progression.

3.2 Paper II: No evidence for rare *TRAP1* mutations influencing the risk of idiopathic Parkinson's disease

This short report was a response to an article by Fitzgerald et al[260], where they report a novel homozygous loss-of-function mutation in *TRAP1* in a late onset PD patient. They go on to show enrichment of rare, predicted-to-be damaging variants in *TRAP1* in controls when compared with PD, using whole exome sequencing data from the Parkinson Progression Markers Initiative (PPMI).

Replicating the parameters of the association analyses, we did not find an association between *TRAP1*-mutations and PD in either direction in our Norwegian whole exome sequencing cohort. Furthermore, we show that by performing a stricter individual and variant quality control in the PPMI cohort, the rare variant enrichment signal detected by Fitzgerald et al disappears.

3.3 Paper III: Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease

GWAS have identified numerous genetic variants affecting the risk of developing PD, but they still, collectively, explain only a minor percentage of the total estimated heritability. Mitochondrial function plays an important role in the pathogenesis of PD. Many genes responsible for monogenic PD encode proteins that have functions in relation to mitochondria, and disruptions of the mitochondrial respiratory chain has been documented in neurons of PD patients.

We hypothesized that mutations in genes important for mitochondrial function could affect the risk of developing PD. Using whole exome sequencing data from two independent PD cohorts, the Norwegian ParkWest and the American PPMI cohort, we explored the impact of rare genetic variation in PD.

For the region-based enrichment analyses, we restricted our analyses to rare (MAF < 1%) mutations that were classified as either missense or predicted to be damaging based on multiple prediction algorithms. Genome-wide analyses of single

variants and single gene enrichment did not yield any statistically significant results after correcting for multiple testing. A power analysis showed that a cohort of approximately 8,000 – 11,000 individuals is needed in order to achieve 80% power in a genome-wide single gene enrichment analysis.

We manually curated a total of 28 mitochondrial pathways, each consisting of genes that were considered to be part of the same biological pathway. Using the sequence kernel association test, we detected a statistically significant enrichment of rare missense mutations in the mitochondrial DNA homeostasis pathways, surviving multiple testing correction in the PPMI cohort and the meta-analysis. There was also an enrichment in the calcium homeostasis pathway, but it was only statistically significant in the meta-analysis.

3.4 Paper IV: Meta-analysis of whole-exome sequencing data from two independent cohorts finds no evidence for rare variant enrichment in Parkinson disease associated loci

GWAS have identified several risk loci for PD, but associating them with genes is challenging since most are in non-coding regions. We hypothesized that enrichment of rare, coding variants is likely to be found in regions tagged by GWAS, and that this could help identify pathologically relevant genes. Using results from the most recent GWAS, we identified 303 genes of interest around the associated SNPs.

Two whole exome sequencing cohorts were used in the analysis, the ParkWest and PPMI, both in single cohort-analyses and in a meta-analysis. In addition, we used a chip-genotyped dataset, NeuroX, as a replication cohort. In total, 190 genes were available for analysis when restricting variants to rare (MAF < 1%) missense, stopgain, stoploss and splicing mutations. We additionally performed gene-set analyses to identify possible enrichment of these variants across the complete set of genes, as well as a subset of rare LoF-variants.

After correcting for multiple comparisons using FDR, there were no statistically significant associations for either the single gene analyses or the gene-set analyses. Three genes, *GALC*, *PARP9* and *SEC23IP*, were nominally significant (uncorrected $p < 0.05$) in either each of the two WES cohorts or in the meta-analysis, but none

replicated across all three cohorts. Our study does not support a major role of rare variants in genes implicated by GWAS, but cannot, due to being underpowered, rule out a role for rare variants with small effect sizes.

4. Materials and methods

4.1 Study populations

4.1.1 ParkWest – Paper I, II, III, IV

The ParkWest study is a prospective longitudinal cohort study of incident PD, and sough to include all newly diagnosed cases from four Norwegian counties between 1 November 2004 and 31 August 2006, with follow-up still ongoing as of 2020. A detailed description of the study population and methodology has been published previously[261]. The diagnosis of PD was made on the basis of the Gelb criteria[262], and cases which later turned out to have atypical parkinsonism or other disorders rather than PD (19 in total) were continuously removed from the study. One case with a *LRRK2*-mutation (G2019S) was also excluded. In total, 192 cases with validated PD were used in our analyses.

For Paper I, age- and sex-matched controls were recruited in the same timeframe as the cases and followed up with the same frequency. 205 controls were originally recruited, but 2 later developed PD themselves and were removed from the study. 10 controls were also close relatives of some of the cases, and were therefore not used given the research question of Paper I.

For the genetic analyses of Paper II, III and IV, controls (n = 219) were provided by two previously sequenced in-house datasets. 167 were patients with testis cancer, and 52 were patients with acoustic neuroma. They had all been recruited at our hospital and had showed no clinical signs of any neurological or neurodegenerative disorders.

4.1.2 Parkinson Progression Markers Initiative – Paper II, III, IV

The Parkinson Progression Markers Initiative, or PPMI, is an observational clinical study that consists of several different longitudinal cohorts, including a de novo PD cohort, a monogenic PD cohort and a prodromal cohort[263]. Patients are primarily from the United States, but recruitment centers are also located in Europe and Australia. Whole exome sequencing data was available from the de novo PD cohort, comprising a total of 459 cases and 181 controls.

4.1.3 NeuroX – Paper IV

The NeuroX dataset is managed by the International Parkinson's Disease Genomics Consortium and consists of genotype data from 11,402 unrelated individuals (5,540 cases and 5,862 controls) of European ancestry. The dataset is a combination of five previously genotyped cohorts from the United States, France, Germany and the UK[264], and has been re-genotyped using the NeuroX platform[265]. The NeuroX exome array was designed to investigate neurodegenerative diseases and contains, in addition to around 240,000 standard Illumina chip variants, approx. 24,000 custom variants focusing on neurological disorders.

4.2 ParkWest clinical data (paper I)

4.2.1 Longitudinal data

Longitudinal data were available up to and including 7 years after baseline. Motor function was examined every 6 months by way of the Unified Parkinson's Disease Rating Scale (UPDRS)[266], Hoehn & Yahr scale[267] and the Schwab & England activities of daily living scale[268]. Cognitive function was measured using the mini mental state examination (MMSE) at baseline and after 1, 3, 5 and 7 years. All measurements were of patients in the ON state. Using data from the baseline visit and an established algorithm[269], patients were subclassified into three groups: tremor dominant (TD), postural instability and gait difficulties (PIGD) or intermediate/mixed.

4.2.2 Family data

At baseline, study participants were given a simplified questionnaire that consisted of a yes/no question on whether they had any first- or second-degree relatives with PD. Three years later, a more thorough follow-up questionnaire was introduced. This questionnaire was a Norwegian translation of a validated questionnaire with an estimated sensitivity of 95.5% and specificity of 96.2% for detecting relatives with PD[270]. It consisted of a variety of questions, e.g. whether the interviewee had any relatives with a shuffling gait or tremor, and if they were ever examined by a neurologist. Any family member with a positive history was then classified as either

"possible", "probable" or "definite" PD according to a diagnostic algorithm. The questionnaire is available in English from the original publication[270], and the Norwegian version can be found in Appendix I of this thesis.

4.3 Genetic data (paper II, III, IV)

4.3.1 Sequencing and genotyping

ParkWest (Paper II, III and IV)

DNA was extracted from blood by routine procedures and sequenced at HudsonAlpha institute for Biotechnology (Huntsville, Alabama) on the Illumina HiSeq platform using Roche-NimbleGen Sequence Capture EZ Exome v2 (173 controls) and v3 (all PD and 46 controls) capture kits and paired-end 100 bp sequencing. Reads were mapped to the hg19 (GRCh37) reference genome using BWA v0.6.2[271], polymerase chain reaction duplicates removed with Picard v1.118[272] and the alignment refined using the Genome Analysis Toolkit (GATK) v3.3.0[273] applying base quality score recalibration and realignment around indels recommended in the GATK Best Practices workflow[274,275]. Variants were called in all samples using the GATK HaplotypeCaller[273] with default parameters. Variant quality score recalibration was performed using 99.9% sensitivity threshold[273]. Using BEDtools[276] and VCFtools[277], the remaining variants were filtered against the intersection of capture targets from the two capture kits. Variants with a total depth below 10X were marked as unknown (no-call) using BCFtools[278]. Indel calls were excluded from downstream analysis, as they were found to be less reliable than single nucleotide variant calls.

PPMI (paper II, III and IV)

Sequencing and variant calling was performed by the PPMI, and the following information was provided by Dr. D.G. Hernandez and Dr. J.R. Gibbs, National Institute on Aging, Laboratory of Neurogenetics. DNA was extracted from blood and sequenced using Nextera Rapid Capture Expanded Exome Kit on the Illumina HiSeq 2500 platform using 2x100 bp paired-end read cycles. FASTQ files (reads) were aligned using BWA[271] against the hg19 reference human genome. Duplicate read removal, format conversion and indexing were performed with Picard[272]. GATK[273-275]

was used to recalibrate base quality scores and perform local realignments around indels for the aligned sequencing reads. The GATK HaplotypeCaller was used for variant calling and genotype likelihood generation. GATK CombineGVCFs and GenotypeGVCFs were used to perform joint genotyping for the cohort from the set of per subject genomic VCF files.

NeuroX (paper IV)

The NeuroX dataset was obtained through dbGaP (dbGaP Study Accession: phs000918.v1.p1). Individuals were genotyped on the NeuroX array, which is a custom array developed specifically for use in neurodegenerative diseases[265], designed by a National Institute of Neurological Disorders and Stroke consortium. It consists of a core of standard Illumina exome variants from the Illumina HumanExome array v1.1 (242,901 variants), and an additional set of custom variants (24,706) with a particular focus on neurologic disorders. The custom variants were selected based on results from previous GWAS, within-consortium sequencing pilot studies and systematic review of the literature, focusing on Alzheimer Disease, PD, multiple system atrophy, progressive supranuclear palsy, amyotrophic lateral sclerosis, multiple sclerosis, frontotemporal dementia, myasthenia gravis and Charcot-Marie-Tooth.

4.3.2 Individual and variant quality control

The quality control procedures followed were identical for both the ParkWest, PPMI and NeuroX datasets. All quality control procedures were performed using PLINK v1.90[279], R[280] and Eigensoft[281,282].

First, genetic data in variant call format (VCF) was recoded into binary PLINK format after indel removal where applicable. Individuals were excluded if they had a genotypic missing rate of > 2%, abnormal heterozygosity (± 3 standard deviations, calculated for rare and common variants separately), conflicting sex assignment, cryptic relatedness (identity by descent > 0.2) or non-European ancestry. Population stratification was studied using multidimensional scaling against the HapMap populations[283].

Variants were excluded if the genotyping rate was < 98%, or if the genotype distribution departed from the Hardy-Weinberg equilibrium ($p < 10^{-5}$). Variants with different call rates in cases and controls ($p < 0.02$) were excluded. Sex chromosomes were removed in all datasets, as well as monomorphic and multiallelic variants. Principal component analysis was carried out using Eigensoft using standard settings (5 iterations, 10 principal components, sigma 6). ANOVA of the first 10 principal components was performed with the significance level set to $p < 0.01$. Significant principal components were used as covariates in all downstream analyses.

4.3.3 Variant annotation and filtering

For all papers and datasets, variants were annotated using ANNOVAR[284] according to the RefSeq gene transcripts[285].

Paper II

As paper II was a replication study of the findings from Fitzgerald et al[260], we followed their methods for variant filtering and classification. Rare variants were defined as variants with a MAF < 1% in the non-Finnish European ExAC dataset[150]. A total of six variant subsets were created: synonymous, nonsynonymous (i.e. missense, Sequence Ontology: 0001583) and four subsets with predicted-to-be-damaging variants of varying severity. Phred-like Combined Annotation Dependent Depletion (CADD) scores were used to predict each variant's disruptive potential, and the four variant groups defined as CADD > 10, CADD > 15, CADD > 20 and CADD > 30. In this scoring system, variants are ranked according to their predicted disruptive potential compared to all other variants. A score > 10 means that the variant is predicted to be among the top 10% of disruptive variants, a score > 20 among the top 1% etc.

Paper III

Two subsets of variants were extracted for use in downstream analyses. The first comprised all exonic variants defined as nonsynonymous, i.e. missense mutations. The second subset consisted of predicted-to-be-damaging variants, as defined by the

variants having a deleterious score in all of the following prediction algorithms: PolyPhen2 HumDiv[43], PolyPhen2 HumVar[43], MutationTaster[286], SIFT[287] and LRT[288]. Only rare variants were considered. In ParkWest, rare variants were defined as having a MAF < 1%. In the PPMI dataset there was an imbalance in the number of cases and controls, and rare variants were therefore defined as having a MAF < 1% in either cases or controls as to avoid any bias.

Paper IV

Two sets of variants were used in paper IV, one consisting of rare variants classified as either nonsynonymous (missense), stop-gain, stop-loss or splicing, another consisting of rare LoF-variants (stop-gain, stop-loss and splicing mutations). Rare variants were defined as having a MAF < 1% in the non-Finnish European population in the Genome Aggregation Database (gnomAD)[289].

4.3.4 Regions used in collapsing analyses

Pathway curation (paper III)

Pathways focused on various aspects of mitochondrial biology, and were defined as groups of genes encoding proteins with a functional and/or structural link to mitochondrial biochemistry. The pathways were manually curated based on data from MitoCarta v2.0[290], a database of all known proteins with strong support for mitochondrial localization. After initial curation, the pathways were expanded by using STRING[291] to identify genes encoding additional pathway-relevant proteins without known mitochondrial localization (and therefore not in MitoCarta). Specifically, we compiled a list for each pathway of additional candidate proteins ranked by the number and strength (STRING combined score) of STRING interactions with the original pathway. The resulting lists were manually inspected, and the original pathways supplemented with additional genes encoding proteins with known involvement in mitochondrial function, but without established mitochondrial localization. In total, 28 pathways focused on different aspects of mitochondrial function were generated. The complete unedited MitoCarta database was also included as a 29th pathway.

Genes of interest (paper IV)

The aim of the study was to perform a focused genetic association analysis of genes implicated by GWAS in PD by using LD to identify nearby genes of interest. The most recent GWAS identified 90 PD-related SNPs[24], and we defined genes of interest as any gene containing a variant in LD within a 2 megabase window around any of these. The threshold was set to $R^2 > 0.5$, and if the variant in LD was located in an intergenic region the nearest gene was included. LD calculations were available from supplemental material of the original GWAS, and included a total of 303 genes that fit the inclusion criteria.

4.4 Statistical analyses

All statistical analyses were performed using either SPSS v22, PLINK v1.9[279] or R/RStudio[280,292]. In R/RStudio, the specific packages used were SKAT[293], MetaSKAT[61] and metap[294], in addition to more general data management packages.

4.4.1 Paper I

The statistical tests used were Pearson's χ^2, Fisher's exact test, independent-samples *t*-test and Mann-Whitney U. Generalized estimating equations (GEEs) were used to analyze differences in disease progression between PD patients with a positive and negative family history. Depending on the distribution of the dependent variable, a gamma or linear model was used together with an exchangeable correlation structure. All regression models were adjusted for age and sex. P-values of < 0.05 were considered statistically significant. All analyses were performed using SPSS v22.

4.4.2 Paper II

Single variant association tests were performed for rare variants (see definition under section 4.2.3) using PLINK v1.9[279]. For ParkWest, single variants were analyzed using logistic regression with significant principal components as covariates (--logistic in PLINK). In PPMI, there were no significant principal components, and single variants

were analyzed using either χ^2 or Fisher's exact test (--assoc or --fisher in PLINK). In this paper, rare variants were defined as having MAF < 1% in an external dataset (ExAC[150]). Therefore, some of the included variants had a within-cohort MAF > 1%. χ^2 was used for all variants with a within-cohort MAF >1%, and Fisher's exact test for those with MAF <1%.

A targeted rare variant enrichment analysis of *TRAP1* was carried out using a weighted burden test, the sequence kernel association test (SKAT) and the optimal sequence kernel association test (SKAT-O). All analyses were done in R[280] using the SKAT R package v1.3.2[293] with default settings. Six variant subsets were analyzed in the ParkWest and PPMI datasets separately, with significant principal components as covariates. The p-values reported were not corrected for multiple testing.

4.4.3 Paper III

Single variant analyses were performed for common variants (MAF > 1%) using χ^2 (--assoc in PLINK) for PPMI and logistic regression (--logistic in PLINK) with significant principal components as covariates for ParkWest using PLINK v1.9[279]. The significance threshold used for single variant associations was the broadly accepted genome-wide standard of $p < 5 \times 10^{-8}$.

Rare variant enrichment analyses were performed on all genes and pathways containing ≥ 2 variants using a weighted burden test and the SKAT. Significant principal components were used as covariates, and all analyses were carried out in R[280] using the SKAT package v1.3.2[293] with default settings. Using resampling, the minimum achievable p value (MAP) was calculated for all genes and pathways, and any gene/pathway with MAP above a Bonferroni-corrected threshold for statistical significance were excluded to reduce statistical noise. After the analyses, Bonferroni correction was applied to the p-values from the single gene analyses to correct for multiple comparisons. For pathway analyses, p-value thresholds for significance were calculated using the minP/maxT method[295]. Here, phenotypes are null-permuted (10,000 times for ParkWest and 100,000 times for PPMI), and for each permuted dataset the minimum p-value is extracted to form a minimum p-value distribution. The

significance threshold is then found at the 0.05-quantile of this distribution, representing a significance level of $p < 0.05$.

For pathways, meta p-values were calculated using the optimally weighted Z test[64], as implemented by the metap R package[294]. A power analysis for single gene associations were performed using the built-in functionality of the SKAT R package. By default, the package uses simulated data and simulated genes (genetic regions of random lengths). We modified the algorithms and used our ParkWest dataset instead of the simulated dataset, as well as the RefSeq gene transcripts instead of random regions[285]. The assumed prevalence of PD was set to 1.5%, and power calculations were carried out for different percentages of causal SNPs (10%, 25%, 50%, 75% and 100%) using default weights. Only missense mutations with MAF < 1% were considered, and a genome-wide significance level of 2.5×10^{-6} was implemented (which corresponds to Bonferroni-correcting for 20,000 genes).

4.4.4 Paper IV

In both the single gene and gene-set analyses, rare variant enrichment analyses were performed using a weighted burden and the SKAT as implemented by the SKAT R package v.1.3.2.1[293]. Statistically significant principal components were used as covariates, and default settings were otherwise used. For the single gene analyses, only genes with at least two or more variants across the two whole exome sequencing cohorts combined (ParkWest and PPMI) were analyzed.

Meta-analysis was performed by using the Meta-SKAT R package v0.60[61], and the same weighted burden test and SKAT as in the single cohort analyses. We hypothesized that genetic effects should be homogenous across studies, meaning the same mutation should have the same direction of effect in both cohorts, and did not perform a version of the test that allows for heterogenous genetic effects. All p-values were corrected for multiple comparisons by using FDR-correction[296].

4.6 Ethical considerations

These studies (paper I-IV) were approved by the Regional Committee for Medical and Health Research Ethics, Western Norway (REK 131/04). This thesis does not contain any information that could be used to identify specific individuals.

5. Methodological considerations

5.1 Study populations

Throughout all four papers, we have utilized data from the Norwegian ParkWest study, a longitudinal cohort of incident PD. All participants had been extensively followed up since 2004. This has ensured that the diagnostic certainty for these patients is very high. During the first few years, several patients were removed from the study as they were found to have another diagnosis than PD. Similarly, some of the controls also developed PD during the course of the study and were excluded. The remaining cohort is therefore of particularly high quality, and because all patients underwent genetic sequencing, we know that they do not have any known monogenic forms of PD (one case was found to harbor the G2019S *LRRK2* mutation and was removed from all studies). This is especially important when studying the familial aggregation of idiopathic PD in paper I.

One clear disadvantage of the genetic studies (paper II-IV) is that the genetic controls differ from those in the ParkWest longitudinal cohort. Only PD patients from the original cohort underwent whole-exome sequencing, and controls were sourced from other, already sequenced, in-house datasets. 167 controls had been diagnosed with testis cancer, and 52 were patients with acoustic neuroma. All controls were diagnosed at our hospital, and none had shown any sign of neurodegenerative disease. Still, one must assume that some of these controls will or would go on to develop PD as they age. Some studies have found an increased risk of brain tumors in PD[297], but crucially this relationship was not detected when the tumor diagnosis came first[298]. Similarly, there is, to our knowledge, no known relationship between testis cancer and PD[299]. The rate of "eventual PDs" in our control sample would therefore be roughly equal to the lifetime prevalence of PD (2.0% for men and 1.3% for women[7]). Given the sex distribution of our controls, this equates to approx. 4 individuals.

As the majority of our genetic controls came from a cohort of testis cancer, male individuals greatly outnumber females in our control group (approx. 87% males). While male individuals also make up the majority of our PD cases, they do so to a lesser extent (approx. 61%). Even though our analyses are limited to the autosome

(sex-chromosomes were removed during individual and variant quality control), a more equal ratio in cases vs controls would have been preferred. We do not believe, however, that it has had any substantial impact on our results. One could argue that since males are at greater risk of PD than females, such a high proportion of males could have led to more "eventual PDs" in our control sample. The prevailing theory of why women are less at risk for PD than men revolves around the neuroprotective effects of estrogen, i.e. an environmental exposure for all intents and purposes[300]. Assuming then that the 0.7% difference in lifetime risk of PD is due to females being protected by this environmental factor, the proportion of genetically predisposed individuals due to variation in the autosome would be equal for a female and male control sample.

5.2 Statistical analysis of longitudinal data (paper I)

For the regression analysis of the longitudinal data, we chose to use GEEs instead of a mixed model approach. GEEs are considered to be particularly robust to misspecification of the correlation matrix, but with reduced efficiency as a potential tradeoff[301,302]. A mixed model analysis could therefore potentially have been able to detect associations with a higher degree of statistical certainty.

5.3 Genetic sequencing and quality control (paper II, III, IV)

5.3.1 Genetic sequencing

Sequencing/genotyping of the PPMI and NeuroX datasets were performed by their respective study groups, and the genetic data provided to us have been used extensively in published research. For the ParkWest dataset, all individuals were sequenced on the Illumina HiSeq platform, but with different versions of the Roche-NimbleGen Sequence Capture EZ Exome (v2 for 176 controls and v3 for all PDs and 46 controls). Because the two versions offer slightly different capture targets, only the intersecting regions were used in downstream analyses. Still, it is possible that this could introduce batch effects in our ParkWest data. We addressed this by strictly following GATK Best Practices guidlines[274,275] for the variant calling, and using rigorous individual and variant quality control procedures.

5.3.2 Individual and variant quality control

For individual and variant quality control, we followed established procedures for genetic association studies[303]. Principal component analysis was performed for all cohorts, and principal components with $p < 0.01$ in an ANOVA of the first 10 principal components were included as covariates in all downstream analyses. As outlined in section 5.3.1, we were aware of the possibility of batch effects in our ParkWest cohort due to differences in capture kit versions between cases and controls. After quality control, however, we did not observe any p-value inflation in our quantile-quantile plots or clustering in our multidimensional scaling plots (see Figure 1 in paper III). We therefore considered any potential batch effects to be sufficiently controlled.

We elected to use fairly conservative thresholds for both call-rate per person and call-rate per SNP (>98%). For genotype data, the threshold is usually between 95% and 99%[303]. For whole exome data, on the other hand, less strict thresholds have generally been considered acceptable. Published studies have used thresholds ranging from 80% to 98%[304,305]. Rare variant enrichment analyses have been shown to be susceptible to type I errors when variant call-rates are low and differ between cases and controls, and we therefore elected to be conservative in order to reduce false positive signals[306]. Given that our cohort is relatively small, and that these strict thresholds further reduce power by removing samples and variants, an argument could certainly be made for a more inclusive approach.

5.4 Variant filtering and annotation (paper II, III, IV)

5.4.1 Rare variant definition

Because GWAS largely focus on common genetic variation, rare variants have been proposed as a possible explanation for the missing heritability of many complex diseases, including PD[307]. Defining what constitutes a rare variant is, however, not set in stone. The upper MAF threshold for rare variants is typically considered to be somewhere between 0.5% and 5%[308]. Some also suggest a lower MAF threshold so as to distinguish rare variants from private mutations[309]. For our planned analyses,

focusing on collapsing multiple rare variants and looking for enrichment, a lower threshold would not be meaningful, as both rare and private mutations are equally relevant and available for analysis. As for the upper threshold, we elected to go with the commonly used figure of MAF < 1%[309,310]. There is an established inverse relationship between MAF and disruptive potential, where, on average, variants with lower MAFs are more likely to be classified as damaging by prediction tools such as CADD[311]. In selecting our rare variant threshold, we aimed to include as many biologically relevant mutations as possible, while limiting non-relevant variation.

In addition to selecting a rare variant threshold, it was also necessary to decide whether to apply it to within-cohort MAFs or use an external public database. Both approaches have been used in whole-exome studies in PD and have their advantages and disadvantages[135,146]. The overall aim of the rare variant filtering is to maximize the inclusion of variants that affect disease, while limiting the inclusion of non-relevant variation. Public datasets are typically very large, e.g. gnomAD contains > 141.000 whole exomes/genomes[289], and variants that are common here are therefore less likely to be disease-relevant. However, when using cohorts from relatively homogeneous populations, harmless variants that are rare or absent in public databases can still be common here due to founder effects. When whole exome or whole genome sequencing is used for diagnostic purposes, variants are typically filtered both against a public database and an in-house database of local genomes/exomes to remove as many non-pathogenic variants as possible[312]. In paper II, we replicated the methodologies used by Fitzgerald et al[260], who used the publicly available ExAC dataset to filter rare variants[150]. In paper III we used within cohort MAFs to define rare variants, as our ParkWest dataset is fairly homogenous and we wanted to prevent the inclusion of non-relevant variation with high MAFs in our analysis. Finally, in paper IV we used MAFs in the publicly available gnomAD database (non-Finnish European subset), as this approach has become more popular among rare variant enrichment studies and avoids potential bias when there is a large discrepancy between the number of cases and controls in the dataset.

In our ParkWest dataset, the number of cases and controls is fairly equal. However, in the PPMI dataset, the ratio between cases and controls is 2.5:1. This can

cause an inclusion/exclusion bias in that risk variants more prevalent in cases are more likely to be excluded than protective variants more prevalent in controls when only cohort-wide MAFs are considered. We addressed this issue by including any variant with MAF < 1% in either the cases or controls separately. Another possible solution would be to include only variants with MAF < 1% in both cases and controls, but this was deemed to be overly conservative in our limited sample as it resulted in the exclusion of a high number of variants.

5.4.2 Variant filtering

In addition to restricting our analysis to rare variants, we also filtered variants based on type and function in order to limit the inclusion of non-relevant variants. Synonymous variants do not, by definition, alter the amino acid of the resulting protein, and are therefore generally considered to be much less likely to impact disease development than for example missense mutations. While there is increasing evidence that synonymous mutations can impact biological processes by affecting transcription[313] and mRNA stability[314], they are generally viewed as neutral variants and not included in rare variant enrichment analyses[315]. Here, studies tend to focus on either missense/nonsynonymous[29,134,135,137,146], predicted-to-be-damaging[29,135,146] or loss-of-function[135,146] variants.

In paper II, following the methodology of Fitzgerald et al[260], we performed enrichment analyses on nonsynonymous, synonymous and predicted-to-be-damaging variants. Phred-like CADD scores were used to group variants into four subgroups of increasingly deleterious mutations.

In paper III, we modeled our approach on the methodologies of Purcell et al[29] in their paper on the polygenic burden of disruptive mutations in schizophrenia. Here, variants were grouped into three categories: disruptive, predicted-to-be-damaging (strict) and predicted-to-be-damaging (broad). Disruptive mutations included nonsense, splice-cite mutations and frameshift mutations. Five prediction algorithms were used to evaluate the disruptive potential of missense mutations: PolyPhen2 HumDiv[43], PolyPhen2 HumVar[43], MutationTaster[286], SIFT[287] and LRT[288]. For the strict category, variants had to have a deleterious score in all five prediction

algorithms, while in the broad category, variants needed only to have a deleterious score in at least one. Because our datasets were substantially smaller than that used by Purcell et al[29], using the same categories would be overly strict and result in too few variants to be able to do any meaningful analysis. We therefore opted to use two variant categories, one with all rare missense mutations, and one with variants with deleterious scores in all five previously mentioned prediction algorithms. Indels were not included, as these had been removed during quality control because they were found to be less reliable than single variant calls. Also not included in the missense category were nonsense and splice-cite mutations. The reason for this was that we treated our two groups as separate categories rather than one as a subset of the other. Retrospectively, it would perhaps have been more appropriate to have included these mutations in our missense category, which we go on to do in paper IV. There, we performed the analyses using two categories, one containing all rare missense, nonsense and splice-cite mutations, and a subset containing only nonsense and splice-cite variants. The latter category was only used in the gene-set analyses, as the number of included variants was very low.

5.5 Pathway selection (paper III)

The major focus of paper III was the pathway analyses, and our aim was to investigate pathways relevant for mitochondrial function. Since our datasets were of limited size, the pathway selection process was of vital importance in order to focus the analyses and limit the number of "off-target" pathways. We originally examined already available pathway databases but found that they were not particularly well suited for our needs. For example, one of our eventual pathways was mitochondrial DNA maintenance. Looking at available pathways in the GO database[128,129] (accessed through the Molecular Signatures Database v7.0[316,317]), there are three pathways related to this process: mitochondrial DNA repair (GO:0043504), mitochondrial DNA replication (GO:0006264) and mitochondrial DNA metabolic process (GO:0032042). Overall, they contain a limited number of genes, and several genes that we believe are biologically relevant are not listed. They do not, for example, contain the genes *NEIL1*, *NEIL2* or *SUCLA2*, all of which have been linked to mitochondrial DNA

homeostasis[318,319]. Similarly, other databases (e.g. KEGG[130] and Reactome[131]) contain very few pathways specifically targeted at mitochondrial function. We therefore elected to generate our own pathways based on the Mitocarta database[290] and STRING tool[291] and published the complete list as a supplement to paper III. The original allocation of Mitocarta genes into separate pathways was done by manual curation, relying on personal expertise and review of the published literature. As this "human approach" has weaknesses in terms of personal biases and knowledge gaps, we introduced an additional computational step with STRING to compensate for these and append our pathways with additional genes. Given our rather specific research question in paper III, we believe this was the best approach available, and using custom gene sets in rare variant enrichment analyses has some precedence[29]. Using pathways from established databases is still the most prevalent approach, and an alternative would have been to test all mitochondrial pathways in GO.

5.6 Enrichment analysis (paper III, IV)

5.6.1 Choice of test

As outlined in the Introduction (section 1.2.1), there are numerous rare variant enrichment analyses available. We elected to follow guidelines outlined by Do et al[125], and implement both a traditional burden test and a test that allows for opposite directions of effect within the same genetic region. Both the C-alpha test[48], SSU test[49]and SKAT[41] are examples of variance-component tests that can accommodate the presence of both risk and protective alleles. We chose to use the SKAT, as it is considered to be the most flexible of the three[28]. It is also a well-established test used in many rare variant enrichment publications[127,320-323]. For the burden test, the SKAT R package used in our analyses includes a weighted burden test that was deemed suitable. An adaptive burden test was considered, but ultimately not used. The main reason was that it would have had to be added as a third test and thus greatly increased the number of tests to correct for in our already low-powered dataset.

In recent years, omnibus tests, and particularly the SKAT-O, has become increasingly popular. We considered using SKAT-O in our analyses instead of the SKAT and burden tests, with the main advantage being a reduced number of tests

needed to do multiple comparisons correction for. However, it is potentially less powerful than the burden test or SKAT under certain conditions[28]. In addition, we considered the p-values from SKAT-O too be more difficult to interpret than presenting p-values for both burden and SKAT, as it would not be clear of the association is largely dependent on uni- or bidirectional effects. Still, given that our sample size was small, reducing the number of tests would have been advantageous.

In our analyses, we utilized prediction algorithms to filter variants into subgroups. One alternative would have been to use these prediction algorithms to assign individual variant weights, which has been utilized by some studies[324]. However, different algorithms can generate conflicting results[325], and we therefore decided against using them in this way.

5.6.2 Multiple testing correction

The Bonferroni method of correcting for multiple comparisons is the most widely used in genetic association studies, likely due to its ease of use and strong control of the type I error rate[326]. It is, however, known to be overly conservative in situations where the tests are not independent. In GWAS for example, because variants are in LD with neighboring genomic regions, correcting for 10^6 tests is deemed sufficient even though the number of actual tests can be much higher[326]. In our pathway analysis (paper III), many genes are present in multiple pathways, and Bonferroni correction was therefore considered to be too conservative. We opted to use the Westfall & Young minP/maxT method[295], where a distribution of minimum p-values is created from case-control permutated datasets. The critical p-value, for which p-values below are considered statistically significant, is then found at the 0.05-quantile in this distribution. This method adapts to the true correlation structure of the data, and avoids the worst-case assumption of the Bonferroni method (that tests are completely independent)[326]. One disadvantage is that it is computationally intensive, and the Bonferroni correction was therefore applied to the individual gene results.

In paper IV we implemented false discovery rate (FDR) correction[296]. FDR offers increased power compared to the Bonferroni method in situations with a large

number of false hypotheses, and would therefore also have been a viable option in paper III[326].

5.6.3 Meta-analysis

We implemented meta-analyses for our whole-exome datasets (ParkWest and PPMI) in both paper III and paper IV. In paper III we used the optimally weighted Z-test[64] to combine results for our pathway analyses. In retrospect, other methods could possibly have yielded better results. Combining p-values restricts the meta-analysis to combining information at the gene level, while meta-analyses that are based on combining tests statistics can incorporate information at the variant level. The latter type of meta-analysis has shown increased power when compared to the former[327]. In paper IV we therefore opted for the Meta-SKAT analysis suite[61], where score statistics for each variant are combined in a meta-analysis.

5.7 Genes of interest (paper IV)

The basis of our analyses in paper IV was the most recently published GWAS by Nalls et al[24], identifying 90 SNPs associated with PD. Using our whole-exome sequencing datasets, we wanted to explore nearby genes for rare variant enrichment. Nalls et al had already performed an LD-analysis of their 90 hits, identifying variants with $R^2 > 0.5$ in a 2 megabase window around the SNPs and the closest gene to all variants in LD. A list of all tagged variants and genes was included in their supplemental materials and included a total of 303 unique genes, and we decided to use those genes as our genes of interest. This was partly because the sample size in the Nalls et al study provided more accurate LD estimates than what we could achieve using publicly available data, and partly because using LD estimates from the same dataset as the GWAS meant that if a variant in LD was the true cause of the association it would be identified. The latter would not necessarily be the case when using LD estimates from a different population.

An alternative approach we considered was to use a Prix fixe-strategy[126]. Here, genes implicated by LD-calculations for all GWAS hits are analyzed, and one gene per GWAS-hit defined loci is selected as the most likely causal gene based on its

functional similarity to genes in the other loci. The underlying assumption is that there exist a few highly relevant biological pathways in PD, and that GWAS hits are likely identifying genes that are part of these few pathways. The Prix-fixe strategy aims to identify the genes that have the assumed highest likelihood of being part of the same pathways. This approach would have reduced the number of tested genes in our analyses to, at most, 90. We ultimately decided against using it, because it would likely prioritize genes in large pathways and therefore maybe miss associated genes in smaller pathways.

6. General discussion

6.1 The familial aggregation of PD in the ParkWest cohort is comparable to previously published estimates (paper I)

In this work we show that the relative risk of having a first-degree relative with PD was 2.0 for PD patients in our ParkWest cohort. This estimate is slightly lower compared to previous studies, generally suggesting an overall RR of 2.8[111]. There are several potential reasons for this discrepancy. The studies included in the meta-analysis used various methodologies for validating the diagnosis of PD among relatives. Some incorporated direct examination or information from death certificates, while others relied solely on the reported family history. There is a known bias of family history data in PD, where PD patients are more likely to report false positive family histories than controls[328]. In our study, we used a validated questionnaire for taking the family history, which has been shown to have comparable sensitivity and specificity to direct examination[270]. In addition, our cohort consists of a true population-based sample where all known cases of monogenic PD were excluded. Some of the studies included in the meta-analysis used patients from specialized movement disorder or PD clinics, or included only young-onset cases, both of which increases the likelihood of including cases of monogenic PD in the analysis. Therefore, we believe that our estimates are highly accurate for idiopathic PD in our population.

6.2 A family history of PD is associated with altered clinical progression (paper I)

In our longitudinal analysis, we show that there is a small, but statistically significant, difference in the rate of decline/progression between PD patients with and without a positive family history. The effect was present in measurements of both cognitive (MMSE) and motor function (UPDRS II), and patients with a family history of PD had slightly lower rates of progression. Younger age of onset has been shown to be associated with less severe phenotypes and rates of progression[329-331], and familial aggregation is especially pronounced among young-onset cases[113,332]. This is in

agreement with our finding that familial aggregation is associated with a slightly milder phenotype, even though there was no association between age of onset and family history in our data. It should be noted that the ParkWest cohort is population-based. The average age of onset was 65.6 years, with only a single case < 40 years. Considering the limited number of young onset cases, our sample was therefore not particularly well suited to investigate the relationship between age of onset and familial aggregation. Other than MMSE and UPDRS II, progression along the Schwab & England activities of daily living scale and the Hoehn & Yahr scale were also tested for association with family history, but there were no detectable differences. We also investigated whether motor subtype was impacted by familial aggregation, but the results were negative.

In our study, family history serves as a proxy for what is most likely genetic effects. Multiple studies have found that genetic variation can influence clinical progression in PD[333-337], but results are not unequivocal[338]. Nevertheless, based on both available evidence in idiopathic PD and knowledge gained from observing the phenotypic variation in monogenic PD, it is highly plausible that genetic variation can contribute to the phenotypic heterogeneity seen in sporadic PD[339,340].

6.3 No evidence for rare *TRAP1* mutations influencing the risk of idiopathic Parkinson's disease (paper II)

The basis for paper II was a study by Fitzgerald et al[260], where they reported depletion of low-impact, as predicted by CADD-scores, variation in the gene *TRAP1* in PD patients. We wanted to see if we could replicate their findings in a separate cohort, the ParkWest dataset. Following the same methods for variant selection and analysis, we did not find any association between *TRAP1*-variation and PD. In their study, Fitzgerald et al detected the *TRAP1*-association using whole exome sequencing data from the PPMI cohort. In our reading of their methods, we questioned certain aspects of their individual and variant quality control. Specifically, we had concerns regarding their threshold for variant call-rate (>90%) and the methods employed to control for population stratification. Because PPMI is a publicly available dataset, we were able to replicate the analyses by Fitzgerald et al using the same cohort, but with stricter

quality control. We used a variant call-rate threshold of > 98% and removed outliers iteratively (5 iterations instead of the single-step procedure used in the original study). With these stricter quality control measures, we failed to detect any association between *TRAP1* and PD. Our analysis showcases that the thresholds and methods employed in the individual and variant quality control can greatly impact analyses that utilize collapsing of rare variants. In particular, controlling for population stratification is likely of high importance. Rare variants are typically highly segregated between even closely related populations[341], and type I error rates can therefore be inflated if cases and controls are composed of different ethnicities.

6.4 There is enrichment of rare missense variation in genes important for mitochondrial DNA maintenance in PD (paper III)

In paper III, we show an association between rare, protein-altering mutations in the pathway of mtDNA maintenance and PD. The association was present in both the ParkWest and PPMI cohort and was the result of variation in multiple genes within this pathway. mtDNA encodes 13 proteins that all are part of the respiratory chain[342]. Maintenance of mtDNA is therefore important to maintain respiratory chain function, which is affected by both single point mutations and larger deletions[254,255]. Previous studies have found evidence for impaired mtDNA homeostasis in PD, with depletion of wild-type mtDNA[256,257]. Our analysis indicates that this impairment may be partly due to genetic variation in genes encoding the mtDNA maintenance pathway.

The rare variant enrichment was detected using SKAT, which indicates that the effect is due to both risk-increasing and protective mutations. This is not surprising, given the complexity of biological pathways. It is conceivable that altered protein function may not only impair, but also enhance mtDNA maintenance, thereby increasing the pathways resilience to insults. Moreover, certain variants may be deleterious in one respect while beneficial in another. For instance, a mutation that impairs the proofreading capabilities of POLG, causing a premature ageing phenotype in mice, has been shown to also trigger a neuroprotective mechanism increasing mtDNA copy number. Interestingly, this increased mtDNA biogenesis renders the

dopaminergic *substantia nigra* of the animals resistant to PD-related neurotoxins such as MPTP[343].

In addition to mtDNA maintenance, our analysis also implicated the pathway of mitochondrial calcium homeostasis which reached statistical significance in the meta-analysis. Mitochondria contribute to the buffering of cytosolic calcium, and the calcium concentration in the mitochondrial matrix regulates several key aspects of mitochondrial function[344]. Higher concentrations enhance the activity of enzymes in the Krebs cycle, and excessive mitochondrial calcium accumulation can trigger apoptosis[344,345]. Deficiency of PINK1 has been shown to decrease mitochondrial calcium efflux, thereby inducing cell death[346]. Furthermore, inhibiting the mitochondrial calcium uptake machinery was found to be neuroprotective in PINK1-deficient neurons[347]. Because the enrichment of the pathway was not significant in our individual cohort analyses, we consider our findings tentative, but still supportive of a role for mitochondrial calcium homeostasis in PD.

Since the publication of our paper, other genetic studies of mitochondrial function in PD have been published. Using mitochondria-specific polygenic risk scores, Billingsley et al used GWAS data to show that variation in genes spanning the entirety of mitochondrial function is associated with PD risk[348]. There is therefore ample evidence to suggest that genetic variation in mitochondrial genes affects the risk of PD and explains some of the observed missing heritability.

6.5 Detecting single gene associations on an exome-wide level will likely require sample sizes in the order of 8,000 to 11,000 individuals (paper III)

Using real world data, we show that the number of individuals needed to achieve acceptable power (80%) is around 8,000 individuals for SKAT and 11,000 individuals for burden (assuming an equal number of cases and controls, and that approximately 50% of variants actually impact disease risk) in a genome-wide whole exome sequencing study of rare variant enrichment for PD. The largest, to date, whole exome sequencing dataset to appear in a peer-reviewed publication has been the study by Blauwendraat et al[349], with 2,440 PD patients. It is therefore not surprising that no

studies identifying significantly enriched genes at a genome-wide level have yet been published in PD.

Power analyses for rare variant enrichment are highly complex, as they require several assumptions about the underlying genetic architecture and variant effects. Most statistical tests for rare variant enrichment include a power analysis in the original publication, where the proposed test is compared to other available methods under certain conditions. Because of the high degree of complexity generated by all of the variables (study design, type of tests compared, variant effects, frequency of variants, MAF cutoff, variant weights etc.), the results of these analyses are often not easily transferrable to any specific disease or research question. Our analysis therefore serves as a guide in terms of the required sample sizes needed for rare variant enrichment studies in PD specifically. Some studies have attempted to perform more easily transferrable power analyses, and their results are not dissimilar to ours. Moutsianas et al explored the power of gene-based rare variant enrichment methods for complex diseases, and found approximately 60% power with 10,000 individuals[350].

6.6 An analysis of 303 genes tagged by GWAS does not support a major role for rare variant enrichment in these genes (paper IV)

In paper IV, we investigated 303 genes implicated in PD through GWAS results for rare variant enrichment. We hypothesized that rare variant enrichment in some of these genes could either be responsible for the nearby GWAS hit through LD, or that rare variant enrichment could, in addition to a common SNP association detected by GWAS, also impact disease risk in relevant genes. Our study is not the first to use this approach[126], but a recently published GWAS in PD greatly expanded the number of known PD-associated loci[351]. We therefore considered further exploration of this hypothesis to be of value. In addition to a single gene analysis, we also incorporated a gene-set analysis to look for rare variant enrichment across the combined set of genes of interest.

Our analysis did not yield any statistically significant results after FDR-correction for multiple comparisons, in neither the single gene analyses nor gene-set analyses. We did observe low raw p-values ($p < 0.05$) for three genes across multiple

datasets. *GALC* and *SEC23IP* were nominally significant in both of the whole exome sequencing cohorts (ParkWest and PPMI), while *PARP9* was nominally significant in the meta-analysis and the genotyped cohort (NeuroX). However, none replicated across all three cohorts. While our study was likely underpowered to detect minor effects, our results do not support a major role for rare variant enrichment in genes tagged by GWAS.

Studies with approaches similar to ours have previously been published, with mixed results. In one study, *LRRK2* was found to be enriched with rare missense variation[137], while another found minor enrichment of common and rare missense variation across a combined GWAS-implicated gene set[134]. Finally, Jansen et al detected rare variant enrichment in both LRRK2, STDB1 and SPATA19[146]. Another approach has been to combine results from GWAS with data on expression quantitative trait loci (eQTL), thereby nominating candidate genes whose expression is regulated by PD-associated SNPs[352-356]. Additional discoveries have been made by also incorporating data on splicing QTL (sQTL) and methylation QTL (mQTL)[357,358]. Future discoveries will most likely be made by integrating multiple types of data (eQTL, sQTL, mQTL) with sequencing data.

7. Conclusions

- Familial aggregation of PD is present in a population-based cohort of incidental PD, but to a lesser extent than what would be expected based on previous estimates. This is likely due to a combination of more accurate family history data, exclusion of monogenic PD cases and a relatively high proportion of late-onset cases in our cohort.

- A positive family history of PD among first-degree relatives seems to modulate disease progression, as measured by UPDRS II and MMSE, resulting in a slightly milder phenotype.

- A family history of PD does not impact disease subtype (tremor dominant vs PIGD) or age of onset.

- *TRAP1*-mutations are not associated with PD in the ParkWest cohort, and the reported association in the PPMI cohort is questionable, as stricter thresholds in the individual and variant quality control procedures renders it undetectable.

- Rare, protein-altering mutations in genes encoding proteins involved in the repair and synthesis of mtDNA are associated with PD. Impaired mtDNA maintenance likely plays an important role in the pathogenesis of PD, and some of this impairment may be due to inherited genetic variation.

- Successful genome-wide, gene-based rare variant enrichment studies in PD will likely require sample sizes in excess of 8,000 to 11,000 individuals.

- A targeted analysis of rare variant enrichment in genes suspected of being involved in PD pathophysiology through their proximity to GWAS hits does not support a major role for rare variant enrichment in these genes.

8. Errata

Paper I: In the results, both the sensitivity and positive predictive value of the baseline questionnaire is reported as being 76.6% when compared to the extended questionnaire. This is incorrect and should be 69.4%.

9. References

1. Parkinson J. An essay on the shaking palsy. 1817. *J Neuropsychiatry Clin Neurosci.* 2002;14(2):223-236; discussion 222.
2. Kalia LV, Lang AE. Parkinson's disease. *Lancet.* 2015;386(9996):896-912.
3. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* 2015;30(12):1591-1601.
4. de Rijk MC, Launer LJ, Berger K, et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology.* 2000;54(11 Suppl 5):S21-23.
5. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord.* 2014;29(13):1583-1590.
6. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. *Lancet Neurol.* 2016;15(12):1257-1272.
7. Elbaz A, Bower JH, Maraganore DM, et al. Risk tables for parkinsonism and Parkinson's disease. *J Clin Epidemiol.* 2002;55(1):25-31.
8. Hamilton J, The Michael J. Fox Foundation for Parkinson's Research, Yang W, The Lewin Group. The Economic Burden of Parkinson’s Disease. Published 2019. Accessed September 6, 2019.
9. Dorsey ER, Elbaz A, Nichols E, et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology.* 2018;17(11):939-953.
10. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. *BMC Res Notes.* 2019;12(1):106.
11. Francis WR, Wörheide G. Similar Ratios of Introns to Intergenic Sequence across Animal Genomes. *Genome Biol Evol.* 2017;9(6):1582-1598.
12. Piovesan A, Caracausi M, Antonaros F, Pelleri MC, Vitale L. GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database (Oxford).* 2016;2016.
13. The 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
14. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science.* 2001;291(5507):1304-1351.
15. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.
16. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics.* 2016;107(1):1-8.
17. Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. *Nat Rev Genet.* 2013;14(11):794-806.
18. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet.* 2002;3(4):299-309.
19. International HapMap C. A haplotype map of the human genome. *Nature.* 2005;437(7063):1299-1320.
20. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101(1):5-22.

21. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012.
22. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019;20(8):467-484.
23. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.
24. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology.* 2019;18(12):1091-1102.
25. Auer PL, Reiner AP, Wang G, et al. Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project. *American journal of human genetics.* 2016;99(4):791-801.
26. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) www.genome.gov/sequencingcostsdata. Accessed September 5, 2019.
27. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell.* 2013;155(1):27-38.
28. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014;95(1):5-23.
29. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature.* 2014;506(7487):185-190.
30. Flannick J, Mercader JM, Fuchsberger C, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. 2019;570(7759):71-76.
31. Nicolas A, Kenna KP, Renton AE, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron.* 2018;97(6):1268-1283 e1266.
32. Fritsche LG, Igl W, Bailey JN, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet.* 2016;48(2):134-143.
33. Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nature Genetics.* 2016;48:1162.
34. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med.* 2002;21(1):35-50.
35. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics.* 2016;24(8):1202-1205.
36. Etymologia: Bonferroni correction. *Emerg Infect Dis.* 2015;21(2):289-289.
37. Nicolae DL. Association Tests for Rare Variants. *Annu Rev Genomics Hum Genet.* 2016;17:117-130.
38. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res.* 2007;615(1-2):28-56.
39. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American journal of human genetics.* 2008;83(3):311-321.

40. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 2009;5(2):e1000384.
41. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 2011;89(1):82-93.
42. Price AL, Kryukov GV, de Bakker PIW, et al. Pooled association tests for rare variants in exon-resequencing studies. *American journal of human genetics.* 2010;86(6):832-838.
43. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature Methods.* 2010;7(4):248-249.
44. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics.* 2014;46:310.
45. Han F, Pan W. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered.* 2010;70(1):42-54.
46. Lin D-Y, Tang Z-Z. A general framework for detecting disease associations with rare variants in sequencing studies. *American journal of human genetics.* 2011;89(3):354-367.
47. Hoffmann TJ, Marini NJ, Witte JS. Comprehensive approach to analyzing rare genetic variants. *PloS one.* 2010;5(11):e13584-e13584.
48. Neale BM, Rivas MA, Voight BF, et al. Testing for an unusual distribution of rare variants. *PLoS genetics.* 2011;7(3):e1001322-e1001322.
49. Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genetic epidemiology.* 2009;33(6):497-507.
50. Basu S, Pan W. Comparison of statistical tests for disease association with rare variants. *Genetic epidemiology.* 2011;35(7):606-619.
51. Derkach A, Lawless JF, Sun L. Robust and Powerful Tests for Rare Variants Using Fisher's Method to Combine Evidence of Association From Two or More Complementary Tests. *Genetic Epidemiology.* 2013;37(1):110-121.
52. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet.* 2012;91(2):224-237.
53. Sun J, Zheng Y, Hsu L. A unified mixed-effects model for rare-variant association in sequencing studies. *Genetic epidemiology.* 2013;37(4):334-344.
54. Lee J, Kim YJ, Lee J, et al. Gene-set association tests for next-generation sequencing data. *Bioinformatics.* 2016;32(17):i611-i619.
55. Chen LS, Hsu L, Gamazon ER, Cox NJ, Nicolae DL. An exponential combination procedure for set-based association tests in sequencing studies. *American journal of human genetics.* 2012;91(6):977-986.
56. Ionita-Laza I, Buxbaum JD, Laird NM, Lange C. A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS genetics.* 2011;7(2):e1001289-e1001289.
57. Dutta D, Scott L, Boehnke M, Lee S. Multi-SKAT: General framework to test for rare-variant association with multiple phenotypes. *Genet Epidemiol.* 2019;43(1):4-23.
58. Chen L, Wang Y, Zhou Y. Association analysis of rare and common variants with multiple traits based on variable reduction method. *Genetics Research.* 2018;100:e2.

59. Chung R-H, Kang C-Y. A Powerful Gene-Based Test Accommodating Common and Low-Frequency Variants to Detect Both Main Effects and Gene-Gene Interaction Effects in Case-Control Studies. *Front Genet.* 2018;8:228-228.
60. Chen Z, Boehnke M, Fuchsberger C. Combining sequence data from multiple studies: Impact of analysis strategies on rare variant calling and association results. *Genetic Epidemiology.* 2019;0(0).
61. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet.* 2013;93(1):42-53.
62. Fisher RA. Statistical methods for research workers. 1930.
63. Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams Jr RM. The american soldier: Adjustment during army life.(studies in social psychology in world war ii), vol. 1. 1949.
64. Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol.* 2011;24(8):1836-1841.
65. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics.* 2014;30(19):2828-2829.
66. Ka J, Lee J, Kim Y, Oh B, Consortium TDG, Park T. Meta-Qtest: meta-analysis of quadratic test for rare variants. *BMC Med Genomics.* 2019;12(Suppl 5):102-102.
67. Gowers WR. *A manual of diseases of the nervous system.* London: J. & A. Churchill; 1886.
68. Mjönes H. Paralysis agitans: a clinical and genetic study. *Acta Psychiatr Neurol Scand.* 1949;54:1-195.
69. Duvoisin RC. On heredity, twins, and Parkinson's disease. *Annals of Neurology.* 1986;19(4):409-411.
70. Duvoisin RC, Eldridge R, Williams A, Nutt J, Calne D. Twin study of Parkinson disease. *Neurology.* 1981;31(1):77.
71. Marttila RJ, Kaprio J, Koskenvuo M, Rinne UK. Parkinson disease in a nationwide twin cohort. *Neurology.* 1988;38(8):1217.
72. Tanner CM, Ottman R, Goldman SM, et al. Parkinson disease in twins: an etiologic study. *JAMA.* 1999;281(4):341-346.
73. Wirdefeldt K, Gatz M, Schalling M, Pedersen NL. No evidence for heritability of Parkinson disease in Swedish twins. *Neurology.* 2004;63(2):305-311.
74. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol Aging.* 2011;32(10):1923 e1921-1928.
75. Noyce AJ, Lees AJ, Schrag A-E. The prediagnostic phase of Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry.* 2016;87:871.
76. Piccini P, Burn DJ, Ceravolo R, Maraganore D, Brooks DJ. The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. *Ann Neurol.* 1999;45(5):577-582.
77. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the α-Synuclein Gene Identified in Families with Parkinson's Disease. *Science.* 1997;276(5321):2045.
78. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol.* 2019(1474-4465 (Electronic)).
79. Trinh J, Zeldenrust FMJ, Huang J, et al. Genotype-phenotype relations for the Parkinson's disease genes SNCA, LRRK2, VPS35: MDSGene systematic review. *Movement Disorders.* 2018;33(12):1857-1870.

80. Kasten M, Hartmann C, Hampf J, et al. Genotype-Phenotype Relations for the Parkinson's Disease Genes Parkin, PINK1, DJ1: MDSGene Systematic Review. *Movement Disorders.* 2018;33(5):730-741.
81. Shojaee S, Sina F, Banihosseini SS, et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. *Am J Hum Genet.* 2008;82(6):1375-1384.
82. Fonzo AD, Dekker MCJ, Montagna P, et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology.* 2009;72(3):240.
83. Najim Al-Din AS, Wriekat A, Mubaidin A, Dasouki M, Hiari M. Pallido-pyramidal degeneration, supranuclear upgaze paresis and dementia: Kufor-Rakeb syndrome. *Acta Neurologica Scandinavica.* 1994;89(5):347-352.
84. Ramirez A, Heimbach A, Gründemann J, et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nature Genetics.* 2006;38(10):1184-1191.
85. Edvardson S, Cinnamon Y, Ta-Shma A, et al. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. *PloS one.* 2012;7(5):e36458-e36458.
86. Koroglu C, Baysal L, Cetinkaya M, Karasoy H, Tolun A. DNAJC6 is responsible for juvenile parkinsonism with phenotypic variability. *Parkinsonism Relat Disord.* 2013;19(3):320-324.
87. Krebs CE, Karkheiran S, Powell JC, et al. The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. *Human mutation.* 2013;34(9):1200-1207.
88. Olgiati S, De Rosa A, Quadri M, et al. PARK20 caused by SYNJ1 homozygous Arg258Gln mutation in a new Italian family. *neurogenetics.* 2014;15(3):183-188.
89. Kirola L, Behari M, Shishir C, Thelma BK. Identification of a novel homozygous mutation Arg459Pro in SYNJ1 gene of an Indian family with autosomal recessive juvenile Parkinsonism. *Parkinsonism Relat Disord.* 2016;31:124-128.
90. Rauschendorf M-A, Jost M, Stock F, et al. Novel compound heterozygous synaptojanin-1 mutation causes l-dopa-responsive dystonia-parkinsonism syndrome. *Movement Disorders.* 2017;32(3):478-480.
91. Lesage S, Drouet V, Majounie E, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *Am J Hum Genet.* 2016;98(3):500-513.
92. Morgan NV, Westaway SK, Morton JEV, et al. PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. *Nature Genetics.* 2006;38(7):752-754.
93. Paisan-Ruiz C, Bhatia KP, Li A, et al. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. *Annals of Neurology.* 2009;65(1):19-23.
94. Karkheiran S, Shahidi GA, Walker RH, Paisán-Ruiz C. PLA2G6-associated Dystonia-Parkinsonism: Case Report and Literature Review. *Tremor Other Hyperkinet Mov (N Y).* 2015;5:317-317.
95. Shen T, Hu J, Jiang Y, et al. Early-Onset Parkinson's Disease Caused by PLA2G6 Compound Heterozygous Mutation, a Case Report and Literature Review. *Front Neurol.* 2019;10:915-915.

96. Funayama M, Ohe K, Amo T, et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet Neurol.* 2015;14(3):274-282.
97. Shi CH, Mao CY, Zhang SY, et al. CHCHD2 gene mutations in familial and sporadic Parkinson's disease. *Neurobiol Aging.* 2016;38:217 e219-217 e213.
98. Palin EJH, Paetau A, Suomalainen A. Mesencephalic complex I deficiency does not correlate with parkinsonism in mitochondrial DNA maintenance disorders. *Brain.* 2013;136(8):2379-2392.
99. Tzoulis C, Tran GT, Schwarzlmuller T, et al. Severe nigrostriatal degeneration without clinical parkinsonism in patients with polymerase gamma mutations. *Brain.* 2013;136(Pt 8):2393-2404.
100. Miguel R, Gago MF, Martins J, Barros P, Vale J, Rosas MJ. POLG1-related levodopa-responsive parkinsonism. *Clin Neurol Neurosurg.* 2014;126:47-54.
101. Konno T, Ross OA, Teive HAG, Sławek J, Dickson DW, Wszolek ZK. DCTN1-related neurodegeneration: Perry syndrome and beyond. *Parkinsonism & related disorders.* 2017;41:14-24.
102. Sullivan R, Yau WY, O'Connor E, Houlden H. Spinocerebellar ataxia: an update. *Journal of Neurology.* 2019;266(2):533-544.
103. Paulson HL. The spinocerebellar ataxias. *J Neuroophthalmol.* 2009;29(3):227-237.
104. Schöls L, Reimold M, Seidel K, et al. No parkinsonism in SCA2 and SCA3 despite severe neurodegeneration of the dopaminergic substantia nigra. *Brain.* 2015;138(11):3316-3326.
105. Choy KR, Watters DJ. Neurodegeneration in ataxia-telangiectasia: Multiple roles of ATM kinase in cellular homeostasis. *Developmental Dynamics.* 2018;247(1):33-46.
106. Carelli V, Musumeci O, Caporali L, et al. Syndromic parkinsonism and dementia associated with OPA1 missense mutations. *Annals of Neurology.* 2015;78(1):21-38.
107. Deng H, Wang P, Jankovic J. The genetics of Parkinson disease. *Ageing Res Rev.* 2018;42:72-85.
108. Healy DG, Falchi M, O'Sullivan SS, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol.* 2008;7(7):583-590.
109. Marder K, Wang Y, Alcalay RN, et al. Age-specific penetrance of LRRK2 G2019S in the Michael J. Fox Ashkenazi Jewish LRRK2 Consortium. *Neurology.* 2015;85(1):89-95.
110. Lee AJ, Wang Y, Alcalay RN, et al. Penetrance estimate of LRRK2 p.G2019S mutation in individuals of non-Ashkenazi Jewish ancestry. *Mov Disord.* 2017;32(10):1432-1438.
111. Thacker EL, Ascherio A. Familial aggregation of Parkinson's disease: a meta-analysis. *Mov Disord.* 2008;23(8):1174-1183.
112. Savica R, Cannon-Albright LA, Pulst S. Familial aggregation of Parkinson disease in Utah: A population-based analysis using death certificates. *Neurology Genetics.* 2016;2(2):e65-e65.
113. Barrett MJ, Hac NE, Yan G, Harrison MB, Wooten GF. Relationship of age of onset and family history in Parkinson disease. *Movement Disorders.* 2015;30(5):733-735.
114. Lücking CB, Dürr A, Bonifati V, et al. Association between Early-Onset Parkinson's Disease and Mutations in the Parkin Gene. *New England Journal of Medicine.* 2000;342(21):1560-1567.

115. Fung HC, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 2006;5(11):911-916.
116. Pankratz N, Wilk JB, Latourelle JC, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Human Genetics.* 2009;124(6):593-605.
117. Simón-Sánchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genetics.* 2009;41(12):1308-1312.
118. Billingsley KJ, Bandres-Ciga S, Saez-Atienzar S, Singleton AB. Genetic risk factors in Parkinson's disease. *Cell Tissue Res.* 2018;373(1):9-20.
119. Halperin A, Elstein D, Zimran A. Increased incidence of Parkinson disease among relatives of patients with Gaucher disease. *Blood Cells, Molecules, and Diseases.* 2006;36(3):426-428.
120. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *The New England journal of medicine.* 2009;361(17):1651-1661.
121. Chang D, Nalls MA, Hallgrimsdottir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017;49(10):1511-1516.
122. Zimprich A, Benet-Pagès A, Struhal W, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet.* 2011;89(1):168-175.
123. Vilariño-Güell C, Wider C, Ross OA, et al. VPS35 mutations in Parkinson disease. *Am J Hum Genet.* 2011;89(1):162-167.
124. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res.* 2018;27(2):e1608-e1608.
125. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Human Molecular Genetics.* 2012;21(R1):R1-R9.
126. Jansen IE, Gibbs JR, Nalls MA, et al. Establishing the role of rare coding variants in known Parkinson's disease risk loci. *Neurobiol Aging.* 2017;59:220 e211-220 e218.
127. Nuytemans K, Bademci G, Inchausti V, et al. Whole exome sequencing of rare variants in EIF4G1 and VPS35 in Parkinson disease. *Neurology.* 2013;80(11):982-989.
128. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nature Genetics.* 2000;25(1):25-29.
129. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research.* 2018;47(D1):D330-D338.
130. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Research.* 2018;47(D1):D590-D595.
131. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Research.* 2017;46(D1):D649-D655.
132. Rouillard AD, Gundersen GW, Fernandez NF, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database.* 2016;2016.

133. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research.* 2018;47(D1):D419-D426.
134. Sandor C, Honti F, Haerty W, et al. Whole-exome sequencing of 228 patients with sporadic Parkinson's disease. *Sci Rep.* 2017;7:41188.
135. Robak LA, Jansen IE, van Rooij J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain.* 2017;140(12):3191-3203.
136. Gaare JJ, Nido GS, Sztromwasser P, et al. Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease. *Mov Disord.* 2018;33(10):1591-1600.
137. Foo JN, Tan LC, Liany H, et al. Analysis of non-synonymous-coding variants of Parkinson's disease-related pathogenic and susceptibility genes in East Asian populations. *Human Molecular Genetics.* 2014;23(14):3891-3897.
138. Mencacci NE, Isaias IU, Reich MM, et al. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain : a journal of neurology.* 2014;137(Pt 9):2480-2492.
139. Quadri M, Yang X, Cossu G, et al. An exome study of Parkinson's disease in Sardinia, a Mediterranean genetic isolate. *neurogenetics.* 2015;16(1):55-64.
140. Petersen MS, Guella I, Bech S, Gustavsson E, Farrer MJ. Parkinson's disease, genetic variability and the Faroe Islands. *Parkinsonism Relat Disord.* 2015;21(1):75-78.
141. Chen KH, Wu RM, Lin HI, Tai CH, Lin CH. Mutational analysis of SYNJ1 gene (PARK20) in Parkinson's disease in a Taiwanese population. *Neurobiol Aging.* 2015;36(10):2905 e2907-2908.
142. Simon-Sanchez J, Heutink P, Gasser T, International Parkinson's Disease Genomics C. Variation in PARK10 is not associated with risk and age at onset of Parkinson's disease in large clinical cohorts. *Neurobiol Aging.* 2015;36(10):2907 e2913-2907.
143. Farlow JL, Robak LA, Hetrick K, et al. Whole-Exome Sequencing in Familial Parkinson Disease. *JAMA Neurology.* 2016;73(1):68-75.
144. Nuytemans K, Maldonado L, Ali A, et al. Overlap between Parkinson disease and Alzheimer disease in ABCA7 functional variants. *Neurol Genet.* 2016;2(1):e44.
145. Lubbe SJ, Escott-Price V, Brice A, et al. Rare variants analysis of cutaneous malignant melanoma genes in Parkinson's disease. *Neurobiology of aging.* 2016;48:222.e221-222.e227.
146. Jansen IE, Ye H, Heetveld S, et al. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol.* 2017;18(1):22.
147. Siitonen A, Nalls MA, Hernandez D, et al. Genetics of early-onset Parkinson's disease in Finland: exome sequencing and genome-wide association study. *Neurobiol Aging.* 2017.
148. Giri A, Mok KY, Jansen I, et al. Lack of evidence for a role of genetic variation in TMEM230 in the risk for Parkinson's disease in the Caucasian population. *Neurobiology of aging.* 2017;50:167.e111-167.e113.
149. Ylönen S, Siitonen A, Nalls MA, et al. Genetic risk factors in Finnish patients with Parkinson's disease. *Parkinsonism & related disorders.* 2017;45:39-43.
150. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.
151. Guo J-f, Zhang L, Li K, et al. Coding mutations in NUS1 contribute to Parkinson's disease. *Proceedings of the National Academy of Sciences.* 2018;115(45):11567.

152. Chew EGY, Liany H, Tan LCS, et al. Evaluation of novel Parkinson's disease candidate genes in the Chinese population. *Neurobiology of Aging.* 2019;74:235.e231-235.e234.
153. Siitonen A, Kytövuori L, Nalls MA, et al. Finnish Parkinson's disease study integrating protein-protein interaction network data with exome sequencing analysis. *Scientific reports.* 2019;9(1):18865-18865.
154. Rudakou U, Ruskey JA, Krohn L, et al. Analysis of common and rare VPS13C variants in late-onset Parkinson disease. *Neurology Genetics.* 2020;6(1):385-385.
155. Gialluisi A, Reccia MG, Tirozzi A, et al. Whole Exome Sequencing Study of Parkinson Disease and Related Endophenotypes in the Italian Population. *Front Neurol.* 2020;10:1362-1362.
156. Oluwole OG, Kuivaniemi H, Abrahams S, et al. Targeted next-generation sequencing identifies novel variants in candidate genes for Parkinson's disease in Black South African and Nigerian patients. *BMC Med Genet.* 2020;21(1):23-23.
157. Bobbili DR, Banda P, Krüger R, May P. Excess of singleton loss-of-function variants in Parkinson's disease contributes to genetic risk. *Journal of Medical Genetics.* 2020:jmedgenet-2019-106316.
158. Gaare JJ, Nido G, Dolle C, et al. Meta-analysis of whole-exome sequencing data from two independent cohorts finds no evidence for rare variant enrichment in Parkinson disease associated loci. *PLoS One.* 2020;15(10):e0239824.
159. Guzman JN, Sánchez-Padilla J, Chan CS, Surmeier DJ. Robust pacemaking in substantia nigra dopaminergic neurons. *J Neurosci.* 2009;29(35):11011-11019.
160. Cheng H-C, Ulane CM, Burke RE. Clinical progression in Parkinson disease and the neurobiology of axons. *Annals of Neurology.* 2010;67(6):715-725.
161. Giguère N, Burke Nanni S, Trudeau L-E. On Cell Loss and Selective Vulnerability of Neuronal Populations in Parkinson's Disease. *Front Neurol.* 2018;9:455.
162. Baumuratov AS, Antony PMA, Ostaszewski M, et al. Enteric neurons from Parkinson's disease patients display ex vivo aberrations in mitochondrial structure. *Scientific Reports.* 2016;6(1):33117.
163. Goedert M, Spillantini MG, Del Tredici K, Braak H. 100 years of Lewy pathology. *Nature Reviews Neurology.* 2013;9(1):13-24.
164. Spillantini MG, Schmidt ML, Lee VMY, Trojanowski JQ, Jakes R, Goedert M. α-Synuclein in Lewy bodies. *Nature.* 1997;388(6645):839-840.
165. Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proceedings of the National Academy of Sciences of the United States of America.* 1998;95(11):6469-6473.
166. Wang C, Zhao C, Li D, et al. Versatile Structures of α-Synuclein. *Frontiers in Molecular Neuroscience.* 2016;9:48.
167. Dettmer U, Newman AJ, Soldner F, et al. Parkinson-causing α-synuclein missense mutations shift native tetramers to monomers as a mechanism for disease initiation. *Nature Communications.* 2015;6(1):7314.
168. Chartier S, Duyckaerts C. Is Lewy pathology in the human nervous system chiefly an indicator of neuronal protection or of toxicity? *Cell and Tissue Research.* 2018;373(1):149-160.

169. Wakabayashi K, Tanji K, Odagiri S, Miki Y, Mori F, Takahashi H. The Lewy Body in Parkinson's Disease and Related Neurodegenerative Disorders. *Molecular Neurobiology.* 2013;47(2):495-508.
170. Dickson DW. Neuropathology of Parkinson disease. *Parkinsonism Relat Disord.* 2018;46 Suppl 1:S30-S33.
171. Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging.* 2003;24(2):197-211.
172. Braak H, Ghebremedhin E, Rüb U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. *Cell and Tissue Research.* 2004;318(1):121-134.
173. Hawkes CH, Del Tredici K, Braak H. Parkinson's disease: a dual-hit hypothesis. *Neuropathology and Applied Neurobiology.* 2007;33(6):599-614.
174. Jellinger KA. A critical reappraisal of current staging of Lewy-related pathology in human brain. *Acta Neuropathologica.* 2008;116(1):1.
175. Parkkinen L, Pirttilä T, Alafuzoff I. Applicability of current staging/categorization of α-synuclein pathology and their clinical relevance. *Acta Neuropathologica.* 2008;115(4):399-407.
176. Rietdijk CD, Perez-Pardo P, Garssen J, van Wezel RJA, Kraneveld AD. Exploring Braak's Hypothesis of Parkinson's Disease. *Front Neurol.* 2017;8:37.
177. Parkkinen L, O'Sullivan SS, Collins C, et al. Disentangling the relationship between lewy bodies and nigral neuronal loss in Parkinson's disease. *J Parkinsons Dis.* 2011;1(3):277-286.
178. Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. *Nature Reviews Neuroscience.* 2017;18:101.
179. Li J-Y, Englund E, Holton JL, et al. Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation. *Nature Medicine.* 2008;14(5):501-503.
180. Kordower JH, Chu Y, Hauser RA, Freeman TB, Olanow CW. Lewy body–like pathology in long-term embryonic nigral transplants in Parkinson's disease. *Nature Medicine.* 2008;14(5):504-506.
181. Volpicelli-Daley LA, Luk KC, Patel TP, et al. Exogenous alpha-synuclein fibrils induce Lewy body pathology leading to synaptic dysfunction and neuron death. *Neuron.* 2011;72(1):57-71.
182. Luk KC, Kehm V, Carroll J, et al. Pathological α-Synuclein Transmission Initiates Parkinson-like Neurodegeneration in Nontransgenic Mice. *Science.* 2012;338(6109):949.
183. Killinger BA, Kordower JH. Spreading of alpha-synuclein – relevant or epiphenomenon? *Journal of Neurochemistry.* 2019;150(5):605-611.
184. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: Review of the literature. *Movement Disorders.* 2017;32(11):1504-1523.
185. Pasanen P, Myllykangas L, Siitonen M, et al. Novel alpha-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol Aging.* 2014;35(9):2180 e2181-2185.
186. Kiely AP, Ling H, Asi YT, et al. Distinct clinical and neuropathological features of G51D SNCA mutation cases compared with SNCA duplication and H50Q mutation. *Molecular Neurodegeneration.* 2015;10(1):41.

187. Vilas D, Sharp M, Gelpi E, et al. Clinical and neuropathological features of progressive supranuclear palsy in Leucine rich repeat kinase (LRRK2) G2019S mutation carriers. *Movement Disorders.* 2018;33(2):335-338.
188. Johansen KK, Torp SH, Farrer MJ, Gustavsson EK, Aasly JO. A Case of Parkinson's Disease with No Lewy Body Pathology due to a Homozygous Exon Deletion in Parkin. *Case Rep Neurol Med.* 2018;2018:6838965-6838965.
189. Tzoulis C, Tran GT, Coxhead J, et al. Molecular pathogenesis of polymerase gamma–related neurodegeneration. *Annals of Neurology.* 2014;76(1):66-81.
190. Reeve A, Meagher M, Lax N, et al. The Impact of Pathogenic Mitochondrial DNA Mutations on Substantia Nigra Neurons. *The Journal of Neuroscience.* 2013;33(26):10790.
191. Palin EJ, Paetau A, Suomalainen A. Mesencephalic complex I deficiency does not correlate with parkinsonism in mitochondrial DNA maintenance disorders. *Brain.* 2013;136(Pt 8):2379-2392.
192. Mishima T, Fujioka S, Tomiyama H, et al. Establishing diagnostic criteria for Perry syndrome. *Journal of Neurology, Neurosurgery & Psychiatry.* 2018;89(5):482.
193. Wider C, Dickson DW, Stoessl AJ, et al. Pallidonigral TDP-43 pathology in Perry syndrome. *Parkinsonism & related disorders.* 2009;15(4):281-286.
194. Verhagen MMM, Martin J-J, van Deuren M, et al. Neuropathology in classical and variant ataxia-telangiectasia. *Neuropathology.* 2012;32(3):234-244.
195. Wider C, Skipper L, Solida A, et al. Autosomal dominant dopa-responsive parkinsonism in a multigenerational Swiss family. *Parkinsonism Relat Disord.* 2008;14(6):465-470.
196. Narendra DP, Isonaka R, Nguyen D, et al. Peripheral synucleinopathy in a DJ1 patient with Parkinson disease, cataracts, and hearing loss. *Neurology.* 2019;92(23):1113.
197. Bras J, Verloes A, Schneider SA, Mole SE, Guerreiro RJ. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. *Human molecular genetics.* 2012;21(12):2646-2650.
198. Konno T, Ross OA, Puschmann A, Dickson DW, Wszolek ZK. Autosomal dominant Parkinson's disease caused by SNCA duplications. *Parkinsonism & related disorders.* 2016;22 Suppl 1(Suppl 1):S1-S6.
199. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003;302(5646):841.
200. Ferese R, Modugno N, Campopiano R, et al. Four Copies of SNCA Responsible for Autosomal Dominant Parkinson's Disease in Two Italian Siblings. *Parkinson's disease.* 2015;2015:546462-546462.
201. Burré J. The Synaptic Function of α-Synuclein. *J Parkinsons Dis.* 2015;5(4):699-713.
202. Zhang G, Xia Y, Wan F, et al. New Perspectives on Roles of Alpha-Synuclein in Parkinson's Disease. *Frontiers in aging neuroscience.* 2018;10:370-370.
203. Vicario M, Cieri D, Brini M, Calì T. The Close Encounter Between Alpha-Synuclein and Mitochondria. *Frontiers in Neuroscience.* 2018;12:388.
204. Betarbet R, Canet-Aviles RM, Sherer TB, et al. Intersecting pathways to neurodegeneration in Parkinson's disease: effects of the pesticide rotenone on DJ-1, alpha-synuclein, and the ubiquitin-proteasome system. *Neurobiol Dis.* 2006;22(2):404-420.
205. Ruan L, Zhou C, Jin E, et al. Cytosolic proteostasis through importing of misfolded proteins into mitochondria. *Nature.* 2017;543:443.

206. Kluss JH, Mamais A, Cookson MR. LRRK2 links genetic and sporadic Parkinson's disease. *Biochemical Society Transactions.* 2019;47(2):651-661.
207. Wallings R, Manzoni C, Bandopadhyay R. Cellular processes associated with LRRK2 function and dysfunction. *The FEBS Journal.* 2015;282(15):2806-2826.
208. Singh A, Zhi L, Zhang H. LRRK2 and mitochondria: Recent advances and current views. *Brain Res.* 2019;1702:96-104.
209. Williams ET, Chen X, Moore DJ. VPS35, the Retromer Complex and Parkinson's Disease. *J Parkinsons Dis.* 2017;7(2):219-233.
210. Tsika E, Glauser L, Moser R, et al. Parkinson's disease-linked mutations in VPS35 induce dopaminergic neurodegeneration. *Human Molecular Genetics.* 2014;23(17):4621-4638.
211. Tang F-L, Erion JR, Tian Y, et al. VPS35 in Dopamine Neurons Is Required for Endosome-to-Golgi Retrieval of Lamp2a, a Receptor of Chaperone-Mediated Autophagy That Is Critical for α-Synuclein Degradation and Prevention of Pathogenesis of Parkinson's Disease. *The Journal of Neuroscience.* 2015;35(29):10613.
212. Zhou Z-D, Saw W-T, Tan E-K. Mitochondrial CHCHD-Containing Proteins: Physiologic Functions and Link with Neurodegenerative Diseases. *Molecular Neurobiology.* 2017;54(7):5534-5546.
213. Bannwarth S, Ait-El-Mkadem S, Chaussenot A, et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain.* 2014;137(8):2329-2345.
214. Aras S, Pak O, Sommer N, et al. Oxygen-dependent expression of cytochrome c oxidase subunit 4-2 gene expression is mediated by transcription factors RBPJ, CXXC5 and CHCHD2. *Nucleic Acids Research.* 2013;41(4):2255-2266.
215. Lee RG, Sedghi M, Salari M, et al. Early-onset Parkinson disease caused by a mutation in CHCHD2 and mitochondrial dysfunction. *Neurology Genetics.* 2018;4(5):e276-e276.
216. Ikeda A, Nishioka K, Meng H, et al. Mutations in CHCHD2 cause α-synuclein aggregation. *Human Molecular Genetics.* 2019.
217. Truban D, Hou X, Caulfield TR, Fiesel FC, Springer W. PINK1, Parkin, and Mitochondrial Quality Control: What can we Learn about Parkinson's Disease Pathobiology? *J Parkinsons Dis.* 2017;7(1):13-29.
218. Hao L-Y, Giasson BI, Bonini NM. DJ-1 is critical for mitochondrial function and rescues PINK1 loss of function. *Proceedings of the National Academy of Sciences.* 2010;107(21):9747.
219. Thomas KJ, McCoy MK, Blackinton J, et al. DJ-1 acts in parallel to the PINK1/parkin pathway to control mitochondrial function and autophagy. *Human Molecular Genetics.* 2010;20(1):40-50.
220. Bonello F, Hassoun S-M, Mouton-Liger F, et al. LRRK2 impairs PINK1/Parkin-dependent mitophagy via its kinase activity: pathologic insights into Parkinson's disease. *Human Molecular Genetics.* 2019;28(10):1645-1660.
221. Di Maio R, Barrett PJ, Hoffman EK, et al. α-Synuclein binds to TOM20 and inhibits mitochondrial protein import in Parkinson's disease. *Science Translational Medicine.* 2016;8(342):342ra378.
222. Park J-S, Blair NF, Sue CM. The role of ATP13A2 in Parkinson's disease: Clinical phenotypes and molecular mechanisms. *Movement Disorders.* 2015;30(6):770-779.

223. Park J-S, Koentjoro B, Veivers D, Mackay-Sim A, Sue CM. Parkinson's disease-associated human ATP13A2 (PARK9) deficiency causes zinc dyshomeostasis and mitochondrial dysfunction. *Human Molecular Genetics.* 2014;23(11):2802-2815.
224. Kong SMY, Chan BKK, Park J-S, et al. Parkinson's disease-linked human PARK9/ATP13A2 maintains zinc homeostasis and promotes α-Synuclein externalization via exosomes. *Human Molecular Genetics.* 2014;23(11):2816-2833.
225. Vidyadhara DJ, Lee JE, Chandra SS. Role of the endolysosomal system in Parkinson's disease. *Journal of Neurochemistry.* 2019;150(5):487-506.
226. Lin G, Lee PT, Chen K, et al. Phospholipase PLA2G6, a Parkinsonism-Associated Gene, Affects Vps26 and Vps35, Retromer Function, and Ceramide Levels, Similar to alpha-Synuclein Gain. *Cell Metab.* 2018;28(4):605-618 e606.
227. Chiu C-C, Lu C-S, Weng Y-H, et al. PARK14 (D331Y) PLA2G6 Causes Early-Onset Degeneration of Substantia Nigra Dopaminergic Neurons by Inducing Mitochondrial Dysfunction, ER Stress, Mitophagy Impairment and Transcriptional Dysregulation in a Knockin Mouse Model. *Molecular Neurobiology.* 2019;56(6):3835-3853.
228. Kinghorn KJ, Castillo-Quan JI, Bartolome F, et al. Loss of PLA2G6 leads to elevated mitochondrial lipid peroxidation and mitochondrial dysfunction. *Brain.* 2015;138(7):1801-1816.
229. Viscomi C, Zeviani M. MtDNA-maintenance defects: syndromes and genes. *Journal of Inherited Metabolic Disease.* 2017;40(4):587-599.
230. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research.* 2018;47(D1):D506-D515.
231. Davis GC, Williams AC, Markey SP, et al. Chronic Parkinsonism secondary to intravenous injection of meperidine analogues. *Psychiatry Res.* 1979;1(3):249-254.
232. Langston JW. The MPTP Story. *J Parkinsons Dis.* 2017;7(s1):S11-S19.
233. Langston JW, Ballard P, Tetrud JW, Irwin I. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science.* 1983;219(4587):979.
234. Langston JW, Irwin I, Langston EB, Forno LS. 1-Methyl-4-phenylpyridinium ion (MPP+): identification of a metabolite of MPTP, a toxin selective to the substantia nigra. *Neurosci Lett.* 1984;48(1):87-92.
235. Ramsay RR, Salach JI, Singer TP. Uptake of the neurotoxin 1-methyl-4-phenylpyridine (MPP+) by mitochondria and its relation to the inhibition of the mitochondrial oxidation of NAD+-linked substrates by MPP+. *Biochem Biophys Res Commun.* 1986;134(2):743-748.
236. Schapira AHV, Cooper JM, Dexter D, Clark JB, Jenner P, Marsden CD. Mitochondrial Complex I Deficiency in Parkinson's Disease. *Journal of Neurochemistry.* 1990;54(3):823-827.
237. Xiong N, Long X, Xiong J, et al. Mitochondrial complex I inhibitor rotenone-induced toxicity and its potential mechanisms in Parkinson's disease models. *Critical Reviews in Toxicology.* 2012;42(7):613-632.
238. Betarbet R, Sherer TB, MacKenzie G, Garcia-Osuna M, Panov AV, Greenamyre JT. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature Neuroscience.* 2000;3(12):1301-1306.
239. Sousa JS, D'Imprima E, Vonck J. Mitochondrial Respiratory Chain Complexes. In: Harris JR, Boekema EJ, eds. *Membrane Protein Complexes: Structure and Function.* Singapore: Springer Singapore; 2018:167-227.

240. Cosi C, Marien M. Decreases in mouse brain NAD+ and ATP induced by 1-methyl-4-phenyl-1, 2,3,6-tetrahydropyridine (MPTP): prevention by the poly(ADP-ribose) polymerase inhibitor, benzamide. *Brain Res.* 1998;809(1):58-67.
241. Pissadaki EK, Bolam JP. The energy cost of action potential propagation in dopamine neurons: clues to susceptibility in Parkinson's disease. *Front Comput Neurosci.* 2013;7:13-13.
242. Muller-Nedebock AC, Brennan RR, Venter M, et al. The unresolved role of mitochondrial DNA in Parkinson's disease: An overview of published studies, their limitations, and future prospects. *Neurochem Int.* 2019;129:104495.
243. Miller FJ, Rosenfeldt FL, Zhang C, Linnane AW, Nagley P. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic Acids Research.* 2003;31(11):e61-e61.
244. Shoubridge EA, Wai T. Mitochondrial DNA and the Mammalian Oocyte. In: *Current Topics in Developmental Biology.* Vol 77. Academic Press; 2007:87-111.
245. Zhang P, Lehmann BD, Samuels DC, et al. Estimating relative mitochondrial DNA copy number using high throughput sequencing data. *Genomics.* 2017;109(5):457-462.
246. Kang I, Chu CT, Kaufman BA. The mitochondrial transcription factor TFAM in neurodegeneration: emerging evidence and mechanisms. *FEBS Lett.* 2018;592(5):793-811.
247. Wanrooij S, Falkenberg M. The human mitochondrial replication fork in health and disease. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* 2010;1797(8):1378-1388.
248. Ruhanen H, Borrie S, Szabadkai G, et al. Mitochondrial single-stranded DNA binding protein is required for maintenance of mitochondrial DNA and 7S DNA but is not required for mitochondrial nucleoid organisation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research.* 2010;1803(8):931-939.
249. Zinovkina LA. Mechanisms of Mitochondrial DNA Repair in Mammals. *Biochemistry (Moscow).* 2018;83(3):233-249.
250. García-Lepe UO, Bermúdez-Cruz RM. *Mitochondrial Genome Maintenance: Damage and Repair Pathways, DNA Repair- An Update.* IntechOpen; 2019.
251. Chocron ES, Munkácsy E, Pickering AM. Cause or casualty: The role of mitochondrial DNA in aging and age-associated disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease.* 2019;1865(2):285-297.
252. Lott MT, Leipzig JN, Derbeneva O, et al. mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Current Protocols in Bioinformatics.* 2013;44(1):1.23.21-21.23.26.
253. Hebert SL, Lanza IR, Nair KS. Mitochondrial DNA alterations and reduced mitochondrial function in aging. *Mechanisms of ageing and development.* 2010;131(7-8):451-462.
254. Sciacco M, Bonilla E, Schon EA, DiMauro S, Moraes CT. Distribution of wild-type and common deletion forms of mtDNA in normal and respiration-deficient muscle fibers from patients with mitochondrial myopathy. *Human Molecular Genetics.* 1994;3(1):13-19.
255. Edgar D, Shabalina I, Camara Y, et al. Random point mutations with major effects on protein-coding genes are the driving force behind premature aging in mtDNA mutator mice. *Cell Metab.* 2009;10(2):131-138.

256. Bender A, Krishnan KJ, Morris CM, et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet.* 2006;38(5):515-517.
257. Dolle C, Flones I, Nido GS, et al. Defective mitochondrial DNA homeostasis in the substantia nigra in Parkinson disease. *Nat Commun.* 2016;7:13548.
258. Lin MT, Cantuti-Castelvetri I, Zheng K, et al. Somatic mitochondrial DNA mutations in early Parkinson and incidental Lewy body disease. *Annals of neurology.* 2012;71(6):850-854.
259. Pinto M, Moraes CT. Mechanisms linking mtDNA damage and aging. *Free Radic Biol Med.* 2015;85:250-258.
260. Fitzgerald JC, Zimprich A, Carvajal Berrio DA, et al. Metformin reverses TRAP1 mutation-associated alterations in mitochondrial function in Parkinson's disease. *Brain.* 2017;140(9):2444-2459.
261. Alves G, Muller B, Herlofson K, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *J Neurol Neurosurg Psychiatry.* 2009;80(8):851-857.
262. Gelb DJ, Oliver E, Gilman S. Diagnostic Criteria for Parkinson Disease. *JAMA Neurology.* 1999;56(1):33-39.
263. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol.* 2011;95(4):629-635.
264. The International Parkinson Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet.* 2011;377(9766):641-649.
265. Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiol Aging.* 2015;36(3):1605 e1607-1612.
266. Fahn S, Elton RL. UPDRS program members. Unified Parkinsons disease rating scale. *Recent developments in Parkinson's disease.* 1987;2:153-163.
267. Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. *Neurology.* 1967;17(5):427-427.
268. Schwab RS, England AC, Gillingham FJ, Donaldson MC. Third symposium on Parkinson's disease. *E And S Livingstone: Edinburgh.* 1969:152-157.
269. Jankovic J, McDermott M, Carter J, et al. Variable expression of Parkinson's disease. *Neurology.* 1990;40(10):1529.
270. Marder K, Levy G, Louis ED, et al. Accuracy of family history data on Parkinson's disease. *Neurology.* 2003;61(1):18.
271. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
272. http://broadinstitute.github.io/picard. Accessed December, 2016.
273. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
274. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498.
275. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:11 10 11-33.

276. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics.* 2014;47:11 12 11-34.
277. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156-2158.
278. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27(21):2987-2993.
279. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
280. The R Project for Statistical Computing. www.r-project.org. Accessed.
281. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-909.
282. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190.
283. International HapMap C. The International HapMap Project. *Nature.* 2003;426(6968):789-796.
284. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
285. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research.* 2016;44(D1):D733-D745.
286. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods.* 2010;7(8):575-576.
287. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research.* 2012;40(W1):W452-W457.
288. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome research.* 2009;19(9):1553-1561.
289. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv.* 2019:531210.
290. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* 2016;44(D1):D1251-1257.
291. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447-452.
292. RStudio Team. RStudio: Integrated Development for R. *RStudio, Inc.* 2015.
293. Lee S, Fuchsberger C, Kim S, Scott L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. *Biostatistics.* 2016;17(1):1-15.
294. Dewey M. metap: meta-analysis of significance values. In. *R package version 0.8*2017.
295. Westfall PH, Young SS. *Resampling-based multiple testing : examples and methods for p-value adjustment.* New York ; Chichester: Wiley; 1993.
296. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological).* 1995;57(1):289-300.

297. Tang CF, Lu MK, Muo CH, Tsai CH, Kao CH. Increased risk of brain tumor in patients with Parkinson's disease: a nationwide cohort study in Taiwan. *Acta Neurologica Scandinavica.* 2016;134(2):148-153.
298. Ye R, Shen T, Jiang Y, Xu L, Si X, Zhang B. The Relationship between Parkinson Disease and Brain Tumor: A Meta-Analysis. *PloS one.* 2016;11(10):e0164388-e0164388.
299. Tacik P, Curry S, Fujioka S, et al. Cancer in Parkinson's disease. *Parkinsonism & related disorders.* 2016;31:28-33.
300. Brann DW, Dhandapani K, Wakade C, Mahesh VB, Khan MM. Neurotrophic and neuroprotective actions of estrogen: basic mechanisms and clinical implications. *Steroids.* 2007;72(5):381-405.
301. Ziegler A, Vens M. Generalized estimating equations. Notes on the choice of the working correlation matrix. *Methods Inf Med.* 2010;49(5):421-425; discussion 426-432.
302. Park T, Shin D-Y. On the use of working correlation matrices in the gee approach for longitudinal data. *Communications in Statistics - Simulation and Computation.* 1999;28(4):1011-1029.
303. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature Protocols.* 2010;5(9):1564-1573.
304. Do R, Stitziel NO, Won H-H, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature.* 2015;518(7537):102-106.
305. Polfus Linda M, Khajuria Rajiv K, Schick Ursula M, et al. Whole-Exome Sequencing Identifies Loci Associated with Blood Cell Traits and Reveals a Role for Alternative GFI1B Splice Variants in Human Hematopoiesis. *The American Journal of Human Genetics.* 2016;99(2):481-488.
306. Auer PL, Wang G, Leal SM. Testing for rare variant associations in the presence of missing data. *Genet Epidemiol.* 2013;37(6):529-538.
307. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics.* 2010;11(6):446-450.
308. Saint Pierre A, Génin E. How important are rare variants in common disease? *Briefings in Functional Genomics.* 2014;13(5):353-361.
309. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics.* 2010;11(6):415-425.
310. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics.* 2009;10(4):241-251.
311. Itan Y, Shang L, Boisson B, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. *Proceedings of the National Academy of Sciences.* 2015;112(44):13615.
312. Maffucci P, Bigio B, Rapaport F, et al. Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis. *Proceedings of the National Academy of Sciences.* 2019;116(3):950.
313. Stergachis AB, Haugen E, Shafer A, et al. Exonic Transcription Factor Binding Directs Codon Choice and Affects Protein Evolution. *Science.* 2013;342(6164):1367.
314. Presnyak V, Alhusaini N, Chen Y-H, et al. Codon Optimality Is a Major Determinant of mRNA Stability. *Cell.* 2015;160(6):1111-1124.

315. Povysil G, Petrovski S, Hostyk J, Aggarwal V, Allen AS, Goldstein DB. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nature Reviews Genetics.* 2019;20(12):747-759.
316. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences.* 2005;102(43):15545.
317. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011;27(12):1739-1740.
318. Elpeleg O, Miller C, Hershkovitz E, et al. Deficiency of the ADP-Forming Succinyl-CoA Synthase Activity Is Associated with Encephalomyopathy and Mitochondrial DNA Depletion. *The American Journal of Human Genetics.* 2005;76(6):1081-1086.
319. Prakash A, Doublié S. Base Excision Repair in the Mitochondria. *J Cell Biochem.* 2015;116(8):1490-1499.
320. Sirkis DW, Bonham LW, Aparicio RE, et al. Rare TREM2 variants associated with Alzheimer's disease display reduced cell surface expression. *Acta Neuropathologica Communications.* 2016;4(1):98.
321. Vojinovic D, Adams HHH, van der Lee SJ, et al. The dystrophin gene and cognitive function in the general population. *European Journal of Human Genetics.* 2015;23(6):837-843.
322. Cornes Belinda K, Brody Jennifer A, Nikpoor N, et al. Association of Levels of Fasting Glucose and Insulin With Rare Variants at the Chromosome 11p11.2-MADD Locus. *Circulation: Cardiovascular Genetics.* 2014;7(3):374-382.
323. Weeke P, Mosley JD, Hanna D, et al. Exome Sequencing Implicates an Increased Burden of Rare Potassium Channel Variants in the Risk of Drug-Induced Long QT Interval Syndrome. *Journal of the American College of Cardiology.* 2014;63(14):1430-1437.
324. Curtis D, Bakaya K, Sharma L, Bandyopadhyay S. Weighted burden analysis of exome-sequenced late onset Alzheimer's cases and controls provides further evidence for a role for PSEN1 and suggests involvement of the PI3K/Akt/GSK-3β and WNT signalling pathways. *bioRxiv.* 2019:596007.
325. Schulz WL, Tormey CA, Torres R. Computational Approach to Annotating Variants of Unknown Significance in Clinical Next Generation Sequencing. *Lab Med.* 2015;46(4):285-289.
326. Goeman JJ, Solari A. Multiple hypothesis testing in genomics. *Statistics in Medicine.* 2014;33(11):1946-1978.
327. Liu L, Sabo A, Neale BM, et al. Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS genetics.* 2013;9(4):e1003443-e1003443.
328. Elbaz A, McDonnell SK, Maraganore DM, et al. Validity of family history data on PD. *Neurology.* 2003;61(1):11.
329. Diamond SG, Markham CF, Hoehn MM, McDowell FH, Muenter MD. Effect of age at onset on progression and mortality in Parkinson's disease. *Neurology.* 1989;39(9):1187-1190.
330. Pagano G, Ferrara N, Brooks DJ, Pavese N. Age at onset and Parkinson disease phenotype. *Neurology.* 2016;86(15):1400-1407.

331. Inzelberg R, Schecthman E, Paleacu D, et al. Onset and progression of disease in familial and sporadic Parkinson's disease. *American Journal of Medical Genetics Part A.* 2004;124A(3):255-258.
332. Elbaz A, Grigoletto F, Baldereschi M, et al. Familial aggregation of Parkinson's disease. *Neurology.* 1999;52(9):1876.
333. Latourelle JC, Beste MT, Hadzi TC, et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. *The Lancet Neurology.* 2017;16(11):908-916.
334. Davis MY, Johnson CO, Leverenz JB, et al. Association of GBA Mutations and the E326K Polymorphism With Motor and Cognitive Progression in Parkinson Disease. *JAMA Neurology.* 2016;73(10):1217-1224.
335. Liu G, Locascio JJ, Corvol J-C, et al. Prediction of cognition in Parkinson's disease with a clinical–genetic score: a longitudinal analysis of nine cohorts. *The Lancet Neurology.* 2017;16(8):620-629.
336. Paul KC, Schulz J, Bronstein JM, Lill CM, Ritz BR. Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. *JAMA Neurology.* 2018;75(3):360-366.
337. Pihlstrøm L, Morset KR, Grimstad E, Vitelli V, Toft M. A cumulative genetic risk score predicts progression in Parkinson's disease. *Movement Disorders.* 2016;31(4):487-490.
338. Lerche S, Liepelt-Scarfone I, Wurster I, et al. Polygenic load: Earlier disease onset but similar longitudinal progression in Parkinson's disease. *Movement Disorders.* 2018;33(8):1349-1353.
339. Angeli A, Mencacci NE, Duran R, et al. Genotype and phenotype in Parkinson's disease: Lessons in heterogeneity from deep brain stimulation. *Movement Disorders.* 2013;28(10):1370-1375.
340. Koros C, Simitsi A, Stefanis L. Chapter Eight - Genetics of Parkinson's Disease: Genotype–Phenotype Correlations. In: Bhatia KP, Chaudhuri KR, Stamelou M, eds. *International Review of Neurobiology.* Vol 132. Academic Press; 2017:197-231.
341. Gravel S, Henn BM, Gutenkunst RN, et al. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences.* 2011;108(29):11983.
342. Taanman J-W. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* 1999;1410(2):103-123.
343. Perier C, Bender A, Garcia-Arumi E, et al. Accumulation of mitochondrial DNA deletions within dopaminergic neurons triggers neuroprotective mechanisms. *Brain.* 2013;136(Pt 8):2369-2378.
344. Paupe V, Prudent J. New insights into the role of mitochondrial calcium homeostasis in cell migration. *Biochemical and Biophysical Research Communications.* 2018;500(1):75-86.
345. Giorgi C, Baldassari F, Bononi A, et al. Mitochondrial Ca2+ and apoptosis. *Cell Calcium.* 2012;52(1):36-43.
346. Ludtmann MHR, Abramov AY. Mitochondrial calcium imbalance in Parkinson's disease. *Neuroscience Letters.* 2018;663:86-90.

347. Soman S, Keatinge M, Moein M, et al. Inhibition of the mitochondrial calcium uniporter rescues dopaminergic neurons in pink1-/- zebrafish. *Eur J Neurosci.* 2017;45(4):528-535.
348. Billingsley KJ, Barbosa IA, Bandrés-Ciga S, et al. Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. *npj Parkinson's Disease.* 2019;5(1):8.
349. Blauwendraat C, Reed X, Kia DA, et al. Frequency of Loss of Function Variants in LRRK2 in Parkinson Disease. *JAMA Neurology.* 2018;75(11):1416-1422.
350. Moutsianas L, Agarwala V, Fuchsberger C, et al. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS genetics.* 2015;11(4):e1005165-e1005165.
351. Nalls MA, Blauwendraat C, Vallerga CL, et al. Parkinson's disease genetics: identifying novel risk loci, providing causal insights and improving estimates of heritable risk. *bioRxiv.* 2018;Version 03/04/2019.
352. Latourelle JC, Dumitriu A, Hadzi TC, Beach TG, Myers RH. Evaluation of Parkinson Disease Risk Variants as Expression-QTLs. *PLOS ONE.* 2012;7(10):e46199.
353. Trabzuni D, Wray S, Vandrovcova J, et al. MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. *Human Molecular Genetics.* 2012;21(18):4094-4103.
354. Trabzuni D, Ryten M, Emmett W, et al. Fine-Mapping, Gene Expression and Splicing Analysis of the Disease Associated LRRK2 Locus. *PLOS ONE.* 2013;8(8):e70724.
355. Pierce S, Coetzee GA. Parkinson's disease-associated genetic variation is linked to quantitative expression of inflammatory genes. *PLOS ONE.* 2017;12(4):e0175882.
356. Murthy MN, Blauwendraat C, Guelfi S, et al. Increased brain expression of GPNMB is associated with genome wide significant risk for Parkinson's disease on chromosome 7p15.3. *neurogenetics.* 2017;18(3):121-133.
357. Li YI, Wong G, Humphrey J, Raj T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nature Communications.* 2019;10(1):994.
358. Kia DA, Zhang D, Guelfi S, et al. Integration of eQTL and Parkinson's disease GWAS data implicates 11 disease genes. *bioRxiv.* 2019:627216.

Appendices

Appendix I. Family history questionnaire used in paper I.

Pasient / kontroll nr:_____

Familieanamnese (ny november 2007)

Tilleggsspørsmål til spørreskjema Generelle opplysninger I

Ved førstegangsundersøkelse i PARKVEST ble du bl.a. spurt om du har noen i slekten din som har Parkinsons sykdom. Vi trenger litt utfyllende informasjon til de svarene du har gitt. **Bare svar for første grads slektninger (mor, far, søsken og barn).**

1. Har du slektninger med Parkinsons sykdom?

   SVAR: ja ☐ nei ☐

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

2. Har du slektninger som skjelver når han/hun er i ro eller sitter stille?

   SVAR: ja ☐ nei ☐

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

3. Har du slektninger som sleper føttene når han/hun går eller som går med veldig små skritt?

   SVAR: ja ☐ nei ☐

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

4. Har du slektninger med lutet holdning?

   SVAR: ja ☐ nei ☐

HVEM (slektsforhold): _____

HVEM (slektsforhold): _____

HVEM (slektsforhold): _____

5. Har du slektninger som svinger lite med armene når han/hun går?

   SVAR: ja ☐ nei ☐

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

6. Har du slektninger som er stiv i muskulaturen eller legger du merke til at han/hun er stiv og langsom når han/hun gjør ting?

   SVAR: ja ☐ nei ☐

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

   HVEM (slektsforhold): _____

Hvis "ja" på et eller flere spørsmål, fyll ut neste skjema med oppfølgingsspørsmål

Oppfølgingsspørsmål (fyll ut ett ark med oppfølgingsspørsmål for hver slektning som skårer "ja" på spørsmål. 1–6):

Dersom du svarer "vet ikke", vær vennlig å føre opp navn og telefonnummer for vedkommende slik at vi kan ta kontakt for å få en oppklaring.

HVEM (slektsforhold): _____

Ble din slektning tilsett av lege for denne tilstanden?

SVAR: ja ☐ nei ☐ vet ikke ☐

I. Hvis ja, ble diagnosen Parkinsons sykdom stilt?

SVAR: ja ☐ nei ☐ vet ikke ☐

Hvor gammel var vedkommende da diagnosen ble stilt?

SVAR: _____

Ble din slektning tilsett av nevrolog for denne tilstanden?
SVAR: ja ☐ nei ☐ vet ikke ☐

II. Hvis ja, ble diagnosen Parkinsons sykdom stilt?

SVAR: ja ☐ nei ☐ vet ikke ☐

Hvor gammel var vedkommende da diagnosen ble stilt?

SVAR: _____

III. Bruker/tar din slektning levodopa (Madopar, Sinemet)?

SVAR: ja ☐ nei ☐ vet ikke ☐

IV. Dersom din slektning er død - ble det gjennomført en obduksjon som indikerte Parkinsons sykdom?

SVAR: ja ☐ nei ☐ vet ikke ☐

V. Hvor gammel var vedkommende da han/hun utviklet symptomer som skjelvinger, bundet gange, balanseproblemer (posturale endringer) og/eller stivhet?

SVAR: _____

VI. Hvilket symptom var det første som oppstod?

SVAR: _____

Pasient / kontroll nr:_____

Diagnostisk algoritme:

Possible PD: Minst tre av fire følgende: 2, 3, 4/5, 6
Eller: en av de ovennevnte og en av: I eller II eller III

Probable PD: a.) 2 eller 6
b.) 3 eller 4 eller 5
c.) I eller II eller III

Definite PD: a.) 2 eller 6
b.) 3 eller 4 eller 5
c.) II og III

Vennligst fyll ut følgende:

1. Hvor mange av pas. første grads slektninger har Parkinsons sykdom?

   SVAR:_____ (antall)

2. Hvem: (fyll ut for alle slektninger som har Parkinsons sykdom)

   (Hvem: 1 = foreldre, 2 = søsken, 3 = barn)
   (Diagnostisk sikkerhet: 1 = possible PD, 2 = probable PD, 3 = definite PD)

   a. Hvem: _____ Diagnostisk sikkerhet: _____

   b. Hvem: _____ Diagnostisk sikkerhet: _____

   c. Hvem: _____ Diagnostisk sikkerhet: _____

   d. Hvem: _____ Diagnostisk sikkerhet: _____

Papers I-IV

II

doi:10.1093/brain/awx378 BRAIN 2018: 141; 1–3 | e16

BRAIN
A JOURNAL OF NEUROLOGY

LETTER TO THE EDITOR

No evidence for rare *TRAP1* mutations influencing the risk of idiopathic Parkinson's disease

Johannes J. Gaare,[1,2] Gonzalo S. Nido,[1,2] Paweł Sztromwasser,[3,4,5] Per M. Knappskog,[3,6] Olav Dahl,[2,7] Morten Lund-Johansen,[2,8] Guido Alves,[9,10,11] Ole-Bjørn Tysnes,[1,2] Stefan Johansson,[3,6] Kristoffer Haugarvoll[1,2] and Charalampos Tzoulis[1,2]

1 Department of Neurology, Haukeland University Hospital, 5021, Bergen, Norway
2 Department of Clinical Medicine, University of Bergen, 5020, Norway
3 Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, 5021, Bergen, Norway
4 Department of Clinical Science, University of Bergen, 5020, Norway
5 Computational Biology Unit, Department of Informatics, University of Bergen, 5020, Norway
6 K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, 5020, Norway
7 Department of Oncology, Haukeland University Hospital 5021, Bergen, Norway
8 Department of Neurosurgery, Haukeland University Hospital, 5021, Bergen, Norway
9 The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway
10 Department of Neurology, Stavanger University Hospital, Stavanger, Norway
11 Department of Mathematics and Natural Sciences, University of Stavanger, 4036 Stavanger, Norway

Correspondence to: Charalampos Tzoulis, MD, PhD
Department of Neurology,
Haukeland University Hospital,
5021 Bergen, Norway
E-mail: charalampos.tzoulis@nevro.uib.no or charalampos.tzoulis@helse-bergen.no

Sir,

In their recent work, Fitzgerald *et al.* (2017) report a novel homozygous *TRAP1* loss-of-function mutation in a patient with late-onset Parkinson's disease. Further, they show an enrichment of two subgroups of rare *TRAP1* variants in controls compared to patients with Parkinson's disease in the Parkinson's Progression Markers Initiative (PPMI) dataset (Parkinson Progression Marker Initiative, 2011). However, these associations are not significant after correction for multiple testing. The enrichment is measured using the burden and SKAT-O (Lee *et al.*, 2012) tests. From this, the authors stipulate that rare, more benign missense *TRAP1* mutations are depleted in patients with Parkinson's disease.

Here, we sought to replicate these findings and investigate the role of *TRAP1* mutations in our exome sequencing dataset, comprising 181 Parkinson's disease cases from the Norwegian ParkWest cohort (Alves *et al.*, 2009) and 196 in-house controls (unpublished results). Following quality control, variants were annotated using ANNOVAR (Wang *et al.*, 2010) according to the RefSeq gene transcripts, dbNSFP v3.3a (Liu *et al.*, 2016) and ExAC (Lek *et al.*, 2016). We identified 21 exonic variants in the *TRAP1* gene, of which 16 were non-synonymous (missense) and five were synonymous. We did not detect the specific p.R47X mutation described by Fitzgerald *et al.*, nor did we find any other nonsense or splice mutations. Two missense variants were present in cases only (in heterozygous form), but they were predominantly predicted to be benign/tolerated across five different prediction algorithms (SIFT, PolyPhen-2 HumVar/HumDiv, LRT and MutationTaster). No single variant association test was significant after correction for multiple testing.

For collapsing tests, we selected variants with minor allele frequency (MAF) < 1% in the non-Finnish European ExAC dataset. We created subsets of variants within *TRAP1* based on synonymy and CADD score similarly to Fitzgerald *et al.* In addition to burden and SKAT-O, we also performed the

Advance Access publication January 24, 2018

Table 1 Region-based analysis of TRAP1 variants

Group	Number of variants	MAC controls	MAC cases	Burden *P*-value	SKAT-O *P*-value	SKAT *P*-value
The Norwegian ParkWest sample						
Non-synonymous	12	18	12	0.407	0.648	0.566
CADD10	11	14	7	0.130	0.221	0.806
CADD15	9	7	4	0.229	0.379	0.751
CADD20	9	7	4	0.229	0.379	0.751
CADD30	2	2	0	0.326	0.489	0.786
Synonymous	2	2	0	0.332	0.746	0.746
The PPMI sample						
Non-synonymous	9	5	7	0.279	0.367	0.255
CADD10	8	5	6	0.259	0.382	0.205
CADD15	8	5	6	0.259	0.382	0.205
CADD20	6	4	5	0.293	0.312	0.259
CADD30	2	1	1	0.799	0.277	0.277
Synonymous	2	0	2	0.338	0.739	0.739

CADD = non-synonymous variants with CADD score >10, 15, 20 and 30, respectively.
MAC = minor allele count.
P-values are uncorrected for multiple testing.

SKAT test (Wu *et al.*, 2011). Collapsing tests were performed using the SKAT R package (Lee *et al.*, 2016). We found no evidence of variant enrichment in *TRAP1*, in any of the tests/models tested in our population. The results are summarized in Table 1.

Upon close examination of the analyses performed by Fitzgerald *et al.* in the PPMI cohort, we raise a few questions regarding aspects of the quality control and collapsing testing. Firstly, the authors use a particularly lax threshold for variant call-rate ($\geqslant$90%). Missing genotypes may be due to genotyping errors, and region-based collapsing tests using rare variants are particularly susceptible to inflated type I error rates if the distribution of missed calls differs between cases and controls in a tested region (Auer *et al.*, 2013). Another crucial aspect when testing for rare variant associations is the control of population stratification. Rare variants display very little sharing between populations (Gravel *et al.*, 2011), and failure to control for this could therefore lead to spurious associations, especially in a heterogeneous sample such as the PPMI. While removing individuals 3 standard deviations (SD) from the mean of the first and second principal component does reduce ethnic heterogeneousness to some degree, a more prudent approach would perhaps have been to remove outliers iteratively, as implemented by Eigensoft (Patterson *et al.*, 2006; Price *et al.*, 2006).

Considering the above limitations, we sought to replicate the findings of the study in the same PPMI dataset, but following a more stringent quality control procedure. Specifically, we used a variant call-rate cut-off of >98% and performed principal component analysis using Eigensoft with standard filtering settings (five iterations, 10 principal components, sigma 6), in addition to removing outliers ($\geqslant$3 SD) across the first and second principal components post-filtering. Rare variants were defined as variants with MAF < 0.5% in the non-Finnish European ExAC dataset to replicate the parameters described by Fitzgerald *et al.* In this robustly quality controlled dataset, we detected no nominally significant variant enrichment in *TRAP1* by either burden, SKAT-O or SKAT tests. The results of our replicative PPMI analyses are summarized in Table 1.

In conclusion, while the reported p.R47X *TRAP1* mutation may indeed be deleterious to mitochondrial function, no definite evidence is provided that this mutation is the cause of Parkinson's disease in the reported case. Moreover, we found no evidence supporting that rare variation enrichment in *TRAP1* influences the risk of Parkinson's disease in two independent populations. We therefore believe that the proposed role of *TRAP1* in Parkinson's disease is unsubstantiated by the data presented in the study.

Funding

This work was supported by grants from the Regional Health Authority of Western Norway (grant no 911903 and 911988) and the Research Council of Norway (grant no 240369/F20). PPMI, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including AbbVie, Avid Radiopharmaceuticals, Biogen Idec, BioLegend, Bristol-Meyers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Eli Lilly and Company, Lundbeck, Merck, Meso Scale Discovery, Pfizer Inc., Piramal Imaging, Roche CNS group, Sanofi Genzyme, Servier, Takeda, Teva, UCB and Golub Capital.

References

Alves G, Muller B, Herlofson K, HogenEsch I, Telstad W, Aarsland D, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. J Neurol Neurosurg Psychiatry 2009; 80: 851–7.

Auer PL, Wang G, Leal SM. Testing for rare variant associations in the presence of missing data. Genet Epidemiol 2013; 37: 529–38.

Fitzgerald JC, Zimprich A, Carvajal Berrio DA, Schindler KM, Maurer B, Schulte C, et al. Metformin reverses TRAP1 mutation-associated alterations in mitochondrial function in Parkinson's disease. Brain 2017; 140: 2444–59.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci USA 2011; 108: 11983–8.

Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012; 91: 224–37.

Lee S, Fuchsberger C, Kim S, Scott L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. Biostatistics 2016; 17: 1–15.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016; 536: 285–91.

Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat 2016; 37: 235–41.

Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol 2011; 95: 629–35.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006; 2: e190.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38: 904–9.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38: e164.

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011; 89: 82–93.

III

RESEARCH ARTICLE

Rare Genetic Variation in Mitochondrial Pathways Influences the Risk for Parkinson's Disease

Johannes J. Gaare, MD,[1,2] Gonzalo S. Nido, PhD,[1,2] Paweł Sztromwasser, PhD,[3,4,5] Per M. Knappskog, PhD,[3,6] Olav Dahl, MD, PhD,[2,7] Morten Lund-Johansen, MD, PhD,[2,8] Jodi Maple-Grødem, PhD,[9,10] Guido Alves, MD, PhD,[9,11,12] Ole-Bjørn Tysnes, MD, PhD,[1,2] Stefan Johansson, PhD,[3,6] Kristoffer Haugarvoll, MD, PhD[1,2] and Charalampos Tzoulis, MD, PhD[1,2*]

[1]*Department of Neurology, Haukeland University Hospital, Bergen, Norway*
[2]*Department of Clinical Medicine, University of Bergen, Bergen, Norway*
[3]*Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway*
[4]*Department of Clinical Science, University of Bergen, Bergen, Norway*
[5]*Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway*
[6]*K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway*
[7]*Department of Oncology, Haukeland University Hospital, Bergen, Bergen, Norway*
[8]*Department of Neurosurgery, Haukeland University Hospital, Bergen, Bergen, Norway*
[9]*The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway*
[10]*Centre for Organelle Research, University of Stavanger, Stavanger, Norway*
[11]*Department of Neurology, Stavanger University Hospital, Stavanger, Norway*
[12]*Department of Mathematics and Natural Sciences, University of Stavanger, Stavanger, Norway*

ABSTRACT: Background: Mitochondrial dysfunction plays a key role in PD, but the underlying molecular mechanisms remain unresolved. We hypothesized that the disruption of mitochondrial function in PD is primed by rare, protein-altering variation in nuclear genes controlling mitochondrial structure and function.

Objective: The objective of this study was to assess whether genetic variation in genes associated with mitochondrial function influences the risk of idiopathic PD.

Methods: We employed whole-exome sequencing data from 2 independent cohorts of clinically validated idiopathic PD and controls, the Norwegian ParkWest cohort (n = 411) and the North American Parkinson's Progression Markers Initiative (n = 640). We applied burden-based and variance-based collapsing methods to assess the enrichment of rare, nonsynonymous, and damaging genetic variants on genes, exome-wide, and on a comprehensive set of mitochondrial pathways, defined as groups of genes controlling specific mitochondrial functions.

Results: Using the sequence kernel association test, we detected a significant polygenic enrichment of rare, nonsynonymous variants in the gene-set encoding the pathway of mitochondrial DNA maintenance. Notably, this was the strongest association in both cohorts and survived multiple testing correction (ParkWest $P = 6.3 \times 10^{-3}$, Parkinson's Progression Markers Initiative $P = 6.9 \times 10^{-5}$, metaanalysis $P = 3.2 \times 10^{-6}$).

Conclusions: Our results show that the enrichment of rare inherited variation in the pathway controlling mitochondrial DNA replication and repair influences the

***Corresponding author:** Dr. Charalampos Tzoulis, Department of Neurology, Haukeland University Hospital, 5021 Bergen, Norway; charalampos.tzoulis@nevro.uib.no

Funding agencies: This work was supported by grants from the Regional Health Authority of Western Norway (Grants 911903 and 911988) and the Research Council of Norway (Grant 240369/F20). Parkinson's Progression Markers Initiative, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including AbbVie, Avid Radiopharmaceuticals, Biogen Idec, BioLegend, Bristol-Meyers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Eli Lilly and Company, Lundbeck, Merck, Meso Scale Discovery, Pfizer Inc., Piramal Imaging, Roche CNS group, Sanofi Genzyme, Servier, Takeda, Teva, UCB, and Golub Capital.

Relevant conflicts of interests/financial disclosures: Nothing to report.

Received: 23 March 2018; **Revised:** 21 May 2018; **Accepted:** 31 May 2018

Published online 5 September 2018 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.64

risk of PD. We propose that this polygenic enrichment contributes to the impairment of mitochondrial DNA homeostasis, thought to be a key mechanism in the pathogenesis of PD, and explains part of the disorder's "missing heritability." © 2018 The Authors. Movement Disorders published by Wiley Periodicals, Inc. on behalf of International Parkinson and Movement Disorder Society

Key Words: Parkinson's disease; genetics; neurodegeneration; whole-exome sequencing; genetic association studies

Parkinson's disease (PD) affects 1.8% of the population aged older than 65 years.[1] Although the etiology of PD remains unknown, mitochondrial dysfunction clearly plays a role. Mutations in most of the genes causing familial PD, including *SNCA*,[2] *LRRK2*,[3] *PINK1, PRKN*,[4] *PARK7* (*DJ-1*),[5] and *VPS35*,[6] have been shown to disrupt mitochondrial quality control. Moreover, impaired mitochondrial DNA (mtDNA) maintenance[7,8] and widespread respiratory chain dysfunction[9] occur in the brains of individuals with idiopathic PD. The molecular etiology underlying mitochondrial impairment in idiopathic PD remains, however, unresolved.

Pathogenic mutations in several nuclear mitochondrial genes involved in mtDNA homeostasis have been shown to cause severe nigrostriatal degeneration of a similar type as the one that occurs in PD.[10–13] Moreover, inherited variation in *POLG*[14] and *TFAM*,[15] both of which are essential for mtDNA maintenance, has been associated with an increased risk of idiopathic PD, although the reported effects were generally weak and have not been reproduced by large genome-wide association studies. Therefore, it remains undetermined whether mitochondrial dysfunction in PD can, to some extent, be attributed to inherited genetic variation.

We hypothesized that the disruption of mitochondrial function in PD is partly caused by synergistic effects of rare genetic variation in nuclear genes controlling mitochondrial function. To test our hypothesis, we employed whole-exome sequence data from 2 independent, prospectively collected PD cohorts: the Norwegian ParkWest study (ParkWest)[16] and the Parkinson's Progression Markers Initiative (PPMI).[17]

Methods

Cohorts and Sequencing

We sequenced all individuals with clinically validated PD (n = 192) from the Norwegian ParkWest study, a prospective population-based cohort of idiopathic PD.[16] Controls (n = 219) were provided from cohorts of previously sequenced individuals with testis cancer (n = 167) or acoustic neuroma (n = 52) who had been recruited and examined at our hospital and had no clinical signs of neurodegenerative or other neurological disorders. DNA was extracted from blood by routine procedures and sequenced at HudsonAlpha Institute for Biotechnology (Huntsville, Alabama) using Roche-NimbleGen Sequence Capture EZ Exome v2 (173 controls) and v3 (all PD and 46 controls) kits (Roche, Brussels, Switzerland) and paired-end 100 bp sequencing on the Illumina HiSeq platform (Illumina, San Diego, USA). The reads were mapped to the hg19 reference genome using BWA v0.6.2,[18] polymerase chain reaction duplicates removed with Picard v1.118,[19] and the alignment refined using the Genome Analysis Toolkit v3.3.0[20] applying base quality score recalibration and realignment around indels recommended in the GATK Best Practices workflow.[21,22] Variants were called in all samples using the GATK HaplotypeCaller[20] with default parameters. Next, variant quality score recalibration was performed using 99.9% sensitivity threshold.[20] The remaining variants were filtered against the intersection of capture targets (v2 and v3) using BEDtools[23] and VCFtools.[24] Variants with total depth below 10X were marked as unknown genotype (no-call) using BCFtools.[25] Indel calls, which were found to be less reliable than single-nucleotide variants, were excluded from downstream analyses.

Additional data used in the preparation of this article were obtained from the PPMI database.[17] Whole-exome sequence data was available for 640 individuals (459 cases, 181 controls). Sequencing had been performed using the Illumina Nextera Rapid Capture Expanded Exome Kit and paired-end 100 bp reads on the Illumina HiSeq 2500 (Illumina, San Diego, USA). Calling and alignment had been performed by the PPMI. Indels were removed prior to variant quality control (QC) using VCFtools.[24]

Variant Filtering and QC

Whole-exome sequence data were recoded into binary PLINK input format, and QC of individual and single nucleotide polymorphism (SNP) data was performed on the ParkWest and PPMI datasets individually using PLINK v1.90.[26] Individuals were excluded if their genotypic data showed a missing rate >2%, abnormal heterozygosity (±3 standard deviations, calculated for common and rare variants separately), conflicting sex assignment, cryptic relatedness (identity by descent > 0.2), or divergent ancestry (non-European). Population stratification was studied using multidimensional scaling against the HapMap populations.[27]

SNPs were excluded as a result of genotyping rate less than 98%, different call rates between cases and controls ($P < .02$) or departure from the Hardy-Weinberg equilibrium ($P < 10^{-5}$). Only autosomes were considered. Monomorphic and multiallelic variants were removed. The transition-transversion ratio was calculated before and after QC. Principal component analysis was performed using Eigensoft[28,29] with standard filtering settings (5 iterations, 10 principal components, sigma 6). Analysis of variance of the first 10 principal components was performed with significance level cutoff set to $P < .01$. For ParkWest, there were 2 significant principal components (3 and 6) that were corrected for in downstream analyses. There were no significant principal components for PPMI. All QC analyses were performed using PLINK v1.90[26] and R[30] unless otherwise specified.

Annotation and Subset Filtering

Datasets were annotated using ANNOVAR[31] according to the RefSeq gene transcripts, and 2 variant subsets were extracted for further analyses. These were defined as "nonsynonymous" and "damaging" according to ANNOVAR. Nonsynonymous variants comprised missense changes (Sequence Ontology: 0001583). Damaging variants comprised all single nucleotide changes that had a deleterious score in all of the following 5 prediction algorithms: PolyPhen2 HumDiv, PolyPhen2 HumVar, MutationTaster, SIFT, and LRT. In the ParkWest dataset, rare variants were defined as having a minor allele frequency (MAF) of < 1%. Because the PPMI cohort had a substantially uneven number of cases and controls, rare variants were defined as having a MAF of <1% in either cases or controls to avoid a unidirectional variant inclusion bias.

Pathway Curation

The mitochondrial pathways were defined as groups of genes encoding functionally and/or structurally linked proteins directly involved in mitochondrial function. Pathways were manually curated from Mitocarta v2.0,[32] a collection of all known proteins with strong support for mitochondrial localization. The pathways were subsequently expanded using STRING[33] to include genes encoding additional proteins that are involved in the pathways, but have no proven mitochondrial localization. Specifically, for each pathway we compiled a list of additional candidate proteins ranked by the number and strength (STRING combined score) of STRING interactions with the original pathway. The resulting candidate lists were manually inspected, and the original pathways were supplemented with additional genes encoding proteins with a known involvement in mitochondrial pathways, but not established mitochondrial localization (Supplemental Table S1).

Genetic Association Analyses

Single-point association was performed for common variants (MAF > 1%) using the chi-square test for PPMI and logistic regression with significant principal components as covariates for ParkWest. For collapsing tests, variants were annotated using ANNOVAR and analyzed using a weighted burden[34] and the sequence kernel association test (SKAT)[35] as implemented by the SKAT R package v1.3.2.[36] In these analyses, variants within a specified region (ie, a gene or pathway) were combined to a single genetic score that was subsequently tested for association in a logistic regression framework. The burden test assumes that all variants in a specified region are causal and influence the phenotype in the same direction, that is, either increase or decrease the risk of PD. Consequently, it suffers from a substantial loss of power when these premises are not valid. SKAT aggregates variants within a specified region without considering the direction of effect for individual variants. Therefore, SKAT is better suited for detecting associations where both risk and protective variants and/or numerous noncausal variants are present. For our analyses, standard weights were used,[35] and only rare variants were considered. Different methods of per-hypothesis resampling were applied as described by Lee and colleagues.[36] Genes with only 1 variant were excluded from the gene-based analyses as this would be representative of single SNP associations rather than true enrichment. Using resampling, the minimum achievable P value (MAP) was determined for each gene and pathway.[36] Genes/pathways with minimum achievable P values above the Bonferroni-corrected threshold for statistical significance were excluded to reduce statistical noise. In total, 3,441 gene tests were performed in the ParkWest cohort, and 13,034 in the PPMI cohort (across all tests and subsets). Meta P values for pathways were calculated using the optimally weighted Z test[37] as implemented by the metap R package.[38]

Multiple Testing Correction

Single-variant gene and pathway meta-analysis associations were corrected using the Bonferroni method. To control the family-wise error in individual cohort pathways, the minP/maxT method[39] was implemented using 10,000 null-permuted phenotype samples for ParkWest and 100,000 for PPMI (different number of permutations as a result of the difference in dataset size).

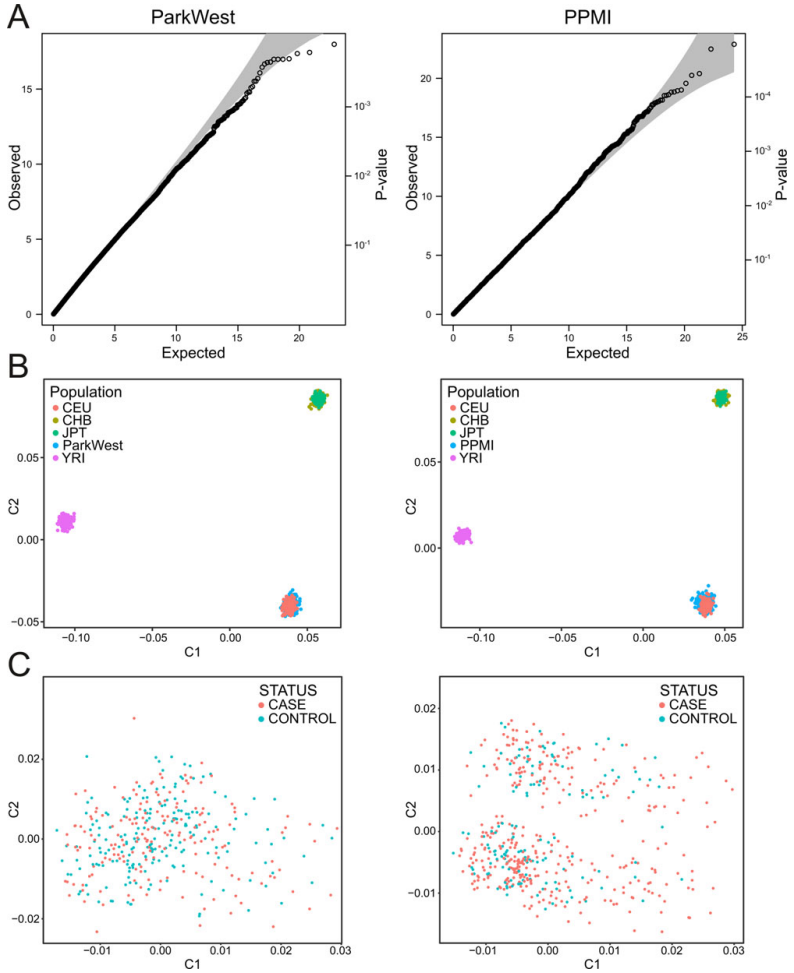

FIG. 1. Quality control. (A) Quantile-quantile-plots of association for common variants using chi-square for Parkinson's Progression Markers Initiative and logistic regression with significant principal components (3 and 6) as covariates for ParkWest. (B) Population stratification (multidimensional scaling) against HapMap populations (CEU, Utah residents with Northern and Western European ancestry; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan; YRI, Yoruba in Ibadan, Nigeria). (C) Multidimensional scaling plots show no stratification between cases and controls in any of the cohorts. All plots are based on data after quality control. C1, first principal component; C2, second principal component.

Power Analysis

The average statistical power of the gene-based SKAT and burden test were calculated using the SKAT R package v1.3.2.[36] The disease model assumed a prevalence of PD of 0.015 and used the empirical MAFs of the ParkWest cohort. Power calculations were carried out using different assumptions of the percentage of causal SNPs (10%, 25%, 50%, 75%, and 100%). The original function from the SKAT R package was modified to calculate the average power of the exonic regions instead of the default random genomic regions.

Standard Protocol Approvals, Registrations, and Patient Consents

These studies were approved by the Regional Committee for Medical and Health Research Ethics, Western Norway (REK 131/04). Written, informed consent was obtained from all participants.

Results

After alignment, variant calling and QC the final ParkWest dataset comprised 377 individuals (181 cases and 196 controls). Mean depth per individual was

TABLE 1. Top single gene results

Nonsynonymous rare variants							
Burden				SKAT			
ParkWest		PPMI		ParkWest		PPMI	
Gene	*P* value	Gene	*P* value	Gene	*P* value	Gene	*P* value
ZNF76	3.98×10^{-4}	*CHRM2*	2.46×10^{-4}	*TSR1*	2.57×10^{-4}	*CHRM2*	1.16×10^{-5}
KIF20B	4.74×10^{-4}	*UGT1A4*	4.41×10^{-4}	*COL24A1*	7.87×10^{-4}	*MC4R*	2.64×10^{-4}
KANK1	2.87×10^{-3}	*MC4R*	5.27×10^{-4}	*WDFY4*	1.52×10^{-3}	*RECQL4*	2.80×10^{-4}
BIRC6	4.03×10^{-3}	*GLTSCR1*	6.76×10^{-4}	*LRRN2*	2.99×10^{-3}	*CEP131*	5.22×10^{-4}
NUMA1	4.75×10^{-3}	*TG*	1.05×10^{-3}	*ABCC11*	3.50×10^{-3}	*KRTAP10-7*	5.52×10^{-4}
DNAJC17	4.80×10^{-3}	*CACNA1H*	1.34×10^{-3}	*FAM214A*	5.14×10^{-3}	*SETD4*	6.35×10^{-4}
ZNFX1	4.88×10^{-3}	*HORMAD2*	1.41×10^{-3}	*GEMIN5*	5.87×10^{-3}	*STEAP4*	6.41×10^{-4}
CP	5.02×10^{-3}	*BRPF3*	1.46×10^{-3}	*USP6*	6.05×10^{-3}	*MRPL4*	7.64×10^{-4}
MKI67	5.78×10^{-3}	*ITGA2B*	1.47×10^{-3}	*LRRC75B*	6.52×10^{-3}	*FER1L6*	8.60×10^{-4}
CFTR	6.52×10^{-3}	*ADGRG7*	1.58×10^{-3}	*DFNB31*	7.71×10^{-3}	*NDUFA9*	9.26×10^{-4}
Damaging rare variants							
Burden				SKAT			
ParkWest		PPMI		ParkWest		PPMI	
Gene	*P* value	Gene	*P* value	Gene	*P* value	Gene	*P* value
DFNB31	5.25×10^{-4}	*ATP8B4*	1.75×10^{-3}	*DNAH2*	2.79×10^{-3}	*ADAMTS14*	1.70×10^{-3}
ACACB	4.31×10^{-2}	*NRAP*	6.72×10^{-3}	*DFNB31*	2.95×10^{-3}	*PPP2R3A*	2.85×10^{-3}
LOXHD1	6.22×10^{-2}	*ATP4A*	1.04×10^{-2}	*POLG*	5.02×10^{-3}	*HADH*	4.99×10^{-3}
CFTR	6.51×10^{-2}	*TEP1*	1.09×10^{-2}	*KIAA0196*	2.24×10^{-2}	*LRP1B*	5.61×10^{-3}
LAMA1	1.56×10^{-1}	*DNAH11*	1.13×10^{-2}	*DNAH3*	2.63×10^{-2}	*PCDHB1*	7.08×10^{-3}
PITRM1	1.64×10^{-1}	*CSMD1*	1.47×10^{-2}	*ACACB*	3.82×10^{-2}	*INADL*	9.07×10^{-3}
DNAH2	1.73×10^{-1}	*PCDHB1*	1.50×10^{-2}	*RELN*	5.96×10^{-2}	*CEL*	1.10×10^{-2}
FAT2	1.92×10^{-1}	*BSN*	1.66×10^{-2}	*PPFIBP2*	6.20×10^{-2}	*UBXN11*	1.10×10^{-2}
TRIM63	2.26×10^{-1}	*CPXM1*	2.08×10^{-2}	*SYNE2*	1.19×10^{-1}	*RGS11*	1.40×10^{-2}
QRSL1	2.49×10^{-1}	*HHAT*	2.10×10^{-2}	*MASP2*	1.19×10^{-1}	*HHAT*	1.60×10^{-2}

Nominal *P* values given, no single gene association was statistically significant after multiple testing correction. PPMI, Parkinson's Progression Markers Initiative; SKAT, sequence kernel association test.

80.2 and 92% of targeted bases were covered at ≥20-fold. In the ParkWest dataset we observed 130,500 variants, of which 61.1% (79,763) were rare. We identified 56,429 nonsynonymous variants and 8,646 damaging variants. The transition-transversion ratio-ratio for exonic variants was 3.28 (3.23 before QC). The final PPMI dataset comprised 513 individuals (371 cases and 142 controls). We observed 380,423 variants of which 73.6% (280,099) were rare. We identified 107,130 nonsynonymous variants and 16,644 damaging variants. The transition-transversion ratio for exonic variants was 3.13 (3.02 before QC). Synonymous and nonsynonymous variants had a similar distribution in the 2 cohorts (Supplemental Table S2) as well as between cases and controls (Supplemental Table S3). Singleton variants comprised 37.4% of the ParkWest dataset and 51.1% of the PPMI dataset (Supplemental Table S4).

Population stratification analysis with HapMap confirmed that both cohorts consisted primarily of individuals of Western and Northern European descent. Quantile-quantile plots of common variant associations after QC showed no inflation of test statistics, with $\lambda = 1.02$ and $\lambda = 1.03$ for PPMI and ParkWest, respectively. Multidimensional scaling showed no stratification between cases and controls (Fig. 1). No single variants produced exome-wide significant associations. The lowest *P* value was for the SNP rs57859638 ($P = 1.2 \times 10^{-4}$), odds ratio (OR) = 0.47, MAF = 0.219) in *IQCF1* for ParkWest, and for rs543304 ($P = 1.1 \times 10^{-5}$), OR = 0.49, MAF = 0.197) in *BRCA2* in PPMI. The complete results are available in Supplemental Tables S5 and S6.

Gene-Based Analyses

Gene-based enrichment analyses for rare nonsynonymous or damaging variants by burden test or SKAT showed no significant associations after Bonferroni correction. Moreover, none of the top 10 nominally significant genes replicated across both the ParkWest and PPMI cohorts (Table 1). The results for all tested genes are shown in Supplemental Tables S7 to S14.

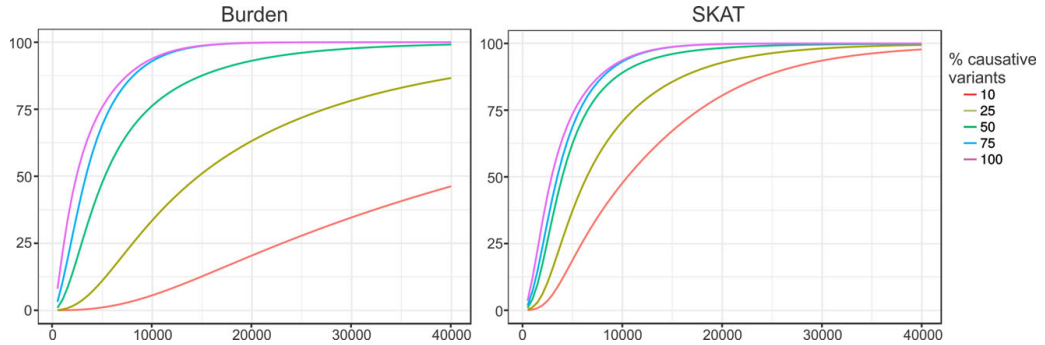

FIG. 2. Power analysis. Power estimation for sequence kernel association test (SKAT) and weighted burden test using the rare variant matrix from the ParkWest dataset as basis for the simulation. Genome-wide significance levels were set to $\alpha = 2.5 \times 10^{-6}$, assuming a total of 20,000 genes. The *y*-axis shows the average power across all genes, and the *x*-axis shows total sample size. Simulations were made for different percentages of causal variants.

Power analyses showed that our sample size was underpowered for detecting single gene associations. Assuming that 50% of the nonsynonymous variants in a gene influence the phenotype in any direction (ie, deleterious or protective), our calculations suggest that a sample size of approximately 8,000 for the SKAT and 11,000 for the weighted burden test are required to reach an average power of 80% across all genes at a significance cut-off of $\alpha = 0.05$ and surviving Bonferroni correction for 20,000 genes (corrected $\alpha = 2.5 \times 10^{-6}$; Fig. 2).

Pathway-Based Analyses

Starting with Mitocarta, we curated a total of 28 pathways with known molecular function, comprising a total of 868 unique genes (including 100 non-Mitocarta genes added after expansion with functionally relevant genes lacking evidence of mitochondrial localization). The remaining Mitocarta proteins could not be confidently assigned to a pathway because of limited information regarding their function. The entire noncurated Mitocarta list was also included as a separate pathway. A complete description of the pathways is provided in Supplemental Table S1. After excluding pathways with minimum achievable *P* value above the threshold for statistical significance, we analyzed 26/18 pathways in ParkWest and 29/26 pathways in PPMI for nonsynonymous/damaging variants, respectively.

Mitochondrial pathway analyses by burden test showed no evidence of enrichment in PD. SKAT detected a significantly skewed distribution of rare, nonsynonymous variants between individuals with PD and controls. The 2 most significant associations were found for the pathways of mtDNA maintenance and mitochondrial calcium homeostasis (Table 2) in both the ParkWest and PPMI cohorts. The strongest association was for mtDNA maintenance (ParkWest$P = 6.3 \times 10^{-3}$, PPMI $P = 6.9 \times 10^{-5}$, meta $P = 3.2 \times 10^{-6}$), which survived multiple testing correction. mtDNA maintenance was also among the top results for damaging variants in both datasets, but did not survive multiple testing correction. None of the other mitochondrial pathways showed significant associations that survived multiple testing correction or had significant nominal *P* values in both cohorts, except from the meta-analysis of mitochondrial calcium homeostasis (meta $P = 6.7 \times 10^{-5}$). Detailed pathway results are available in the supplemental data (Supplemental Table S15). Repeating the pathway analyses with rare synonymous variants only yielded no significant results, indicating that the observed enrichment is specific for protein-altering variation and does not reflect a generally skewed distribution of rare variants in the material.

Gene-based SKAT analysis revealed no specific mono- or oligogenic signals sufficient to drive the observed association in both cohorts, suggesting multiple genes contribute to the observed signal (Supplemental Table S16). This is also visualized in Figure 3, which shows the amount of variance contributed by each gene in the pathway. mtDNA maintenance is a complex process requiring a well-orchestrated synergy of several biological pathways controlled by overlapping, but functionally distinct protein groups. To assess whether the observed association was primarily driven by a particular functional subnetwork, we subdivided the mtDNA maintenance pathway into mtDNA replication, mtDNA repair, and nucleotide homeostasis. None of these subpathways alone was sufficient to drive the overarching pathway signal and observed association with PD (Supplemental Table S17).

Discussion

We show for the first time that idiopathic PD is associated with a significant enrichment of rare, protein-

TABLE 2. SKAT analyses of pathways

Pathway	Nonsynonymous			Damaging		
	ParkWest	PPMI	Meta	ParkWest	PPMI	Meta
Amino acid metabolism	3.35×10^{-1}	6.17×10^{-1}	4.79×10^{-1}	5.08×10^{-1}	5.01×10^{-1}	5.06×10^{-1}
Apoptosis	5.08×10^{-1}	5.52×10^{-1}	5.44×10^{-1}	5.74×10^{-1}	7.40×10^{-1}	7.29×10^{-1}
Dopamine metabolism	NA	4.41×10^{-1}	NA	NA	NA	NA
Fatty acid metabolism	3.88×10^{-1}	4.18×10^{-1}	3.66×10^{-1}	5.91×10^{-1}	6.41×10^{-2}	1.58×10^{-1}
Glycolysis/gluconeogenesis	7.43×10^{-1}	3.92×10^{-1}	5.86×10^{-1}	5.65×10^{-1}	3.65×10^{-1}	4.38×10^{-1}
Heat production	5.82×10^{-1}	7.32×10^{-1}	7.27×10^{-1}	NA	NA	NA
Heme metabolism	8.43×10^{-2}	8.86×10^{-2}	2.74×10^{-2}	NA	7.34×10^{-1}	NA
Iron homeostasis	7.52×10^{-1}	4.12×10^{-1}	6.08×10^{-1}	NA	9.11×10^{-1}	NA
Iron-sulfur cluster building	NA	7.76×10^{-1}	NA	NA	6.94×10^{-1}	NA
PPAR signaling	9.60×10^{-2}	8.68×10^{-2}	2.99×10^{-2}	5.57×10^{-1}	2.17×10^{-1}	3.08×10^{-1}
Krebs cycle	9.75×10^{-2}	1.57×10^{-2}	6.62×10^{-3}	5.45×10^{-1}	4.56×10^{-2}	1.13×10^{-1}
Mitocarta	3.31×10^{-1}	1.01×10^{-1}	1.05×10^{-1}	1.80×10^{-1}	6.06×10^{-2}	3.82×10^{-2}
Mitochondrial acetylation	8.67×10^{-1}	6.26×10^{-1}	8.33×10^{-1}	NA	5.75×10^{-1}	NA
Mitochondrial calcium homeostasis	2.44×10^{-3}	4.41×10^{-3}	**6.67×10^{-5}**[a]	NA	1.63×10^{-3}	NA
Mitochondrial dynamics and quality control	6.98×10^{-2}	2.62×10^{-1}	7.42×10^{-2}	3.08×10^{-2}	5.52×10^{-1}	1.32×10^{-1}
Mitochondrial ribosome	2.97×10^{-2}	1.56×10^{-1}	2.30×10^{-2}	2.40×10^{-2}	3.87×10^{-1}	6.62×10^{-2}
Mitochondrial transcription	5.59×10^{-1}	6.31×10^{-1}	6.37×10^{-1}	NA	9.81×10^{-1}	NA
Mitochondrial translation	3.00×10^{-1}	4.42×10^{-1}	3.26×10^{-1}	3.56×10^{-1}	6.20×10^{-1}	4.97×10^{-1}
Mitochondrial transport	9.68×10^{-1}	3.76×10^{-1}	8.32×10^{-1}	9.03×10^{-1}	9.55×10^{-1}	9.84×10^{-1}
Mitochondrial tRNA homeostasis	2.93×10^{-1}	1.29×10^{-1}	1.12×10^{-1}	2.74×10^{-1}	5.36×10^{-1}	3.73×10^{-1}
mtDNA maintenance	6.35×10^{-3}	**6.95×10^{-5}**[a]	**3.17×10^{-6}**[a]	1.29×10^{-2}	1.91×10^{-2}	1.25×10^{-3}
NAD metabolism	NA	6.27×10^{-1}	NA	NA	4.22×10^{-1}	NA
One carbon and folate metabolism	2.30×10^{-1}	6.15×10^{-3}	8.63×10^{-3}	3.30×10^{-1}	1.92×10^{-2}	3.16×10^{-2}
Oxidative phosphorylation	8.24×10^{-1}	1.47×10^{-2}	1.47×10^{-1}	4.24×10^{-1}	3.03×10^{-3}	1.36×10^{-2}
Pyruvate metabolism	3.80×10^{-2}	1.11×10^{-1}	1.87×10^{-2}	2.34×10^{-2}	3.72×10^{-2}	4.05×10^{-3}
ROS metabolism	4.82×10^{-1}	2.06×10^{-1}	2.57×10^{-1}	4.56×10^{-1}	6.25×10^{-2}	1.08×10^{-1}
Steroid metabolism	5.61×10^{-1}	2.30×10^{-1}	3.22×10^{-1}	NA	3.33×10^{-1}	NA
Sulfur metabolism	7.87×10^{-1}	1.00	1.00	9.70×10^{-1}	9.71×10^{-1}	9.96×10^{-1}
Urea cycle	9.30×10^{-1}	7.90×10^{-2}	4.55×10^{-1}	NA	NA	NA

Nominal *P* values given. mtDNA, mitochondrial DNA; NA, not analyzed; NAD, nicotinamide adenine dinucleotide; PPAR, peroxisome proliferator-activated receptor; PPMI, Parkinson's Progression Markers Initiative; ROS, reactive oxygen species; SKAT, sequence kernel association test; tRNA, transfer ribonucleic acid.
[a]Statistically significant after multiple testing correction.

altering genetic variants in the pathway of mtDNA replication and repair. This association was significant in both the ParkWest and PPMI cohorts and survived multiple testing correction. Moreover, this association was driven by the combined effects of multiple variants and genes spanning the mtDNA homeostasis network, consistent with a true polygenic signal.

The discrepancy between burden and SKAT in our results suggests that protein-altering variation in the genes controlling mtDNA maintenance can both be deleterious or protective for PD. Biological pathways generally have a high degree of complexity, such that genetically determined variation in their function may be phenotypically neutral or affect associated disease traits in both directions (ie, increase or decrease the risk). This may particularly apply to late-onset disorders that are not likely to be affected by selection pressure, such as PD. SKAT is a variance-component test, which integrates pathway-level information without consideration of the directionality of single-variant effects and is therefore particularly powerful in the presence of both protective and risk variants or numerous noncausal variants.

Our results are highly consistent with prior knowledge of mitochondrial involvement in PD. The disruption of mtDNA maintenance as a result of mutations in genes controlling mtDNA replication and repair or mitochondrial nucleotide homeostasis causes nigrostriatal degeneration with or without clinical parkinsonism.[11–13] Moreover, we and others have shown that mtDNA maintenance is impaired in idiopathic PD, resulting in the accumulation of somatic damage and progressive depletion of the wild-type mtDNA population.[7,40] Based on our results, we propose that this impairment may be partly determined by inherited variation in the genes encoding the pathway of mtDNA homeostasis.

The observation that genetic variation in the mtDNA maintenance pathway appears to influence the risk of PD in both directions is intriguing, but hardly surprising. Maintaining quantitative and qualitative mtDNA integrity in aging neurons requires a balanced interplay of multiple factors including replication processivity,

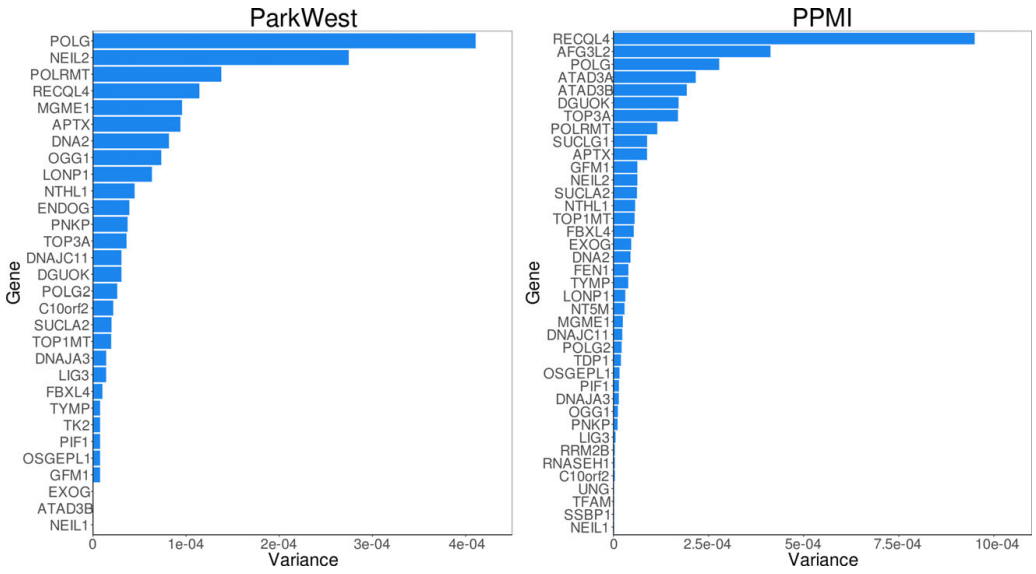

FIG. 3. Gene-based variance in the mitochondrial DNA (mtDNA) maintenance pathway. Sequence kernel association test (SKAT)-based gene-based variance for the mtDNA maintenance pathway in the ParkWest and Parkinson's Progression Markers Initiative datasets for rare, nonsynonymous variation. Variance is a measure of SKAT-based enrichment (ie, the degree of skewed variant distribution between cases and controls) and is defined as the sum of squared difference in minor allele frequency (MAF) between cases and controls for all variants within each gene, $V = \sum_{i=0}^{N} (MAF^{i}_{ctr} - MAF^{i}_{PD})^{2}$. A high variance score for a gene indicates that variants within that gene show a highly uneven distribution between cases and controls. [Color figure can be viewed at wileyonlinelibrary.com]

fidelity, nucleotide metabolism, repair mechanisms, and the underlying mtDNA sequence. It is conceivable that variation in the multitude of genes controlling any and each of these processes may inhibit or enhance mtDNA maintenance thus increasing or decreasing the risk of PD, respectively. It is known for instance that mtDNA haplogroup J, which is linked to higher rates of mtDNA replication and transcription, is associated with a decreased risk for PD.[41] Intriguingly, even apparently pathogenic mutations may have additional beneficial effects. The murine *POLG* mutation p. D257A impairs the enzymes proofreading activity, resulting in increased somatic mtDNA mutagenesis and an aged phenotype in mice. At the same time, however, this defect triggers a neuroprotective compensatory mechanism increasing mtDNA copy number in substantia nigra neurons, rendering them resistant to neurodegeneration.[42] Moreover, rare protective variants have been shown in other complex disorders including autoimmunity,[43] cancer,[44] ischemic vascular disease,[45] and Alzheimer's disease.[46]

The pathway of mitochondrial calcium homeostasis also reached nominal significance in both cohorts, but only survived multiple testing correction in the meta-analysis. Functional evidence suggests that calcium metabolism may be involved in neurodegeneration in PD,[47,48] and this could be partly genetically determined. As this pathway did not survive multiple testing correction in either cohort, however, these results should be interpreted with caution.

Gene-based tests, by either burden or SKAT, produced no exome-wide significant associations. Power analyses based on our Norwegian cohort estimated that a substantially larger sample, probably in the order of 8,000 to 15,000, would be required to detect exome-wide significant genic associations in idiopathic PD by variance- or burden-based tests. Notably, this estimate is based on a clinically homogeneous cohort with validated phenotype, coming from a homogeneous founder population, which is ideal for studying rare variants. Recruiting a PD sample of several thousand individuals will require the combination of multiple cross-sectional materials across several populations. Such material will unavoidably be more phenotypically and genetically heterogeneous. It is therefore likely that even higher numbers will be required in a real-life experiment compared to our simulated estimates. A recent study employing a similar rare-variant collapsing approach in a sample of ~5,000 schizophrenia cases and controls also failed to produce genic associations at exome-wide significance.[49] Given the fact that schizophrenia has a higher estimated heritability than idiopathic PD, these results corroborate our estimates that samples more than 5,000 individuals will be

required to confidently detect gene-based effects in idiopathic PD (Fig. 2).

In conclusion, our data show that large samples are required to identify rare genetic variation associated with PD at the gene level. In contrast, we show that pathway-based analysis of rare genetic variation is a powerful tool for deciphering the genetic susceptibility to PD, even in moderately sized samples. Our results suggest that rare, nonsynonymous variation in the genes encoding the complex pathway of mtDNA maintenance influences the risk for idiopathic PD in 2 independent populations. We propose that this variation underlies part of the observed impairment of mtDNA maintenance in the dopaminergic substantia nigra of individuals with PD and may explain part of the disorder's "missing heritability." The replication of these findings in other sequencing cohorts of idiopathic PD will be essential to further validate these findings.

Acknowledgments: We are grateful to the patients and their families for participating in this study. We thank the Parkinson's Progression Markers Initiative for providing access to their exome data. For up-to-date information on the study, visit www.ppmi-info.org/. We are grateful to Dr Tetyana Zayats for the valuable discussions.

References

1. de Rijk MC, Launer LJ, Berger K, et al. Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 2000;54(11 suppl 5):S21-S23.
2. Kamp F, Exner N, Lutz AK, et al. Inhibition of mitochondrial fusion by alpha-synuclein is rescued by PINK1, Parkin and DJ-1. EMBO J 2010;29(20):3571-3589.
3. Wang X, Yan MH, Fujioka H, et al. LRRK2 regulates mitochondrial dynamics and function through direct interaction with DLP1. Hum Mol Genet 2012;21(9):1931-1944.
4. Park J, Lee G, Chung J. The PINK1-Parkin pathway is involved in the regulation of mitochondrial remodeling process. Biochem Biophys Res Commun 2009;378(3):518-523.
5. Thomas KJ, McCoy MK, Blackinton J, et al. DJ-1 acts in parallel to the PINK1/parkin pathway to control mitochondrial function and autophagy. Hum Mol Genet 2011;20(1):40-50.
6. Sugiura A, McLelland GL, Fon EA, McBride HM. A new pathway for mitochondrial quality control: mitochondrial-derived vesicles. EMBO J 2014;33(19):2142-2156.
7. Dolle C, Flones I, Nido GS, et al. Defective mitochondrial DNA homeostasis in the substantia nigra in Parkinson disease. Nat Commun 2016;7:13548.
8. Grunewald A, Rygiel KA, Hepplewhite PD, Morris CM, Picard M, Turnbull DM. Mitochondrial DNA depletion in respiratory chain-deficient Parkinson disease neurons. Ann Neurol 2016;79(3):366-378.
9. Flones IH, Fernandez-Vizarra E, Lykouri M, et al. Neuronal complex I deficiency occurs throughout the Parkinson's disease brain, but is not associated with neurodegeneration or mitochondrial DNA damage. Acta Neuropathol 2018;135:409-425.
10. Luoma P, Melberg A, Rinne JO, et al. Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: clinical and molecular genetic study. Lancet 2004;364(9437):875-882.
11. Tzoulis C, Schwarzlmuller T, Biermann M, Haugarvoll K, Bindoff LA. Mitochondrial DNA homeostasis is essential for nigrostriatal integrity. Mitochondrion 2016;28:33-37.
12. Tzoulis C, Tran GT, Schwarzlmuller T, et al. Severe nigrostriatal degeneration without clinical parkinsonism in patients with polymerase gamma mutations. Brain 2013;136(Pt 8):2393-2404.
13. Palin EJ, Paetau A, Suomalainen A. Mesencephalic complex I deficiency does not correlate with parkinsonism in mitochondrial DNA maintenance disorders. Brain 2013;136(Pt 8):2379-2392.
14. Balafkan N, Tzoulis C, Muller B, et al. Number of CAG repeats in POLG1 and its association with Parkinson disease in the Norwegian population. Mitochondrion 2012;12(6):640-643.
15. Gaweda-Walerych K, Safranow K, Maruszak A, et al. Mitochondrial transcription factor A variants and the risk of Parkinson's disease. Neurosci Lett 2010;469(1):24-29.
16. Alves G, Muller B, Herlofson K, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. J Neurol Neurosurg Psychiatry 2009;80(8):851-857.
17. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol 2011;95(4):629-635.
18. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754-1760.
19. Picard. http://broadinstitute.github.io/picard. Accessed December 2016.
20. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297-1303.
21. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43(5):491-498.
22. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43:11 10 11-33.
23. Quinlan AR. BEDTools: the Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 2014;47:11.12.1-34.
24. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27(15):2156-2158.
25. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27(21):2987-2993.
26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.
27. International HapMap Consortium. The International HapMap Project. Nature 2003;426(6968):789-796.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904-909.
29. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2(12):e190.
30. The R Project for Statistical Computing. https://www.r-project.org. Accessed August 2017.
31. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164.
32. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res 2016;44(D1):D1251-D1257.
33. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43(Database issue):D447-D452.
34. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91(2):224-237.
35. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89(1):82-93.
36. Lee S, Fuchsberger C, Kim S, Scott L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. Biostatistics 2016;17(1):1-15.

37. Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J Evol Biol 2011;24(8): 1836-1841.

38. Dewey M. metap: meta-analysis of significance values. R package version 082017. https://cran.r-project.org/web/packages/metap/index.html. Accessed August 2017.

39. Westfall PH, Young SS. *Resampling-Based Multiple Testing : Examples and Methods for p-Value Adjustment*. Chichester: Wiley; 1993.

40. Bender A, Krishnan KJ, Morris CM, et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. Nat Genet 2006;38(5):515-517.

41. Hudson G, Nalls M, Evans JR, et al. Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. Neurology 2013;80(22):2042-2048.

42. Perier C, Bender A, Garcia-Arumi E, et al. Accumulation of mitochondrial DNA deletions within dopaminergic neurons triggers neuroprotective mechanisms. Brain 2013;136(Pt 8):2369-2378.

43. Momozawa Y, Mni M, Nakamura K, et al. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. Nat Genet 2011;43(1):43-47.

44. Fejerman L, Ahmadiyeh N, Hu D, et al. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. Nat Commun 2014;5:5260.

45. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. N Engl J Med 2014;371(1):32-41.

46. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 2012;488(7409):96-99.

47. Guzman JN, Sanchez-Padilla J, Wokosin D, et al. Oxidant stress evoked by pacemaking in dopaminergic neurons is attenuated by DJ-1. Nature 2010;468(7324):696-700.

48. Soman S, Keatinge M, Moein M, et al. Inhibition of the mitochondrial calcium uniporter rescues dopaminergic neurons in pink1-/- zebrafish. Eur J Neurosci 2017;45(4):528-535.

49. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506(7487):185-190.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

IV

PLOS ONE

RESEARCH ARTICLE

Meta-analysis of whole-exome sequencing data from two independent cohorts finds no evidence for rare variant enrichment in Parkinson disease associated loci

Johannes Jernqvist Gaare[1,2], Gonzalo Nido[1,2], Christian Dölle[1,2], Paweł Sztromwasser[3,4,5], Guido Alves[6,7], Ole-Bjørn Tysnes[1,2], Kristoffer Haugarvoll[1,2], Charalampos Tzoulis[1,2]*

1 Department of Neurology, Haukeland University Hospital, Bergen, Norway, **2** Department of Clinical Medicine, University of Bergen, Bergen, Norway, **3** Department of Clinical Science, University of Bergen, Bergen, Norway, **4** Computational Biology Unit, Institute of Informatics, University of Bergen, Bergen, Norway, **5** Department of Biostatistics and Translational Medicine, Medical University of Lodz, Lodz, Poland, **6** The Norwegian Centre for Movement Disorders and Department of Neurology, Stavanger University Hospital, Stavanger, Norway, **7** Department of Chemistry, Bioscience and Environmental Engineering, University of Stavanger, Stavanger, Norway

* charalampos.tzoulis@nevro.uib.no, charalampos.tzoulis@helse-bergen.no

Citation: Gaare JJ, Nido G, Dölle C, Sztromwasser P, Alves G, Tysnes O-B, et al. (2020) Meta-analysis of whole-exome sequencing data from two independent cohorts finds no evidence for rare variant enrichment in Parkinson disease associated loci. PLoS ONE 15(10): e0239824. https://doi.org/10.1371/journal.pone.0239824

Editor: Giuseppe Novelli, Universita degli Studi di Roma Tor Vergata, ITALY

Received: June 1, 2020

Accepted: September 15, 2020

Published: October 1, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0239824

Data Availability Statement: The PPMI dataset can be obtained from the study website (www.ppmi-info.org) upon application. The NeuroX

Abstract

Parkinson disease (PD) is a complex neurodegenerative disorder influenced by both environmental and genetic factors. While genome wide association studies have identified several susceptibility loci, many causal variants and genes underlying these associations remain undetermined. Identifying these is essential in order to gain mechanistic insight and identify biological pathways that may be targeted therapeutically. We hypothesized that gene-based enrichment of rare mutations is likely to be found within susceptibility loci for PD and may help identify causal genes. Whole-exome sequencing data from two independent cohorts were analyzed in tandem and by meta-analysis and a third cohort genotyped using the NeuroX-array was used for replication analysis. We employed collapsing methods (burden and the sequence kernel association test) to detect gene-based enrichment of rare, protein-altering variation within established PD susceptibility loci. Our analyses showed trends for three genes (*GALC*, *PARP9* and *SEC23IP*), but none of these survived multiple testing correction. Our findings provide no evidence of rare mutation enrichment in genes within PD-associated loci, in our datasets. While not excluding that rare mutations in these genes may influence the risk of idiopathic PD, our results suggest that, if such effects exist, much larger sequencing datasets will be required for their detection.

Introduction

Parkinson disease (PD) is a complex disorder influenced by the crosstalk between genetic and environmental factors [1]. Monogenic causes account for a small fraction of cases, whereas the

dataset can be obtained through dbGaP (dbGaP Study Accession: phs000918.v1.p1). The ParkWest dataset is currently not publicly available due to limitations set by the regional ethical board approval and study consent form, but will be made available from the Neuromics Lab upon request (https://neuromics.org/contact/).

Funding: CT, Grant number 240369/F20 from the Norwegian Research Council (www.nfr.no). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. CT, grant number 911903 from the Regional Health Authority of Western Norway (www.helse-vest.no). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. JJG, grant number 911988 from the Regional Health Authority of Western Norway (www.helse-vest.no). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

vast majority of patients have idiopathic disease. While genome-wide association studies (GWAS) have revealed several susceptibility *loci* for idiopathic PD, these collectively explain only a fraction of the disorder's estimated heritability, and most have not been linked to pathways which can be targeted by therapies [2]. This is partly due to the uncertainty regarding which genes actually drive the GWAS signals.

The associated variants in GWAS are typically located in noncoding regions of the genome and assumed to be in linkage disequilibrium (LD) with causative variants in nearby genes [3]. Methods to identify candidate genes from GWAS range from simply choosing the closest gene to more sophisticated algorithms [4], but all are, in essence, inferential by nature. Next generation sequencing technologies have enabled us to investigate the impact of rare genetic variation, which is theorized to explain parts of the "missing heritability" in complex diseases [5]. In PD, rare variants have been implicated in sporadic disease both at the gene- [6] and pathway level [7,8]. Whether rare variants can explain GWAS signals in PD, remains, however, unknown.

We hypothesized that gene-based enrichment of rare, protein-altering variation is likely to be found in regions tagged by single nucleotide polymorphisms (SNPs) associated with PD in GWAS, and may help identify the causal genes driving these associations. To test our hypothesis, we selected genes with variants in LD with associated SNPs from the most recent GWAS meta-analysis[9], and tested for enrichment of rare, protein-altering variants in whole-exome sequencing data from two independent cohorts.

Methods

Cohorts and sequencing

The Norwegian whole-exome sequencing (WES) cohort comprised 191 patients with PD from the Norwegian ParkWest study [10] and 219 controls. The control group consisted of individuals with testis cancer (n = 167) and acoustic neuroma (n = 52) who had been recruited and examined at our hospital and had no clinical signs of neurodegenerative- or other neurological disorders. DNA was extracted from blood by routine procedures and sequenced at HudsonAlpha Institute for Biotechnology (Huntsville, Alabama) on the Illumina HiSeq platform using paired-end 100 bp sequencing and Roche-NimbleGen Sequence Capture EZ Exome v2 (173 controls) and v3 (all PD and 46 controls) capture kits. Reads were mapped to the hg19 (GRCh37) reference genome using BWA v0.6.2 [11], PCR duplicates removed with Picard v1.118 [12], and the alignment refined using Genome Analysis Toolkit (GATK) v3.3.0 [13] applying base quality score recalibration and realignment around indels recommended in the GATK Best Practices workflow [14,15]. Variants were called in all samples using GATK HaplotypeCaller [13] with default parameters. Next, Variant Quality Score Recalibration (VQSR) was performed using 99.9% sensitivity threshold [13]. The remaining variants were filtered against the intersection of capture targets (v2 and v3) using BEDtools [16] and VCFtools [17]. Variants with total depth below 10X were marked as unknown genotype (no-call) using BCFtools [18]. In addition, we used a cutoff of at least 6 reads supporting each variant (alternated allele). Indels were removed prior to downstream analyses. The depth distribution for all variants and variants of interest is shown in S1 Fig.

Additional whole-exome sequencing data was obtained from the Parkinson Progression Markers Initiative (PPMI) [19]. WES data was available from 640 individuals (459 cases and 181 controls). Control subjects were individuals without PD 30 years or older, without first degree relatives with PD. Sequencing was performed on the Illumina HiSeq 2500 platform using the Illumina Nextera Rapid Capture Expanded Exome Kit and paired-end 100 bp reads.

Calling and alignment were performed by the PPMI. Indels were removed prior to variant quality control using VCFTools [17].

SNP-chip data was obtained from the International Parkinson's Disease Genomics Consortium (IPDGC) (dbGaP Study Accession: phs000918.v1.p1). The dataset consisted of 11,402 individuals (5,540 cases and 5,862 controls) genotyped on the NeuroX array, comprising approximately 240,000 standard Illumina exome variants and 24,000 custom variants focusing on neurological diseases [20,21].

Individual and variant quality control

Sequencing and genotype data were recoded into binary PLINK input format, and quality control of individual and SNP data was performed for all three cohorts separately. Individuals were excluded if they had an individual genotype missingness rate of $> 2\%$, heterozygosity outside +/- 3 standard deviations (calculated for common and rare variants separately), cryptic relatedness (IBD > 0.2), conflicting sex assignment or non-European ancestry. Population stratification was studies using multi-dimensional scaling against the HapMap populations [22]. Variants were removed if they had a genotyping rate $< 98\%$, different call rates in cases and controls ($p > 0.02$) or departure from the Hardy-Weinberg equilibrium ($p < 10^{-5}$). Only autosomes were kept for downstream analyses. Principal component analysis was performed using Eigensoft [23,24] with standard filtering settings. ANOVA of the 10 first principal components was performed with the significance level set to $p < 0.01$. Significant principal components were included as covariates in all downstream analyses. Outside of the principal component analysis, all quality control procedures were performed using PLINK v1.90 [25] and R [26].

Annotation and subset filtering

The datasets were annotated using ANNOVAR [27] according to the RefSeq gene transcripts, and variants classified as nonsynonymous, stop-gain, stop-loss or splicing were extracted for further analysis. Rare variants were defined as having a minor allele frequency (MAF) of $< 1\%$ in the non-Finnish European population in gnomAD [28].

Selection of genes of interest

Genomic regions associated with PD where extracted from the largest and most recent, to date, meta-analysis of GWASes, which identified 90 SNPs associated with PD at genome-wide significance level [9]. We defined genes of interest as any gene containing a variant in LD within a 2 megabase window around any of these 90 SNPs, with the threshold of LD set to $R^2 > 0.5$. If a variant in LD was localized in an intergenic region, the nearest gene was included. LD calculations were available from the supplementary material of the original study [9], and a total of 303 genes fit the inclusion criteria (S1 Table).

Genetic association analyses

For each cohort, genes of interest were analyzed by two different tests: the burden test and the sequence kernel association test (SKAT) [29], using the SKAT R package v1.3.2.1 [30] with default settings. Statistically significant principal components, as determined by an ANOVA of the first 10 principal components with significance level cutoff set to $p < 0.01$, were added as covariates to all downstream analyses. Only genes with variants in both WES cohorts (Park-West and PPMI) and at least two or more variants across cohorts were included. The meta-analysis was performed using the MetaSKAT R package v0.60 [31], using the same burden test

and SKAT as described above in a meta-analysis framework. For the meta-analysis, we hypothesized that genetic effects should be homogenous across studies, meaning that the same mutation should have the same direction of effect in both cohorts. NeuroX was used as a replication cohort for the results from the WES analyses, and analyzed using the same methods. Only variants defined as rare were included. All p-values were corrected for multiple comparisons using FDR (Benjamini-Hochberg) [32].

Gene set analyses

In addition to the single gene analyses, enrichment of genetic variation across all genes of interest was explored in a gene set analysis. Only rare variants were included, and the combined gene set was analyzed by burden and SKAT tests using the same methods and statistical tools as for the single gene analyses. A subset of loss-of-function (LoF) variants (containing only splicing, stop-loss and stop-gain variants) was extracted and similarly analyzed.

Ethical considerations

These studies were approved by the Regional Committee for Medical and Health Research Ethics, Western Norway (REK 131/04), and all subjects gave written, informed consent. All research was performed in accordance with the relevant guidelines and regulations.

Results

Using the inclusion criteria outlined previously, 168 genes of interest were analyzed in the single gene analyses, comprising a total of 543 rare nonsynonymous, stop-gain, stop-loss or splicing variants in the ParkWest cohort, and 1135 in the PPMI cohort. 160 of these genes were available for replication analysis in the NeuroX dataset, comprising a total of 1380 variants. For the gene set analysis, the number of included variants was 554 in the ParkWest, 1341 in the PPMI and 1534 in the NeuroX cohorts. A total of 14 LoF variants were identified in the ParkWest cohort, 17 in the PPMI cohort and 40 in the NeuroX cohort.

Gene-based analyses indicated three genes with nominally significant p-values (uncorrected $p < 0.05$) across multiple cohorts: *GALC*, *SEC23IP* and *PARP9*. However, no gene reached statistical significance surviving multiple testing correction in either of the cohorts or the meta-analysis (see S2 Table). Similarly, there were no statistically significant results in the gene set analyses (see S3 Table). The top results of the gene enrichment analyses, ranked by nominal p-value in the meta-analysis, are shown in the Tables 1 and 2.

Discussion

Our analyses revealed no statistically significant enrichment of rare variants in genes implicated by previous GWAS in PD. Three genes (*GALC*, *SEC23IP* and *PARP9*) showed trends across multiple cohorts, but none survived multiple testing correction. Nalls et al [9] conducted rare variant burden analysis for *SEC23IP* finding no enrichment signal. Thus, it is highly unlikely that *SEC23IP* is involved in PD. The variant tagging *PARP9* (rs55961674) is a weak expression quantitative trait loci (eQTL) for *PARP9* in some tissues (nerve and thyroid) [33]. However, it is also a strong splicing QTL (sQTL) for *KPNA1*, suggesting that this is a more likely candidate gene. Finally, the variant tagging *GALC* (rs979812) is a strong eQTL for *GALC*, supporting a potential role in PD [33]. *GALC* encodes the enzyme galactocerebrosidase, and mutations in this gene cause Krabbe disease, a lysosomal storage disorder [34]. Current evidence suggests that lysosomal dysfunction plays a key role in PD [35], and rare mutations in a broad range of genes causing lysosomal storage disorders have been associated with PD

Table 1. Top results for burden-based gene enrichment analyses.

Gene	ParkWest			PPMI			Meta			NeuroX		
	Variants	P-value	FDR	Variants	P-value	FDR	Variants	P-value	FDR	Variants	P-value	FDR
SEC23IP	3	0.0276	0.9035	9	0.0332	0.9191	11	0.0040	0.6669	10	0.6819	0.9037
PARP9	3	0.0819	0.9035	9	0.0730	0.9191	11	0.0110	0.7058	7	0.0908	0.6353
GALC	4	0.8111	0.9035	5	0.0032	0.5335	8	0.0210	0.7058	7	0.2607	0.8180
NFKB2	1	0.0180	0.9035	7	0.2543	0.9191	7	0.0333	0.7058	6	0.7477	0.9037
ATP2A1	2	0.0518	0.9035	10	0.2232	0.9191	12	0.0416	0.7058	7	0.6897	0.9037
PBXIP1	1	0.3142	0.9035	8	0.0656	0.9191	9	0.0457	0.7058	10	0.6148	0.9037
CASR	1	0.6461	0.9035	4	0.0276	0.9191	4	0.0457	0.7058	6	0.5906	0.9037
ITGA8	2	0.1473	0.9035	10	0.1905	0.9191	12	0.0533	0.7058	18	0.6969	0.9037
VPS13C	26	0.2533	0.9035	24	0.1276	0.9191	42	0.0557	0.7058	49	0.0636	0.5420
CTSB	3	0.9743	0.9801	15	0.0279	0.9191	17	0.0688	0.7058	10	0.4903	0.9037

Genes are ranked by p-value in the meta-analysis. The FDR-column contains p-values after applying false discovery rate-correction.

https://doi.org/10.1371/journal.pone.0239824.t001

[7]. Mutations of *GBA* in particular, the gene encoding the enzyme glucosylcerebrosidase that carries out a very similar reaction to that of galactocerebrosidase, are the most common genetic risk factor for PD and this association is driven by both common [36] and rare variants [37]. A role for *GALC* in α-synucleinopathies is therefore not farfetched [38].

Taken together, our findings provide no evidence of rare mutation enrichment in PD GWAS loci, in our datasets. These results do not support our initial hypothesis that gene-based enrichment of rare mutations can be helpful in identifying causal genes in PD-associated loci. It should be stressed that these findings do not disprove the hypothesis that rare mutations in these genes may influence the risk of idiopathic PD. They do, however, suggest that if such effects exist, much larger sequencing datasets will be required for their detection.

A few studies with similar approaches to ours have previously been published, using older GWAS data. Foo et al [39] probed 39 genes implicated in PD by GWAS and described enrichment of rare missense variation in *LRRK2*. Sandor et al [40] investigated 329 genes located within GWAS loci, and detected a possible enrichment of missense variation, including both common and rare mutations in their analysis, across the complete gene set. Finally, Jansen et al [41] used a Prix fixe strategy to select one candidate gene per GWAS locus, and detected

Table 2. Top results from SKAT-based gene enrichment analyses.

Gene	ParkWest			PPMI			Meta			NeuroX		
	Variants	P-value	FDR	Variants	P-value	FDR	Variants	P-value	FDR	Variants	P-value	FDR
CASR	1	0.6461	0.9008	4	0.0012	0.2089	4	0.0029	0.4900	6	0.7926	0.9880
PARP9	3	0.1087	0.9008	9	0.1820	0.9393	11	0.0215	0.8873	7	0.0311	0.5528
GALC	4	0.0460	0.9008	5	0.0366	0.9393	8	0.0311	0.8873	7	0.8539	0.9880
NFKB2	1	0.0180	0.9008	7	0.9714	0.9944	7	0.0381	0.8873	6	0.2684	0.9257
SEC23IP	3	0.1313	0.9008	9	0.1113	0.9393	11	0.0469	0.8873	10	0.3037	0.9320
SCARB2	2	0.1983	0.9008	3	0.4048	0.9393	4	0.0829	0.8873	11	0.5043	0.9613
BTNL2	1	0.7423	0.9008	8	0.1152	0.9393	9	0.1085	0.8873	14	0.3768	0.9494
CTSB	3	0.4239	0.9008	15	0.1788	0.9393	17	0.1153	0.8873	10	0.0664	0.7225
PAM	5	0.1454	0.9008	13	0.2638	0.9393	17	0.1188	0.8873	11	0.8316	0.9880
TUFM	1	0.3609	0.9008	2	0.5208	0.9393	2	0.1205	0.8873	1	0.1826	0.9257

Genes are ranked by p-value in the meta-analysis. The FDR-column contains p-values after applying false discovery rate-correction.

https://doi.org/10.1371/journal.pone.0239824.t002

rare variation association signals in *LRRK2*, *STBD1* and *SPATA19*. While we could not replicate enrichment for any of these genes in our datasets, it should be noted that our sample size (n = 1050) is smaller than that of Jansen et al [6,41].

In addition to rare variant enrichment analyses, several other methodologies have been employed to nominate causal genes from GWAS loci. eQTL studies integrate genotype and gene expression date, to identify genes whose expression is regulated by PD associated SNPs [42–46]. The effect of non-coding genetic variation on splicing of pre-mRNA (splicing QTLs or sQTLs) has also recently been highlighted and used to further explore possible causal genes in PD [47]. Finally, epigenetic quantitative trait loci, such as DNA methylation (mQTL), have also been used in combination with GWAS and eQTL data with variable success [48].

PD is a complex disease of heterogeneous etiology. While there is a clear genetic component, as evidenced by twin studies [49], known risk loci are primarily common mutations which, collectively, only explain a fraction of the total estimated heritability [9]. As for other complex disorders, much of the unexplained heritability is believed to be caused by rare variants [50]. Multiple studies have linked common mutations, either through the use of polygenic risk scores [51] or machine learning algorithms [42], to motor progression and cognitive decline. In addition, common genetic variation has also been shown to impact drug responsiveness in PD [52]. Similar applications of rare variants could potentially increase the predictive precision of these models and provide clinicians with a powerful tool to individualize treatment and follow-up for PD patients.

In conclusion, our results indicate that rare variant enrichment alone is unlikely to be helpful in identifying causal risk genes for PD in small to moderately sized cohorts. Larger studies are needed to determine if rare variant enrichment with small effect sizes are present in these genes. Future studies will likely need to integrate multiple types of data, including GWAS, sequencing and various forms of QTL analyses as well as functional experiments in order to better characterize the effects of rare coding variation in PD and identify novel genes and biological pathways.

Supporting information

S1 Fig. Variant depth distribution. A) Depth distribution of all variants called across all samples. B) Depth distribution for the subset of variants called within the predefined regions of interest across all samples. Red bars represent heterozygous variants (0/1), and blue bars represent homozygous (1/1) variants. The vertical dashed line represents the cutoff of minimum 10 reads employed in the analyses.
(PDF)

S1 Table. Genes of interest.
(PDF)

S2 Table. Complete results from gene-based rare variant enrichment analyses.
(PDF)

S3 Table. Gene set analyses.
(PDF)

Acknowledgments

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org.

Author Contributions

Conceptualization: Charalampos Tzoulis.

Data curation: Paweł Sztromwasser.

Formal analysis: Johannes Jernqvist Gaare.

Funding acquisition: Charalampos Tzoulis.

Methodology: Johannes Jernqvist Gaare, Gonzalo Nido, Charalampos Tzoulis.

Project administration: Charalampos Tzoulis.

Supervision: Kristoffer Haugarvoll, Charalampos Tzoulis.

Writing – original draft: Johannes Jernqvist Gaare, Charalampos Tzoulis.

Writing – review & editing: Johannes Jernqvist Gaare, Gonzalo Nido, Christian Dölle, Guido Alves, Ole-Bjørn Tysnes, Kristoffer Haugarvoll, Charalampos Tzoulis.

References

1. Deng H, Wang P, Jankovic J. The genetics of Parkinson disease. Ageing Res Rev. 2018; 42:72–85. Epub 2017/12/31. https://doi.org/10.1016/j.arr.2017.12.007 PMID: 29288112.
2. Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simon-Sanchez J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. Hum Mol Genet. 2012; 21 (22):4996–5009. https://doi.org/10.1093/hmg/dds335 PMID: 22892372.
3. Zhang X, Bailey SD, Lupien M. Laying a solid foundation for Manhattan—'setting the functional basis for the post-GWAS era'. Trends Genet. 2014; 30(4):140–9. Epub 2014/03/26. https://doi.org/10.1016/j.tig.2014.02.006 PMID: 24661571.
4. Tasan M, Musso G, Hao T, Vidal M, MacRae CA, Roth FP. Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. Nat Methods. 2015; 12(2):154–9. Epub 2014/12/23. https://doi.org/10.1038/nmeth.3215 PMID: 25532137.
5. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biol. 2017; 18(1):77. Epub 2017/04/30. https://doi.org/10.1186/s13059-017-1212-4 PMID: 28449691.
6. Jansen IE, Ye H, Heetveld S, Lechler MC, Michels H, Seinstra RI, et al. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. Genome Biol. 2017; 18(1):22. https://doi.org/10.1186/s13059-017-1147-9 PMID: 28137300.
7. Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. Brain. 2017; 140(12):3191–203. Epub 2017/11/16. https://doi.org/10.1093/brain/awx285 PMID: 29140481.
8. Gaare JJ, Nido GS, Sztromwasser P, Knappskog PM, Dahl O, Lund-Johansen M, et al. Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease. Mov Disord. 2018; 33 (10):1591–600. Epub 2018/09/27. https://doi.org/10.1002/mds.64 PMID: 30256453.
9. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. bioRxiv. 2019:388165. https://doi.org/10.1101/388165
10. Alves G, Muller B, Herlofson K, HogenEsch I, Telstad W, Aarsland D, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. J Neurol Neurosurg Psychiatry. 2009; 80(8):851–7. https://doi.org/10.1136/jnnp.2008.168211 PMID: 19246476.
11. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168.
12. http://broadinstitute.github.io/picard [cited 2016 December].
13. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199.
14. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8. https://doi.org/10.1038/ng.806 PMID: 21478889.

15. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 43:11 0 1–33. https://doi.org/10.1002/0471250953.bi1110s43 PMID: 25431634.

16. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014; 47:11 2 1–34. https://doi.org/10.1002/0471250953.bi1112s47 PMID: 25199790.

17. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27(15):2156–8. https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522.

18. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27(21):2987–93. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627.

19. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol. 2011; 95(4):629–35. https://doi.org/10.1016/j.pneurobio.2011.09.005 PMID: 21930184.

20. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet. 2014; 46 (9):989–93. Epub 2014/07/30. https://doi.org/10.1038/ng.3043 PMID: 25064009.

21. Nalls MA, Bras J, Hernandez DG, Keller MF, Majounie E, Renton AE, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol Aging. 2015; 36 (3):1605 e7–12. Epub 2014/12/03. https://doi.org/10.1016/j.neurobiolaging.2014.07.028 PMID: 25444595.

22. International HapMap C. The International HapMap Project. Nature. 2003; 426(6968):789–96. Epub 2003/12/20. https://doi.org/10.1038/nature02168 PMID: 14685227.

23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8):904–9. https://doi.org/10.1038/ng1847 PMID: 16862161.

24. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2(12): e190. Epub 2006/12/30. https://doi.org/10.1371/journal.pgen.0020190 PMID: 17194218.

25. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4:7. https://doi.org/10.1186/s13742-015-0047-8 PMID: 25722852.

26. The R Project for Statistical Computing. www.r-project.org.

27. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38(16):e164. https://doi.org/10.1093/nar/gkq603 PMID: 20601685.

28. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536(7616):285–91. Epub 2016/08/19. https://doi.org/10.1038/nature19057 PMID: 27535533.

29. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011; 89(1):82–93. https://doi.org/10.1016/j.ajhg.2011.05.029 PMID: 21737059.

30. Lee S, Fuchsberger C, Kim S, Scott L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. Biostatistics. 2016; 17(1):1–15. https://doi.org/10.1093/biostatistics/kxv033 PMID: 26363037.

31. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. Am J Hum Genet. 2013; 93(1):42–53. Epub 2013/06/19. https://doi.org/10.1016/j.ajhg.2013.05.010 PMID: 23768515.

32. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57(1):289–300. https://doi.org/10.2307/2346101

33. Consortium GTEx. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013; 45 (6):580–5. https://doi.org/10.1038/ng.2653 PMID: 23715323.

34. Ferreira CR, Gahl WA. Lysosomal storage diseases. Transl Sci Rare Dis. 2017; 2(1–2):1–71. Epub 2017/11/21. https://doi.org/10.3233/TRD-160005 PMID: 29152458.

35. Moors T, Paciotti S, Chiasserini D, Calabresi P, Parnetti L, Beccari T, et al. Lysosomal Dysfunction and alpha-Synuclein Aggregation in Parkinson's Disease: Diagnostic Links. Mov Disord. 2016; 31(6):791–801. Epub 2016/03/01. https://doi.org/10.1002/mds.26562 PMID: 26923732.

36. Gegg ME, Schapira AHV. The role of glucocerebrosidase in Parkinson disease pathogenesis. FEBS J. 2018. Epub 2018/02/01. https://doi.org/10.1111/febs.14393 PMID: 29385658.

37. Davis AA, Andruska KM, Benitez BA, Racette BA, Perlmutter JS, Cruchaga C. Variants in GBA, SNCA, and MAPT influence Parkinson disease risk, age at onset, and progression. Neurobiol Aging. 2016; 37:209 e1–e7. Epub 2015/11/26. https://doi.org/10.1016/j.neurobiolaging.2015.09.014 PMID: 26601739.

38. Marshall MS, Bongarzone ER. Beyond Krabbe's disease: The potential contribution of galactosylceramidase deficiency to neuronal vulnerability in late-onset synucleinopathies. J Neurosci Res. 2016; 94 (11):1328–32. Epub 2016/09/18. https://doi.org/10.1002/jnr.23751 PMID: 27638614.

39. Foo JN, Tan LC, Liany H, Koh TH, Irwan ID, Ng YY, et al. Analysis of non-synonymous-coding variants of Parkinson's disease-related pathogenic and susceptibility genes in East Asian populations. Human Molecular Genetics. 2014; 23(14):3891–7. https://doi.org/10.1093/hmg/ddu086 PMID: 24565865

40. Sandor C, Honti F, Haerty W, Szewczyk-Krolikowski K, Tomlinson P, Evetts S, et al. Whole-exome sequencing of 228 patients with sporadic Parkinson's disease. Sci Rep. 2017; 7:41188. https://doi.org/10.1038/srep41188 PMID: 28117402.

41. Jansen IE, Gibbs JR, Nalls MA, Price TR, Lubbe S, van Rooij J, et al. Establishing the role of rare coding variants in known Parkinson's disease risk loci. Neurobiol Aging. 2017; 59:220 e11–e18. Epub 2017/09/05. https://doi.org/10.1016/j.neurobiolaging.2017.07.009 PMID: 28867149.

42. Latourelle JC, Beste MT, Hadzi TC, Miller RE, Oppenheim JN, Valko MP, et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. The Lancet Neurology. 2017; 16(11):908–16. https://doi.org/10.1016/S1474-4422(17)30328-9 PMID: 28958801

43. Trabzuni D, Ryten M, Emmett W, Ramasamy A, Lackner KJ, Zeller T, et al. Fine-Mapping, Gene Expression and Splicing Analysis of the Disease Associated LRRK2 Locus. PLOS ONE. 2013; 8(8): e70724. https://doi.org/10.1371/journal.pone.0070724 PMID: 23967090

44. Trabzuni D, Wray S, Vandrovcova J, Ramasamy A, Walker R, Smith C, et al. MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. Human Molecular Genetics. 2012; 21(18):4094–103. https://doi.org/10.1093/hmg/dds238 PMID: 22723018

45. Pierce S, Coetzee GA. Parkinson's disease-associated genetic variation is linked to quantitative expression of inflammatory genes. PLOS ONE. 2017; 12(4):e0175882. https://doi.org/10.1371/journal.pone.0175882 PMID: 28407015

46. Murthy MN, Blauwendraat C, Guelfi S, Hardy J, Lewis PA, Trabzuni D, et al. Increased brain expression of GPNMB is associated with genome wide significant risk for Parkinson's disease on chromosome 7p15.3. neurogenetics. 2017; 18(3):121–33. https://doi.org/10.1007/s10048-017-0514-8 PMID: 28391543

47. Li YI, Wong G, Humphrey J, Raj T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. Nature Communications. 2019; 10(1):994. https://doi.org/10.1038/s41467-019-08912-9 PMID: 30824768

48. Kia DA, Zhang D, Guelfi S, Manzoni C, Hubbard L, Reynolds RH, et al. Integration of eQTL and Parkinson's disease GWAS data implicates 11 disease genes. bioRxiv. 2019:627216. https://doi.org/10.1101/627216

49. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. Neurobiol Aging. 2011; 32(10):1923 e1–8. Epub 2011/04/13. https://doi.org/10.1016/j.neurobiolaging.2011.02.017 PMID: 21482443.

50. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–53. Epub 2009/10/09. https://doi.org/10.1038/nature08494 PMID: 19812666.

51. Paul KC, Schulz J, Bronstein JM, Lill CM, Ritz BR. Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. JAMA Neurology. 2018; 75(3):360–6. https://doi.org/10.1001/jamaneurol.2017.4206 PMID: 29340614

52. Miller NS, Chou KL, Bohnen NI, Müller MLTM, Seidler RD. Dopaminergic polymorphisms associated with medication responsiveness of gait in Parkinson's disease. Parkinsonism & related disorders. 2018; 48:54–60. Epub 2017/12/12. https://doi.org/10.1016/j.parkreldis.2017.12.010 PMID: 29249680.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230868478 (print)
9788230850213 (PDF)