A closer look into Parkinson's Disease GWAS genes: Targeted next-generation sequencing approach

By
Uladzislau Rudakou
Department of Human Genetics
Faculty of Medicine
McGill University
Montreal, Quebec, Canada

June 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree
Of
Master of Science

© Uladzislau Rudakou, 2020

Abstract

It is difficult to find someone who has never heard of Parkinson's Disease (PD). This is not surprising because it is the second most common neurodegenerative disorder after Alzheimer's Disease that affects up to 10 million people worldwide. The classic image of the PD patient is an older man that is slightly bent, has shaking hands and a shuffling gait. Indeed, older age is the most important risk factor, but PD can develop in patients below thirty years of age. Men are more likely to suffer from PD, but women are not safe as they make up one third of the patients. The motor symptoms and changes in posture are more visible to the public, but the patients often suffer from debilitating symptoms that are not as easy to see: fatigue, sleep disturbance, gastrointestinal problems, pain, anxiety, depression and cognitive impairments. PD creates a burden to the patients, their families and society as a whole. The impact of PD will keep growing because our population is aging, and the life expectancy is increasing.

PD was described over two hundred years ago, yet the underlying mechanisms of the disease are poorly understood and there is no cure. In the past two decades the research of the genetic background of PD revealed that genetic variants can cause PD, increase or decrease its risk, and have an effect on the progression of the disease. Genome-wide association studies proved to be a great tool to identify novel genetic regions that are associated with PD. However, there is a need for a more in-depth analysis to identify which genes in the GWAS regions are responsible for the signal and then to pinpoint the variants within the gene that drive the association with the disease. The general objective of this thesis was to have a closer look of the genes discovered through PD GWAS using targeted next-generation sequencing.

Chapter 2 in this thesis describes the work focused on investigation of the GTP Cyclohydrolase 1 (*GCH1*) which encodes an essential enzyme in the dopamine synthesis pathway. *GCH1* is one of the genes found in a locus associated with PD in GWASs. Our burden analysis showed that three rare *GCH1* variants that were previously reported to be pathogenic in dopamine-responsive dystonia (DRD) are also associated with PD (p=0.024). Additionally, a common variant rs841 was found to be associated with reduced risk of PD (OR=0.71, 95% CI=0.61-0.83, p= 1.24×10^{-4}).

Chapter 3 describes the project focused on the Vacuolar Protein Sorting 13 Homolog C (*VPS13C*), also a gene within a PD GWAS locus. Furthermore, it was suggested in the literature

that rare bi-allelic variants of *VPS13C* lead to early-onset PD-like disorder with severe clinical presentation. We studied the role of rare bi-allelic variants, as well as rare heterozygous and common variants in our large cohort of late-onset PD. Our results showed that rare bi-allelic or heterozygous variants have no major role in PD without early-onset. However, the analysis of the common variants identified a haplotype including the p.R153H-p.I398I-p.I1132V-p.Q2376Q variants that was associated with reduced the risk of PD (OR=0.48, 95% CI=0.28-0.82, p=0.0052) and is not in linkage disequilibrium (LD) with the top GWAS hits in this locus.

Chapter 4 describes a larger study that examined at 32 genes from 25 loci that were previously reported to be associated with PD risk in GWASs. Rare and common variants were analyzed in three cohorts totaling 2,657 PD patients and 3,647 unrelated controls with both European and Ashkenazi Jewish ancestry. After correcting for multiple testing, the burden of rare variants in three genes was found to be associated with PD risk: *SYT11, FGF20* and *GCH1*. The association of *SYT11* was mainly driven by a rare 3' UTR variant (rs945006601) and was independent of *GBA* variants (p=5.23x10^{-5} after exclusion of all *GBA* variant carriers). A variant located in the promoter region (rs1034608171) was the main driver of the association of *FGF20.* As previously described in Chapter 2, the rare *GCH1* variants found to be associated with PD are also reported to be pathogenic in DRD. We have also identified a few rare variants in *LRRK2* and *SCARB2* that were carried only by controls and might have a protective effect. The analysis of common variants was performed using adjusted logistic regression and was followed up by a meta-analysis of the three cohorts. Common variants in *MAPT, TMEM175, BST1, SNCA* and *GPNMB* were found to be associated with PD risk, and all identified variants were in strong or complete LD with their respective GWAS hits. A potentially novel association was identified for *PM20D1*. A common coding variant (p.Ile148Val) was associated with reduced risk of PD (OR=0.73, 95% CI=0.60-0.89, p=1.161 x10^{-3}) and is not in LD with the top GWAS hits in its region. As this association is nominal, further investigation in a larger cohort should be performed to validate this finding.

The work described in this thesis has demonstrated the benefit of fine mapping the genes discovered through GWAS. Our results provide additional evidence for the role of the identified genes and specific variants in PD risk, making them potential targets for future functional studies and for therapeutics development.

Resumé

Il est difficile de trouver quelqu'un qui n'a jamais entendu parler de la maladie de Parkinson (MP). Cela est loin d'être surprenant, la MP étant la seconde maladie neurodégénérative la plus commune derrière la maladie d'Alzheimer. Elle affecte plus de 10 millions de personnes dans le monde. L'image que l'on associe classiquement au patient de la MP est un homme âgé, recroquevillé, les mains tremblantes, la démarche lourde et difficile. En effet, l'âge est le facteur de risque le plus important, malgré le fait que la MP peut aussi se développer chez des patients de moins de 30 ans. Les hommes sont le plus souvent affecté, mais les femmes ne sont pas à l'abris et constitue le tiers des patients. Les symptômes liés à la motricité at aux changements dans la posture sont plus évidents visuellement mais les patients souffrent souvent de troubles débilitants qui sont parfois peu visibles: fatigue, troubles de sommeil, troubles gastro-intestinaux, douleurs, anxiété, dépression et troubles cognitifs. L'impact de cette maladie ne fait qu'augmenter, notre population étant vieillissante dans une société ou l'espérance de vie s'allonge.

La MP a été décrite pour la première fois il y a au-delà de 200 ans. Par contre, les mécanismes contribuant au développement de la maladie sont très peu connus et il n'existe pas de remède. Les 20 dernières années de recherche sur l'aspect génétique de la maladie ont révélé des marqueurs génétiques associés avec une diminution ou une augmentation du risque de développer la maladie, ainsi qu'un effet sur sa progression. Les études génomiques se sont avérées un outil de choix pour le dépistage de régions génétiques qui pourraient être associés à la MP. Il y a par contre un besoin d'étudier plus en profondeur quels gènes de la région d'études d'association pangénomiques (GWAS) sont responsables pour le signal pour ensuite identifier les variants du gène lui-même qui explique le lien avec la maladie. L'objectif général de la thèse est de porter un regard plus approfondi sur les gênes découverts par l'entremise des GWAS, à l'aide de séquençage de nouvelle génération.

Le deuxième chapitre de la thèse décrit le cheminement spécifique de l'investigation du GTP Cyclohydrolase 1 (*GCH1*), qui encode une enzyme essentielle dans le processus de synthèse de la dopamine. *GCH1* est l'un des gènes qui se trouve dans le locus associé avec la MP dans les études GWAS. La partie la plus significative de notre analyse démontre que le fardeau de trois variants rares du *GCH1* qui ont été reportées comme étant pathogéniques pour la dystonie sensible

à la dopamine (DRD) sont associés avec la MP (p=0.024). De plus, un risque moindre en lien à la MP (OR=0.71, 95% CI=0.61-0.83, p=1.24x10^{-4}) a été trouvé dans le variant commun rs841.

Le chapitre 3 comprend et décrit le Vacuolar Protein Sorting 13 Homolog C (*VPS13C*) qui se trouve aussi être un gène en lien avec la MP dans les GWAS. De plus, la littérature suggère que les rares variants bi-alléliques de *VPS13C* mène à une maladie similaire à la MP mais avec début précoce et avec une présentation clinique sévère. Nous avons étudié le rôle des variants rares bi-alléliques et hétérozygotes ainsi que des variants communs dans notre grande cohorte avec apparitions tardives des symptômes de la MP. Nos résultats démontrent que les variants rares bi-alléliques et hétérozygotes n'ont pas d'impact majeur sur la MP sans l'apparition précoce de symptômes. Cependant, l'analyse des variants communs ont permis d'identifier un haplotype incluant p.R153H-p.I398I-p.I1132V-p.Q2376Q qui est associé avec la réduction du risque de la MP (OR=0.48, 95% CI=0.28-0.82, p=0.0052) sans engendrer un déséquilibre de liaison (LD) avec le signal le plus dominant de GWAS dans ce locus.

Le Chapitre 4 décrit une étude plus vaste qui comporte l'analyse de 32 gènes de 25 loci qui ont été associés avec le risque de la MP dans la GWAS. De variants rares et communs ont été analysés dans 3 cohortes. Celles-ci comprenaient 2,657 patients de la MP et 3,647 cas de control non-reliés dans des communautés européennes et de descendance juifs ashkénazes.

Après la correction de tests multiples, les fardeaux de variants rares dans trois gènes ont été associés avec risques de la MP : *SYT11, FGF20* et *GCH1*. L'association du *SYT11* a été principalement déterminée par une variant rare 3' UTR (rs945006601) et indépendante des variants *GBA* (p=5.23 x10^{-5} après l'exclusion de tous les porteurs de variants *GBA*). Un variant trouvé dans la région du promoteur du gène a été principal déterminant de l'association de *FGF20.* Tel que décrit dans le chapitre 2, les variants rare associés à la MP ont aussi été reportés comme pathogènes dans le DRD. Nous avons aussi identifié quelques variants rares de *LRRK2,* qui ont été portées seulement par les contrôles et qui pourraient avoir un effet protecteur. L'analyse des variants communs a été faite en appliquant la régression logistique ajustée. Une méta-analyse des 3 cohortes a suivi.

Un lien avec le risque de la MP a été trouvé dans les variants communs de *MAPT, TMEM175, BST1, SNCA* et *GPNMB,* et tous les variants identifiés étaient très proches ou complètement dans la LD avec leurs signaux GWAS respectives. Une association potentielle inédite a été identifiée pour *PM20D1*. Un variant commun non-synonyme (p.Ile148Val) a été

associée avec un risque réduit de la MP (OR=0.73, 95% CI=0.60-0.89, p=1.161 x10^{-3}) et n'est pas dans la LD en lien avec le signal le plus significatif de GWAS de la région. Étant donné que cette association est nominale, une investigation sur une cohorte plus grande serait nécessaire pour valider ses résultats.

Cette thèse établit les avantages de la cartographie fine de gènes trouvés à travers la GWAS. Nos résultats fournissent des preuves supplémentaires qui identifient le rôle de ces gènes et des variants spécifiques dans le risque de la MP qui les catégoriseraient comme cibles potentielles pour des futures études fonctionnels ainsi que pour des développements thérapeutiques.

Table of Contents

List of Abbreviations

AAO : age at onset

AJ : Ashkenazi Jewish

bp : base pair

CADD : combined annotation dependent depletion

CCCP : Carbonyl cyanide m-chlorophenyl hydrazone

CF : cystic fibrosis

CI : confidence interval

CNS : central nervous system

CNV : copy-number variation

CRP : C-reactive protein

DA : dopamine

DOPA : dihydroxyphenylalanine

DP : depth of coverage

EDS : excessive daytime sleepiness

EOPD : early onset Parkinson's Disease

eQTL : expression quantitative trait loci

ExAC : exome aggregation consortium

GATK : genome analysis toolkit

Gcase : glucocerebrosidase

GnomAD : genome aggregation database

GTEx : genotype-tissue expression

GWAS : genome-wide association study

hg19 : human genome version 19

LB : Lewy bodies

LD : linkage disequilibrium

LOF : loss-of-function

LOPD : late onset Parkinson's Disease

LP : Lewy pathology

MAF : minor allele frequency

MCI : mild congnitive impairment

MDS : movement disorders society

MIPs : molecular inversion probes

MODY : muaturity-onset diabetes of the young

MR : Mendelian randomization

MSA : multiple system atrophy

NGS : next-generation sequencing

NMS : non-motor symptoms

NS, Nonsyn : non-synonymous

OR : odds ratio

PCR : polymerase chain reaction

PD : Parkinson's Disease

PRS : polygenic risk score

PSP : progressive supranuclear palsy

QC : quality control

RBD : REM sleep behavior disorder

REM : rapid eye movement

SD : standard deviation

SKAT-O : optimized sequence kernel association test

SNc : substantia nigra pars compacta

SNP : single nucleotide polymorphism

SNV : single nucleotide variation

sQTL : splicing quantitative trait loci

T1/2DM : type 1/2 diabetes mellitus

UPS : ubiquitin-protasomal system

UTR : untranslated region

vPSG : video polysomnography

WES : whole exome sequencing

WGS : whole genome sequencing

yo : years of age

αSyn : α-Synuclein

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank all the patients and controls for participating in the studies.

I would like to express my deepest appreciation to my supervisor Dr. Ziv Gan-Or for giving me the opportunity to study in his lab on many exciting projects. Your work ethic and supportive attitude are unparalleled. As the first graduate student produced by your lab, I would like to especially thank you on your remarkable dedication to your students' development and success as scientists, as well as your sincere care for their well-being.

I would also like to thank my supervisory committee members, Dr. Alain Dagher and Dr. Jean-Baptiste Riviere for their time and insight.

I was fortunate to work alongside other members of the lab. I am especially thankful to Jennifer Ruskey for showing me the ropes when I first started in the lab. Our bioinformaticians, Dan Spiegelman and Alex Dionne-Laporte for being patient in introducing me to the world of bioinformatics and spending time helping me with various problems, Lynne Krohn for sharing the struggles of a fresh grad student together since day one and Eric Yu for your assistance with debugging and improving the scripts. I am also grateful to all other members of the Gan-Or lab 'family' both past and present (in no particular order): Mehrdad, Etienne, Khair, Konstantin, Yuri, Prabhjyot, Farnaz, Army, Kate, Catie and Richard. Thank you for not only being my coworkers, but my friends, and making it a pleasure to work alongside you, as well as for all the memorable moments outside of the lab.

Je voudrais remercier mes amis francophones Kelly Malka et Etienne Leveille pour leur aide avec la version française du résumé.

To our extensive list of collaborators, thank you for the insightful comments on the manuscripts and the critique, they are all much better for it.

I am also very thankful to the graduate students/post docs/staff from other labs for their helpful advice and discussions over the years, namely Dr. Bouchra Ouled Amar Bencheikh, Jay Ross, Sandra Laurent and Daniel Rochefort from Dr. Guy Rouleau's lab.

I would like to thank Ross MacKay for all the support throughout my journey as a graduate student in the department of Human Genetics. You were always approachable and your efficiency was astounding.

I wish to acknowledge the generous funding of the McGill Human Genetics department, the International RBD Study Group, and the Montreal Neurological Institute for the travel awards that allowed me to present my research at international meetings and conferences.

In addition to all those who have in some way contributed either directly or indirectly to the research presented in this thesis, there are others who have played a significant part outside of the lab. I am tremendously grateful to my friends for their encouragement and support.

Finally, I would like to express my gratitude to my family. To my dear parents whose hard work and sacrifices provided me the opportunity the get an education and do what makes me happy. To my brother Slava for always finding ways to boost my morale and take my mind off problems. To my beloved wife Polina, for your constant effort to keep me healthy and happy, with delicious food and timely reminders to take time off.

Format of the thesis

This thesis has been written in the manuscript-based format according to the Thesis Preparation Guidelines provided by the McGill Graduate and Postdoctoral studies website.
This thesis contains 7 chapters. Chapter 1 is an introduction to the genetics of Parkinson Disease and describes the relevant background to this thesis. Chapters 2-4 are preceded by a preface that links the thesis together.

Chapter 2 is a manuscript published in *Neurobiology of Aging*, Jan 2019 ;73:231.e1-231.e6. (PMID: 30314816 / DOI: 10.1016/j.neurobiolaging.2018.09.008).

Chapter 3 is a manuscript that was published in *Neurology: Genetics;* 2020 Jan 9; 6(1):385. (PMID:X 32042909 / DOI: 10.1212/NXG.0000000000000385). Electronic supplementary material can be accessed through the journal or on the BioRxiv: https://www.biorxiv.org/content/10.1101/705533v2.supplementary-material

Chapter 4 is a manuscript that was submitted for publication and is currently under review. Electronic supplementary material can be accessed on MedRxiv: https://www.medrxiv.org/content/10.1101/2020.05.29.20116111v1.supplementary-material

Chapters 5 and 6 include a global overview and discussion of the findings in the described studies, conclusions and future directions.

Permissions were obtained for reproducing all published figures.

Contribution of Authors

Chapter 2: **Common and rare *GCH1* variants are associated with Parkinson disease**

List of authors: Uladzislau Rudakou, Bouchra Ouled Amar Bencheikh, Jennifer A. Ruskey, Lynne Krohn, Sandra B. Laurent, Dan Spiegelman, Christopher Liong, Edward Fon, Yves Dauvilliers, Roy N. Alcalay, Nicolas Dupré, Ziv Gan-Or

Contributions:

EF, YD, RNA, and ND Recruited patients

JAR Performed DNA capture. Sequencing was performed at the Genome Quebec and McGill University Genome Center

UR, JAR Sanger sequencing validations

UR, DS, BOAB Bioinformatic processing

UR, BOAB, LK Statistical analysis

UR Writing of the manuscript

ZGO Designed the experiments

ZGO supervised the work, provided financial support and contributed to the writing of the manuscript

All authors revised the manuscript

Chapter 3: **Analysis of common and rare *VPS13C* variants in late onset Parkinson disease**

List of authors: Uladzislau Rudakou, Jennifer A. Ruskey, Lynne Krohn, Sandra B. Laurent, Dan Spiegelman, Lior Greenbaum, Gilad Yahalom, Alex Desautels, Jacques Y. Montplaisir, Stanley Fahn, Cheryl H. Waters, Oren Levy, Caitlin M. Kehoe, Sushma Narayan, Yves Dauvilliers, Nicolas Dupré, Sharon Hassin-Baer, Roy N. Alcalay, Guy A. Rouleau, Edward A. Fon, Ziv Gan-Or

Contributions:

LG, GY, SF, CHW, OL, YD, ND, SHB, RNA Recruited patients

JAR Performed DNA Performed DNA capture. Sequencing was performed at the Genome Quebec and McGill University Genome Center UR Sanger sequencing validations

UR, DS Bioinformatic processing

UR, LK, Statistical analysis

UR Writing of the manuscript

ZGO Designed the experiments

ZGO supervised the work, provided financial support and contributed to the writing of the manuscript

All authors revised the manuscript

Chapter 4: **Targeted sequencing of Parkinson's disease loci genes highlights *FGF20* and other associations**

List of authors: Uladzislau Rudakou, Eric Yu, Lynne Krohn, Jennifer A. Ruskey, Yves Dauvilliers, Dan Spiegelman, Lior Greenbaum, Stanley Fahn, Cheryl H. Waters, Nicolas Dupré, Guy A. Rouleau, Sharon Hassin-Baer, Edward A. Fon, Roy N. Alcalay, Ziv Gan-Or, MD, PhD

Contributions:

YD, LG, SF, CHW, ND, SHB, EF, RNA Recruited patients

UR, JAR, Performed DNA capture. Sequencing was performed at the Genome Quebec and McGill University Genome Center

UR, EY, DS Bioinformatic processing

UR, LK Statistical analysis

UR Writing of the manuscript

ZGO Designed the experiments

ZGO supervised the work, provided financial support and contributed to the writing of the manuscript

All authors revised the manuscript

CHAPTER 1: Introduction and literature review

1.1 Overview of Parkinson's Disease

1.1.1 Epidemiology and etiology

A beautiful metaphor, Silver Tsunami, is often used to describe the surge in the number of elderly people that will happen in the near future. According to the U.S. Census Bureau, come 2034, around 77 million people will be over age 65 in the United States alone[1]. That means that it is imperative to start preparing for this inevitable demographic shift by increasing efforts in the research of aging-related diseases. Parkinson's Disease (PD) deserves a place on that list as it is the second most common neurodegenerative disorder with a prevalence of 1-2% in the population over 60 years old[2] and up to 4% in individuals over 80 years old[3]. Considering that age is the most important risk factor and the population is progressively aging, it was estimated that the number of PD patients will double by 2030[4]. The disease risk is also higher for males, with incidence being 1.5 higher than for females[5]. Due to its long, slowly progressive course, PD creates a major burden on patients, their caregivers, as well as society as a whole, and with the arrival of the Silver Tsunami, the burden created by PD will keep increasing. Debilitating motor (bradykinesia, resting tremor, rigidity, postural instability, freezing) and non-motor symptoms (depression, anxiety, sleep disorders, fatigue, pain, constipation, cognitive and behavioural symptoms) progressively worsen and in later stages lead to a partial or complete loss of autonomy[6]. The estimated economic loss due to health care and reduced employment costs exceeds $US 22,000 per patient annually[2] making PD not only a medical problem, but also an important socio-economic issue.

The multifactorial nature of PD makes it challenging to understand the underlying mechanisms of the disease. It is a complicated interplay between environmental exposures, genetics and the normal process of aging. Various environmental factors have been linked with the risk of developing PD. The increased incidence of PD in rural areas was linked with the use of pesticides and herbicides, as well as with consumption of well water[7]. The studies on pesticide exposure for 5 and 10 years reported increasing the risk of PD by 5% and 11% respectively[8]. Some professions were reported to be associated with increased risk of PD which can be explained by exposure to neurotoxins in workplace: e.g. carpenters (OR=3.9) and cleaners (OR=6.7)[9]. Mild to moderate head trauma has also been associated with PD, and interestingly variants in *SNCA* were

reported to modify the risk[10]. On the other hand, there are reports of cigarette smoking and caffeine consumption associated with reduced risks of PD[11,12]. Current smokers have been reported to have up to 30% less risk of developing PD[13]. However, there are conflicting results regarding the directionality of causation, i.e. whether smoking reduces the risk of PD or individuals at preclinical/prodromal stages of PD[14]. Similar association with reduced risk was observed in male coffee drinkers, while the effect in females was less consistent[11]. Studies on the genes *CYP1A2* and *ADORA2A* involved in caffeine metabolism and encoding an adenosine receptor that is blocked by caffeine, have not yet produced results that could lead to any conclusions about this potential gene-environment interaction[15]. Some other factors were also reported to reduce the risk of PD, including use of nonsteroidal anti-inflammatory agents[16], calcium channel blockers[17], hypolipidemic agents[18] as well as physical activity and Mediterranean diet[19].

The discovery of *SNCA* mutations that cause PD showed that genetics have a role in the disease and it is not a strictly environmental problem as it was once thought. While the majority of PD cases are sporadic, and only ~10%-20% of patients have a family history of PD[20], we now know of many genes that are associated with risk of PD, rate of progression and specific clinical presentations. Expanding our knowledge of genetic influence on the disease can help understanding the underlying mechanisms, identifying targets for future basic and clinical studies, improving clinical trials by stratifying the cohorts by the genetic profiles of the participants, tailoring the treatments to the individual patient and estimating the prognosis.

1.1.2 Clinical presentation of PD

PD is an extremely heterogenous disorder with important interpatient variability in clinical presentation, including age at onset, which ranges from the thirties (or rarely earlier) to over eighty years old[21]. The timeline of PD development (Figure 1.1) can be broken down to different stages. The clinical stage begins at the diagnosis of the disease, when the patient presents with motor symptoms. These symptoms include progressively worsening resting tremor, bradykinesia (slowed movement) and muscle rigidity. Posture and balance impairments, freezing gait, as well as cognitive and psychiatric symptoms may appear in the later stages of typical PD[22]. It is estimated that up to 70% of nigrostriatal dopaminergic neurons are degenerated by the time the motor symptoms appear[23].

The clinical diagnosis of PD is preceded by a prodromal stage, during which non-motor symptoms (NMS) such as REM Sleep Behavior Disorder (RBD), hyposmia, constipation and

others may start appearing[24], but it is still not enough to diagnose the patient with PD. The nonmotor symptoms often develop years before the onset of motor dysfunction and can go unnoticed by the patient for a long time. The preclinical stage refers to the time when the neurodegeneration has already begun, but it has not yet progressed to the point of causing any noticeable motor or non-motor symptoms[25].

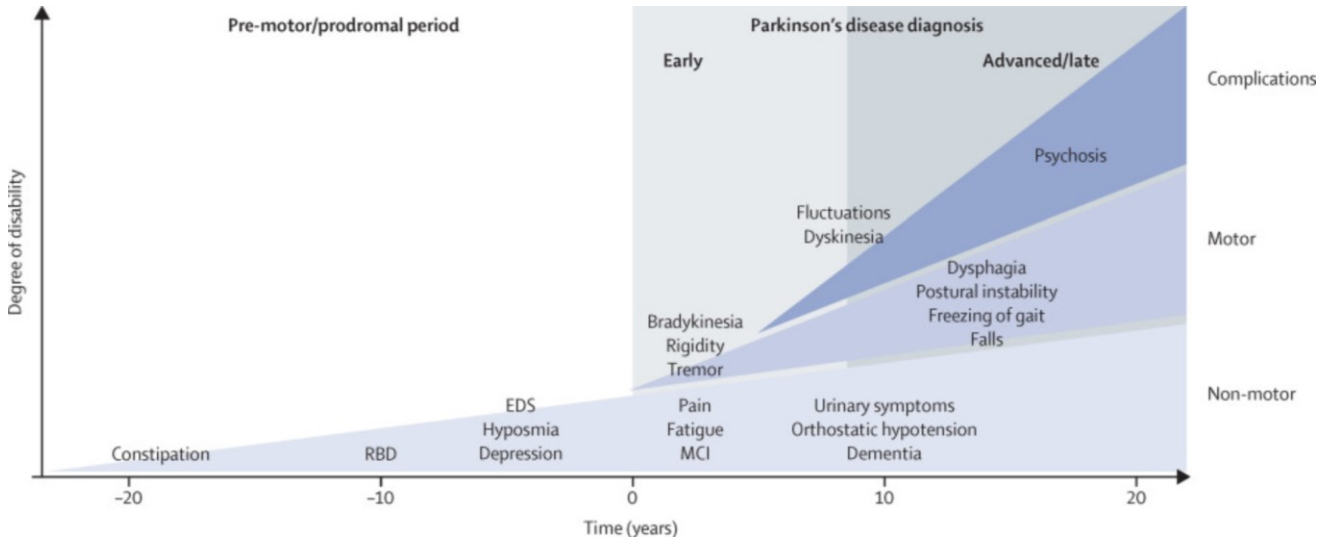


Figure 1.1: Clinical symptoms and time course of Parkinson's Disease.

Legend: The zero point of the x-axis represents the time of diagnosis; negative direction represents the period before the diagnosis. The y-axis represents the measure of the disability by the PD symptoms. EDS: excessive daytime sleepiness; MCI: mild cognitive impairment(adapted with permission from Kalia and Lang, 2015 [26]).

Considering that PD is characterised as a movement disorder, it is not surprising that the NMS were overlooked in the context of PD. Because of their major contribution to disability, poor quality of life and increased disease burden, the assessment of NMS started to be considered a requirement in order to deliver a holistic care to the patients in 2006[27]. PD is a very heterogenous disease and not all patients will develop all of the symptoms discussed in this section. Table 1.1 contains the prevalence of some NMS from a large multicentre study[27].

Cognitive impairment is the most serious nonmotor symptoms that can appear in the later stages of the disease. The prevalence of dementia varies from ~30% and up to ~80% depending on the characteristics of the study population and diagnostic criteria used[28]. Rapid eye movement

(REM) sleep behavior disorder (RBD) is a parasomnia characterized by dream enactment during the REM-phase of sleep due to absence of muscle atonia. The only way to be certain of RBD and avoid misdiagnosis of another parasomnia is to perform a video polysomnography (vPSG)[29], which complicates its use as a biomarker for screening in the general population. Hyposmia is found in up to 90% of PD patients and typically appears years before the motor symptoms, making it a potential biomarker of prodromal PD[30], although it is not specific.

Table 1.1: **Nonmotor symptoms of PD**

Nonmotor symptom	**Cumulative prevalence**	**Reference**
Anxiety	30-40%	[21]
Apathy	30-40%	[21]
Depression	15-50%	[21]
Psychosis	22-42%	[31]
Dementia	30-80%	[28]
Pains	24-83%	[32]
Fatigue	50%	[33]
Constipation	25-63%	[34]
Hyposmia	70-90%	[30]
Dream enactment (possible RBD)	25-50%	[35]
Nocturia	35-78%	[36]
Sleep disturbances	80%	[37]

Differential diagnosis of PD is largely based on clinical assessment. Meta-analysis of 28 studies[38] reported that the UK Brain Bank diagnostic clinical criteria, used as gold standard for a long time, have a 90.8% sensitivity and a 34% specificity, which leads to common misdiagnoses of other tremor disorders, atypical parkinsonian conditions, secondary parkinsonism and other dementias[38]. The phenotypic variability of PD, the overlap of symptoms with other disorders and the increasing probability of co-pathology with advanced age make differential diagnosis challenging. The more recent PD diagnostic criteria by the MDS[25] are independent of dementia

timeline, allowing dementia to be considered a comorbid condition at the time of the diagnosis which is still debated by the experts.

1.1.3 Neuropathology

The current knowledge of PD neuropathology is incomplete. Multiple mechanisms have been suggested as contributors to the underlying pathology. The complex interplay of known and unknown pathways leads to the development of PD and produces the heterogeneity of disease course, symptoms and response to treatments. Just like in the Anna Karenina principle, physiology of healthy individuals is alike while every PD patient is sick in their own way. Any deviation from normal physiological function, due to environmental exposure or genetic variations or both, can trigger a cascade of consequences affecting multiple pathways and producing variability in the resulting disease.

The pathological hallmark of PD typically found in patients with clinical PD is the degeneration of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNc)[39]. These neurons innervate the motor portions of the basal ganglia, the putamen, the subthalamic nucleus, thalamus and the cortex. The disruption of this physiological circuit results in the motor symptoms of PD[40]. It was shown by studies in patients with both short disease durations and prodromal PD, that degeneration of dopaminergic neurons occurs very early[41,42]. Hence to prevent the neurodegeneration, the treatment needs to be provided before the clinical symptoms are manifested, which means that there is a need for biomarkers that would allow early detection.

Another pathological hallmark, Lewy pathology (LP), is typical in PD. however, autopsy studies on patients with some genetic forms showed absence of LP[43,44]. Even in sporadic PD, the localization pattern, the quantity and the spreading speed of LP is highly variable among patients, and some patients do not have LP at all [21]. The relationship between the degeneration of dopaminergic neurons and LP is yet to be elucidated. Lewy bodies (LB) are formed by aggregation of misfolded proteins. More than 90 molecules have been found inside the LBs and one of them is suspected to be a major contributor – α-synuclein (αSyn)[45]. αSyn is a soluble protein made of 140 amino acids and encoded by the *SNCA* gene. Under certain conditions αSyn starts forming dimers, oligomers, fibrils and aggregates[46], and the oligomers are considered to be neurotoxic. It is suggested that αSyn plays a role in vesicular transportation and recycling in the nigrostriatal presynaptic terminals, as well as storage and compartmentalization of dopamine[47], yet its exact function is still unclear.

The dopaminergic neurons of the SNc are not the first target of the αSyn aggregates and Lewy bodies/neurites. At the earlier stages of PD, they can be found in the olfactory bulbs, lower brain stem structures, Meissner's plexus (gastrointestinal nerves) and dorsal motor nucleus of the vagus nerve[22]. This early systemic spread of the pathology may explain some of the earlier onset of nonmotor symptoms such as hyposmia and gastrointestinal symptoms. These observations also lead to a concept of pathological progression proposed by Braak and colleagues (Figure 1.2). In this hypothesis, the LP is propagating to the medulla oblongata and the olfactory system from the periphery via the vagus nerve or olfactory nerves causing autonomic and olfactory disturbances. LP then progresses into the brainstem causing the prodromal sleep disturbances and initial motor symptoms. Finally, limbic system and neocortical regions are affected last, leading to cognitive and behaviour impairments associated with the late stages of the disease. This hypothesis of stepwise progression of LP was supported by findings about the ability of αSyn to be transferred cell-to-cell[48]. What olfactory bulb and Meissner's plexus have in common is their immediate proximity to the outside world, which could be the access point for exogenous agents, such as toxins, pathogens etc. to trigger the pathology.

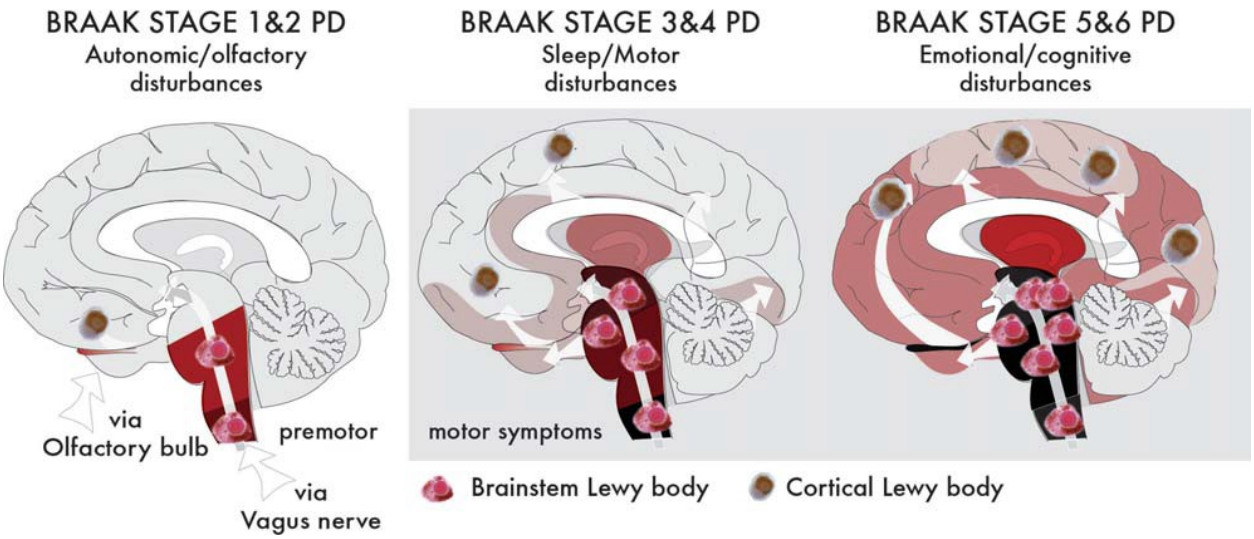


Figure 1.2: **Representation of Braak stages for PD**

Adapted with permission from (Halliday et al., 2011)[49].

Braak's hypothesis laid foundation for another hypothesis involving αSyn as the main culprit – the prion hypothesis. Prion transmission implies that a misfolded protein infects nearby healthy protein and causes it to become misfolded too. Observations supporting this hypothesis

were made in patients receiving neural transplants for PD[50]and in animal experiments of inoculation of animal model brains with fibrillar αSyn or proteins taken from human LP[51,52]. In both cases the spread of LP was observed, however there is no reported evidence that this happens in living humans[53].

If the main mechanism of molecular pathogenesis of PD is misfolded protein, why isn't it taken care of by any of the quality control mechanisms that our cells have? Protein misfolding is a normal occurrence in physiology of eukaryotic cells. Chaperones are responsible for the quality control by preventing misfolding and aggregation of proteins, refolding or, if necessary, directing them towards degradation[54]. The misfolded proteins can be cleared by two different main ways: ubiquitin-proteasome system (UPS)[55] and autophagy pathway involving delivery to the lysosome[56]. Genetic mutations affecting both of these systems have been reported in PD[57,58], offering explanation for accumulation of misfolded and aggregated αSyn in PD patients.

Multiple observations linked another factor with the neurodegeneration of the dopaminergic neurons – mitochondrial impairment. Mutations in some genes responsible for familial forms of PD are involved in mitochondrial function and quality control[59]. Considering that physiological functioning of neurons is energy demanding, it is reasonable that impairments to the energy producing organelles may lead to serious consequences.

There are indications that neuroinflammation has an important role in pathogenesis of neurodegenerative diseases, including PD[60]. It was also suggested that oligomers and aggregates of αSyn can trigger neuroinflammation in PD[61]. On the other hand, neuroinflammation could have a role in the spread of αSyn pathology following the prion hypothesis[62]. Studies have shown that inflammatory reactions, infiltration of T-cells (not found in the central nervous system (CNS) at physiological conditions) and activation of glial cells can be found in PD patients, as well as in animal models of parkinsonism, playing a role in the degeneration of the dopamine neurons[63]. Multiple studies have reported signs of oxidative stress in PD patients. It is however hard to establish whether it is a trigger or a consequence of other neuropathological mechanisms[64,65].

1.1.4 Treatment

There is currently no cure, nor disease modifying treatments for PD. The symptomatic relief provided by levodopa is temporary and has side effects that worsen with time. Dyskinesias and motor fluctuations are common levodopa-induced motor complications[66].

The slow progress in drug development can be partially attributed to the lack of clear understanding of the underlying biological mechanisms of PD and to its heterogenous nature.
A surgical treatment called deep brain stimulation (DBS) is one of the most effective treatments for parkinsonian motor symptoms[67]. During this procedure, electrodes are inserted in the specific areas of the brain (subthalamic nucleus or globus pallidus interna), to transmit an electrical impulse generated by a special pacemaker. It can be a good alternative for patients who did not respond to levodopa treatment or suffered from side effects of the medications.

Symptomatic relief is also provided for the nonmotor symptoms. Depression can be treated with tricyclic antidepressants[68] and selective serotonin reuptake inhibitors[69]. There is a variety of antipsychotic treatments available but all of them are associated with increased mortality[21]. Cognitive impairments are treated with cognitive behaviour therapy with some efficacy[70].

1.2 The genetics of Parkinson's Disease

For over two decades, major advancements have been made in the understanding of the genetic component of PD, transforming our notion of PD from being a sporadic disease into being a disorder considerably affected by various genetic factors. PD is the net result of an elaborate interplay of several genes, modifying effects of susceptibility alleles, environmental exposures, gene-environment interactions and the normal process of aging.

With the development of new genetic methods, we now know of numerous genes, genetic loci and specific variants that cause or affect the risk of PD. An important distinction can be made between the two groups: gene variants that cause PD and those that affect the risk. The variants with Mendelian inheritance are further classified to autosomal-dominant (one copy of mutation is sufficient to cause the disease) and autosomal-recessive (two copies of the mutation needed to cause the disease). It is important to note that some genes do not cause PD but can have variants that confer high risk for the disease.

1.2.1 Familial genes in Parkinson's Disease and Parkinsonism

The gene that put PD on the map as a disease with genetic contribution was the αSyn gene – *SNCA*. From the first report as a causal gene in an Italian family with autosomal-dominant form of parkinsonism[71] to the latest findings in GWASs, it is clear that *SNCA* plays an important role in PD. Single nucleotide variants (SNVs), duplications and triplications can be found in *SNCA* and

are associated with autosomal-dominant pattern of inheritance (with a few cases of incomplete penetrance)[72]. There is a multitude of SNVs reported in *SNCA*: p.A53T[73], p.A30P[74], p.E46K[75], p.A18T, p.A29S, p.G51D[76]; yet they are extremely rare in the general population. The p.H50Q variant that was initially reported to cause PD has been recently refuted[77]. Most of the time, the carriers of *SNCA* SNVs were diagnosed with juvenile / early onset parkinsonism with cognitive impairment, sometimes with dementia. Duplications and triplications are more common than SNVs but are still considered rare in PD. Carriers of *SNCA* duplications are mostly phenotypically similar to sporadic PD patients, however the disease onset is earlier, and they are more prone to developing cognitive symptoms. Triplication of *SNCA* typically leads to juvenile parkinsonism with cognitive impairment and psychiatric symptoms[78].

LRRK2 was discovered as another autosomal-dominant cause of PD. The most common variant in *LRRK2* is p.G2019S, which is enriched in certain North African Arabs (30-40%) and Ashkenazi Jews (10-30%) due to founder effect[79-81]. The frequency of this variant is much lower in the general population, for example in patients of European descent, the frequency is around 1% in sporadic PD and 4% in familial PD[82]. *LRRK2* encodes the protein leucine-rich repeat kinase (dardarin) which is an important kinase of Rab proteins. Additionally, it has been suggested to play a role in protein processing, control of vesicular transport, growth and branching of axons and functioning of Golgi apparatus, lysosomes and mitochondria[83]. Variants in *LRRK2* are thought to cause PD via increased kinase activity of the protein. Clinically, *LRRK2*-parkinsonism is similar to sporadic PD with late onset. While the autopsies of the majority of p.G2019S carriers (65%) showed LB containing neurons, most carriers of other *LRRK2* variants (70%) did not have that PD hallmark pathology[26]. Penetrance of *LRRK2* variants is age-dependent and was estimated to be around 42.5% at age 80[84].

PRKN, PINK1 and *DJ-1* are the three autosomal recessive genes that cause early-onset PD (EOPD). A systematic review was carried out to estimate the prevalence of variants in these three genes in EOPD[85]. Variants in *PRKN* were identified in 4.3% in sporadic EOPD cases, up to 15.5% in familial cases of EOPD. Variants in *PINK1* were found more frequently in Asian EOPD patients than in white or Latin American patients: 13.5%, 0.6% and 0.9% respectively with weighted pooled frequency of pathogenic *PINK1* variants of 3.7%. The frequency of mutations in *DJ-1* is much lower than the previous two genes with 0.8% in familial EOPD cases and 0.4% in sporadic EOPD. All three genes take part in mitochondrial quality control via mitophagy[86].

GBA encoding glucocerebrosidase (GCase) is a perfect example of a gene that confers risk of PD. Homozygous variants in *GBA* cause Gaucher disease, an autosomal-recessive disorder. Clinical observations of comorbidities and increased frequency of PD in families affected by Gaucher disease led to the identification of *GBA* association with PD[87]. Currently, more than 300 *GBA* variants have been identified, including missense, nonsense, indels and recombinations with the *GBA* pseudogene, leading to loss or reduction of GCase activity. Most of the variants are located in exons 8 to 11, and only half of them are relevant in Gaucher disease[88]. Detection of variants in *GBA* is challenging because of the pseudogene *GBAP* that has 96% homology. In a multicentre study, frequency of *GBA* variants in PD was 7%, compared to 1% in controls, identifying *GBA* as an important PD risk factor. The most common variants that also cause Gaucher disease were found to be N370S and L444P[89,90]. N370S is the most common among the two, and has a milder effect, leading to a slight reduction of GCase activity. L444P is considered to be stronger risk factor. More recently, the E326K and T369M variants, which do not cause Gaucher disease, have also been shown to mildly increase the risk of PD[91].There is a significant enrichment in *GBA* variants among Ashkenazi Jews with 19.2.% in patients and 6.4% in controls[92] compared to 6.7% and 1% among Europeans in a study of similar size[93]. Clinically, *GBA*-associated PD has a faster disease progression and higher frequencies of cognitive impairment, RBD and hyposmia than idiopathic PD, with the severe L444P variant having a stronger effect on the rate of progression than the milder N370S[94]. Although the exact mechanism for *GBA* involvement in PD is still not clear, several mechanisms have been suggested, including: reduced αSyn degradation in the lysosome leading to αSyn accumulation, misfolding of GCase leading to accumulation in the endoplasmic reticulum and oxidative stress, effect on mitophagy, and others[95].

Table 1.2 summarizes the genes that were identified in the familial forms of PD which accounts for less than 10% of PD cases[96]. Additionally, *ATP13A2, PLA2G6, FBXO7, DNAJC6, SYNJ1, VPS13C* and others were reported in familial forms of atypical parkinsonism, mostly with juvenile or early onset of disease[96]. *RAB39B* is the only X-linked Mendelian gene in parkinsonism and is associated with early-onset intellectual disability[97]. Multiple other genes have been suggested to cause familial forms of PD or atypical parkinsonism in the past, but were since refuted or remain controversial: *UCHL1, HTRA2, EIF4G1, DNAJC13*, *GIGYF2, TMEM230, CHCHD2, LRP10* and others [97].

Table 1.2: **Known genes of familial form of PD**

Type of inheritance	PARK designation	Gene	Phenotype
Autosomal dominant	PARK1 PARK4	*SNCA*	SNVs: early onset, cognitive impairment, dementia; Duplications: typical PD, often early onset; Triplications: Juvenile onset, cognitive impairment, psychiatric symptoms
	PARK8	*LRRK2*	Typical PD (incomplete penetrance in some variants)
	PARK17	*VPS35*	Typical PD
Autosomal recessive	PARK2	*PRKN*	Early onset PD
	PARK6	*PINK1*	Early onset PD
	PARK7	*DJ-1*	Early onset PD

1.2.2 Genome Wide Association Studies of Parkinson's Disease

A GWAS is based on the hypothesis of common-disease common-variant and seeks to test a large number of single nucleotide polymorphisms (SNPs) randomly distributed across the genome. In a GWAS, risk allele frequencies in cases are compared against healthy controls to test if the SNPs are associated with the disease susceptibility[98]. GWASs proved to be a great tool for identification novel research targets that can lead to expansion of the knowledge of PD biology.

The latest PD GWAS analyzed 7.8 million SNPs in 37.7 thousand cases, 18.6 thousand UK Biobank proxy-cases, and 1.4 million controls[99]. The study identified 90 independent associations in 78 loci, including 38 independent signals in 37 novel loci (Figure 1.3).

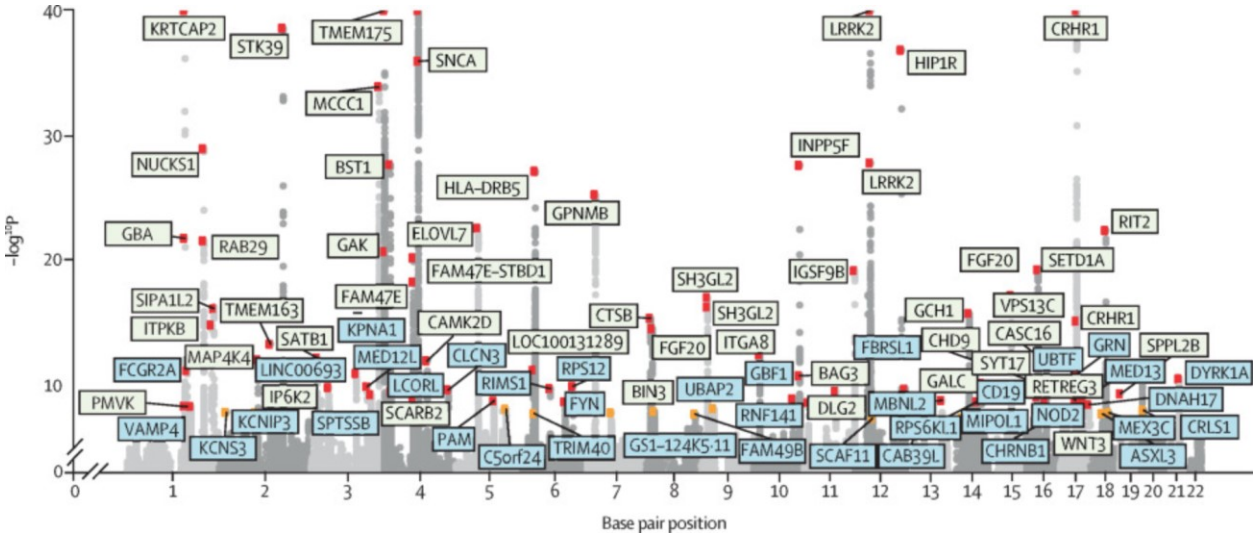


Figure 1.3: **Manhattan plot for significant associations with PD.**

Adapted with permission from Nalls et al. 2019[99]

Legend: The nearest gene to each of the 90 significant variants is labelled in green for previously identified loci and in blue for novel loci. Variant points are color-coded red and orange, with orange representing significant variants at p<5E-08 and red representing significant variants at p<5E-09. The x-axis represents the base pair position of variants per chromosome (1-22), only autosomes were included in the analysis.

Further analysis combined data on methylation and expression with Mendelian randomization (MR) for functional and causal inference. Seventy genes that are possibly causal were identified and suggested for follow-up functional studies. The identified genes were found to be enriched in ten biological pathways related to vacuolar function, calcium transporters, kinase signaling, lysosomal function, endocytosis and dopamine metabolism. The analysis of protein-protein interaction highlighted networks related to chemical signaling pathways or response to a stressor[99].

Polygenic risk score (PRS) is calculated by combining of the effects of multiple variants associated with the disease into a statistical model. The output can be an estimation of the additive effect of all the tested variants on disease risk, age at onset, rate of progression or risk of specific symptoms of the disease, etc. The reliability of the estimation depends on the level of knowledge about the genetic variants that are associated with the tested trait and the completeness of the tested set of variants. Association studies using PRS of PD based on different sets of variants reported

PRS to be correlating with PD age at onset, disease status, rate of motor and cognitive decline[100-102]. The latest PD GWAS also reported that their PRS based on 1805 variants explained 26% of PD heritability in normal prevalence population and up to 36% of heritability in a high-risk population. Unfortunately, the ratio of false positives to true positives of 14 to 1 would be unacceptably high for usage in population-wide screening[99].

1.3 Advantages and limitations of genome-wide association studies

GWASs have been shown to be effective in identifying novel variant–trait associations. More than 3,500 GWASs have been published, with more than 71,000 SNV–trait associations at the genome wide significance threshold[103]. What is even more important than the sheer number of the identified associations is the replicability of the associations found which increased substantially since the invention of the GWAS[104]. GWASs have transformed the field of complex disease genetics. The associations identified through GWAS enabled discoveries of novel disease-causing genes and mechanisms, which in turn led to identification of new drug targets and disease biomarkers that can potentially be used in clinical care. That was already shown in diseases such as Type 2 diabetes mellitus (T2DM), rheumatoid arthritis, psoriasis, osteoporosis, schizophrenia and dyslipidemia[105,106]. GWASs will continue to produce more new loci and improve our understanding of PD biology as the sample size of studies increases, as was suggested in the latest meta-analysis of PD GWAS, which identified 37 novel independent genome-wide significant risk signals almost doubling the number of known risk signals (Figure 1.3)[99].

Testing large number of SNPs comes at a price of an important multiple testing, which forces the need for adopting a stringent statistical significance threshold. Larger sample sizes are required to overcome correction for multiple testing, but it is not always possible, especially in the case of less common diseases or smaller isolated populations. Stringent significance threshold cause SNPs with modest effects to be missed which is not negligible. After all, the value of biological insight gained from GWAS is not proportional to the strength of association. This was clearly demonstrated by multiple drugs that were approved by the US Food and Drug Administration (FDA), which used molecular targets identified by GWAS with less than 1% of phenotypic variation explained by the variant[105]. The missing heritability in GWAS can be partially explained by these biologically important variants with modest effect that are overlooked due to the lack of power. Other variants that are challenging to identify and may contribute to

missing heritability, are rare variants with very low allele frequency, which will require a huge sample size to be detected through GWAS.

Association signals identified through GWASs are often mapped to non-coding regions of the genome, making it challenging to know which gene is responsible for the observed association[107]. Pinpointing causal variants is even more challenging, especially in haplotypes with extended LD[108]. Functional studies suggest that only about one third of causal genes are located closest to the GWAS signal[109,110]. Regulatory variants make it even more complex as they can have an effect on a gene through a long-range functional connection, as in the example of an intronic variant of *FTO* (GWAS locus) that was found to interact with the promoter of *IRX3* located several hundred kilobases away[111]. Additionally, partial LD of common non-causal SNPs with one or multiple rare variants of large effect could lead to a phenomenon called synthetic association. However, it is suggested that those occur rarely and involvement of rare and common variants at a single locus is mostly independent[112,113].

Most GWASs still use data from SNP arrays which are a reliable genotyping technology that provides accuracy while being cost-effective. Also, there has been significant improvement of the original arrays that targeted exclusively common genetic variants, and many known low-frequency and rare variants are now also genotyped. However, to avoid missing very rare variants, GWAS would need to be based on high-depth whole genome sequencing (WGS), which is cost prohibitive at this time. Therefore, one of the more cost-effective follow up options for GWASs is to do a targeted next-generation sequencing (NGS).

Relying on genotyping by SNP arrays brings another limitation – dependence on reference panels for imputation[114,115]. This limitation will leave rare variants undetected, can create false haplotypes, and is especially problematic for samples with ethnicities that do not have a reliable reference panel.

Even though GWAS provides practical means to discover valuable associations, in order to explain more of the heritability of complex traits, further steps need to be taken to identify the disease related genes within the GWAS loci, for example by resequencing full candidate genes and performing functional studies.

1.4 Next generation sequencing for gene discovery

Molecular scientists and translational researchers have been focusing on the identification of genetic variants associated with various diseases since the rise of technology that allowed them to do so. Quantitative PCR, Sanger sequencing and microarray technology were the core approaches and had critical roles in genetic research. The arrival of NGS revolutionized the field. With NGS, it became possible to examine large regions of the human genome in large cohorts of samples with one experiment. On top of that, the price of NGS is relatively affordable, making it accessible for a wider research community. The use of NGS can be divided in three main strategies: WGS, whole-exome sequencing (WES) and panel or targeted sequencing[116].

WGS, as its name suggests, is a process of determining the entire DNA sequence of the genome. While it is a good way to capture all of the genetic variation, including the noncoding regions, the sequencing, processing and storing the massive raw files require a considerable budget. WES is an enrichment method that targets most protein-coding regions (the "exome") which effectively capture exonic and splice-site variants in most of the genes. This approach has been proven to be very successful in finding the causing mutations in monogenic diseases[117]. While the price of WGS and WES continues to decrease, it can still be cost prohibitive for many researchers, especially for the study of large cohorts. Targeted NGS is a solution which can be more cost-effective. In targeted sequencing, a specific set of genes or genetic regions are selected to be isolated and enriched. The targeted approach allows to investigate the regions of the genome most relevant to the disease with faster processing and simpler data analysis requirements. Panel sequencing can be used in clinical settings for genetic screening of cases that do not carry the usual variants that are found on genotyping chips. In research, the ability to detect rare variants can enable identification of novel functional variants, biomarkers, and clinically relevant targets for therapies. The reliable use of targeted NGS in PD has already been demonstrated earlier, showing a great potential of this approach for Mendelian and complex genetics[118]. However, an important limitation of targeted sequencing (as well as WES) is its limited ability to identify copy-number variations (CNVs). Also, NGS methods may identify a large number of rare genetic variants which can create challenges with interpretations. In order to process the massive raw data, a number of bioinformatic filters and tools need to be used. In many cases one has to rely on the databases to estimate pathogenicity, functional role, levels of expression in different tissues, etc. of the identified variants.

The method of targeted capture used to produce the work described in this thesis is called molecular inversion probes (MIPs) designed by O'Roak and colleagues[119] (Figure 1.4). It is a variation of padlock probe technology that allows a reliable capture of regions of interest for the fraction of WES cost. Each probe has the same structure, 70bp in length: a common 30bp linker flanked by an extension arm of 16-20 bp and a ligation arm of 20-24bp. The unique, target specific arms of each probe target a specific genomic region of 112bp. Probes that are activated by phosphorylation hybridize to the target region, the gap between specific arms is filled, and the construct circularizes. The unreacted probes stay linear and are degraded by the exonuclease treatment. PCR amplification is performed using the primers corresponding to the common linker. During the amplification step, Illumina sequencing adaptors and sample-specific barcodes are attached.

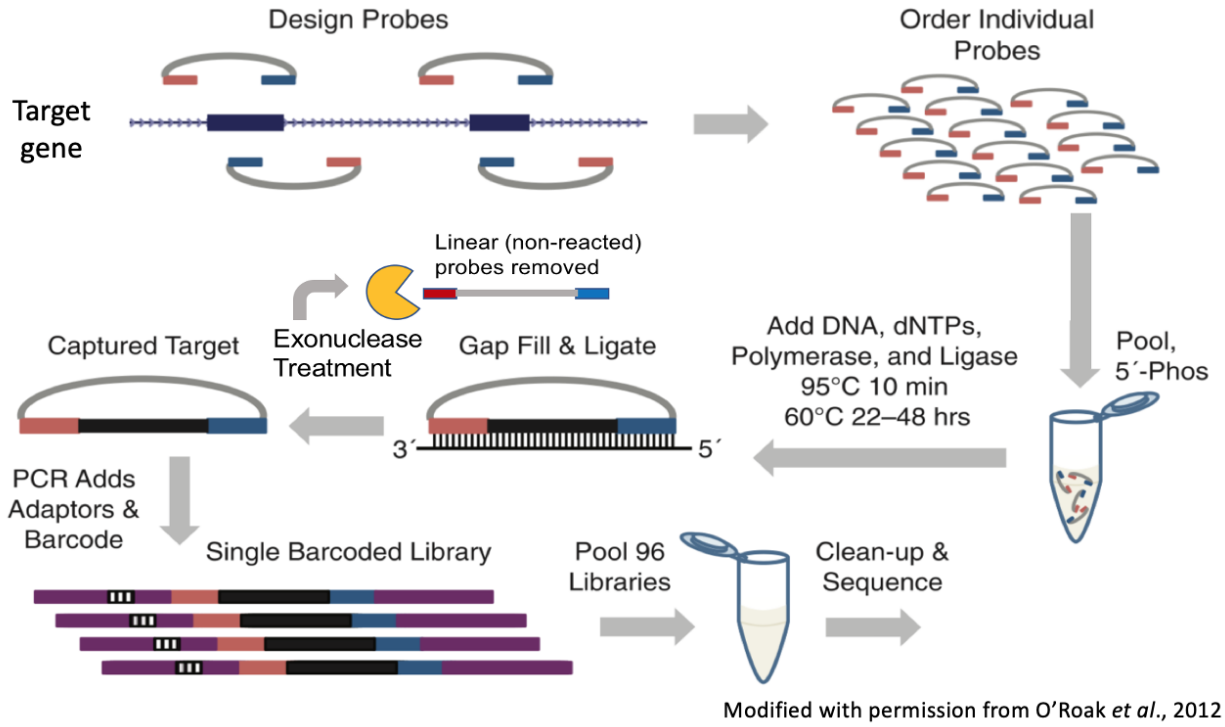


Figure 1.4:**Targeted capture by molecular inversion probes (MIPs)**

Legend: Schematic showing design and general workflow of the targeted capture method using MIPs.

1.5 Aims and rationale

General aim

The aim of this thesis is to study the role of common and rare variants in genes within PD GWAS loci.

Specific aims and rationale

Aim 1 (Chapter 2). To study the role of common and rare *GCH1* variants in PD. *Rationale: GCH1 is located within a GWAS locus, and rare variants in GCH1 may cause dopamine responsive dystonia and were also reported in PD. However, the role of GCH1 variants in PD is still not clear, and GCH1 is not generally considered as a PD-associated gene. GCH1 codes GTP cyclohydrolase 1, an enzyme important in dopamine metabolism. We hypothesize that rare and common variants in GCH1 are associated in PD.*

Aim 2 (Chapter 3). To examine whether *VPS13C* variants are associated with typical PD. *Rationale: Rare biallelic mutations have been reported to cause early onset, atypical parkinsonism, and VPS13C is located within a GWAS locus. However, the role of VPS13C have not been thoroughly studied in PD. We hypothesize that rare monoallelic or biallelic variants in VPS13C, as well as common variants, are associated with PD.*

Aim 3 (Chapter 4). To perform fine mapping of 32 genes from 25 PD GWAS loci. *Rationale: While many PD GWAS loci have been identified, the disease-associated gene within each locus is mostly unknown. We hypothesize that by fine mapping using targeted sequencing we will be able to identify some of the disease-associated genes within the tested loci.*

Preface to Chapter 2

GCH1 is one of the genes identified within a locus associated with PD in GWASs[120,121]. The fact that it encodes GTP-cyclohydrolase 1, an enzyme that is essential in the pathway of dopamine synthesis in nigrostriatal cells[122], provides a plausible biological mechanism for its role in PD. Additionally, mutations in *GCH1* are known to be the major cause of another neurological disorder – dopamine-responsive dystonia (DRD), and there were reports of possible role of DRD pathogenic variants of *GCH1* in PD[123]. However, *GCH1* is still not considered a PD-associated gene by the research community (for example, it is not mentioned in many reviews of PD genetics)[124-126]. The work described in Chapter 2 focused on studying common and rare *GCH1* variants identified through targeted next-generation sequencing, and their association with PD.

CHAPTER 2 : Common and rare GCH1 variants are associated with Parkinson disease

Uladzislau Rudakou[1,2], Bouchra Ouled Amar Bencheikh[2,3], Jennifer A. Ruskey[2,4], Lynne Krohn[1,2], Sandra B. Laurent[2,4], Dan Spiegelman[2,4], Christopher Liong[5], Stanley Fahn, MD[5], Cheryl Waters, MD,[5], Oury Monchi[6,7], Edward A. Fon[2,8], Yves Dauvilliers[9], Roy N. Alcalay[5,10], Nicolas Dupré[11,12], Ziv Gan-Or[1,2,4].

Affiliations: 1. Department of Human Genetics, McGill University, Montréal, QC, Canada; 2. Montreal Neurological Institute, McGill University, Montréal, QC, Canada; 3. Centre de Recherche, Centre Hospitalier de l'Universite de Montreal, Montreal, QC H2X 0A9, Canada; 4. Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada; 5. Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA; 6. Department of Clinical Neurosciences and Department of Radiology, University of Calgary, Calgary, Alberta, Canada; 7. Cumming School of Medicine, Hotchkiss Brain Institute, Calgary, Alberta, Canada; 8. McGill Parkinson Program and Neurodegenerative Diseases Group, Montreal Neurological Institute, McGill University, Montréal, QC, Canada; 9. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France; 10. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA; 11. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, Canada; 12. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, Canada.

This paper was published in *Neurobiology of Aging.* Jan 2019 ;73:231.e1-231.e6. (PMID: 30314816 / DOI: 10.1016/j.neurobiolaging.2018.09.008)[127].

Keywords: Parkinson disease, Dopa-responsive dystonia (DRD), GTP Cyclohydrolase 1 deficiency, *GCH1*.

2.1 Abstract

GCH1 encodes the enzyme GTP cyclohydrolase 1, essential for dopamine synthesis in nigrostriatal cells, and rare mutations in *GCH1* may lead to Dopa-responsive dystonia (DRD). While *GCH1* is implicated in genome-wide association studies in Parkinson disease (PD), only a few studies examined the role of rare *GCH1* variants in PD, with conflicting results. In the current study, *GCH1* and its 5' and 3' untranslated regions were sequenced in 1,113 PD patients and 1,111 controls. To examine the association of rare *GCH1* variants with PD, burden analysis was performed. Three rare *GCH1* variants, which were previously reported to be pathogenic in DRD, were found in five PD patients and not in controls (SKAT, p=0.024). A common haplotype, tagged by rs841, was associated with a reduced risk for PD (OR= 0.71, 95% CI=0.61-0.83, $p= 1.24x10^{-4}$), and with increased *GCH1* expression in brain regions relevant for PD (www.gtexportal.org). Our results support a role for rare, DRD-related variants, and common *GCH1* variants in the pathogenesis of PD.

2.2 Introduction

The genetics of Parkinson disease (PD) is a rapidly evolving field, which may help identifying patients with specific variants that will be eligible for future, specific precision medicine. Genetic studies from recent years reported conflicting results on the involvement of rare and common *GCH1* variants in PD. GTP-cyclohydrolase 1, encoded by *GCH1,* controls the first, rate-limiting step of the biosynthesis of tetrahydrobiopterin (BH_4), which is an essential cofactor for synthesis of dopamine in nigrostriatal cells[1]. Loss-of-function mutations in *GCH1* have been shown to cause two rare disorders: autosomal dominant DOPA-responsive dystonia (DRD) and autosomal recessive GCH-deficient hyperphenylalaninemia (HPA)[2]. Co-occurrence of DRD and Parkinsonism has been reported in families with *GCH1* mutations[3,4], and a study on sporadic PD patients demonstrated an increased frequency of pathogenic *GCH1* mutations that were previously reported to cause DRD[5]. Three subsequent studies also supported an association between rare *GCH1* variants and PD in different populations[4,6,7]. However, other studies did not provide convincing evidence for association of rare *GCH1* DRD-causing mutations and PD[3,8,9].

Some conflicting results were also reported on common variants in the *GCH1* locus. Large genome wide association studies (GWAS) identified common variants near *GCH1*, associated with risk for PD[10,11]. These were replicated is several studies[12,13], but not in others[14-16], possibly due to the different sizes and ethnicities of the population studied. Understanding whether rare and common *GCH1* variants have a role in PD is of major importance, as GTP-cyclohydrolase 1 can become a target for PD drug development in the upcoming era of precision medicine.

To further examine the potential role of rare and common *GCH1* variants in PD, we sequenced its entire coding regions, as well as the 5' and 3' untranslated regions and the intronic regions around the exon-intron boundaries in two cohorts of PD patients and controls.

2.3 Subjects and methods

2.3.1 Study population

Two cohorts with a total of 1,113 PD patients and 1,111 controls were included: A cohort composed of French and French-Canadian unrelated PD patients (n=538) and controls (n=831), recruited in Quebec (Canada) and in France. Average patient age was 65.7±10.0 years, with 62.9% men. The control population of this cohort included 2 groups, elderly controls (n =201, average age at enrollment of 62.7 ± 8.2 years) and young controls (n = 619, average age at enrollment of 35.4 ± 6.5 years, data on age were not available for 11 controls). There was no significant difference in *GCH1* variant frequencies between the 2 groups, which allowed us to combine them for the analysis (average age of 41.9 ± 13.6 years with 51.7% men). The second cohort was recruited in New York (Columbia University) and included 575 PD patients (average age 66.3±10.55 years, 64% men) and 280 unrelated controls (average age 65.0 ±9.7 years, 35.4% men). As detailed below, due to the differences in age and sex, statistical analysis was adjusted. All PD patients were diagnosed by movement disorder specialists according to the UK brain bank criteria[17], without excluding patients with family history of PD. All patients signed informed consent before entering the study, and the institutional review boards approved the study protocols.

2.3.2 DNA extraction and *GCH1* sequencing

DNA was extracted using a standard salting out protocol. The coding sequence and regulatory regions of *GCH1* were targeted using molecular inversion probes (MIPs), designed as previously described[18]. MIPs were selected based on their predicted coverage quality and overlap. All MIPs used to sequence *GCH1* in the present study are detailed in Supplementary Table 1. Targeted DNA capture and amplification was done as previously described[19], and the full protocol is available upon request. The library was sequenced using Illumina HiSeq 2500 platform at the McGill University and Genome Quebec Innovation Centre. Sequence processing was done by Burrows-Wheeler Aligner for alignment[20], the Genome Analysis Toolkit (GATK, v3.8) for post-alignment cleanup and variant calling[21], and ANNOVAR for annotation[22]. Data on the frequency of each *GCH1* variant were extracted from the public database Exome Aggregation Consortium (ExAC). Only variants with high coverage (>30x) and read quality were included in the analysis. Pathogenicity of variants was examined in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and through specific searches in PubMed. Since exon 1 of *GCH1* was not covered well by the targeted

sequencing, Sanger sequencing of exon 1 was also done in all samples, using the following primers: forward 5' – GAGGCAACTCCGGAAACT – 3', reverse 5' – GCTCATTCCGCAATAAGTGG – 3'.

2.3.3 QC steps

During QC filtration using the PLINK software, we excluded SNPs that failed to follow Hardy-Weinberg equilibrium, set at 0.001 threshold, and SNPs with genotyping rate of less than 90%. Same genotyping rate cut-off was used for individual samples. Threshold for missingness between cases and controls was set at 0.05. After the QC, 1,082 patients and 1,110 controls were included in the analysis. The final genotype call rate post QC filtration was greater than 99%.

2.3.4 Statistical Analysis

The association between common *GCH1* variants and PD was analyzed using a logistic regression with the status (patient or control) as a dependent variable. Since there were differences in age and sex between patients and controls, and since the different recruitment sites recruited patients with different ethnical background, age, sex and recruitment site were used as covariates. To analyse all the rare variants (minor allele frequency [MAF] <1%), optimised sequence Kernel association test (SKAT-O, R package) was performed[23]. Association of presumed pathogenic variants was tested using burden analysis (R package SKAT)[23], since the direction of the association was presumed as pathogenic prior to the test. All other statistical analysis was performed using the SPSS software, version 24 (IBM Inc).

2.4 Results

Table 2.1 details the identified nonsynonymous *GCH1* variants, and Figure 2.1 depicts the location of mutations in *GCH1* reported in PD patients in the current and previous studies on PD. A total of 11 rare variants (MAF <1%) and one less-frequent variant P23L (MAF 1-5%) were identified. There were 6 novel variants that were not found in public databases (2 in patients and 4 in controls). Three variants (p.R184C, p.M221T and p.K224R), that were previously reported to be pathogenic in DRD[2,5,6,24,25] were found in five (~0.5%) PD patients, and none in controls (SKAT burden test p=0.024). The other variants are of unknown significance. When analyzing all nonsynonymous variants, regardless of their contribution to DRD, no association was found (SKAT-O, p=0.223).

One variant (p.V204I) that is reported in the ClinVar database to have conflicting interpretation of pathogenicity in DRD was found in 3 controls (~0.3%) and in no PD patients.

When examining common variants, logistic regression showed a common SNP, rs841, to be associated with reduced PD risk in our data (OR= 0.73, 95% CI=0.62-0.86, $p= 1.24x10^{-4}$). The rs841 SNP was also associated with reduced risk for PD in the PD GWAS portal (www.pdgene.org) (Meta OR=0.91, 95%CI=0.87-0.95, $p=1.00x10^{-6}$)[11,26]. This SNP is in partial linkage disequilibrium (LD) with the GWAS tagging SNP rs11158026[10,11] (D'=0.98, r^2=0.5, due to lower frequency of rs841). To examine whether this SNP is associated with *GCH1* expression, we accessed the Genotype-Tissue Expression portal (GTEx, www.gtexportal.org). Supplementary Figure 2.1 demonstrates that this SNP is significantly associated with increased expression of *GCH1* in multiple brain tissues, including the substantia nigra and other basal ganglia. The rs841 SNP was not associated with age at onset in 947 patients for which data was available.

Clinical presentation of PD patients with DRD-causing *GCH1* mutation

Partial clinical data was available for four of the five patients carrying a DRD-causing *GCH1* mutation. Patient 1 (p.K224R) is a male patient that was diagnosed with idiopathic PD at age 58, presenting with right hand tremor, that later progressed to bi-lateral hand tremor. The patient also presented with rigidity in both knees as well as facial freezing and tremors at rest of lower limbs. The patient had good response to L-DOPA. At age 66, he developed mild cognitive impairment. Patient2 (p.K224R) is a male PD patient with age at onset of 64 years old. He is not treated with L-DOPA and information about dystonia is not available. His unified PD rating scale (UPRDS) part III was 15.5 and his Montreal cognitive assessment (MoCA) was 28 in his last examination. Clinical data was not available for the third patient with the p.K224R mutation. Patient 4 (p.R184C) is a male patient with age at onset of 48 years old and family history of PD (father). His initial clinical presentation was mainly left-sided rigidity, yet later tremor became more predominant. He is treated with L-DOPA medication with good response, however, he suffered from dyskinesia and on/off fluctuations. This patient also suffers from anxiety, hallucinations and probable REM sleep behavior disorder, based on a questionnaire. In 2013 the patient went through a successful deep brain stimulation surgery which led to improvement of motor symptoms. Patient 5 (p.M221T) is a female patient with age at onset of 61 years old. She had an excellent response to L-DOPA medication. Her UPDRS and MoCA scores are 11 and 27, respectively.

2.5 Discussion

The current study provides further support to previous reports suggesting that rare DRD-causing *GCH1* variants may also cause PD, and that common *GCH1* variants are associated with a small effect on PD risk, perhaps through regulation of *GCH1* expression. The negative results previously reported for common and rare *GCH1* variants in PD[3,14,16,27,28] may be due to small sample sizes, differential sequencing techniques, or population-specific effects (Table 2.2 details results from previous sequencing and genotyping studies on *GCH1* in PD).

A total of 20 rare *GCH1* variants have been reported in PD / Parkinsonism (Figure 2.1). However, it is still not clear if all 20 variants indeed have a pathogenic role in PD, as some of them may be rare benign variants that were randomly found in PD patients. For example, there is conflicting evidence regarding the role of p.V204I; this variant was reported in three PD patients[5], but was also found in compound heterozygosity with another pathogenic *GCH1* variant. If the p. V204I was indeed pathogenic as well, this patient should have had the infant-onset severe phenotype[29]. In our study, this variant was found in three controls, and none in patients, further supporting lack of pathogenicity. Furthermore, in ExAC it is found in about 1:500 individuals of South Asian origin (http://exac.broadinstitute.org/variant/14-55312502-C-T), which is a somewhat high frequency for a disease-causing variant. Similarly, we have identified two PD patients with the p. P69L variant, which was previously reported in PD patients with dystonia[30], but was also reported as a benign variant by others[5]. However, this mutation is slightly less common, found in about 1:1000 Europeans. At this point, the pathogenicity of these variants, or alternatively, their role as risk factors with reduced penetrance (as seen with some *GBA* variants for example[31]), cannot be ruled out, and larger genetic studies or functional studies are required to examine their pathogenicity. Other variants such as p.S80N[7,9,32] p.R184C[25,33], p.M221T[2], p.K224R[5,6] are more rare and repeatedly reported in DRD and PD, and thus can be considered as pathogenic for both.

Large GWASs identified a risk locus that includes *GCH1*[10,11], suggesting that common, possibly regulatory variants in *GCH1* may affect the susceptibility for PD. The common SNP rs841, which was identified in our study, seems to be associated with reduced risk of PD. In the GTEx portal (www.gtexportal.org), this SNP is associated with increased *GCH1* expression in brain regions important in PD, including the substantia nigra (Supplementary Figure 2.1). Of note, since this SNP is in LD with the GWAS tagging SNP rs11158026, it is possible that the rs841 SNP

is responsible for the association in this locus, however, functional studies are required to determine whether this or other SNPs in this locus drive the association. Considering that rare, probably loss-of-function (LOF) mutations in *GCH1* cause DRD or PD, that the biologic function of GTP-cyclohydrolase 1 is in the synthesis of dopamine, together with the protective effect of common variants that are likely to increase expression of *GCH1*, may suggest that increasing the expression and/or the activity of *GCH1*/GTP-cyclohydrolase 1 could be an attractive target for drug development for sporadic PD as well.

Clinically, late onset DRD may present as Parkinsonism[26], which may suggest that patients diagnosed with PD who carry a presumed pathogenic *GCH1* mutation, actually have a rare phenotype of DRD which presents similarly to PD[3-5]. Alternatively, it is possible that *GCH1* mutations may predispose to both, and whether a carrier will develop DRD or PD is dependent on other genetic or environmental factors. For example, in a family with a novel pathogenic *GCH1* variant, p.E2G, it seemed to cause both PD and DRD phenotypes in different members of the studied family[4]. Further supporting this notion, neuropathological studies of DRD patients with *GCH1* mutations demonstrated that in most DRD patients there was an absence of Lewy bodies (LB) pathology, while a subset of patients with late-onset Parkinsonism was positive for LB pathology[34]. Chart reviews of two of our patients from NY confirmed the clinical diagnosis of PD.

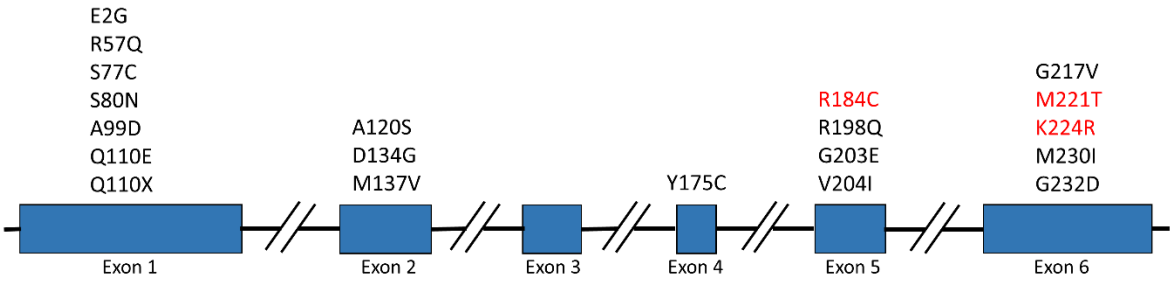
Our study has several limitations. First, in the French-Canadian and French controls, some of the controls are significantly younger than the patients. It is important to note that this is a bias towards the null hypothesis, which only means that our results could be more significant had we used an age-matched control group. Furthermore, we accounted for the age and sex differences by demonstrating that there was no difference in frequencies of rare or common variants between the young and elderly controls, and by adjusting the regression model with age and sex as covariates. Lastly, despite having a relatively large cohort, since many of the *GCH1* variants are rare, a larger study, or rather a meta-analysis of multiple studies, will be required to determine the role of some of the variants which are still questionable.

Overall, our results support a role for rare *GCH1* variants in PD, and for common variants as modifiers of risk for PD. While larger genetic studies, as well as functional studies are still warranted, *GCH1* should already be considered as a target for drug development.

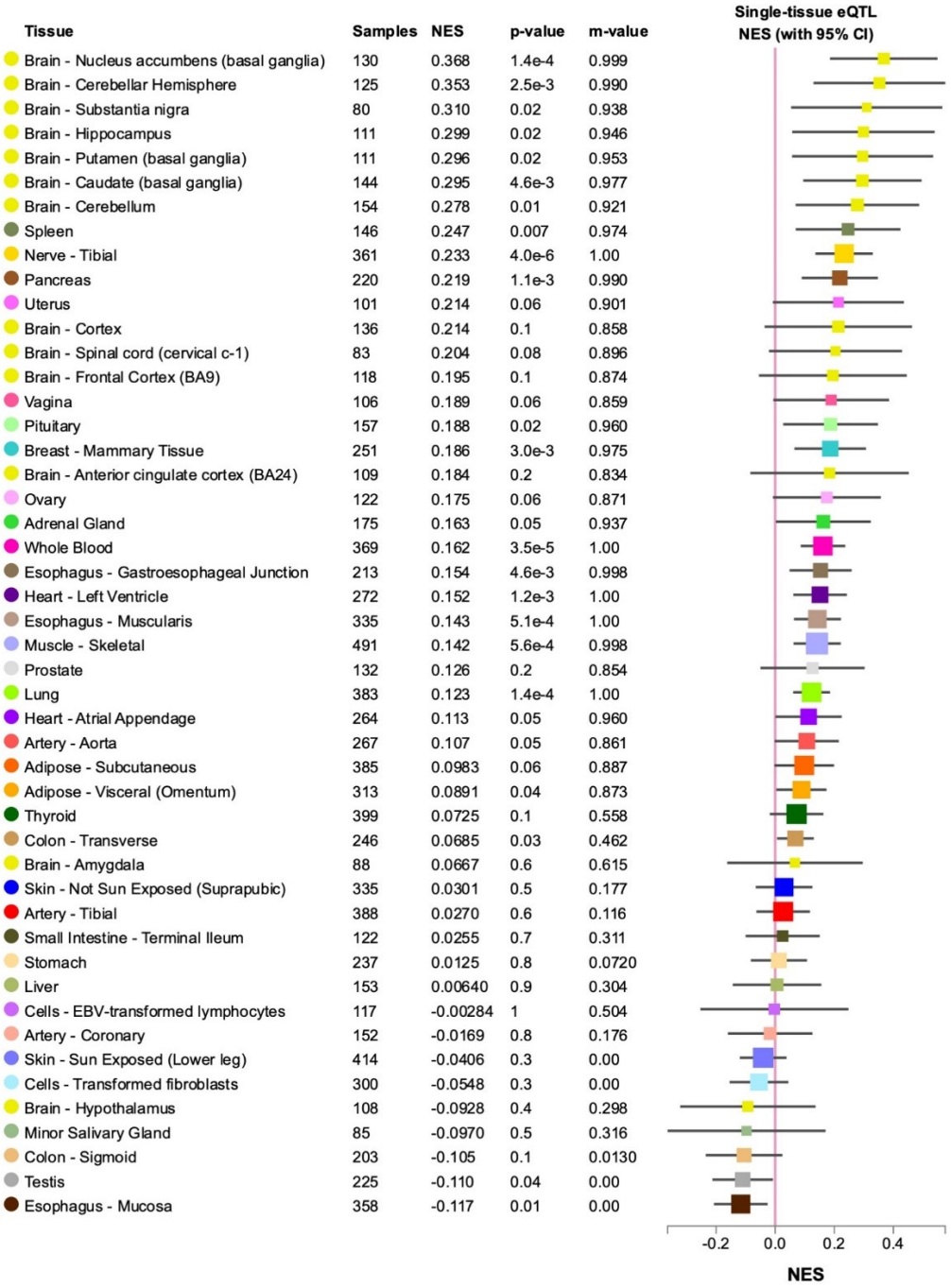
2.6 Figures and Tables

Figure 2.1: Mutations in the guanosine triphosphate cyclohydrolase I (*GCH1*) gene detected in patients with Parkinson's Disease.

The mutations found in this study are indicated highlighted in red.



Supplementary Figure 2.1: **eQTL of rs841 in the GCH1 locus in different tissues**



The rs841 SNP is associated with increased expression of GCH1 in several brain regions, including those involved in PD such as the substantia nigra and other basal ganglia (top part of figure, in yellow). Figure was extracted from the GTEx portal (www.gtexportal.org).

Table 2.1: **Summary of all non-synonymous variants detected in the current study**

dbSNP	Position	Substitution	PD N=1082	Controls N=1110	ExAC MAF in Europeans
rs41298432	14:55369314	p.P23L	3	8	1.502E-02
rs1030068813	14:55369276	p.A36S	0	1	-
-	14:55369231	p.D51N	0	1	-
-	14:55369219	p.G55S	0	1	-
rs56127440	14:55369176	p.P69L	2	0	5.934E-04
-	14:55312562	p.R184C*	1	0	-
rs200891969	14:55312502	p.V204I	0	3	1.798E-04
-	14:55310569	p.N215K	0	1	-
-	14:55310570	p.N215I	0	1	-
-	14:55310841	p.R216Q	1	0	-
rs104894434	14:55310826	p.M221T*	1	0	4.495E-05
rs41298442	14:55310817	p.K224R*	3	0	4.195E-04
Total			11 (1.02%)	16 (1.44%)	

*Variants reported to be pathogenic in DRD.

Table 2.2: **Role of *GCH1* in PD - Summary of previous reports**

Article	Role in PD	#controls	#cases	Diagnosis of cases	Ethnicity	GCH1 findings
Hertz et al., 2006	No	0	87	EOPD	Danish	No pathogenic GCH1 variants found
Cobb et al., 2009	No	0	53	familial EOPD, 21 with EOPD+dystonia	North-American Caucasian	No coding changes/CNV
Momma et al., 2009	Yes	96	2	EOPD	Chinese	1 rare mutation found in patients
Mencacci et al., 2014	Yes	5935	1318	PD	North-American of European descent, Estonians	11 different heterozygous variants at low frequency, 4 of them associated with DRD
Nalls et al., 2014	Yes	95282	13708	PD	European ancestry	GWAS signal
Newman et al., 2014	No	862	1105	PD/dystonia	Australian	No association between PD and the analyzed SNPs
Weissbach et al., 2014	Yes	0	15	PD/Parkinsonism/dystonia	N/A	GCH1 mutation carriers with parkinsonism and idiopathic PD (one had dystonia)
Guella et al., 2015	Yes	290	528	361PD/167 atypical Parkinsonism+DLB+MSA+PSP	N/A	Rare heterozygous nonsynonymous substitutions found in patients
Lewthwaite et al., 2015	Yes	6	6	2EOPD, 1 Parkinsonism, 3DRD	Caucasian	1 novel heterozygous substitution found in a very conserved region
Bandres-Ciga et al., 2016	No	0	134	97LOPD/28EOPD/9FPD	South Spanish	No mutation carriers for GCH1
Chen et al., 2016	Yes	553	528	PD	Taiwanese	rs11158026 increased the risk of developing PD
Rengmark et al., 2016	No	230	509	LOPD	Norwegian/Swedish	No pathogenic GCH1 variants found
Safaralizadeh et al., 2016	Yes	1200	600	PD (excluded EOPD,FPD)	Iranian	Replicated the association of rs11158026 with PD
Chang et al., 2017	Yes	302042	6476	PD	European ancestry	GWAS signal

Xu et al., 2017	Yes	1565	1758	PD	Chinese	7 rare heterozygous non-synonymous mutations in patients
Yang et al., 2017	No	634	589	sporadic PD (FPD excluded)	Han Chinese	No association of rs11158026 with PD
Yan et al., 2018	Yes	438	421	170EOPD/251LOPD(FPD excluded)	Han Chinese	1 LOPD patient (maybe +dystonia) with rare GCH1 mutation(+1 found earlier)
Zou et al., 2018	No	624	579	sporadic PD (FPD excluded)	East Asians	No association of rs11158026 with PD

CNV, copy number variation; DLB, dementia with Lewy Bodies; DRD, dopa-responsive dystonia; EOPD, early-onset PD; FPD, familial PD; GWAS, genome wide association study; LOPD, late onset PD; MSA, multiple system atrophy; PSP, progressive supranuclear palsy

Acknowledgements

We thank the patients and control subjects for their participation in this study. This work was financially supported by the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the Canadian Glycomics Network (GlycoNet). This research was also undertaken thanks in part to funding from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health [K02NS080915, and UL1 TR000040] and the Brookdale Foundation. GAR holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences. ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award. The access to part of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (http://rpq-qpn.ca/en/). We thank Jay Ross, Daniel Rochefort, Helene Catoire, Cathy Mirarchi and Vessela Zaharieva for their assistance. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 05/15/18.

References

1. Kurian, M.A., Gissen, P., Smith, M., Heales, S., Jr. & Clayton, P.T. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. *Lancet Neurol* **10**, 721-33 (2011).
2. Furukawa, Y. *et al.* Dystonia with motor delay in compound heterozygotes for GTP-cyclohydrolase I gene mutations. *Ann Neurol* **44**, 10-6 (1998).
3. Rengmark, A., Pihlstrom, L., Linder, J., Forsgren, L. & Toft, M. Low frequency of GCH1 and TH mutations in Parkinson's disease. *Parkinsonism Relat Disord* **29**, 109-11 (2016).
4. Lewthwaite, A.J. *et al.* Novel GCH1 variant in Dopa-responsive dystonia and Parkinson's disease. *Parkinsonism Relat Disord* **21**, 394-7 (2015).
5. Mencacci, N.E. *et al.* Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain* **137**, 2480-92 (2014).
6. Guella, I. *et al.* Parkinsonism in GTP cyclohydrolase 1 mutation carriers. *Brain* **138**, e349 (2015).
7. Xu, Q. *et al.* Rare GCH1 heterozygous variants contributing to Parkinson's disease. *Brain* **140**, e41 (2017).
8. Bandres-Ciga, S. *et al.* Analysis of the genetic variability in Parkinson's disease from Southern Spain. *Neurobiol Aging* **37**, 210 e1-210 e5 (2016).
9. Yan, Y.P. *et al.* Study of GCH1 and TH genes in Chinese patients with Parkinson's disease. *Neurobiol Aging* (2018).
10. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511-1516 (2017).
11. Nalls, M.A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-93 (2014).
12. Chen, C.M. *et al.* Association of GCH1 and MIR4697, but not SIPA1L2 and VPS13C polymorphisms, with Parkinson's disease in Taiwan. *Neurobiol Aging* **39**, 221 e1-5 (2016).
13. Safaralizadeh, T. *et al.* SIPA1L2, MIR4697, GCH1 and VPS13C loci and risk of Parkinson's diseases in Iranian population: A case-control study. *J Neurol Sci* **369**, 1-4 (2016).

14. Yang, X. *et al.* Polymorphism in MIR4697 but not VPS13C, GCH1, or SIPA1L2 is associated with risk of Parkinson's disease in a Han Chinese population. *Neurosci Lett* **650**, 8-11 (2017).
15. Zou, M. *et al.* Association analyses of variants of SIPA1L2, MIR4697, GCH1, VPS13C, and DDRGK1 with Parkinson's disease in East Asians. *Neurobiol Aging* (2018).
16. Newman, J.R., Todorovic, M., Silburn, P.A., Sutherland, G.T. & Mellick, G.D. Lack of reproducibility in re-evaluating associations between GCH1 polymorphisms and Parkinson's disease and isolated dystonia in an Australian case--control group. *Parkinsonism Relat Disord* **20**, 668-70 (2014).
17. Hughes, A.J., Daniel, S.E., Kilford, L. & Lees, A.J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* **55**, 181-4 (1992).
18. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619-22 (2012).
19. Ross, J.P. *et al.* Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. *Neurobiol Aging* **45**, 212 e13-212 e17 (2016).
20. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
21. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
22. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
23. Lee, S. *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* **91**, 224-37 (2012).
24. Leuzzi, V. *et al.* Autosomal dominant GTP-CH deficiency presenting as a dopa-responsive myoclonus-dystonia syndrome. *Neurology* **59**, 1241-3 (2002).
25. Chenbhanich, J., Sringean, J. & Bhidayasiri, R. Beyond the Classic Segawa Disease, GCH1-Associated Neurodegenerative Parkinsonism: Practical Considerations for Physicians. *J Mov Disord* **10**, 102-104 (2017).
26. Lill, C.M. Genetics of Parkinson's disease. *Mol Cell Probes* **30**, 386-396 (2016).

27. Cobb, S.A. *et al.* GCH1 in early-onset Parkinson's disease. *Mov Disord* **24**, 2070-5 (2009).
28. Hertz, J.M. *et al.* Low frequency of Parkin, Tyrosine Hydroxylase, and GTP Cyclohydrolase I gene mutations in a Danish population of early-onset Parkinson's Disease. *Eur J Neurol* **13**, 385-90 (2006).
29. Weissbach, A. & Klein, C. Hereditary dystonia and parkinsonism: two sides of the same coin? *Brain* **137**, 2402-4 (2014).
30. Furukawa, Y., Filiano, J.J. & Kish, S.J. Amantadine for levodopa-induced choreic dyskinesia in compound heterozygotes for GCH1 mutations. *Mov Disord* **19**, 1256-8 (2004).
31. Hernandez, D.G., Reed, X. & Singleton, A.B. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J Neurochem* **139 Suppl 1**, 59-74 (2016).
32. Cao, L. *et al.* Four novel mutations in the GCH1 gene of Chinese patients with dopa-responsive dystonia. *Mov Disord* **25**, 755-60 (2010).
33. Dobricic, V. *et al.* GCH1 mutations are common in Serbian patients with dystonia-parkinsonism: Challenging previously reported prevalence rates of DOPA-responsive dystonia. *Parkinsonism Relat Disord* **45**, 81-84 (2017).
34. Schneider, S.A. & Alcalay, R.N. Neuropathology of genetic synucleinopathies with parkinsonism: Review of the literature. *Mov Disord* **32**, 1504-1523 (2017).

Preface to Chapter 3

The project described in Chapter 3 also focused on a gene that was identified within a PD GWAS locus – the vacuolar protein sorting 13C gene (*VPS13C*)[120,121]. Homozygous and compound heterozygous mutations in this gene were identified as a rare cause of early onset atypical Parkinsonism with rapid progression and severe clinical presentation[128]. Since full sequencing studies of *VPS13C* were never reported in late onset PD, we decided to examine if variants of this gene have a role in typical PD. The inclusion of participants with European and Ashkenazi Jewish ancestry allowed us to assess the possible differences due to ethnicity.

CHAPTER 3: Analysis of common and rare *VPS13C* variants in late onset Parkinson disease

Uladzislau Rudakou, BSc[1,2], Jennifer A. Ruskey, MSc[2,3], Lynne Krohn, BSc[1,2], Sandra B. Laurent, BSc[2,3], Dan Spiegelman, MSc[2,3], Lior Greenbaum, MD, PhD[4,5,6], Gilad Yahalom, MD[6,7,8], Alex Desautels, MD, PhD[9,10,] Jacques Y. Montplaisir, MD, PhD[9,11], Stanley Fahn, MD[12], Cheryl H. Waters, MD[12], Oren Levy, MD, PhD[12] , Caitlin M. Kehoe, MA[12], Sushma Narayan, MS[12], Yves Dauvilliers, MD, PhD[13], Nicolas Dupré, MD, FRCP[14,15], Sharon Hassin-Baer, MD[5,6,7,8], Roy N. Alcalay, MD, MS[11,16], Guy A. Rouleau, MD, PhD, FRCPC, FRSC[1,2,3], Edward A. Fon, MD, FRCPC[2,3], Ziv Gan-Or, MD, PhD[1,2,3].

Affiliations: 1. Department of Human Genetics, McGill University, Montréal, QC, H3A 1A1, Canada; 2. Montreal Neurological Institute, McGill University, Montréal, QC, H3A 1A1, Canada; 3. Department of Neurology and Neurosurgery, McGill University, Montréal, QC, H3A 1A1, Canada; 4. The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer, Ramat Gan , Israel; 5. The Joseph Sagol Neuroscience Center, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel; 6. Sackler school of medicine, Tel-Aviv University, Tel-Aviv 69978, Israel; 7. Department of Neurology, Sheba Medical Center, Tel Hashomer, 52621 Ramat-Gan, Israel; 8. Movement Disorders Institute, Sheba Medical Center, Tel Hashomer, 52621 Ramat Gan, Israel; 9. Centre d'Études Avancées en Médecine du Sommeil, Hôpital du Sacré-Cœur de Montréal, Montréal, QC, H4J 1C5, Canada; 10. Department of Neurosciences, Université de Montréal, Montréal, QC, H3T 1J4, Canada; 11. Department of Psychiatry, Université de Montréal, Montréal, QC, H3T 1J4, Canada; 12. Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA; 13. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France; 14. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, G1V 0A6, Canada; 15. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, G1V 0A6, Canada;; 16. Taub Institute for Research on Alzheimer's Disease and the Aging Brain,

College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA;

This paper was published in *Neurology: Genetics;* 2020 Jan 9; 6(1):385. (PMID:X 32042909 / DOI: 10.1212/NXG.0000000000000385)[129].

Keywords: Parkinson's disease; genetic variants; *VPS13C*; association study

Publication History: A preprint of this manuscript was previously published in bioRxiv: BIORXIV/2019/705533

Study Funding:

This work was financially supported by the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the Canadian Glycomics Network (GlycoNet). This research was also undertaken thanks in part to funding from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health [K02NS080915, and UL1 TR000040] and the Brookdale Foundation.

3.1 Abstract

Objective: We aimed to study the role of coding *VPS13C* variants in a large cohort of late-onset PD (LOPD) patients.

Methods: *VPS13C* and its untranslated regions were sequenced using targeted next-generation sequencing in 1,567 PD patients and 1,667 controls from 3 cohorts. Association tests of rare potential homozygous and compound heterozygous variants and burden tests for rare heterozygous variants were performed. Common variants were analyzed using logistic regression adjusted for age and sex in each of the cohorts, followed by a meta-analysis.

Results: No bi-allelic carriers of rare *VPS13C* variants were found among patients and two carriers of compound heterozygous variants were found in two controls. There was no statistically significant burden of rare (MAF<1%) or very rare (MAF<0.1%) coding *VPS13C* variants in PD. A *VPS13C* haplotype including the p.R153H-p.I398I-p.I1132V-p.Q2376Q variants was nominally associated with a reduced risk for PD (meta-analysis of the tagging SNP p.I1132V (OR=0.48, 95%CI=0.28-0.82, p=0.0052). This haplotype was not in linkage disequilibrium (LD) with the known genome-wide association study (GWAS) top hit.

Conclusions: Our results do not support a role for rare heterozygous or bi-allelic *VPS13C* variants in LOPD. Additional genetic replication and functional studies are needed to examine the role of the haplotype identified here associated with reduced risk for PD.

3.2 Introduction

The vacuolar protein sorting 13C (*VPS13C*) gene is located within a risk locus for Parkinson Disease (PD), reported in large genome-wide association studies (GWAS) of European population[1,2]. The SNP reported in the GWAS (rs2414739, chr15.hg19:g.61994134G>A) was also studied in several Asian populations and in Iranians, with conflicting results[3-7], possibly due to ethnicity-related differences.

Subsequently, homozygous and compound-heterozygous *VPS13C* mutations were identified as a rare cause of early onset PD (EOPD) characterized by rapid progression and early cognitive dysfunction. It was demonstrated that *VPS13C* is partially localized at the mitochondrial membrane, and its silencing led to mitochondrial dysfunction and increased PINK1/Parkin-dependent mitophagy[8]. A follow-up study in 80 EOPD patients identified an additional patient with compound-heterozygous mutations with similar clinical features to the previously reported patients[9]. In another recent study, a homozygous deletion in *VPS13C* was reported to be the probable cause of an early onset parkinsonism in one patient[10]. Thus far, full sequencing studies of *VPS13C* have not been reported in late onset PD (LOPD).

To further study the potential role of *VPS13C* variants in PD, we sequenced its coding and regulatory regions using targeted next-generation sequencing in three cohorts of PD (predominantly LOPD) and in controls. We examined the association of common, rare and bi-allelic *VPS13C* variants on the risk for PD. We further tested whether any coding variant or variants in the untranslated regions of *VPS13C* are in linkage disequilibrium (LD) with the top *VPS13C*-associated GWAS hit, to determine if any of these variants can explain the GWAS association of this locus.

3.3 Methods

3.3.1 Study population

Three cohorts, with a total of 1,567 unrelated PD patients and 1,667 controls, were included in this study, detailed in Table 3.1. First cohort was composed of French and French-Canadian participants recruited in Quebec (Canada) and in France. This cohort was previously genotyped using the GWAS OmniExpress array, the ethnicity was confirmed using principal component analysis, and samples that were of different ethnicities were not included in this study. The second

cohort was recruited in New York - Columbia University, and was previously described[11]. The majority of participants from New York are of European descent and 38% are Ashkenazi Jewish (AJ), 40% of PD patients and 35% of controls. This difference was not statistically significant, yet we adjusted for ethnicity when analyzing this cohort, to avoid effects of ethnicity on the results. The third cohort was recruited in Israel (Sheba Medical Center) and all participants included in this study from the Israeli cohort are of full AJ origin (all four grandparents are full AJ). All patients were consecutively recruited through the clinics, and they represent the typical LOPD patient population with AAO of about 60 (Table 3.1), as opposed to the studies published so far on *VPS13C* in EOPD. As detailed below, due to the differences in age and sex (Table 3.1), statistical analysis was adjusted and included age and sex as co-variates. To account for different ethnicities in the New York cohort, an ethnicity covariate was also introduced in this cohort (GWAS data was not available for this cohort, therefore the reported ethnicity was used and not principal components). All three cohorts were sequenced in the same lab (McGill University), following the same protocol. All PD patients were diagnosed by movement disorder specialists according to the UK brain bank criteria[12], without excluding patients with family history of PD, since it is now known that there are familial cases of PD, so patients who reported family history of PD were included. However, it is important to emphasize that in the current study only unrelated patients were included, there were no multiple cases from the same family.

3.3.2 Standard Protocol Approvals, Registrations, and Patient Consents

The institutional review board (McGill University Health Center Research Ethics Board - MUHC REB) approved the study protocols (reference number IRB00010120). Informed consent was obtained from all individual participants before entering the study.

3.3.3 DNA extraction and *VPS13C* sequencing

DNA was extracted using a standard salting out protocol. The coding sequence and regulatory regions of *VPS13C* were targeted using molecular inversion probes (MIPs), that were designed as previously described[13]. MIPs were selected based on their predicted coverage, quality and overlap. All MIPs used to sequence *VPS13C* in the present study are included in Table e-1. Targeted DNA capture and amplification was done as previously described[14], and the full protocol is available upon request. The library was sequenced using Illumina HiSeq 2500 platform at the McGill University and Genome Quebec Innovation Centre. Reads were mapped to the human reference

genome (hg19) with Burrows-Wheeler Aligner[15]. Genome Analysis Toolkit (GATK, v3.8) was used for post-alignment quality control and variant calling[16] , and ANNOVAR was used for annotation[17]. Data on the frequency of each *VPS13C* variant were extracted from the public database Genome Aggregation Database (GnomAD)[18]. Validation of the tagging variant p.I1132V was performed using Sanger sequencing, with the following primers: forward 5' – CCGGGAAGGTAATGACAAAA – 3', reverse 5' – CCCCTGATTGAAAAGTCACA– 3'

3.3.4 Quality control

During quality control (QC) filtration using PLINK software v1.9[19], SNPs with genotyping rate lower than 90% were excluded. Genotyping rate cut-off for individuals was 90%, and individuals with a lower genotyping rate were excluded. SNPs that deviated from Hardy-Weinberg equilibrium set at p=0.001 threshold were filtered out. Threshold for missingness difference between cases and controls was set at p=0.05 and the filtration script adjusted it with Bonferroni correction. After these QC steps, cohort composition was as described in Table 3.1. To be included in the analysis, minimum quality score (GQ) was set to 30. Rare variants (minor allele frequency, MAF<0.01 or 0.001) had to have a minimal coverage of >50x to be included, and common variants had to have a minimal coverage of >15x to be included in the analysis.

3.3.5 Statistical Analysis

The association between common *VPS13C* variants and PD was examined by using logistic regression models using PLINK v1.9, with the status (patient or control) as a dependent variable, age and sex as covariates in all cohorts, and AJ ancestry as an additional covariate in the New York cohort. To analyze rare variants (MAF < 0.01) and very rare variants (MAF < 0.001), an optimised sequence Kernel association test (SKAT-O, R package) was performed[20]. In addition, we examined using SKAT-O the burden of predicted pathogenic variants with Combined Annotation Dependent Depletion (CADD) score of ≥ 12.37 representing the top 2% of potentially deleterious variants. The effects of SNP genotypes on the AAO was tested using analysis of variance (ANOVA; in R software). Meta-analysis of common variants in the three cohorts was performed using Metafor Package in R software[21]. Linkage disequilibrium in our data was examined by PLINK v1.9 and LD between discovered SNPs and the GWAS top hit rs2414739 was tested using LDlink application, selecting all non-Finish Europeans[22].

3.3.6 Data Availability Statement

Anonymized data is available upon request by any qualified investigator.

3.4 Results

3.4.1 Rare *VPS13C* variants and homozygous or compound heterozygous *VPS13C* variants are not associated with late onset PD.

The average coverage of *VPS13C* with the MIPs was 94% of nucleotides covered at > 10x, and 90% covered at > 50x. This coverage, while not ideal, is better than the whole exome sequencing coverage reported in the original paper on *VPS13C* in PD, and better than the whole exome and whole genome sequencing coverage of this specific gene in gnomAD (https://gnomad.broadinstitute.org/). There were no differences in coverage between the cohorts and between patients and controls. A total of 60 rare variants that are either nonsynonymous, stop variants or potentially affect a splicing site were identified in the three cohorts and are detailed in Table e-2.

In order to examine whether rare homozygous or compound heterozygous *VPS13C* variants may cause LOPD, and since patients with *VPS13C* bi-allelic mutations are very rare, we included in this analysis only rare variants with allele frequency < 0.001. Only two carriers of two heterozygous variants were identified, and both were controls (Table e-3), suggesting that bi-allelic mutations are not common in LOPD. Of note, we did not examine whether these two variants were on the same allele or different allele (compound heterozygous), since they were found only in controls, which suggests that rare bi-allelic variants are not involved in LOPD in our cohorts.

To further study a potential role for rare (allele frequency < 0.01) or very rare (allele frequency < 0.001) *VPS13C* nonsynonymous or splice variants in LOPD, a SKAT-O was performed on the variants detailed in Table e-2. In the French and French Canadian cohort, 33 (6.6%) PD patients carried a rare variant compared to 49 (6.2%) in controls. In the NY cohort, 43 PD patients (8.6%) carried a rare variant compared to 29 (11.8%) controls. In the Israeli cohort 57 (11.9%) PD patients carried a rare variant compared to 59 (12.1%) among controls. There was no association between rare variants (French/French Canadian cohort p=0.44, New York cohort p=0.34, Israel cohort p=0.91) or very rare variants (French/French Canadian cohort p=0.17, New York cohort p=0.85, Israel cohort p=0.89) and PD. We further examined whether rare variants

that are predicted to be deleterious based on CADD score ≥ 12.37 are enriched in PD (the variants included in this analysis are detailed in Table e-4), and no association was found (French/French Canadian cohort p=0.58, New York cohort p=0.39, Israel cohort p=0.40).

3.4.2 A *VPS13C* haplotype including the p.R153H-p.I398I-p.I1132V-p.Q2376Q coding variants is nominally associated with reduced risk for PD.

We have identified 14 common coding variants in our cohort of French and French Canadians, and 13 such variants in each of NY and Israeli cohorts. More details on the number of carriers and frequencies can be found in Table e-5. To test whether common coding variants in *VPS13C* are associated with LOPD, logistic regression models adjusted for age and sex were performed, and additional adjustment for ethnicity was included in the New York cohort (see methods). A nominal association was observed in four variants (p.R153H [rs12595158, chr15.hg19:g.62316035C>T], p.I398I [rs9635356, chr15.hg19:g.62299603T>G], p.I1132V [rs3784635, chr15.hg19: g.62254989T>C] and p.Q2376Q [rs17238189, chr15.hg19: g.62212781T>C]) with reduced risk for PD in the New York cohort (Table 3.2). These remained nominally significant with and without including adjustment for ethnicity, suggesting that ethnicity has no role in this association, and only one of them, p. I1132V (OR 0.28, 95% CI 0.12-0.64, p=0.0025), remained statistically significant after correction for multiple comparisons. In the two other cohorts, these variants showed the same directionality as in the New York cohort, but did not reach statistical significance (Table 3.2). These four variants, p.R153H, p.I398I, p.I1132V and p.Q2376Q, are in strong or even complete LD (Table e-6). The most significant tagging SNP of this haplotype, p. I1132V, was validated using Sanger sequencing in all three cohorts. Meta-analysis of all the common coding variants showed an association of these four linked SNPs with reduced risk for PD (Figure 3.1), with similar directionality across all cohorts. To examine whether this haplotype may affect AAO of PD, ANOVA with the status of the p. I1132V variant was performed. This variant was not associated with AAO in all three cohorts (French/French Canadian cohort p=0.65, New York cohort p=0.34, Israel cohort p=0.99).

None of these variants were in LD with the known GWAS top hit, rs2414739, and therefore these associations do not explain the GWAS hit in the *VPS13C* locus. Of all the other common coding variants, only one variant, p.E2008D (rs78071599, chr15.hg19: g.62223303C>G), was in some LD with the GWAS top hit (D'=0.808, r^2=0.006, Table e-7), but this variant was not associated with LOPD, and therefore it also cannot explain the GWAS hit in our populations.

Interestingly, an intronic variant, rs78530361 (chr15.hg19: g.62214265C>T), was in strong LD with the top GWAS SNP (D'=1, r^2=0.003). Of note, r^2 is low since this intronic SNP has a much lower allele frequency than the top GWAS hit, but every time the rs78530361 SNP is found, it is on an allele which harbors the top GWAS hit rs2414739. However, in our cohorts, this variant was not associated with PD (p=0.66), likely due to its very low r^2 with the GWAS hit.

3.5 Discussion

Our study, which included full sequencing of *VPS13C* in three cohorts, identified 60 rare *VPS13C* variants (MAF<0.01) that are nonsynonymous or affect splicing, and 18 common variants (MAF>0.01) in coding regions of the gene and splice sites. Our results suggest that rare homozygous and compound heterozygous variants are rare in LOPD and probably have importance mainly in early onset PD, as previously described. We have identified a potentially protective haplotype, which includes four variants, two of which are substitutions of amino acids, p.R153H and p.I1132V. The association was driven by the NY cohort mainly, but the two other cohorts demonstrated similar directionality of effect and effect size, and also contributed to the association. It is unlikely that differences in ethnicity in the NY cohort drove the association, since analysis was with adjustment for ethnicity, and analysis without adjustment yielded nearly identical results, ruling out effect of ethnicity. The other two cohorts were of homogeneous ethnicities, therefore in these cohorts too, ethnicity could not have affected the results. Interestingly, this haplotype is not in LD with the top GWAS SNP, rs2414739, suggesting that this may be a secondary association in the *VPS13C* locus, which was not identified in previous studies. However, this should be considered as preliminary results and needs to be examined in additional cohorts in order to conclude whether this haplotype is indeed associated with reduced risk for PD. Since the disease-causing mutations reported in *VPS13C* are loss-of-function mutations, a protective effect could occur for example due to gain of function or overexpression. In GTEx (https://www.gtexportal.org/home/) these variants were not associated with increased expression or affect splicing. The two nonsynonymous variants of this haplotype, p.R153H and p.I1132V, have high CADD scores (22.8 and 13.7, respectively), suggesting that they may affect the protein structure or function. Whether there is such effect and whether it is associated with gain-of-function will need to be examined in follow-up studies. Furthermore, the full sequencing analysis did not identify any coding variant that is in LD with the original GWAS hit that can explain the

GWAS association in this locus. This may suggest that the variant that has the main effect on the risk for PD in the *VPS13C* is outside of the coding and untranslated regions of the gene, likely being a regulatory element.

Previous studies on the top GWAS hit in the *VPS13C* locus demonstrated conflicting results. While significant associations of the top GWAS hit in this locus (rs2414739) with PD were found in Iranian[5] and East Asian[3] populations, negative results were reported in Taiwanese[4] and Han Chinese[6,7] populations. Therefore, it is possible that the association of *VPS13C* with PD is population dependent. Of note, that the association in the current study was mainly driven by populations enriched in Ashkenazi Jews, while in the French/French Canadian cohort the differences between patients and controls were much smaller and not statistically significant (Table 3.2). This may suggest that this haplotype association is population specific, and additional studies in different populations are required to answer this question.

In our three cohorts, we did not find any very rare homozygous variants with MAF<0.001 and found only two carriers of two heterozygous variants, both of which were controls (Table e-3). We were unable to determine if the variants were in cis or trans, but since no carriers of two variants was found among patients, it is clear that *VPS13C* bi-allelic mutations do not contribute to PD in our cohorts. Previously reported cases of PD in carriers of compound heterozygous or homozygous mutations all shared a specific clinical presentation of PD: early onset, rapid progression and early cognitive dysfunction[8,9]. In one study, the patient's AAO was 39, and disease progression was moderately severe with psychiatric symptoms and impaired cognition[9]. In another study, the three patients showed severe phenotypes: AAO of 25, 33, and 46 years, severe and early cognitive dysfunction, and became bedridden at 31, 43, and 58 years, respectively. Considering the young age of one of the compound heterozygous *VPS13C* variant carriers in our cohort (30 years), it is still possible, although unlikely that this individual will develop PD in the future. The negative results of the SKAT-O analyses demonstrate that rare heterozygous variants in *VPS13C* do not have an important role in PD in our cohorts.

Our study has several limitations. The differences between PD patients and controls in sex and mainly age are significant in some of our cohorts. To address this limitation, we included age and sex as covariates in the regression models. Therefore, if the association of the protective haplotype was related to age and not to disease status, we would likely not observe an association in the adjusted model. Furthermore, the association with the haplotype had the same directionality

and similar effect size in the NY cohort in which the controls were older (and the association was statistically significant), and in the other two cohorts where the controls were younger, likely ruling out effect of age. There was no significant difference in the percentage of Ashkenazi Jews between patients and controls in the New York cohort, yet we still performed the regression model with and without ethnicity as a covariate, and in both analyses the results remained significant. Nevertheless, the fact that our populations are enriched with relatively homogeneous populations such as Ashkenazi Jews and French Canadians requires additional studies in other populations. Another potential limitation is that we could not analyze the effect of CNVs in *VPS13C* with our data. Loss of function and exonic deletions/duplications are rare in gnomAD, found in only 3 individuals at a heterozygous state, and therefore not likely to have a major contribution in PD. However, future studies to examine the potential role of *VPS13C* in PD are required. Furthermore, as no functional experiments were performed in the current study, the potential effects of variants that we report here should be examined in additional studies.

In conclusion, our results suggest that *VPS13C* variants have a limited role in late-onset PD. The potentially PD protective haplotype located within *VPS13C,* which requires additional replications, may suggest that *VPS13C* could be a future target for PD therapeutic development. If naturally occurring genetic variants may reduce PD risk, it is conceivable that drugs that can mimic their effects could be developed. Additional genetic and functional studies will be required to determine if *VPS13C* may be a viable target for PD drug development.

3.6 Figures and Tables

Figure 3.1: **Forest plots – meta-analyses of four *VPS13C* coding variants associated with reduced risk for Parkinson Disease.**

The forest plots depict the effects of the four VPS13C coding variants that create the haplotype associated with reduced PD risk in the three cohorts studied, and their meta-analyzed effect on risk for PD. The results in the random effect model are nearly identical. CI, confidence interval; FE, fixed effect

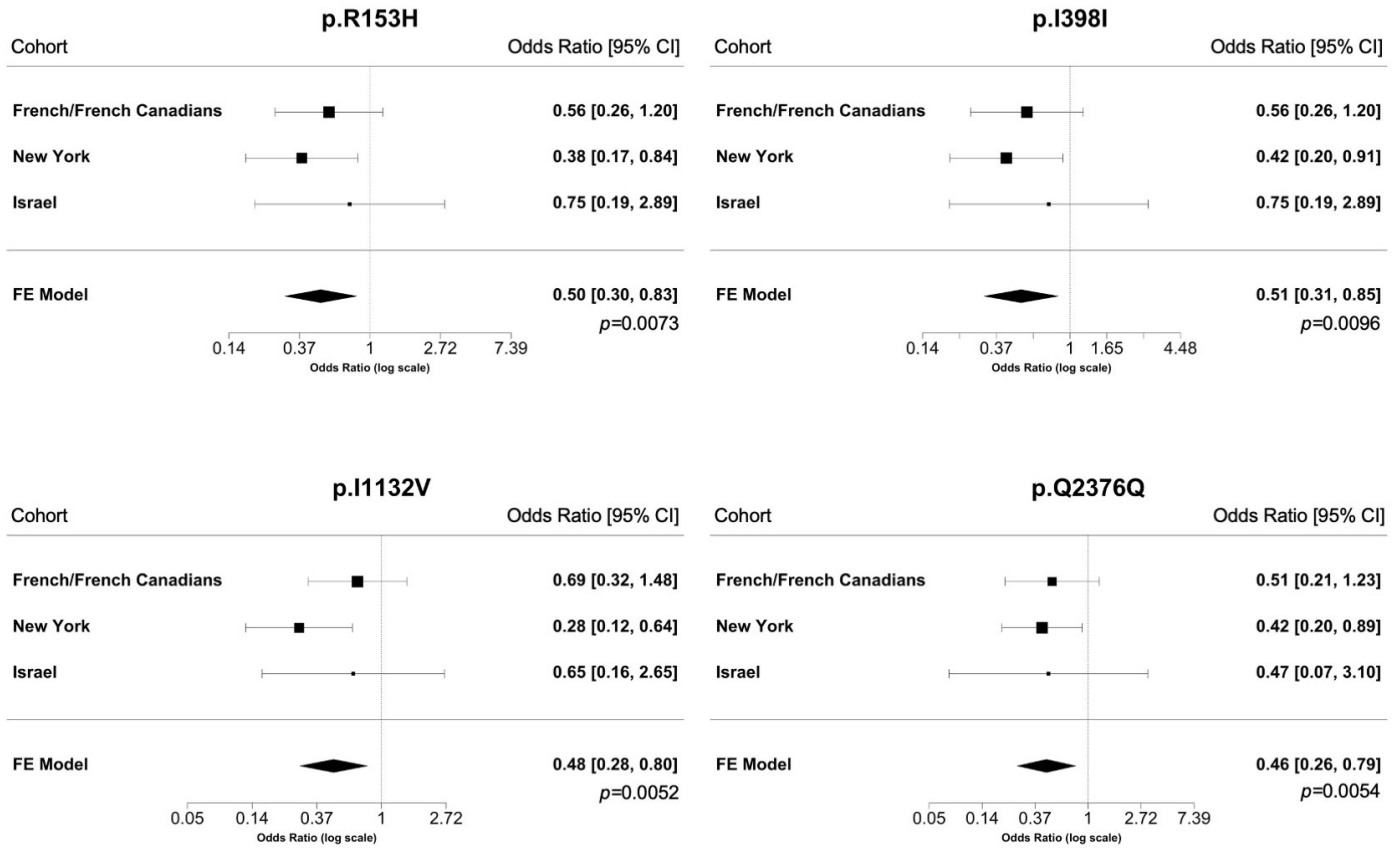


Table 3.1: **Study population details**

Sequenced		Cohort	Min Depth of Coverage	Analyzed						
				Cases			Controls			Genotyping call rate %
N (cases)	N (controls)			N	Mean Age (SD) y	%Male	N	Mean age (SD) y	%Male	
543	866	French/French-Canadian	15x	534	59.7(11.4)	63.7	858	42.0(13.6)	51.6	99.6
			50x	498	59.8(11.2)	63.5	785	42.3(13.5)	52.0	99.2
533	270	New York	15x	520	59.3(11.8)	64.0	262	65.0(9.8)	34.0	99.9
			50x	502	59.3(11.8)	64.1	245	65.4(9.6)	32.7	99.5
491	531	Israel	15x	482	60.6(11.7)	61.8	488	33.9(7.2)	57.8	99.3
			50x	478	60.7(11.7)	61.7	487	33.9(7.2)	57.7	99.3

N, number; Min, minimum; SD, standard deviation; y, year

Table 3.2: **Four variants forming the protective haplotype found in *VPS13C* and the results of the logistic regression in three cohorts**

Variant	rs number	nt-substitution	Cohort	N (cases)	N (controls)	Affected MAF	Unaffected MAF	GnomAD MAF†	OR (95%CI)	P
p.R153H	rs12595158	c.G458A	F/FC	15	27	0.0143	0.0157	0.0734	0.56 (0.26-1.20)	0.1358
			NY	14	16	0.0135	0.0324		0.38 (0.17-0.84)	0.0172
			Israel	16	22	0.0166	0.0225	0.0236	0.70 (0.19-2.6)	0.5940
p.I398I	rs9635356	c.A1194C	F/FC	15	28	0.0143	0.0163	0.0737	0.56 (0.26-1.20)	0.1330
			NY	16	16	0.0154	0.0324		0.42 (0.20-0.91)	0.0274
			Israel	15	19	0.0156	0.0195	0.0214	0.75 (0.19-2.89)	0.6752
p.I1132V	rs3784635	c.A3394G	F/FC	15	26	0.0143	0.0152	0.0726	0.69 (0.32-1.48)	0.3404
			NY	11	18	0.0106	0.0363		0.28 (0.12-0.64)	0.0025
			Israel	14	19	0.0145	0.0195	0.0233	0.65 (0.16-2.65)	0.5444
p.Q2376Q	rs17238189	c.A7128G	F/FC	10	24	0.0095	0.0140	0.0698	0.51 (0.21-1.23)	0.1344
			NY	16	17	0.0154	0.0344		0.42 (0.20-0.90)	0.0245
			Israel	16	20	0.0166	0.0215	0.0212	0.47 (0.07-3.10)	0.4350

nt: nucleotide; N, number; MAF, minor allele frequency; GnomAD, Genome Aggregation Database; OR, odds ratio; CI, confidence interval; P, p value; F/FC, French/French-Canadian; NY, New York.

†GnomAD_ASJ frequencies were used for the Israeli cohort; all samples are Ashkenazi Jewish.

Acknowledgments

We thank the patients and control subjects for their participation in this study. This work was financially supported by the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the Canadian Glycomics Network (GlycoNet). This research was also undertaken thanks in part to funding from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health [K02NS080915, and UL1 TR000040] and the Brookdale Foundation. GAR holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences. EAF is supported by a Foundation Grant from the Canadian Institutes of Health Research (FDN grant – 154301). ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award, and by the Young Investigator Award by Parkinson Canada. The access to part of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (http://rpq-qpn.ca/en/). We thank Armaghan Alam, Daniel Rochefort, Helene Catoire, Clotilde Degroot and Vessela Zaharieva for their assistance.

References

1. Nalls, M.A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-93 (2014).
2. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511-1516 (2017).
3. Zou, M. *et al.* Association analyses of variants of SIPA1L2, MIR4697, GCH1, VPS13C, and DDRGK1 with Parkinson's disease in East Asians. *Neurobiol Aging* (2018).
4. Chen, C.M. *et al.* Association of GCH1 and MIR4697, but not SIPA1L2 and VPS13C polymorphisms, with Parkinson's disease in Taiwan. *Neurobiol Aging* **39**, 221 e1-5 (2016).
5. Safaralizadeh, T. *et al.* SIPA1L2, MIR4697, GCH1 and VPS13C loci and risk of Parkinson's diseases in Iranian population: A case-control study. *J Neurol Sci* **369**, 1-4 (2016).
6. Wang, L. *et al.* Association of four new candidate genetic variants with Parkinson's disease in a Han Chinese population. *Am J Med Genet B Neuropsychiatr Genet* **171B**, 342-7 (2016).
7. Yang, X. *et al.* Polymorphism in MIR4697 but not VPS13C, GCH1, or SIPA1L2 is associated with risk of Parkinson's disease in a Han Chinese population. *Neurosci Lett* **650**, 8-11 (2017).
8. Lesage, S. *et al.* Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *Am J Hum Genet* **98**, 500-513 (2016).
9. Schormair, B. *et al.* Diagnostic exome sequencing in early-onset Parkinson's disease confirms VPS13C as a rare cause of autosomal-recessive Parkinson's disease. *Clin Genet* **93**, 603-612 (2018).
10. Darvish, H. *et al.* Identification of a large homozygous VPS13C deletion in a patient with early-onset Parkinsonism. *Mov Disord* **33**, 1968-1970 (2018).
11. Alcalay, R.N. *et al.* SCARB2 variants and glucocerebrosidase activity in Parkinson's disease. *NPJ Parkinsons Dis* **2**(2016).
12. Hughes, A.J., Daniel, S.E., Kilford, L. & Lees, A.J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* **55**, 181-4 (1992).

13. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619-22 (2012).
14. Ross, J.P. *et al.* Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. *Neurobiol Aging* **45**, 212 e13-212 e17 (2016).
15. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
16. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
17. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
18. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
19. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
20. Lee, S. *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* **91**, 224-37 (2012).
21. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* **36**(2010).
22. Machiela, M.J. & Chanock, S.J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-7 (2015).

Preface to Chapter 4

Chapters 2 and 3 described two single-gene analyses that focused on *GCH1*, a gene that plays a role in the synthesis of dopamine and causes DRD if certain mutations occur and *VPS13C* that was reported to cause atypical Parkinsonism with early onset and severe phenotype if contains rare bi-allelic mutations. Both genes were also identified within GWAS loci associated with PD. Hence, the next logical step was to study more genes reported in this way. The latest GWAS identified 90 genes from 78 loci, but our study was designed a few years earlier and based its targets on the data of previous PD GWASs[120,121]. We targeted 32 genes which were picked for the analysis in 3 cohorts adding up to 2,657 PD cases and 3,647 unrelated controls.

Working with this number of genes presented an additional challenge for processing of the data. In order to automate the quality control and the analysis steps, a custom bioinformatic pipeline was designed with the help of our team. Furthermore, receiving additional samples allowed me to get practical experience performing the targeted capture using MIPs and prepare that samples for the sequencing.

CHAPTER 4: Targeted sequencing of Parkinson's disease loci genes highlights SYT11, FGF20 and other associations

Uladzislau Rudakou, BSc[1,2], Eric Yu, BSc[1,2], Lynne Krohn, BSc[1,2], Jennifer A. Ruskey, MSc[2,3], Yves Dauvilliers, MD, PhD[4], Dan Spiegelman, MSc[2,3], Lior Greenbaum, MD, PhD[5,6,7], Stanley Fahn, MD[8], Cheryl H. Waters, MD[8], Nicolas Dupré, MD, FRCP[9,10], Guy A. Rouleau, MD, PhD, FRCPC, FRSC[1,2,3], Sharon Hassin-Baer, MD[7,11,12], Edward A. Fon, MD, FRCPC[2,3], Roy N. Alcalay, MD, MS[10,13], Ziv Gan-Or, MD, PhD[1,2,3].

Affiliations: 1. Department of Human Genetics, McGill University, Montréal, QC, H3A 1A1, Canada; 2. Montreal Neurological Institute, McGill University, Montréal, QC, Canada
3. Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada; 4. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France; 5. The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer, Ramat Gan , Israel; 6. The Joseph Sagol Neuroscience Center, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel; 7. Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 8. Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA; 9. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, G1V 0A6, Canada; 10. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, G1V 0A6, Canada; 11. Department of Neurology, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel; 12. Movement Disorders Institute, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel; 13. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA;

This paper was submitted for publication, and the preprint is available on medRxiv (https://www.medrxiv.org/content/10.1101/2020.05.29.20116111v1).

Keywords (5): Parkinson's disease, genetic fine mapping, risk loci, association study, sequencing

4.1 Abstract

Genome-wide association studies (GWAS) have identified numerous loci associated with Parkinson's disease. The specific genes and variants that drive the associations within the vast majority of these loci are unknown. We aimed to perform a comprehensive analysis of selected genes to determine the potential role of rare and common genetic variants within these loci. We fully sequenced 32 genes from 25 loci previously associated with Parkinson's disease in 2,657 patients and 3,647 controls from three cohorts. Capture was done using molecular inversion probes targeting the exons, exon-intron boundaries and untranslated regions (UTRs) of the genes of interest, followed by sequencing. Quality control was performed to include only high-quality variants. We examined the role of rare variants (minor allele frequency < 0.01) using optimized sequence Kernel association tests (SKAT-O). The association of common variants was estimated using regression models adjusted for age, sex and ethnicity as required in each cohort, followed by a meta-analysis. After Bonferroni correction, we identified a burden of rare variants in *SYT11, FGF20* and *GCH1* associated with Parkinson's disease. Nominal associations were identified in 21 additional genes. Previous reports suggested that the *SYT11* GWAS association is driven by variants in the nearby *GBA* gene. However, the association of *SYT11* was mainly driven by a rare 3' UTR variant (rs945006601) and was independent of *GBA* variants (p=5.23E-05 after exclusion of all *GBA* variant carriers). The association of *FGF20* was driven by a rare 5' UTR variant (rs1034608171) located in the promoter region. The previously reported association of *GCH1* with Parkinson's Disease is driven by rare nonsynonymous variants, some of which are known to cause dopamine-responsive dystonia. We also identified two *LRRK2* variants, p.Arg793Met and p.Gln1353Lys, in ten and eight controls, respectively, but not in patients. We identified common variants associated with Parkinson's disease in *MAPT, TMEM175, BST1, SNCA* and *GPNMB* which are all in strong linkage disequilibrium (LD) with known GWAS hits in their respective loci. A common coding *PM20D1* variant, p.Ile149Val, was nominally associated with reduced risk of Parkinson's disease (OR 0.73, 95% CI 0.60-0.89, p=1.161E-03). This variant is not in LD with the top GWAS hits within this locus and may represent a novel association. These results further demonstrate the importance of fine mapping of GWAS loci, and suggest that *SYT11, FGF20*, and potentially *PM20D1*, *BST1* and *GPNMB* should be considered for future studies as possible Parkinson's disease-related genes.

4.2 Introduction

Genome-wide association studies (GWASs) are an important tool for identifying genetic associations with complex disorders such as Parkinson's Disease[1-3]. GWASs examine the association of multiple common variants across the genome with specific traits. Typically, each associated locus includes multiple genes, and very often GWAS alone cannot identify the specific gene or variant within each locus that drives the association. Furthermore, GWASs generally overlook rare variants, while accumulation of rare variants on a specific common allele may drive some of these associations. Even if not located on a specific common allele, rare variants in GWAS genes may contribute to disease. For example, rare variants in *SNCA*[4], *GBA*[5-8], *LRRK2*[9,10], *VPS13C*[11-15], *GCH1*[16-21], all GWAS loci genes, also cause or increase the risk of Parkinson's disease.

Recently, a large-scale GWAS has identified 78 loci associated with risk of Parkinson's disease[3], yet only for a few of them we have identified the specific genes and variants that drive the association. In some of these loci, there are multiple hits, suggesting that more than one variant, or perhaps more than one gene within the locus, can be associated with Parkinson's disease. Fine-mapping of GWAS loci may help identifying these genes and variants, as was done for *SNCA*[22,23], *VPS13C*[11-15] and *TMEM175*[24]. In the *SNCA* region, fine mapping has identified independent 5' and 3' variants that drive the associations with different synucleinopathies[23]. Similarly, a potentially protective haplotype in *VPS13C*[15] and common nonsynonymous variants in *TMEM175* have been associated with Parkinson's Disease through fine mapping, and in the case of *TMEM175* also by functional studies[24,25].

In the current study, we fully sequenced 32 genes from 25 different Parkinson's disease GWAS loci in a total of 2,657 cases and 3,647 controls. We examined the association of common and rare variants with Parkinson's Disease, including the role of rare bi-allelic variants.

4.3 Materials and Methods

4.3.1 Study Population

The study population included, consecutively recruited, unrelated patients with Parkinson's Disease (n=2,657) and controls (n=3,647) from three cohorts, collected at McGill University (Quebec, Canada and Montpellier, France), Columbia University (New York, NY, USA), and the Sheba Medical Center (Israel). The McGill cohort was recruited in Quebec, Canada (in part with the assistance of the Quebec Parkinson's Network[26]) and in Montpellier, France. It includes 1,026 Parkinson's Disease patients (mean age 61.5±11.3, 62.6% male) and 2,588 controls (mean age 54.5±14.2, 47.4% male), all unrelated of French-Canadian/French ancestry. The Columbia cohort includes 1,026 Parkinson's Disease patients (mean age 59.4±11.7, 64.2% male) and 525 controls (mean age 64.2±10.1, 37.1% male). This cohort has been previously described in detail, and is mainly composed of participants with European descent, but 21.2% of Parkinson's Disease patients and 18.7% controls are of Ashkenazi Jewish (AJ) origin[27]. The Sheba cohort included 605 Parkinson's Disease patients (mean age 60.7±11.9, 62.3% male) and 534 controls (mean age 34.0±7.0, 55.8% male), all of full AJ origin; previously described in more details[28] . Movement disorder specialists diagnosed Parkinson's Disease with UK brain bank criteria[29] or the MDS clinical diagnostic criteria [30].

As detailed in the Statistical Analyses subsection, due to the differences in age and sex, statistical analyses were adjusted and included age and sex as co-variates. To account for ethnical heterogeneity of the Columbia cohort, an ethnicity covariate was also introduced for the analysis of this cohort (GWAS data was not available, therefore the reported ethnicity was used and not principal components). All three cohorts were sequenced in the same lab (McGill University), using the same protocol.

4.3.2 Standard Protocol Approvals, Registrations, and Patient Consents

The institutional review boards approved the study protocols and informed consent was obtained from all individual participants before entering the study.

4.3.3 Target genes

A recent Parkinson's Disease GWAS identified 78 loci, however the current study was designed and performed before its publication and the target genes were taken from earlier literature[1,2]. Our

target genes included: *ACMSD, BST1, CCDC62, DDRGK1, FGF20, GAK, GCH1, GPNMB, HIP1R, INPP5F, ITGA8, LAMP3, LRRK2, MAPT, MCCC1, PM20D1, RAB25, RAB29, RIT2, SCARB2, SETD1A, SIPA1L2, SLC41A1, SNCA, SREBF1, STK39, STX1B, SYT11, TMEM163, TMEM175, USP25* and *VPS13C.* These genes included several known Parkinson's disease related genes (*LRRK2*, *GCH1*, *SNCA*) found in GWAS loci in order to perform full sequencing assessment. Other genes were selected for sequencing based on assessment that included effects of quantitative trait loci, brain expression, potential association with known Parkinson's disease-causing genes and involvement in pathways potentially involved in Parkinson's disease, such the autophagy-lysosomal pathway, mitochondria quality control and endolysosomal recycling.

4.3.4 Sequencing

The coding regions, exon-intron boundaries and the 5' and 3' untranslated regions (UTRs) of the genes listed above were targeted using molecular inversion probes (MIPs), as previously described[31]. All MIPs used to sequence the genes of the study are detailed in Supplementary Table 1. Targeted DNA capture and amplification was done as previously described[32], and the full protocol is available upon request. The library was sequenced using Illumina HiSeq 2500/4000 platform at the McGill University and Genome Quebec Innovation Centre.

4.3.5 Data processing and quality control

Reads were mapped to the human reference genome (hg19) with Burrows-Wheeler Aligner[33]. Post-alignment quality control and variant calling was performed using Genome Analysis Toolkit (GATK, v3.8)[34]. Variant annotation was done with ANNOVAR[35]. Data on the frequency of each variant in different populations were extracted from the public database Genome Aggregation Database (GnomAD)[36].

The quality control pipeline that was used in this study is available on GitHub (https://github.com/gan-orlab/MIPVar.git). All calls with less than 25% reads with the variant have been removed. The remaining quality control (QC) was performed using GATK and PLINK software v1.9[37]. Genotyping quality (GQ) < 30 was used as a threshold for variant exclusion. Different depths of coverage thresholds were used for common and rare variants. Common variants threshold for inclusion was ≥ 15x, while rare variants (minor allele frequency [MAF] < 0.01) thresholds were more stringent and repeated twice, at ≥ 30x and ≥ 50x. Variants with genotype rate lower than 90% were excluded. Variants that deviated from Hardy-Weinberg

equilibrium set at $p \leq 0.001$ were filtered out. Threshold for missingness difference between cases and controls was set at $p = 0.05$ and the filtration script adjusted it with Bonferroni correction. Samples with average genotyping rate of less than 90% were excluded.

4.3.6 Statistical analysis

The burden of rare variants in each of the target genes was tested using the optimized sequence Kernel association test (SKAT-O, R package)[38]. In addition, we examined the burden of specific variant subgroups including: i) Funct.: Potentially functional included all nonsynonymous mutations, stop gain/loss, frameshift mutations, splicing variants located within two base pairs of exon-intron junctions, and variants that were flagged by ENCODE (explained below). ii) NS: nonsynonymous variants. iii) LOF: loss-of-function variants included stop gain/loss, frameshift, and splicing variants located within two base pairs of exon-intron junctions. iv) CADD: variants with the Combined Annotation Dependent Depletion (CADD) score ≥ 12.37; this threshold indicates the top 2% variants predicted to be the most damaging in the genome[39]. v) ENCODE: variants that are predicted to have regulatory activity, such as promoter, enhancer, activator and repressor, based on the chromatin profiling studies[40]. To increase power, rare variants were assessed in the three merged cohorts (since rare variants are less affected by ethnicity). Analysis of the individual cohorts was also performed post-hoc. Similarly to exome-wide burden analyses of rare variants[41], Bonferroni correction for statistical significance was performed based on the number of genes that were tested, and the thresholds are $p < 1.85E-03$ for genes with variant coverage of ≥ 50x and $p < 1.56E-03$ at ≥ 30x

The association between common variants and Parkinson's Disease was examined by logistic regression models using PLINK v1.9, with the status (patient or control) as a dependent variable, age and sex as covariates in all cohorts. Ethnicity as an additional covariate in the Columbia University cohort due to the ethnical heterogeneity. Meta-analysis of our three cohorts was conducted using METAL (R package)[42] with a fixed effect model. Cochran's Q-test was applied to test for residual heterogeneity. LD of common variants and their respective GWAS top marker was examined using the LDlink online tool LDmatrix (https://ldlink.nci.nih.gov/?tab=ldmatrix)[43]. Effects of common and rare variants on expression levels (eQTL) or splicing (sQTL) was evaluated using GTEx Portal online tool (https://www.gtexportal.org/home/).

4.3.7 Data availability

Anonymized data is available upon request by any qualified investigator.

4.4 Results

4.4.1 Coverage and variants identified

The average coverage for all 32 genes included in the current analysis was 642X, with a range of 86-1,064 (median 688) across all genes. An average of 97% of the targeted regions was covered at > 15X, 96% at > 30X, and 92% at >50X. Supplementary Table 4.2 details the average coverage and the percent coverage for each gene. At coverage of > 30X, we found 3,327 rare variants and at > 50X, 1,713 rare were identified (Supplementary Tables 3 and 4). A total of 311 common variants were identified across all genes in all three cohorts (Supplementary Table 5) with a coverage of >15X.

4.4.2 Rare variants in *SYT11, FGF20* and *GCH1* are associated with Parkinson's disease

Table 4.1 details the SKAT-O results of the analysis in the three combined cohorts, and Supplementary Table 6 details the results per cohort. After Bonferroni correction, *SYT11, FGF20* and *GCH1* were associated with Parkinson's Disease. Nominal associations were identified in 21 genes (Table 1, Supplementary Table 6). When considering all variants with MAF < 0.01 in *SYT11*, there was a strong association with Parkinson's Disease (p = 5.3E-06, Table 4.1), mainly driven by a *SYT11* 3' UTR variant (rs945006601) with MAF of 0.004 in Parkinson's Disease patients and 0.00015 in controls (OR = 27.0, 95%CI 3.6-202.6, p = 0.0013). As was previously reported [44], LD analysis showed that this variant is in some LD with the known Parkinson's disease-associated variants in *GBA* p.Glu326Lys (E326K, r^2 = 1.5E-05, D' = 1) and p.Asn370Ser (N370S, r^2 = 4.23E-05, D' = 1). The low r^2 for both variants are due to the different allele frequencies, yet the D' of 1 indicates an LD. To test if the identified association in *SYT11* is driven by these two pathogenic *GBA* variants, we removed the carriers of these variants and reran the *SYT11* SKAT-O analysis (p=5.58E-05). Since we have full sequencing data on *GBA* for all participants, we further excluded carriers of any other *GBA* variant that was ever reported to be potentially pathogenic and the association remained (p=5.23E-05).

When analyzing *FGF20,* a burden of all rare variants was demonstrated (p=0.0002), mainly driven by a rare 5' UTR variant, rs1034608171, which is located in the promoter region of the gene, with MAF of 0.002 in Parkinson's Disease patients and 0.00015 in controls (OR = 13.4, 95% CI 1.7-106.2, p = 0.014). The association of *GCH1* with Parkinson's Disease is driven by rare nonsynonymous variants (Supplementary Tables 3 and 4), some of which are known to cause dopamine responsive dystonia as we have previously reported[21].

After removing the *LRRK2* p.Gly2019Ser variant from the analysis, there was no statistically significant burden of other rare *LRRK2* variants in our cohorts. However, we did identify two variants, p.Arg793Met (rs35173587) and p.Gln1353Lys (rs200526782), in ten and eight controls, respectively, but not in Parkinson's Disease. Neither of these two variants had a statistically significant association after Bonferroni correction.

4.4.3 Association of common variants with potential functional effects in known Parkinson's disease loci

To examine the association of common variants in the 32 genes, we performed logistic regression adjusted for age and sex in all three cohorts, and also adjusted for ethnicity in the Columbia University cohort. Additionally, our three cohorts were meta-analyzed. After Bonferroni correction we found 24 variants from five genes associated with risk of Parkinson's Disease (Table 4.2). Forest plots in Figure 4.1 show the meta-analysis results for the tagging variants of these five genes. We further examined the LD of these variants with the known Parkinson's Disease-associated variants from the most recent Parkinson's Disease GWAS, and their potential functional effects (detailed in Methods) in ENCODE, GTEx and UCSC genome browser. In the *MAPT* locus, 17 common variants at full LD, all represent the *MAPT* H2 haplotype, were associated with reduced risk of Parkinson's Disease (Table 4.2). The variants identified in *TMEM175, BST1, SNCA* and *GPNMB* are all in strong LD with known GWAS hits in their respective loci (Table 4.2). Two *SNCA* 3' UTR variants were in full LD between themselves, and in strong LD with rs356219, a known 3' GWAS hit in the *SNCA* locus. These two variants are located within a strong enhancer region and are associated with altered splicing of *SNCA* in the Cortex (Figure 4.2). Three *GPNMB* variants were identified, one of which (rs5850) is located at the 3' UTR of *GPNMB* and associated with reduced expression of *GPNMB* in numerous brain regions including the substantia nigra, putamen and cortex (Figure 4.3). A coding variant in *PM20D1*, p.Ile149Val (rs1891460), was nominally associated with reduced risk of Parkinson's Disease (OR = 0.73, 95% CI 0.60-0.89, p

= 1.161E-03, specific frequencies in each cohort are in Supplementary Table 5). This variant is located within the *PARK16* region which also includes *RAB29* (previously named *RAB7L1*), *SLC41A1* and other genes. However, this variant is not in LD with the top GWAS hits within this locus ($r^2 = 0.04$, D' = 0.46).

4.5 Discussion

In the current study, we examined the role of common and rare variants in 32 genes located within known Parkinson's Disease risk loci, by performing targeted next generation sequencing of the coding and 5' and 3' untranslated regions, followed by association tests. We found a statistically significant burden, after Bonferroni correction, of rare variants in *SYT11, FGF20* and *GCH1*, and nominal associations ($p<0.05$) with other genes and variants (Table 4.1, Supplementary Tables 3,4,6). In addition, we have identified common coding and putative regulatory variants associated with Parkinson's Disease in *MAPT, TMEM175, BST1, SNCA* and *GPNMB*, including a potential novel association with a nonsynonymous variant in *PM20D1,* p.Ile149Val , which is nominally significant and requires further confirmation.

Three genes, *SYT11*, *FGF20* and *GCH1* were identified with a statistically significant burden of rare variants comparing Parkinson's Disease to controls. *SYT11* is located close to the *GBA* locus, and it was reported that the GWAS signal from *SYT11-GBA* locus in Parkinson's Disease can be attributed to coding *GBA* variants[44]. The *SYT11* variant mainly responsible for the association in our data is in weak LD with two coding *GBA* variants, p.Glu326Lys (E326K) and p.Asn370Ser (N370S). After exclusion of carriers of these variants, and further exclusion of all other *GBA* variants, the association of *SYT11* remained strong. These results suggest that *SYT11* may independently be associated with risk of PD. The Protein encoded by *SYT11*, Synaptotagmin 11, belongs to the synaptotagmin family, known calcium sensors mediating calcium-dependent regulation of membrane trafficking in synaptic transmission[45]. Synaptotagmin 11 is involved in regulation of endocytosis and vesicle recycling process, which is essential for neurotransmission[45,46]. Overexpression of Synaptotagmin 11 in mice led to impairment of DA transmission[47]. Synaptotagmin 11 was also reported to be a physiological substrate for E3 ubiquitin ligase – parkin, playing a critical role in parkin-related neurotoxicity[47]. Involvement of Synaptotagmin 11 in other mechanisms potentially related to Parkinson's Disease such as autophagy-lysosomal pathway[48], phagocytosis of α-synuclein fibrils[49], and repair of injured

astrocytes via lysosome exocytosis regulation[50], suggests that *SYT11* needs to be considered as a target for further research in PD. Fibroblast growth factor 20 (*FGF20*) encodes a neurotrophic factor preferentially expressed in the substantia nigra pars compacta. It regulates central nervous system development and function[51] and plays a role in dopaminergic neurons differentiation and survival[52]. There is a number of conflicting reports about FGF20 overexpression increasing α - synuclein levels in dopaminergic neurons[53-56]. The rare variant (rs1034608171) that drives the association in our data is located in a regulatory region and its potential functional effects would require additional studies. *GCH1* encodes GTP-cyclohydrolase 1, an essential enzyme for the dopamine synthesis pathway in the nigrostriatal cells[57]. Rare *GCH1* variants may cause DOPA-responsive dystonia, and the same variants have also been associated with Parkinson's Disease[16,21]. Since these mutations are rare, they only reached statistical significance in the McGill cohort, as we have previously described[21]. The two *LRRK2* variants, p.Arg793Met and p.Gln1353Lys, that were found only in controls require additional studies to examine their potential protective roles in Parkinson's Disease. *LRRK2* is a well-validated Parkinson's Disease related gene[58].

All 17 *MAPT* variants that were found to be significantly associated with Parkinson's Disease are in perfect LD with each other, and represent the H2 haplotype, known to be associated with reduced risk of Parkinson's Disease[59]. The identified coding variants in *MAPT* are located within exons that are likely not expressed in the CNS[60], suggesting that the association in this locus may be due to regulatory variants. *TMEM175* is located in the fourth most significant GWAS locus in Parkinson's Disease[3], and encodes a transmembrane potassium channel responsible for K^+ conductance in endosomes and lysosomes[61]. TMEM175 regulates lysosomal function through maintenance of lysosomal membrane potential and luminal pH stability, as well as its role in autophagosome-lysosome fusion[62]. In a recent genetic and functional study, two coding *TMEM175* variants, including the variant identified here, were shown to be associated with risk of Parkinson's Disease, risk of REM-sleep behavior disorder, and glucocerebrosidase activity[63].

The *BST1* and *SNCA* variants associated with risk of Parkinson's Disease in the current study are in strong LD with the known GWAS markers in their respective loci. The *BST1* variant is intronic without known function, and the *SNCA* 3 'UTR variants are located within strong *SNCA* enhancers. Whether these *SNCA* variants drive the association with Parkinson's Disease and their potential functional effects require additional studies. *GPNMB* encodes the glycoprotein nonmetastatic melanoma protein B a type 1, which is a transmembrane glycoprotein shown to have

a role in melanoma[64]. Patients with melanoma have an increased risk of Parkinson's Disease[65], and Parkinson's Disease patients have an increased risk for melanoma[66]. GPNMB may have a role in the lysosome[67] and in regulation of microglial inflammation[68,69]. While the expression of GPNMB in brain is overall low, the two variants identified in the current study are associated with reduced GPNMB expression in various brain tissues (Figure 4.3). One of the variants, rs5850, is located within the 3' UTR of *GPNMB*, but there is no clear regulatory function for this variant.

Peptidase M20 Domain Containing 1 (*PM20D1*) encodes a protein that regulates the production of N-fatty-acyl amino acids which have a function of chemical uncouplers of mitochondrial respiration[70]. Mitochondrial quality control is likely important in Parkinson's Disease, and genes involved in mitophagy are involved in familial forms of Parkinson's Disease[71]. *PM20D1* is located within the *PARK16* locus which also contains *SLC45A3, NUCKS1, RAB29* (previously named *RAB7L1*) and *SLC41A1*[3,72,73].Several studies have highlighted *RAB29* as the potential gene associated with Parkinson's Disease in this locus[74],as it was shown to interact with *LRRK2*[75] and affect lysosomal function[76]. In the current study we identified a common nonsynonymous variant in *PM20D1*, p.Ile149Val (rs1891460), which is not in LD with the Parkinson's Disease GWAS marker in this locus. Since this variant is not reported in ClinVar, has a low CADD score (1.01), and the statistical association is of borderline significance when correcting by the number of genes, it is possible that this association was found by chance. Replications in other cohorts will be required to determine whether this variant is associated with Parkinson's Disease.

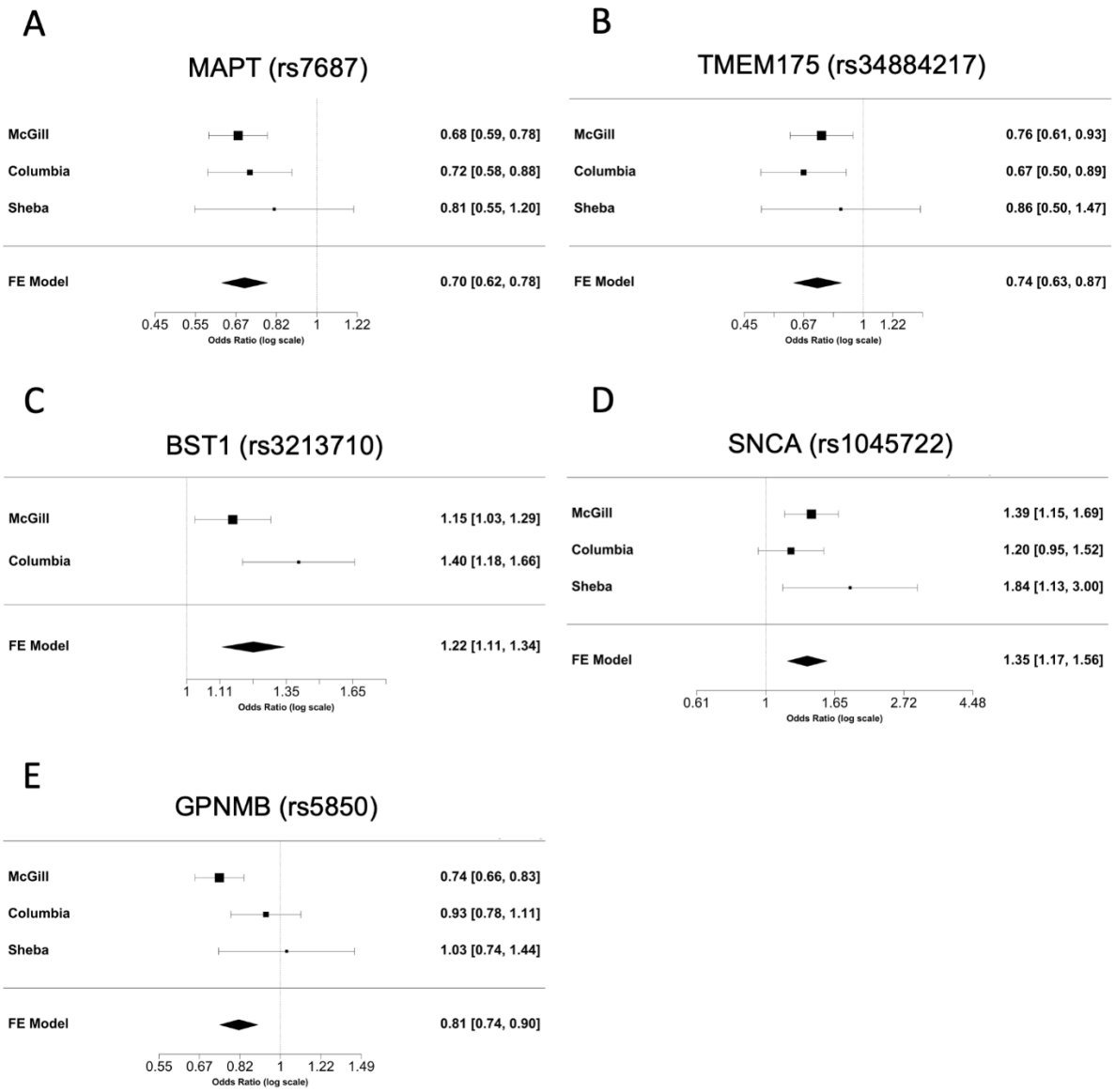
Our study has several limitations. The Parkinson's disease patients and controls are not matched by sex and age, hence we adjusted for these variables in the statistical analysis. We found no significant difference between allele frequencies of young (<40 years old) and old controls, which allowed us to use all of the controls in our analyses. In addition, the coverage of some of the genes we analyzed was suboptimal (Supplementary Table 2), and we did not target intronic regions and other potentially regulatory regions outside the 5' and 3' UTRs, which may also affect gene function. Whole genome sequencing studies will be required to study these regions and overcome the coverage limitations.

Our results, combined with previous studies, highlight the importance of performing fine mapping of GWAS loci to identify the variants and genes within each locus that drive the reported associations. Thus far, the causative genes within most Parkinson's Disease GWAS loci are

unknown, with the exception of *SNCA*, *GBA*, *LRRK2*, and likely *VPS13C*, *MAPT*, *TMEM175* and *GCH1*. The current study suggests that *SYT11, FGF20*, and potentially *PM20D1*, *BST1* and *GPNMB* should be considered as possible Parkinson's Disease-related genes and targets for further research.

4.6 Figures and Tables

Figure 4.1: **Forest plots, meta-analyses of common variants in the three cohorts.**



Each panel presents the individual odds ratio and 95% confidence interval for each cohort, and the results of the meta-analysis for the tagging variant of six loci identified in our study. Meta-analysis *p*-values: A. *MAPT* (*p*=8.28E-10); B. *TMEM175* (*p*=2.11E-04); C. *BST1* (*p*=3.14E-05); D. *SNCA* (*p*=4.08E-05); E. *GPNMB* (*p*=2.19E-05)

Figure 4.2: **Effect of 3'UTR *SNCA* variants on splicing.**

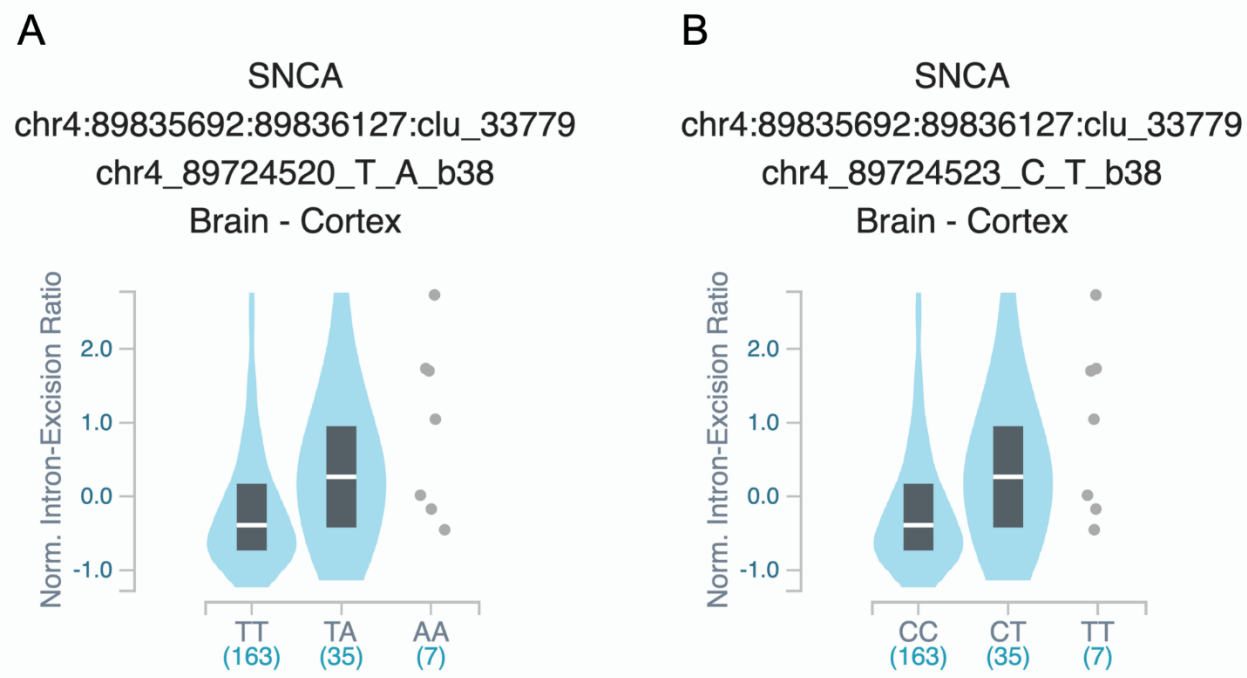


Figure is processed from GTeX (https://www.gtexportal.org), and represent the splicing quantitative trait locus (sQTL) effects of: **A.** rs1045722 in the 3' UTR of *SNCA* and **B.** rs3857053 in the 3' UTR of *SNCA*

Figure 4.3: **The 3'UTR *GPNMB* variant (rs5850) is associated with reduced expression in numerous brain regions.**

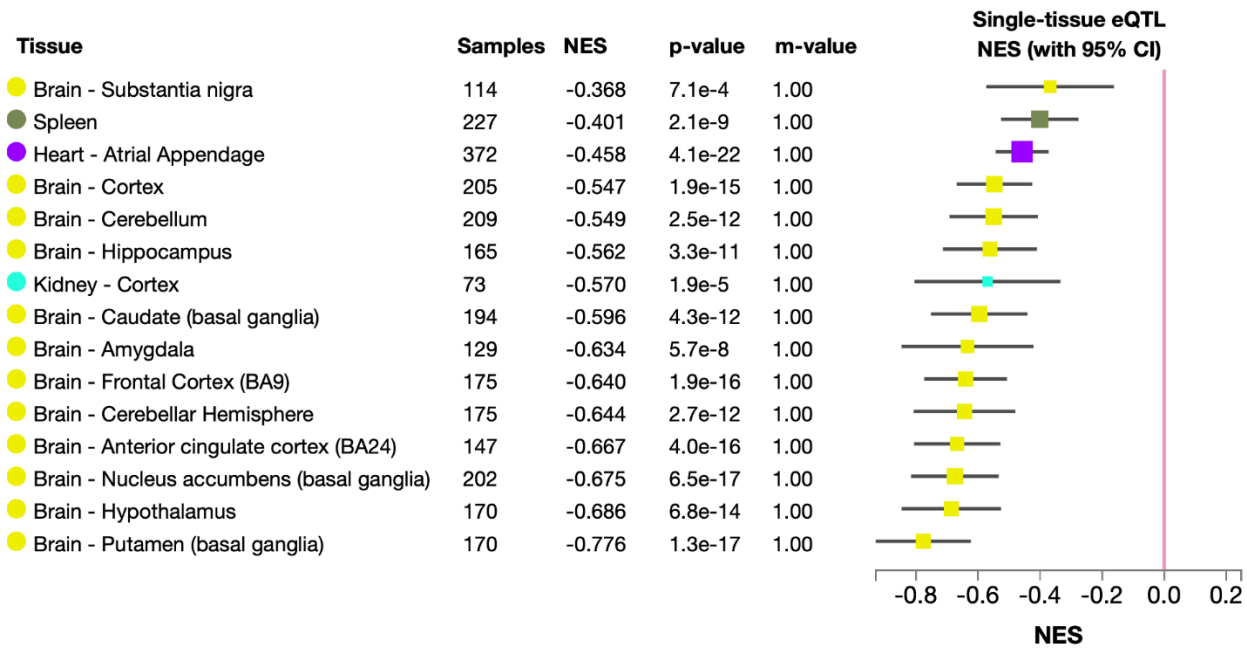


Figure is processed from GTeX (https://www.gtexportal.org), and shows the expression quantitative trait locus (eQTL) effect of the 3' UTR *GPNMB* variant rs5850 on GPNMB expression in different brain regions

Table 4.1: Results of SKAT-O analysis in the merged cohorts at 30x and 50x minimum depth of coverage

	50x						30x					
Gene	**CADD**	**ENCODE**	**Funct**	**LOF**	**Nonsyn**	**All**	**CADD**	**ENCODE**	**Funct**	**LOF**	**Nonsyn**	**All**
SYT11	0.45(6)	0.89(6)	0.66(14)	NA	0.40(8)	5.35E-06(34)**	0.31(13)	0.75(32)	0.75(48)	NA	0.22(16)	0.02(86)*
RAB25	0.20(1)	NA	0.20(1)	NA	0.20(1)	0.55(6)	0.34(6)	0.55(6)	0.52(14)	0.48(4)	0.40(4)	0.31(27)
RAB29	0.51(3)	0.04(11)*	0.03(14)*	NA	0.48(4)	0.09(23)	0.90(3)	0.12(19)	0.13(22)	NA	0.91(4)	0.25(41)
SLC41A1	0.13(5)	NA	0.13(5)	NA	0.13(5)	0.25(30)	0.93(18)	0.86(2)	0.97(22)	0.83(4)	0.65(16)	0.56(109)
PM20D1	0.38(9)	0.97(11)	0.22(23)	0.07(5)	0.11(12)	0.73(58)	0.27(15)	0.97(11)	0.13(32)	0.14(6)	0.09(20)	0.54(80)
SIPA1L2	0.24(33)	0.06(5)	0.18(46)	NA	0.25(42)	0.003(121)*	0.56(38)	0.06(5)	0.36(56)	0.06(3)	0.59(49)	0.01(142)*
TMEM163	0.09(4)	NA	0.09(4)	NA	0.09(4)	0.04(28)*	0.12(5)	NA	0.12(5)	NA	0.12(5)	0.03(44)*
ACMSD	0.26(14)	NA	0.40(18)	0.08(2)	0.30(16)	0.15(29)	0.65(19)	NA	0.68(23)	0.04(3)*	0.58(20)	0.32(45)
STK39	0.42(3)	NA	0.11(4)	NA	0.11(4)	0.11(48)	0.30(8)	NA	0.20(9)	NA	0.20(9)	0.31(70)
MCCC1	0.30(19)	0.83(4)	0.42(24)	0.11(4)	0.43(19)	0.20(50)	0.39(29)	0.82(33)	0.58(63)	0.03(6)*	0.53(30)	0.43(99)
LAMP3	0.76(9)	0.56(13)	0.65(21)	0.34(2)	0.84(12)	0.08(28)	0.43(18)	0.65(40)	0.78(55)	0.34(2)	0.63(24)	0.17(64)
GAK	NA	NA	NA	NA	NA	NA	0.79(14)	0.1(5)	0.32(19)	0.70(1)	0.88(16)	0.24(63)
TMEM175	NA	NA	NA	NA	NA	NA	0.66(4)	0.25(4)	0.34(8)	0.34(1)	0.56(3)	0.59(15)
BST1	0.02(12)*	0.33(9)	0.01(23)*	0.76(2)	0.004(14)*	0.06(46)	0.03(14)*	0.21(9)	0.007(25)*	0.76(2)	0.007(16)*	0.11(63)
SCARB2	0.83(8)	0.87(7)	0.05(17)*	0.27(2)	0.05(10)*	0.003(62)*	0.82(10)	0.76(9)	0.08(20)	0.23(3)	0.07(11)	0.01(86)*
SNCA	0.11(2)	0.99(9)	0.79(11)	NA	0.18(3)	0.55(34)	0.11(2)	0.24(29)	0.21(31)	NA	0.18(3)	0.34(57)
GPNMB	0.05(27)*	0.28(26)	0.08(53)	0.97(8)	0.03(27)*	0.07(80)	0.28(30)	0.20(52)	0.12(80)	1(9)	0.09(31)	0.14(118)
FGF20	NA	NA	NA	NA	NA	0.06(8)	0.76(2)	0.0008(4)**	0.0005(5)**	NA	0.76(2)	0.00017(15)**
ITGA8	0.70(26)	0.34(2)	0.85(30)	0.34(2)	0.93(26)	0.26(103)	0.65(39)	0.39(6)	0.27(52)	0.34(2)	0.27(44)	0.06(144)
INPP5F	0.13(5)	0.91(12)	0.59(16)	NA	0.25(6)	0.02(46)*	0.10(27)	0.23(23)	0.10(62)	0.61(5)	0.08(38)	0.003(152)*
LRRK2	0.07(35)	0.07(13)	0.11(52)	0.68(3)	0.08(37)	0.29(106)	0.03(53)*	0.80(17)	0.16(73)	0.83(4)	0.06(55)	0.31(142)
CCDC62	0.12(17)	NA	0.11(25)	0.46(3)	0.11(22)	0.09(59)	0.04(21)*	0.30(20)	0.23(50)	0.17(3)	0.04(28)*	0.17(91)
HIP1R	NA	NA	NA	NA	NA	NA	0.12(5)	0.14(14)	0.14(14)	0.70(1)	0.16(4)	0.08(23)
GCH1	0.02(4)*	NA	0.02(4)*	NA	0.02(4)*	0.64(13)	0.02(8)*	0.63(6)	0.43(14)	NA	0.02(8)*	0.005(76)*
VPS13C	0.38(34)	0.03(5)*	0.17(48)	0.17(3)	0.31(41)	0.32(159)	0.14(33)	0.03(5)*	0.08(42)	0.17(3)	0.20(35)	0.21(180)
SETD1A	0.01(3)*	NA	0.22(4)	NA	0.22(4)	0.11(13)	0.004(15)*	0.73(17)	0.04(32)*	NA	0.03(18)*	0.14(61)
STX1B	NA	NA	NA	NA	NA	NA	0.40(2)	NA	0.20(3)	NA	0.20(3)	0.07(30)
SREBF1	NA	NA	NA	NA	NA	NA	0.78(1)	NA	0.40(3)	NA	0.40(3)	0.46(4)
MAPT	0.60(3)	NA	0.60(3)	0.71(1)	0.65(2)	0.93(7)	0.27(13)	0.23(5)	0.45(24)	0.70(1)	0.07(19)	0.01(106)*

RIT2	0.71(1)	0.21(1)	0.76(2)	NA	0.71(1)	0.51(7)	0.48(4)	0.09(9)	0.11(14)	0.71(1)	0.61(5)	0.15(23)
DDRGK1	0.18(5)	NA	0.18(5)	NA	0.18(5)	0.09(13)	0.16(11)	NA	0.07(13)	0.17(3)	0.21(10)	0.44(53)
USP25	0.47(17)	NA	0.47(17)	0.76(2)	0.44(15)	0.37(63)	0.75(33)	0.22(15)	0.31(46)	0.12(6)	0.81(28)	0.45(128)

The table shows *p*-values of the SKAT-O analysis in the specified subgroup; the number of tested variants is indicated in the parentheses; *result with nominal significance ($p<0.05$); **result significant after Bonferroni correction for the number of genes CADD subgroup: CADD score ≥ 12.37; ENCODE subgroup: variants in regions with predicted regulatory function; Funct subgroup: potentially functional variants (Methods); LOF subgroup: loss-of-function variants; NA: not applicable - no variants tested; Nonsyn subgroup: non-synonymous; SKAT-O: Sequence Optimized Kernel Association Test

Table 4.2: **The common variants with statistically significant association with Parkinson's Disease in the meta-analysis of three cohorts**

Gene	Chr	Position	Ref Allele	Mutant Allele	SNP	D'	Rsq	CADD	Detailed annotation of the variant	Meta OR(95%CI)	Meta *p*-value
MAPT	17	44049329	C	T	rs75242405	0.99	0.98		NM_016835:intronic	0.72(0.64-0.81)	1.05E-07
MAPT	17	44051846	A	G	rs1800547	0.99	0.98		NM_016835:intronic	0.74(0.62-0.89)	1.24E-03
MAPT	17	44061023	G	A	rs62063786	0.99	0.98	7.7	NM_016835:exon6:c.G853A:p.D285N	0.71(0.63-0.79)	2.42E-09
MAPT	17	44061036	T	C	rs62063787	0.99	0.98	0.001	NM_016835:exon6:c.T866C:p.V289A	0.71(0.64-0.80)	2.66E-09
MAPT	17	44061278	C	T	rs17651549	0.99	0.98	34	NM_016835:exon6:c.C1108T:p.R370W	0.71(0.64-0.80)	3.35E-09
MAPT	17	44067400	T	C	rs10445337	0.99	0.98	9.93	NM_016835:exon8:c.T1339C:p.S447P	0.74(0.62-0.89)	1.04E-03
MAPT	17	44067508	A	G	rs79447161	0.99	0.98		NM_016835:intronic	0.71(0.64-0.80)	6.01E-09
MAPT	17	44071294	T	C	rs62063845	0.99	0.98		NM_016835:intronic	0.72(0.64-0.80)	6.82E-09
MAPT	17	44101563	T	C	rs9468	0.99	0.98		NM_016835:UTR3:c.*26T>C	0.72(0.64-0.81)	8.35E-09
MAPT	17	44102604	T	C	rs1052587	0.99	0.98		NM_016835:UTR3:c.*1067T>C	0.72(0.65-0.81)	9.98E-09
MAPT	17	44102638	A	G	rs1052590	0.99	0.98		NM_016835:UTR3:c.*1101A>G	0.72(0.64-0.81)	9.51E-09
MAPT	17	44102865	A	C	rs17574040	0.99	0.98		NM_016835:UTR3:c.*1328A>C	0.72(0.64-0.80)	4.79E-09
MAPT*	17	44103296	T	C	rs7687	0.99	0.98		NM_016835:UTR3:c.*1759T>C	0.70(0.62-0.78)	8.28E-10
MAPT	17	44103616	C	T	rs17652748	0.99	0.98		NM_016835:UTR3:c.*2079C>T	0.71(0.64-0.80)	3.81E-09
MAPT	17	44103825	T	C	rs75010486	0.99	0.98		NM_016835:UTR3:c.*2288T>C	0.71(0.64-0.80)	4.02E-09
MAPT	17	44104343	A	C	rs2158257	0.99	0.98		NM_016835:UTR3:c.*2806A>C	0.73(0.65-0.81)	2.36E-08
MAPT	17	44104410	TCTC	T	rs568475466	0.99	0.98		NM_016835:UTR3:c.*2874_*2876delCTC	0.72(0.64-0.80)	7.73E-09
TMEM175	4	944210	A	C	rs34884217	1	0.01	0.2	NM_032326:exon4:c.A194C:p.Q65P	0.74(0.63-0.87)	2.11E-04
BST1	4	15717321	G	A	rs3213710	0.94	0.79		NM_004334:intronic	1.22(1.11-1.34)	3.14E-05
SNCA*	4	90645671	T	A	rs1045722	1	0.12		NM_000345:UTR3:c.*2108A>T	1.35(1.17-1.56)	4.08E-05
SNCA	4	90645674	C	T	rs3857053	1	0.12		NM_000345:UTR3:c.*2105G>A	1.35(1.17-1.56)	4.56E-05
GPNMB	7	23300049	A	C	rs199351	0.95	0.89		NM_002510:intronic	0.82(0.75-0.90)	2.70E-05
GPNMB	7	23307634	ATGGG	A	rs2307783	0.95	0.89		NM_002510:intronic	0.83(0.76-0.91)	7.55E-05
GPNMB*	7	23314547	C	T	rs5850	0.95	0.89		NM_002510:UTR3:c.*704C>T	0.81(0.74-0.90)	2.19E-05

*tagging SNP for the haplotype

Chr: chromosome; CI: confidence interval; OR: odds ratio; Ref: reference; Rsq: r-square; SNP: single nucleotide polymorphism

Acknowledgements

We thank the participants for contributing to the study. This work was financially supported by grants from the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives initiative (HBHL), and Parkinson Canada. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health (K02NS080915, and UL1 TR000040) and the Brookdale Foundation. GAR holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences. EAF is supported by a Foundation Grant from the Canadian Institutes of Health Research (FDN grant – 154301). ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award, in collaboration with Parkinson Quebec, and by the Young Investigator Award by Parkinson Canada. The access to part of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (http://rpq-qpn.ca/en/). We thank Daniel Rochefort, Helene Catoire, Sandra B. Laurent, Clotilde Degroot and Vessela Zaharieva for their assistance.

Competing interests

Ziv Gan-Or has received consulting fees from Lysosomal Therapeutics Inc., Idorsia, Prevail Therapeutics, Denali, Ono Therapeutics, Deerfield and Inception Sciences (now Ventus). None of these companies were involved in any parts of preparing, drafting and publishing this review.

Cheryl H. Waters disclosed the following: Research support (Biogen, Roche, Sanofi); consulting fees (Acorda, Amneal, Impel, Kyowa, Mitsubishi, and Neurocrine); speaker's honoraria (Acadia, Acorda, Adamas, Amneal, and US WorldMeds)

Supplementary material

Supplementary Table 1 Detailed information on molecular inversion probes (MIPs)

Supplementary Table 2 Coverage statistics

Supplementary Table 3 All identified rare variants with 30x minimum depth of coverage

Supplementary Table 4 All identified rare variants with 50x minimum depth of coverage

Supplementary Table 5 All identified common variants

Supplementary Table 6 SKAT-O results for individual cohorts

References

1. Nalls, M.A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-93 (2014).
2. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511-1516 (2017).
3. Nalls, M.A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**, 1091-1102 (2019).
4. Polymeropoulos, M.H. *et al.* Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045-7 (1997).
5. Amaral, C.E.M. *et al.* GBA mutations p.N370S and p.L444P are associated with Parkinson's disease in patients from Northern Brazil. *Arq Neuropsiquiatr* **77**, 73-79 (2019).
6. Emelyanov, A.K. *et al.* Mutation analysis of Parkinson's disease genes in a Russian data set. *Neurobiol Aging* **71**, 267 e7-267 e10 (2018).
7. Ran, C. *et al.* Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. *Neurobiol Aging* **45**, 212 e5-212 e11 (2016).
8. Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* **361**, 1651-61 (2009).
9. Paisan-Ruiz, C. *et al.* Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**, 595-600 (2004).
10. Khan, N.L. *et al.* Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: clinical, pathological, olfactory and functional imaging and genetic data. *Brain* **128**, 2786-96 (2005).
11. Lesage, S. *et al.* Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *Am J Hum Genet* **98**, 500-513 (2016).
12. Jansen, I.E. *et al.* Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol* **18**, 22 (2017).
13. Darvish, H. *et al.* Identification of a large homozygous VPS13C deletion in a patient with early-onset Parkinsonism. *Mov Disord* **33**, 1968-1970 (2018).

14. Schormair, B. *et al.* Diagnostic exome sequencing in early-onset Parkinson's disease confirms VPS13C as a rare cause of autosomal-recessive Parkinson's disease. *Clin Genet* **93**, 603-612 (2018).
15. Rudakou, U. *et al.* Analysis of common and rare VPS13C variants in late-onset Parkinson disease. *Neurol Genet* **6**, 385 (2020).
16. Mencacci, N.E. *et al.* Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain* **137**, 2480-92 (2014).
17. Guella, I. *et al.* Parkinsonism in GTP cyclohydrolase 1 mutation carriers. *Brain* **138**, e349 (2015).
18. Lewthwaite, A.J. *et al.* Novel GCH1 variant in Dopa-responsive dystonia and Parkinson's disease. *Parkinsonism Relat Disord* **21**, 394-7 (2015).
19. Xu, Q. *et al.* Rare GCH1 heterozygous variants contributing to Parkinson's disease. *Brain* **140**, e41 (2017).
20. Yoshino, H. *et al.* GCH1 mutations in dopa-responsive dystonia and Parkinson's disease. *J Neurol* **265**, 1860-1870 (2018).
21. Rudakou, U. *et al.* Common and rare GCH1 variants are associated with Parkinson's disease. *Neurobiol Aging* **73**, 231 e1-231 e6 (2019).
22. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* **533**, 95-9 (2016).
23. Krohn, L. *et al.* Fine-mapping of SNCA in REM sleep behavior disorder and overt synucleinopathies. *BioRxiv*, 756528 (2019).
24. Krohn, L. *et al.* Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies. *Ann Neurol* **87**, 139-153 (2020).
25. Jinn, S. *et al.* Functionalization of the TMEM175 p.M393T variant as a risk factor for Parkinson disease. *Hum Mol Genet* **28**, 3244-3254 (2019).
26. Gan-Or, Z. *et al.* The Quebec Parkinson Network: A Researcher-Patient Matching Platform and Multimodal Biorepository. *J Parkinsons Dis* **10**, 301-313 (2020).
27. Alcalay, R.N. *et al.* SCARB2 variants and glucocerebrosidase activity in Parkinson's disease. *NPJ Parkinsons Dis* **2**(2016).
28. Ruskey, J.A. *et al.* Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. *Eur J Med Genet* **62**, 65-69 (2019).

29. Hughes, A.J., Daniel, S.E., Kilford, L. & Lees, A.J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* **55**, 181-4 (1992).
30. Postuma, R.B. *et al.* MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* **30**, 1591-601 (2015).
31. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619-22 (2012).
32. Ross, J.P. *et al.* Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. *Neurobiol Aging* **45**, 212 e13-212 e17 (2016).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
34. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
35. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
36. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
37. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
38. Lee, S. *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* **91**, 224-37 (2012).
39. Amendola, L.M. *et al.* Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res* **25**, 305-15 (2015).
40. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
41. Kenna, K.P. *et al.* NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet* **48**, 1037-42 (2016).
42. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

43. Machiela, M.J. & Chanock, S.J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-7 (2015).
44. Blauwendraat, C. *et al.* Coding variation in GBA explains the majority of the SYT11-GBA Parkinson's disease GWAS locus. *Mov Disord* **33**, 1821-1823 (2018).
45. Xie, Z. *et al.* Molecular Mechanisms for the Coupling of Endocytosis to Exocytosis in Neurons. *Front Mol Neurosci* **10**, 47 (2017).
46. Wang, C. *et al.* Synaptotagmin-11 inhibits clathrin-mediated and bulk endocytosis. *EMBO Rep* **17**, 47-63 (2016).
47. Wang, C. *et al.* Synaptotagmin-11 is a critical mediator of parkin-linked neurotoxicity and Parkinson's disease-like pathology. *Nat Commun* **9**, 81 (2018).
48. Bento, C.F., Ashkenazi, A., Jimenez-Sanchez, M. & Rubinsztein, D.C. The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway. *Nat Commun* **7**, 11803 (2016).
49. Du, C. *et al.* Synaptotagmin-11 inhibits cytokine secretion and phagocytosis in microglia. *Glia* **65**, 1656-1667 (2017).
50. Sreetama, S.C., Takano, T., Nedergaard, M., Simon, S.M. & Jaiswal, J.K. Injured astrocytes are repaired by Synaptotagmin XI-regulated lysosome exocytosis. *Cell Death Differ* **23**, 596-607 (2016).
51. IPDGC *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641-9 (2011).
52. Itoh, N. & Ohta, H. Roles of FGF20 in dopaminergic neurons and Parkinson's disease. *Front Mol Neurosci* **6**, 15 (2013).
53. Wang, G. *et al.* Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *Am J Hum Genet* **82**, 283-9 (2008).
54. Wider, C. *et al.* FGF20 and Parkinson's disease: no evidence of association or pathogenicity via alpha-synuclein expression. *Mov Disord* **24**, 455-9 (2009).

55. Sekiyama, K., Takamatsu, Y., Waragai, M. & Hashimoto, M. Role of genomics in translational research for Parkinson's disease. *Biochem Biophys Res Commun* **452**, 226-35 (2014).
56. Tarazi, F.I., Sahli, Z.T., Wolny, M. & Mousa, S.A. Emerging therapies for Parkinson's disease: from bench to bedside. *Pharmacol Ther* **144**, 123-33 (2014).
57. Kurian, M.A., Gissen, P., Smith, M., Heales, S., Jr. & Clayton, P.T. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. *Lancet Neurol* **10**, 721-33 (2011).
58. Bandres-Ciga, S., Diez-Fairen, M., Kim, J.J. & Singleton, A.B. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiol Dis* **137**, 104782 (2020).
59. Li, J. *et al.* Full sequencing and haplotype analysis of MAPT in Parkinson's disease and rapid eye movement sleep behavior disorder. *Mov Disord* **33**, 1016-1020 (2018).
60. Caillet-Boudin, M.L., Buee, L., Sergeant, N. & Lefebvre, B. Regulation of human MAPT gene expression. *Mol Neurodegener* **10**, 28 (2015).
61. Cang, C., Aranda, K., Seo, Y.J., Gasnier, B. & Ren, D. TMEM175 Is an Organelle K(+) Channel Regulating Lysosomal Function. *Cell* **162**, 1101-12 (2015).
62. Lee, C. *et al.* The lysosomal potassium channel TMEM175 adopts a novel tetrameric architecture. *Nature* **547**, 472-475 (2017).
63. Krohn, L. *et al.* Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies. *Ann Neurol* (2019).
64. Taya, M. & Hammes, S.R. Glycoprotein Non-Metastatic Melanoma Protein B (GPNMB) and Cancer: A Novel Potential Therapeutic Target. *Steroids* **133**, 102-107 (2018).
65. Liu, R., Gao, X., Lu, Y. & Chen, H. Meta-analysis of the relationship between Parkinson disease and melanoma. *Neurology* **76**, 2002-9 (2011).
66. Bertoni, J.M. *et al.* Increased melanoma risk in Parkinson disease: a prospective clinicopathological study. *Arch Neurol* **67**, 347-52 (2010).
67. Tomihari, M., Hwang, S.H., Chung, J.S., Cruz, P.D., Jr. & Ariizumi, K. Gpnmb is a melanosome-associated glycoprotein that contributes to melanocyte/keratinocyte adhesion in a RGD-dependent fashion. *Exp Dermatol* **18**, 586-95 (2009).

68. Dzamko, N., Geczy, C.L. & Halliday, G.M. Inflammation is genetically implicated in Parkinson's disease. *Neuroscience* **302**, 89-102 (2015).
69. Herrero, M.T., Estrada, C., Maatouk, L. & Vyas, S. Inflammation in Parkinson's disease: role of glucocorticoids. *Front Neuroanat* **9**, 32 (2015).
70. Long, J.Z. *et al.* The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. *Cell* **166**, 424-435 (2016).
71. Park, J.S., Davis, R.L. & Sue, C.M. Mitochondrial Dysfunction in Parkinson's Disease: New Mechanistic Insights and Therapeutic Perspectives. *Curr Neurol Neurosci Rep* **18**, 21 (2018).
72. Satake, W. *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet* **41**, 1303-7 (2009).
73. Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* **41**, 1308-12 (2009).
74. Gan-Or, Z. *et al.* Association of sequence alterations in the putative promoter of RAB7L1 with a reduced parkinson disease risk. *Arch Neurol* **69**, 105-10 (2012).
75. MacLeod, D.A. *et al.* RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron* **77**, 425-39 (2013).
76. Eguchi, T. *et al.* LRRK2 and its substrate Rab GTPases are sequentially targeted onto stressed lysosomes and maintain their homeostasis. *Proc Natl Acad Sci U S A* **115**, E9115-E9124 (2018).

CHAPTER 5: General Discussion

The current state of knowledge indicates that PD is not a single entity but rather an agglomeration of a large variety of subgroups, a few that follow Mendelian trait, while most having multifactorial and complex etiology. As described in Chapter 1, genetic studies helped to identify multiple molecular mechanisms potentially underlying the pathogenesis of PD. As GWASs grow in size they will keep discovering more loci associated with PD, however it is often unclear which of the genes in each genetic locus drives the association. The use of fine mapping has proven an effective means to identify the specific genes or even variants that impact the risk of PD[128,130-135].

The purpose of this MSc thesis was to study the role of common and rare variants in genes within PD GWAS loci by using targeted next-generation sequencing approach. Combining the three studies described in Chapters 2-4, *GCH1, VPS13C, SYT11, FGF20* and potentially *PM20D1, BST1* and *GPNMB* were identified as likely PD-related genes and are suggested for further investigation.

The impairment of dopamine release in the striatum is the principal cause of the hypokinetic features of PD[136]. Two genes that were highlighted in this thesis have an effect on dopamine levels by either affecting its synthesis or recycling. *GCH1* encodes GTP cyclohydrolase 1, an essential enzyme for a rate-limiting step in the pathway of dopamine synthesis by the nigrostriatal cells (Figure 5.1)[122]. Our study provides additional evidence to a previous study[123] suggesting that rare variants that are pathogenic in DRD are associated with PD. The effect of these variants on GTP Cyclohydrolase 1 activity is most likely impairment of dopamine synthesis. On the other hand, the common variant found to be associated with reduced risk of PD is also associated with increased expression in the brain regions important in PD. These findings suggest that *GCH1* is a potential target for drug development, as increasing *GCH1* levels or the enzyme activity may lead to higher levels of dopamine, which may lead to improvement or delay of dopamine deficiency-related symptoms. Synaptotagmin 11, encoded by *SYT11* gene, belongs to a protein family involved in regulation of membrane trafficking in synaptic transmission[137]. A series of functional experiments on mice showed that overexpression of *SYT11* has an impairing effect on dopamine release by inhibiting endocytosis and the vesicle recycling process in DA neurons[138]. Unlike the variants found in *GCH1,* the main driver of the association in *SYT11* is a non-coding 3'UTR variant, which makes it more challenging to design functional studies. This variant

probably has a regulatory function and leads to increased expression. Due to its relative rarity, this variant does not appear in GTEx, therefore this hypothesis needs to be tested in relevant models. Previously, the association of *SYT11* with PD was reported to be driven by coding *GBA* variants, by analyzing LD between common *GBA* and *SYT11* variants[139]. Therefore, another interesting aspect of *SYT11* association described in Chapter 4 is the apparent independence of the signal from *GBA*. We have excluded all *GBA* carriers from our analysis (found by full sequencing of *GBA*), and the association of *SYT11* remained strong. Overall, our findings associating *GCH1* and *SYT11* with PD highlight the importance of pathways that can influence DA levels.

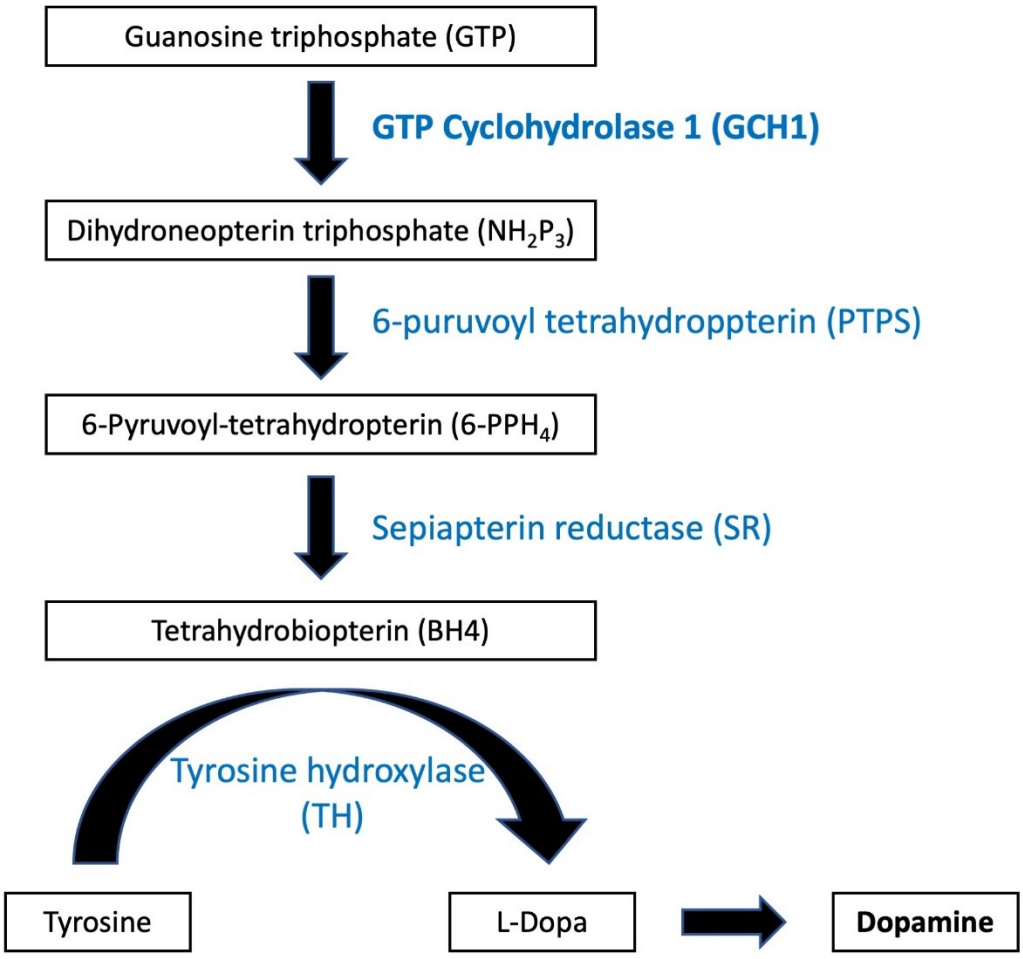


Figure 5.1: Representation of the metabolic pathway of dopamine synthesis in the brain

The expression of *SYT11* is especially high in different brain regions compared to other tissues (Figure 5.2, GTEx). These regions of high expression include areas in which dopaminergic

neurons are less abundant, suggesting that *SYT11* is probably important in other types of cells and in other mechanisms. In addition to the role in endocytosis and vesicle recycling affecting dopamine levels, *SYT11* has been reported having a role in various mechanisms that can be involved in the neurodegeneration of PD: phagocytosis and cytokine secretion in microglia[140-143], membrane repair in injured astrocytes[144], lysosomal function and autophagy[145], synaptic plasticity in neurons[146] and caveolae-mediated endocytosis[147]. Therefore, additional studies are needed to identify the specific mechanism by which *SYT11* affects the risk of developing PD.

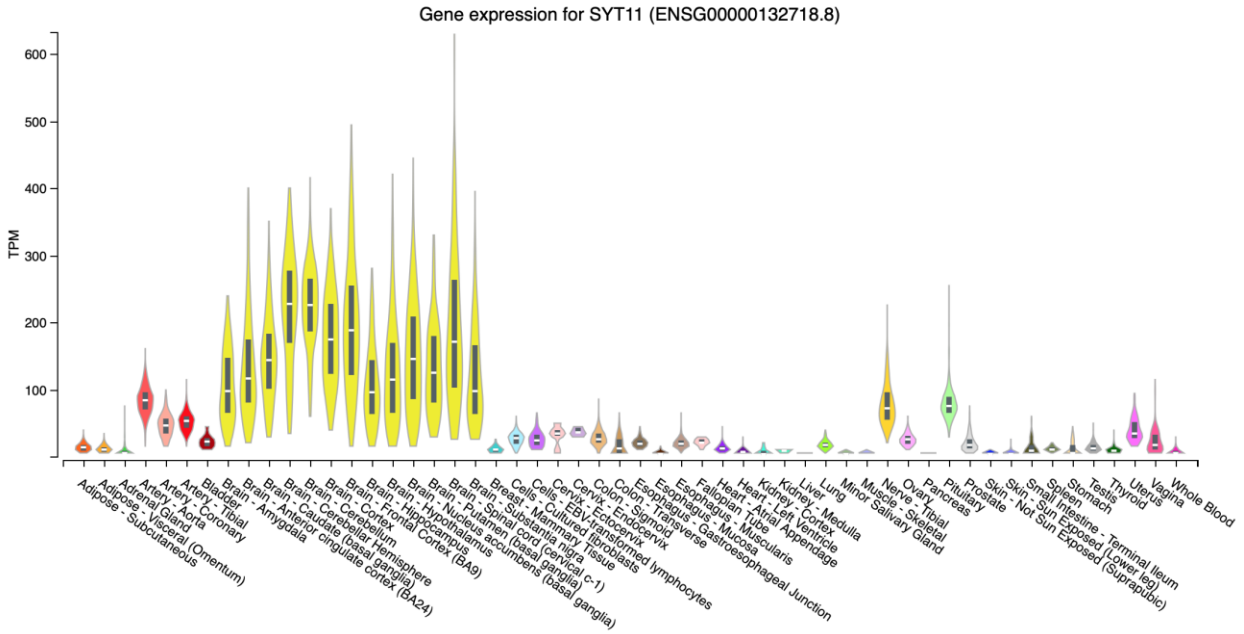


Figure 5.2: Expression levels for *SYT11* in different tissues measured in Transcripts Per Million (TPM); GTEx Portal

In chapters 3 and 4 the biological functions of the identified genes were mentioned only briefly. Considering that our current knowledge PD biology is incomplete, it is worth discussion the roles of these genes that can potentially be related to PD.

The study that identified mutations in *VPS13C* in autosomal recessive, early-onset PD, reported that *VPS13C* is necessary for mitochondrial maintenance[128]. Silencing of the gene in COS-7 and HEK293T cells resulted in perinuclear redistribution of the mitochondria, mitochondrial fragmentation, decrease in mitochondrial membrane potential and increase in maximal respiration rates and respiratory reserve. It was also demonstrated that silencing of *VPS13C* exaggerated PINK1/Parkin-mediated mitophagy triggered by mitochondrial

depolarization induced by protonophore CCCP. Another study of *VPS13C* localized it in the contacts between the endoplasmic reticulum and late endosomes/lysosome and suggested that it may play a role in tethering between these two organelles, which allows lipid transport and maintenance of lipid homeostasis. However, there was no evidence of *VPS13C* localization in the mitochondria[148]. These differences could be due to the type of models used in the different studies. It is also possible that the effect of *VPS13C* on mitochondria was indirect and caused by impairment of the endolysosomal pathway and mitophagy. Nevertheless, both studies suggest potential mechanisms by which *VPS13C* may be involved in PD. Since *VPS13C* is also implicated in sporadic PD through GWAS (unlike other genes associated with autosomal recessive Parkinsonism such as *PRKN*, *PINK1* and *PARK7/DJ-1*), it may serve as a potential target for drug development in sporadic PD, and not just in the rare familial form of the disease.

The neurotrophic factor encoded by *FGF20* has been reported to play a role in dopaminergic neurons differentiation and survival[149], as well as to provide protection from neurodegeneration in animal models[150]. One therapeutic approach involving *FGF20* in PD is to use is as a promoter of differentiation of the transplanted neural and embryonic stem cells into dopaminergic neurons, which showed promising results in animal models[149]. Additional experiments in rats suggested that the neuroprotection of the dopaminergic neurons by the endogenous FGF20 is delivered in a paracrine fashion from the astrocytes of the SNc where this neurotrophic factor is localised[151]. Another potential therapeutic approach involving *FGF20* is boosting the endogenous levels of FGF20 or delivering an exogenous infusion[151-153]. The expression of *FGF20* in the brain is not high but noticeably increased compared to other tissues (Figure 5.3). One hypothesis can be that *FGF20* is more important in brain development, and that reduced levels or activity of *FGF20* may lead to reduced number of dopaminergic neurons in the developing brain. This may lead later on to increased susceptibility to PD. This hypothesis needs to be tested, perhaps by analyzing imaging data comparing carriers and non-carries of *FGF20* variants.

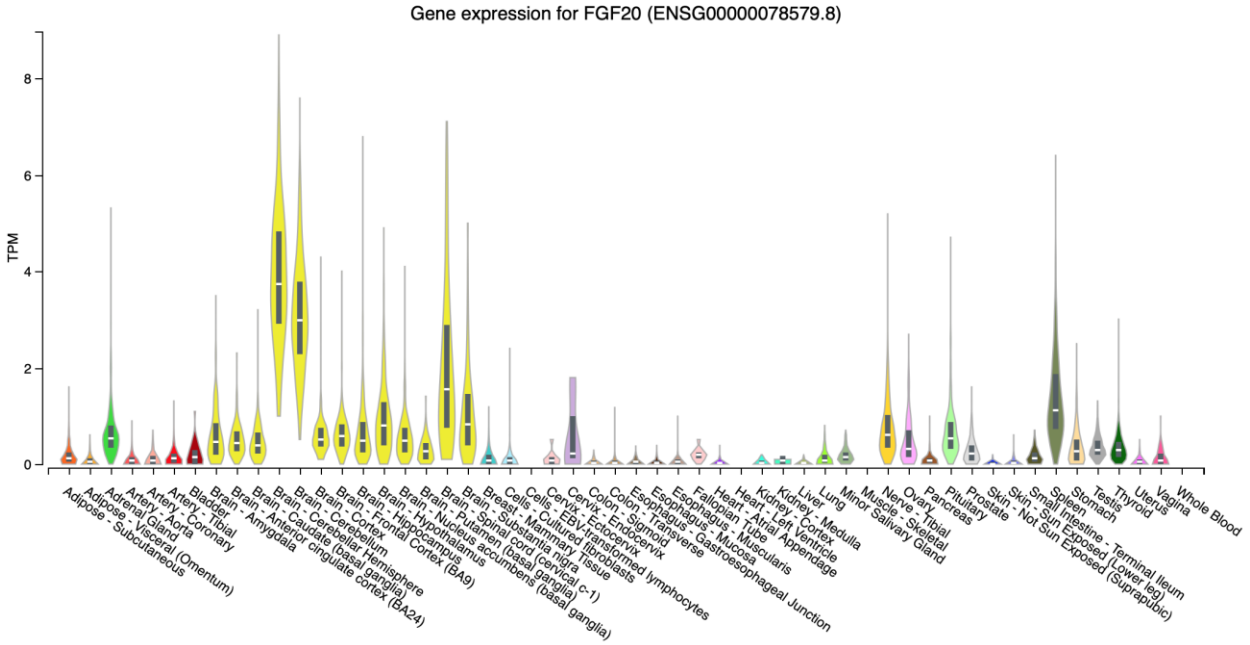


Figure 5.3: Expression levels for *FGF20* in different tissues measured in Transcripts Per Million (TPM); GTEx Portal

BST1 encodes Bone marrow stromal antigen 1 (or CD157). The only proposed mechanisms of its role in PD that were suggest are immune responses and regulation of calcium homeostasis via cyclic ADP-ribose[154]. It has a very low expression in the brain, but high expression in the blood, which may favor involvement in PD through immune response. There is no available data on whether *BST1* is expressed in microglia, therefore its potential in neuroinflammation is still unknown. Neuroinflammation was shown as one of the components of PD neuropathology. In the discussion section of Chapter 4, it was already mentioned that *GPNMB* seems to have a role in regulation of microglial activation and neuroinflammation. Additionally, it has been reported to reduce the inflammatory response in astrocytes[155]. Both GPNMB and its receptor, CD44, were shown to have increased levels in the SN of human PD patients (postmortem) and the striatum of MPTP-treated mice[155]. A study that experimented on mice with globally overexpressed *GPNMB* and treated with MPTP suggested that overexpression of *GPNMB* protects dopaminergic neurons in SN by reducing neuroinflammation and gliosis, yet the exact mechanism of protection remains to be elucidated[156]. However, MPTP treated animals are generally a poor model in terms of translatability in humans, therefore studies in other models are necessary to better understand the role of *GPNMB* in PD.

The work described in this thesis shows that targeted methods, allowing to study a panel of genes, are a viable and scalable approach, which should be considered for fine mapping of novel loci. This approach may help identifying the specific genes or variants responsible for the GWAS signal, which is crucial for the understanding of the disease mechanism, better identifying individuals at risk and identifying novel drug targets. In recent years, the use of polygenic risk scores (PRS) is increasing. In brief, PRS is a numerical value that can be assigned to each individual based on the different risk variants that they carry. It is a sum of the effects of risk of all variants in a specific person, and it has some predictive value in terms of identifying individuals at risk. For example, in PD, individuals at the highest quartile in terms of PRS had 3.7-6.3-fold increase in risk of PD compared to individuals in the lowest quartile[99]. In addition, PD PRS is also associated with clinical features of PD, such as age at onset[101], motor progression and cognitive decline[102]. Augmenting PRS with additional variants, common and rare, could improve the predictive accuracy and make PRS a better tool. It was recently suggested that stratifying by PRS could be used to improve clinical trials. Assuring that all arms of the trial have similar PRS in the pre-trial phase, may reduce the confounding of genetic variance between the groups, and lead to a truer calculation of the therapeutic effect[157].

Chapter 6: Conclusions and Future Directions

In this dissertation, I have pursued to 1) study the role of common and rare *GCH1* variants in PD, 2) examine whether *VPS13C* variants are associated with typical PD, 3) perform fine-mapping of 32 genes from PD GWAS and examine their association with PD.

1) Our study supported the role of rare *GCH1* variants in PD and suggested that a common variant, rs841, was associated with PD risk. Our data provided additional evidence that *GCH1* is a PD-associated gene and should be further investigated. We suggest that *GCH1* should be added to clinical genetic screening panels for PD.
2) Our data suggested that unlike in EOPD, rare monoallelic and biallelic *VPS13C* variants do not have a major role in typical PD. A haplotype including four common *VPS13C* variants was found to be associated with reduced risk of PD in the meta-analysis of the three cohorts, but more thorough investigation in additional cohorts is needed to determine if the apparent protective effect is real.
3) The analysis of the coding and regulatory regions of the 32 genes identified seven PD associated genes. *SYT11, FGF20* and *GCH1* had the burden of rare variants associated with PD. Common variants in *MAPT, TMEM175, BST1, SNCA* and *GPNMB* were found to be significantly associated with PD in the meta-analysis of the three cohorts. A common coding variant in *PM20D1* was associated at a near-significant level.

In summary, this body of work has contributed towards expanding the knowledge of the genetic landscape of PD by providing additional evidence that *GCH1, VPS13C, SYT11, FGF20, BST1 and GPNMB* genes and some specific variants within these genes should be considered as targets for further exploration. The associations we have reported with *MAPT*, *TMEM175* and *SNCA* are already well established[130,134,135,158].

As a complex condition, there remains much to be learnt about the pathogenesis of PD. Considering that the latest PD GWAS identified 38 novel independent associations in 37 novel loci, adding up to 90 associations in 78 loci with the previously known ones[99], one way to expand on our research is to perform targeted study on all genes found in loci associated with PD in GWAS. Another way could be to include CNV analysis, the lack of which was a limitation in our

studies. As follow-up, functional studies would need to assess how these genes and variants can affect the neuropathogenesis of PD. That could be done, for example, by CRISPR screening, systematically overexpressing or knocking out the target genes and measuring PD related markers such as phosphorylation of αSyn, internalization of αSyn into cells, GCase activity and lysosomal function. Of course, with WGS becoming more affordable, future studies can assess all the genes including the coding regions that are currently overlooked and might contain important regulatory functions.

Genetic findings can be applied to disease classification and subtyping. A good example is maturity-onset diabetes of the young (MODY), a genetically heterogeneous monogenic form of non-insulin-dependent type 2 diabetes mellitus (T2DM). Only 1-2% of all diabetes patients have MODY, and it is commonly misdiagnosed as either type 1 diabetes mellitus (T1DM) or T2DM[159]. Rare LOF mutations in *HNF1A* are known to cause a common form of MODY (MODY3)[160]. When it was reported by two independent GWAS that SNPs near *HNF1A* are associated with a marker of inflammation – C-reactive protein (CRP) levels[161,162] it lead to a hypothesis that serum levels of CRP could be used as a biomarker for *HNF1A* mutations, and hence for MODY. It was observed that MODY3 patients had decreased levels of CRP when compared to individuals without diabetes, with T1DM, with T2DM or even non-*HNF1A* subtypes of MODY, validating high sensitivity CRP as a clinical biomarker for MODY3[163,164]. In this example, a genetic discovery through GWAS provided physicians with a tool that will reduce the rate of misdiagnosis, leading to better care.

Genetic subtyping of cystic fibrosis (CF) is a great example of a personalized approach to treatment because CF patients that are carriers of certain mutations can benefit from a mutation-specific treatment. CF is an autosomal-recessive genetic condition caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene that impair the activity of the protein by altering its structure, localization and function[165]. CF can be classified by the *CFTR* mutations and their functional impact on the protein. Class I: p.G542X, p.W1282X, 621+1G>T – no functional protein is made. Class II: p.F508del, p.G85E – protein is misfolded and does not reach the cell membrane. Class III: p.G551D – protein reaches the cell membrane, but chloride channel gate does not open. Class IV: p.R117H, p.R334W, p.R347P – protein reaches the cell membrane, but the efficiency of the chloride channel is decreased. Class V: 3849+10kb C>T, 2789+5 G>A – protein function is normal, but the amount is insufficient. Class VI: 4326del,

4279insTC – protein stability is reduced and the turnover is accelerated[166]. All of these functional abnormalities provide one or more targets for the therapeutic agents that can enable production of functional protein (i.e. Ataluren), promote chloride channel opening (i.e. Ivacaftor) or increase the amount of cell surface protein (i.e. Lumacaftor, Tezacaftor); with patient's class of CF determining the choice of the drug or the combination of drugs needed to achieve therapeutic efficiency[167].

For a disease so heterogenous as PD, having a simple way of subtyping, as described in the examples above, would allow physicians to have a more accurate prognosis and to tailor the treatment to each patient. This personalized approach would not only improve care, but it would alleviate expenses for the healthcare, replacing time consuming trial and error by the efficient matching of the right treatment with the right patient.

CHAPTER 7: Bibliography

1. Bureau, U.S.C. National Population Projections. . Vol. 2020 (2017).
2. Kowal, S.L., Dall, T.M., Chakrabarti, R., Storm, M.V. & Jain, A. The current and projected economic burden of Parkinson's disease in the United States. *Mov Disord* **28**, 311-8 (2013).
3. Nussbaum, R.L. & Ellis, C.E. Alzheimer's disease and Parkinson's disease. *N Engl J Med* **348**, 1356-64 (2003).
4. Dorsey, E.R. *et al.* Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology* **68**, 384-6 (2007).
5. Taylor, K.S., Cook, J.A. & Counsell, C.E. Heterogeneity in male to female risk for Parkinson's disease. *J Neurol Neurosurg Psychiatry* **78**, 905-6 (2007).
6. Lees, A.J., Hardy, J. & Revesz, T. Parkinson's disease. *Lancet* **373**, 2055-66 (2009).
7. Hubble, J.P., Cao, T., Hassanein, R.E., Neuberger, J.S. & Koller, W.C. Risk factors for Parkinson's disease. *Neurology* **43**, 1693-7 (1993).
8. Yan, D., Zhang, Y., Liu, L., Shi, N. & Yan, H. Pesticide exposure and risk of Parkinson's disease: Dose-response meta-analysis of observational studies. *Regul Toxicol Pharmacol* **96**, 57-63 (2018).
9. Fall, P.A., Fredrikson, M., Axelson, O. & Granerus, A.K. Nutritional and occupational factors influencing the risk of Parkinson's disease: a case-control study in southeastern Sweden. *Mov Disord* **14**, 28-37 (1999).
10. Tanner, C.M., Goldman, S.M., Ross, G.W. & Grate, S.J. The disease intersection of susceptibility and exposure: chemical exposures and neurodegenerative disease risk. *Alzheimers Dement* **10**, S213-25 (2014).
11. Palacios, N. *et al.* Caffeine and risk of Parkinson's disease in a large cohort of men and women. *Mov Disord* **27**, 1276-82 (2012).
12. Ritz, B. *et al.* Pooled analysis of tobacco use and risk of Parkinson disease. *Arch Neurol* **64**, 990-7 (2007).
13. Thacker, E.L. *et al.* Temporal relationship between cigarette smoking and risk of Parkinson disease. *Neurology* **68**, 764-8 (2007).
14. Noyce, A.J. *et al.* The Parkinson's Disease Mendelian Randomization Research Portal. *Mov Disord* **34**, 1864-1872 (2019).

15. Yamada-Fowler, N. & Soderkvist, P. Coffee, Genetic Variants, and Parkinson's Disease: Gene-Environment Interactions. *J Caffeine Res* **5**, 3-10 (2015).
16. Rees, K. *et al.* Non-steroidal anti-inflammatory drugs as disease-modifying agents for Parkinson's disease: evidence from observational studies. *Cochrane Database Syst Rev*, CD008454 (2011).
17. Mullapudi, A., Gudala, K., Boya, C.S. & Bansal, D. Risk of Parkinson's Disease in the Users of Antihypertensive Agents: An Evidence from the Meta-Analysis of Observational Studies. *J Neurodegener Dis* **2016**, 5780809 (2016).
18. Bykov, K., Yoshida, K., Weisskopf, M.G. & Gagne, J.J. Confounding of the association between statins and Parkinson disease: systematic review and meta-analysis. *Pharmacoepidemiol Drug Saf* **26**, 294-300 (2017).
19. Cassani, E. *et al.* Dietary habits in Parkinson's disease: Adherence to Mediterranean diet. *Parkinsonism Relat Disord* **42**, 40-46 (2017).
20. Kalinderi, K., Bostantjopoulou, S. & Fidani, L. The genetic background of Parkinson's disease: current progress and future prospects. *Acta Neurol Scand* **134**, 314-326 (2016).
21. Obeso, J.A. *et al.* Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy. *Mov Disord* **32**, 1264-1310 (2017).
22. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* **24**, 197-211 (2003).
23. Cheng, H.C., Ulane, C.M. & Burke, R.E. Clinical progression in Parkinson disease and the neurobiology of axons. *Ann Neurol* **67**, 715-25 (2010).
24. Tolosa, E. & Pont-Sunyer, C. Progress in defining the premotor phase of Parkinson's disease. *J Neurol Sci* **310**, 4-8 (2011).
25. Postuma, R.B. *et al.* MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* **30**, 1591-601 (2015).
26. Kalia, L.V. & Lang, A.E. Parkinson's disease. *Lancet* **386**, 896-912 (2015).
27. Martinez-Martin, P. *et al.* Prevalence of nonmotor symptoms in Parkinson's disease in an international setting; study using nonmotor symptoms questionnaire in 545 patients. *Mov Disord* **22**, 1623-9 (2007).
28. Hanagasi, H.A., Tufekcioglu, Z. & Emre, M. Dementia in Parkinson's disease. *J Neurol Sci* **374**, 26-31 (2017).

29. Boeve, B.F. *et al.* Pathophysiology of REM sleep behaviour disorder and relevance to neurodegenerative disease. *Brain* **130**, 2770-88 (2007).
30. Schapira, A.H.V., Chaudhuri, K.R. & Jenner, P. Non-motor features of Parkinson disease. *Nat Rev Neurosci* **18**, 509 (2017).
31. Ffytche, D.H. *et al.* The psychosis spectrum in Parkinson disease. *Nat Rev Neurol* **13**, 81-95 (2017).
32. Antonini, A. *et al.* Pain in Parkinson's disease: facts and uncertainties. *Eur J Neurol* **25**, 917-e69 (2018).
33. Siciliano, M. *et al.* Fatigue in Parkinson's disease: A systematic review and meta-analysis. *Mov Disord* **33**, 1712-1723 (2018).
34. Stocchi, F. & Torti, M. Constipation in Parkinson's Disease. *Int Rev Neurobiol* **134**, 811-826 (2017).
35. Gan-Or, Z., Alcalay, R.N., Rouleau, G.A. & Postuma, R.B. Sleep disorders and Parkinson disease; lessons from genetics. *Sleep Med Rev* **41**, 101-112 (2018).
36. Smith, M. *et al.* Nocturia in Patients With Parkinson's Disease. *Mov Disord Clin Pract* **3**, 168-172 (2016).
37. Videnovic, A. & Comella, C.L. Sleep disorders in Parkinson's disease. *Handb Clin Neurol* **99**, 997-1010 (2011).
38. Rizzo, G. *et al.* Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. *Neurology* **86**, 566-76 (2016).
39. Fahn, S. The medical treatment of Parkinson disease from James Parkinson to George Cotzias. *Mov Disord* **30**, 4-18 (2015).
40. Picconi, B., Hernandez, L.F., Obeso, J.A. & Calabresi, P. Motor complications in Parkinson's disease: Striatal molecular and electrophysiological mechanisms of dyskinesias. *Mov Disord* **33**, 867-876 (2018).
41. Emre, M. *et al.* Rivastigmine for dementia associated with Parkinson's disease. *N Engl J Med* **351**, 2509-18 (2004).
42. Iacono, D. *et al.* Parkinson disease and incidental Lewy body disease: Just a question of time? *Neurology* **85**, 1670-9 (2015).
43. Doherty, K.M. *et al.* Parkin disease: a clinicopathologic entity? *JAMA Neurol* **70**, 571-9 (2013).

44. Kalia, L.V., Kalia, S.K. & Lang, A.E. Disease-modifying strategies for Parkinson's disease. *Mov Disord* **30**, 1442-50 (2015).
45. Dickson, D.W. *et al.* Neuropathological assessment of Parkinson's disease: refining the diagnostic criteria. *Lancet Neurol* **8**, 1150-7 (2009).
46. Stefanis, L. alpha-Synuclein in Parkinson's disease. *Cold Spring Harb Perspect Med* **2**, a009399 (2012).
47. Scott, D. & Roy, S. alpha-Synuclein inhibits intersynaptic vesicle mobility and maintains recycling-pool homeostasis. *J Neurosci* **32**, 10129-35 (2012).
48. Brundin, P., Li, J.Y., Holton, J.L., Lindvall, O. & Revesz, T. Research in motion: the enigma of Parkinson's disease pathology spread. *Nat Rev Neurosci* **9**, 741-5 (2008).
49. Halliday, G., Lees, A. & Stern, M. Milestones in Parkinson's disease--clinical and pathologic features. *Mov Disord* **26**, 1015-21 (2011).
50. Aarsland, D. *et al.* Risk of dementia in Parkinson's disease: a community-based, prospective study. *Neurology* **56**, 730-6 (2001).
51. Luk, K.C. *et al.* Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **338**, 949-53 (2012).
52. Recasens, A. *et al.* Lewy body extracts from Parkinson disease brains trigger alpha-synuclein pathology and neurodegeneration in mice and monkeys. *Ann Neurol* **75**, 351-62 (2014).
53. McCann, H., Cartwright, H. & Halliday, G.M. Neuropathology of alpha-synuclein propagation and braak hypothesis. *Mov Disord* **31**, 152-60 (2016).
54. Chen, B., Retzlaff, M., Roos, T. & Frydman, J. Cellular strategies of protein quality control. *Cold Spring Harb Perspect Biol* **3**, a004374 (2011).
55. Ciechanover, A. The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J* **17**, 7151-60 (1998).
56. Klionsky, D.J. The molecular machinery of autophagy and its role in physiology and disease. *Semin Cell Dev Biol* **21**, 663 (2010).
57. Moors, T. *et al.* Lysosomal Dysfunction and alpha-Synuclein Aggregation in Parkinson's Disease: Diagnostic Links. *Mov Disord* **31**, 791-801 (2016).
58. Xilouri, M., Brekk, O.R. & Stefanis, L. Autophagy and Alpha-Synuclein: Relevance to Parkinson's Disease and Related Synucleopathies. *Mov Disord* **31**, 178-92 (2016).

59. Giannoccaro, M.P., La Morgia, C., Rizzo, G. & Carelli, V. Mitochondrial DNA and primary mitochondrial dysfunction in Parkinson's disease. *Mov Disord* **32**, 346-363 (2017).
60. Pasqualetti, G., Brooks, D.J. & Edison, P. The role of neuroinflammation in dementias. *Curr Neurol Neurosci Rep* **15**, 17 (2015).
61. Gardai, S.J. *et al.* Elevated alpha-synuclein impairs innate immune cell function and provides a potential peripheral biomarker for Parkinson's disease. *PLoS One* **8**, e71634 (2013).
62. Emanuele, M. & Chieregatti, E. Mechanisms of alpha-synuclein action on neurotransmission: cell-autonomous and non-cell autonomous role. *Biomolecules* **5**, 865-92 (2015).
63. Hirsch, E.C., Vyas, S. & Hunot, S. Neuroinflammation in Parkinson's disease. *Parkinsonism Relat Disord* **18 Suppl 1**, S210-2 (2012).
64. Malkus, K.A., Tsika, E. & Ischiropoulos, H. Oxidative modifications, mitochondrial dysfunction, and impaired protein degradation in Parkinson's disease: how neurons are lost in the Bermuda triangle. *Mol Neurodegener* **4**, 24 (2009).
65. Olivares, D., Huang, X., Branden, L., Greig, N.H. & Rogers, J.T. Physiological and pathological role of alpha-synuclein in Parkinson's disease through iron mediated oxidative stress; the role of a putative iron-responsive element. *Int J Mol Sci* **10**, 1226-60 (2009).
66. Kalinderi, K., Papaliagkas, V. & Fidani, L. Pharmacogenetics and levodopa induced motor complications. *Int J Neurosci* **129**, 384-392 (2019).
67. Beudel, M. & Brown, P. Adaptive deep brain stimulation in Parkinson's disease. *Parkinsonism Relat Disord* **22 Suppl 1**, S123-6 (2016).
68. Menza, M. *et al.* A controlled trial of antidepressants in patients with Parkinson disease and depression. *Neurology* **72**, 886-92 (2009).
69. Richard, I.H. *et al.* A randomized, double-blind, placebo-controlled trial of antidepressants in Parkinson disease. *Neurology* **78**, 1229-36 (2012).
70. Dobkin, R.D. *et al.* Cognitive-behavioral therapy for depression in Parkinson's disease: a randomized, controlled trial. *Am J Psychiatry* **168**, 1066-74 (2011).
71. Polymeropoulos, M.H. *et al.* Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045-7 (1997).

72. Miller, D.W. *et al.* Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. *Neurology* **62**, 1835-8 (2004).
73. Puschmann, A. *et al.* A Swedish family with de novo alpha-synuclein A53T mutation: evidence for early cortical dysfunction. *Parkinsonism Relat Disord* **15**, 627-32 (2009).
74. Seidel, K. *et al.* First appraisal of brain pathology owing to A30P mutant alpha-synuclein. *Ann Neurol* **67**, 684-9 (2010).
75. Zarranz, J.J. *et al.* The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* **55**, 164-73 (2004).
76. Fujioka, S. *et al.* Update on novel familial forms of Parkinson's disease and multiple system atrophy. *Parkinsonism Relat Disord* **20 Suppl 1**, S29-34 (2014).
77. Blauwendraat, C. *et al.* Insufficient evidence for pathogenicity of SNCA His50Gln (H50Q) in Parkinson's disease. *Neurobiol Aging* **64**, 159 e5-159 e8 (2018).
78. Chartier-Harlin, M.C. *et al.* Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167-9 (2004).
79. Benamer, H.T. & de Silva, R. LRRK2 G2019S in the North African population: a review. *Eur Neurol* **63**, 321-5 (2010).
80. Kachergus, J. *et al.* Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. *Am J Hum Genet* **76**, 672-80 (2005).
81. Lesage, S. *et al.* Parkinson's disease-related LRRK2 G2019S mutation results from independent mutational events in humans. *Hum Mol Genet* **19**, 1998-2004 (2010).
82. Healy, D.G. *et al.* Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol* **7**, 583-90 (2008).
83. Berwick, D.C., Heaton, G.R., Azeggagh, S. & Harvey, K. LRRK2 Biology from structure to dysfunction: research progresses, but the themes remain the same. *Mol Neurodegener* **14**, 49 (2019).
84. Lee, A.J. *et al.* Penetrance estimate of LRRK2 p.G2019S mutation in individuals of non-Ashkenazi Jewish ancestry. *Mov Disord* **32**, 1432-1438 (2017).
85. Kilarski, L.L. *et al.* Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. *Mov Disord* **27**, 1522-9 (2012).

86. Larsen, S.B., Hanss, Z. & Kruger, R. The genetic architecture of mitochondrial dysfunction in Parkinson's disease. *Cell Tissue Res* **373**, 21-37 (2018).
87. Gan-Or, Z. *et al.* Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. *Neurology* **70**, 2277-83 (2008).
88. Hruska, K.S., LaMarca, M.E., Scott, C.R. & Sidransky, E. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum Mutat* **29**, 567-83 (2008).
89. Winder-Rhodes, S.E. *et al.* Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. *Brain* **136**, 392-9 (2013).
90. Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* **361**, 1651-61 (2009).
91. Alcalay, R.N. *et al.* Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. *Brain* **138**, 2648-58 (2015).
92. Gan-Or, Z. *et al.* Differential effects of severe vs mild GBA mutations on Parkinson disease. *Neurology* **84**, 880-7 (2015).
93. Lesage, S. *et al.* Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. *Hum Mol Genet* **20**, 202-10 (2011).
94. Gan-Or, Z., Liong, C. & Alcalay, R.N. GBA-Associated Parkinson's Disease and Other Synucleinopathies. *Curr Neurol Neurosci Rep* **18**, 44 (2018).
95. Senkevich, K. & Gan-Or, Z. Autophagy lysosomal pathway dysfunction in Parkinson's disease; evidence from human genetics. *Parkinsonism Relat Disord* **73**, 60-71 (2020).
96. Deng, H., Wang, P. & Jankovic, J. The genetics of Parkinson disease. *Ageing Res Rev* **42**, 72-85 (2018).
97. Lunati, A., Lesage, S. & Brice, A. The genetic landscape of Parkinson's disease. *Rev Neurol (Paris)* **174**, 628-643 (2018).
98. Bush, W.S. & Moore, J.H. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* **8**, e1002822 (2012).
99. Nalls, M.A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**, 1091-1102 (2019).

100. Escott-Price, V. *et al.* Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol* **77**, 582-91 (2015).
101. Ibanez, L. *et al.* Parkinson disease polygenic risk score is associated with Parkinson disease status and age at onset but not with alpha-synuclein cerebrospinal fluid levels. *BMC Neurol* **17**, 198 (2017).
102. Paul, K.C., Schulz, J., Bronstein, J.M., Lill, C.M. & Ritz, B.R. Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. *JAMA Neurol* **75**, 360-366 (2018).
103. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-d1012 (2019).
104. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
105. Hirschhorn, J.N. Genomewide association studies--illuminating biologic pathways. *N Engl J Med* **360**, 1699-701 (2009).
106. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22 (2017).
107. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
108. Altshuler, D., Daly, M.J. & Lander, E.S. Genetic mapping in human disease. *Science* **322**, 881-8 (2008).
109. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
110. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
111. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-5 (2014).

112. Anderson, C.A., Soranzo, N., Zeggini, E. & Barrett, J.C. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol* **9**, e1000580 (2011).
113. Scherag, A. *et al.* Investigation of a genome wide association signal for obesity: synthetic association and haplotype analyses at the melanocortin 4 receptor gene locus. *PLoS One* **5**, e13967 (2010).
114. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
115. Xing, C. *et al.* Evaluation of power of the Illumina HumanOmni5M-4v1 BeadChip to detect risk variants for human complex diseases. *Eur J Hum Genet* **24**, 1029-34 (2016).
116. Warman Chardon, J., Beaulieu, C., Hartley, T., Boycott, K.M. & Dyment, D.A. Axons to Exons: the Molecular Diagnosis of Rare Neurological Diseases by Next-Generation Sequencing. *Curr Neurol Neurosci Rep* **15**, 64 (2015).
117. Bamshad, M.J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* **12**, 745-55 (2011).
118. Pihlstrom, L., Rengmark, A., Bjornara, K.A. & Toft, M. Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease. *Ann Hum Genet* **78**, 243-52 (2014).
119. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619-22 (2012).
120. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511-1516 (2017).
121. Nalls, M.A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-93 (2014).
122. Kurian, M.A., Gissen, P., Smith, M., Heales, S., Jr. & Clayton, P.T. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. *Lancet Neurol* **10**, 721-33 (2011).
123. Mencacci, N.E. *et al.* Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain* **137**, 2480-92 (2014).
124. Delamarre, A. & Meissner, W.G. Epidemiology, environmental risk factors and genetics of Parkinson's disease. *Presse Med* **46**, 175-181 (2017).

125. Mastrangelo, L. The Genetics of Parkinson Disease. *Adv Genet* **98**, 43-62 (2017).
126. Zhang, P.L., Chen, Y., Zhang, C.H., Wang, Y.X. & Fernandez-Funez, P. Genetics of Parkinson's disease and related disorders. *J Med Genet* **55**, 73-80 (2018).
127. Rudakou, U. *et al.* Common and rare GCH1 variants are associated with Parkinson's disease. *Neurobiol Aging* **73**, 231 e1-231 e6 (2019).
128. Lesage, S. *et al.* Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *Am J Hum Genet* **98**, 500-513 (2016).
129. Rudakou, U. *et al.* Analysis of common and rare VPS13C variants in late-onset Parkinson disease. *Neurol Genet* **6**, 385 (2020).
130. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* **533**, 95-9 (2016).
131. Jansen, I.E. *et al.* Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol* **18**, 22 (2017).
132. Darvish, H. *et al.* Identification of a large homozygous VPS13C deletion in a patient with early-onset Parkinsonism. *Mov Disord* **33**, 1968-1970 (2018).
133. Schormair, B. *et al.* Diagnostic exome sequencing in early-onset Parkinson's disease confirms VPS13C as a rare cause of autosomal-recessive Parkinson's disease. *Clin Genet* **93**, 603-612 (2018).
134. Krohn, L. *et al.* Fine-mapping of SNCA in REM sleep behavior disorder and overt synucleinopathies. *BioRxiv*, 756528 (2019).
135. Krohn, L. *et al.* Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies. *Ann Neurol* **87**, 139-153 (2020).
136. Surmeier, D.J., Graves, S.M. & Shen, W. Dopaminergic modulation of striatal networks in health and Parkinson's disease. *Curr Opin Neurobiol* **29**, 109-17 (2014).
137. Xie, Z. *et al.* Molecular Mechanisms for the Coupling of Endocytosis to Exocytosis in Neurons. *Front Mol Neurosci* **10**, 47 (2017).
138. Wang, C. *et al.* Synaptotagmin-11 is a critical mediator of parkin-linked neurotoxicity and Parkinson's disease-like pathology. *Nat Commun* **9**, 81 (2018).
139. Blauwendraat, C. *et al.* Coding variation in GBA explains the majority of the SYT11-GBA Parkinson's disease GWAS locus. *Mov Disord* **33**, 1821-1823 (2018).

140. Arango Duque, G., Fukuda, M. & Descoteaux, A. Synaptotagmin XI regulates phagocytosis and cytokine secretion in macrophages. *J Immunol* **190**, 1737-45 (2013).
141. Arango Duque, G., Fukuda, M., Turco, S.J., Stager, S. & Descoteaux, A. Leishmania promastigotes induce cytokine secretion in macrophages through the degradation of synaptotagmin XI. *J Immunol* **193**, 2363-72 (2014).
142. Du, C. *et al.* Synaptotagmin-11 inhibits cytokine secretion and phagocytosis in microglia. *Glia* **65**, 1656-1667 (2017).
143. Wang, C. *et al.* Synaptotagmin-11 inhibits clathrin-mediated and bulk endocytosis. *EMBO Rep* **17**, 47-63 (2016).
144. Sreetama, S.C., Takano, T., Nedergaard, M., Simon, S.M. & Jaiswal, J.K. Injured astrocytes are repaired by Synaptotagmin XI-regulated lysosome exocytosis. *Cell Death Differ* **23**, 596-607 (2016).
145. Bento, C.F., Ashkenazi, A., Jimenez-Sanchez, M. & Rubinsztein, D.C. The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway. *Nat Commun* **7**, 11803 (2016).
146. Shimojo, M. *et al.* Synaptotagmin-11 mediates a vesicle trafficking pathway that is essential for development and synaptic plasticity. *Genes Dev* **33**, 365-376 (2019).
147. Yan, S. *et al.* Synaptotagmin-11 regulates the functions of caveolae and responds to mechanical stimuli in astrocytes. *FASEB J* **34**, 2609-2624 (2020).
148. Kumar, N. *et al.* VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact sites. *J Cell Biol* **217**, 3625-3639 (2018).
149. Itoh, N. & Ohta, H. Roles of FGF20 in dopaminergic neurons and Parkinson's disease. *Front Mol Neurosci* **6**, 15 (2013).
150. Sleeman, I.J., Boshoff, E.L. & Duty, S. Fibroblast growth factor-20 protects against dopamine neuron loss in vitro and provides functional protection in the 6-hydroxydopamine-lesioned rat model of Parkinson's disease. *Neuropharmacology* **63**, 1268-77 (2012).
151. Boshoff, E.L., Fletcher, E.J.R. & Duty, S. Fibroblast growth factor 20 is protective towards dopaminergic neurons in vivo in a paracrine manner. *Neuropharmacology* **137**, 156-163 (2018).

152. Fletcher, E.J.R., Jamieson, A.D., Williams, G., Doherty, P. & Duty, S. Targeted repositioning identifies drugs that increase fibroblast growth factor 20 production and protect against 6-hydroxydopamine-induced nigral cell loss in rats. *Sci Rep* **9**, 8336 (2019).
153. Niu, J. *et al.* Efficient treatment of Parkinson's disease using ultrasonography-guided rhFGF20 proteoliposomes. *Drug Deliv* **25**, 1560-1569 (2018).
154. Shen, Y.T. *et al.* BST1 rs4698412 allelic variant increases the risk of gait or balance deficits in patients with Parkinson's disease. *CNS Neurosci Ther* **25**, 422-429 (2019).
155. Neal, M.L., Boyle, A.M., Budge, K.M., Safadi, F.F. & Richardson, J.R. The glycoprotein GPNMB attenuates astrocyte inflammatory responses through the CD44 receptor. *J Neuroinflammation* **15**, 73 (2018).
156. Budge, K.M., Neal, M.L., Richardson, J.R. & Safadi, F.F. Transgenic Overexpression of GPNMB Protects Against MPTP-Induced Neurodegeneration. *Mol Neurobiol* (2020).
157. Leonard, H. *et al.* Genetic variability and potential effects on clinical trial outcomes: perspectives in Parkinson's disease. *J Med Genet* **57**, 331-338 (2020).
158. Li, J. *et al.* Full sequencing and haplotype analysis of MAPT in Parkinson's disease and rapid eye movement sleep behavior disorder. *Mov Disord* **33**, 1016-1020 (2018).
159. Shields, B.M. *et al.* Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia* **53**, 2504-8 (2010).
160. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455-8 (1996).
161. Ridker, P.M. *et al.* Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am J Hum Genet* **82**, 1185-92 (2008).
162. Reiner, A.P. *et al.* Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am J Hum Genet* **82**, 1193-201 (2008).
163. Owen, K.R. *et al.* Assessment of high-sensitivity C-reactive protein levels as diagnostic discriminator of maturity-onset diabetes of the young due to HNF1A mutations. *Diabetes Care* **33**, 1919-24 (2010).
164. Thanabalasingham, G. *et al.* A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes. *Diabetologia* **54**, 2801-10 (2011).

165. Elborn, J.S. Cystic fibrosis. *Lancet* **388**, 2519-2531 (2016).
166. Strug, L.J., Stephenson, A.L., Panjwani, N. & Harris, A. Recent advances in developing therapeutics for cystic fibrosis. *Hum Mol Genet* **27**, R173-R186 (2018).
167. Habib, A.R. *et al.* A Systematic Review of the Clinical Efficacy and Safety of CFTR Modulators in Cystic Fibrosis. *Sci Rep* **9**, 7234 (2019).

Appendix

Copyright permissions

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Uladzislau Rudakou

Parkinson's disease

Author: Lorraine V Kalia,Anthony E Lang
Publication: The Lancet
Publisher: Elsevier
Date: 29 August–4 September 2015

Copyright © 2015 Elsevier Ltd. All rights reserved.

Order Completed

Thank you for your order.

This Agreement between Mr. Uladzislau Rudakou ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	4831990365695
License date	May 18, 2020

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	The Lancet
Licensed Content Title	Parkinson's disease
Licensed Content Author	Lorraine V Kalia,Anthony E Lang
Licensed Content Date	29 August–4 September 2015
Licensed Content Volume	386
Licensed Content Issue	9996
Licensed Content Pages	17
Journal Type	HS

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title	A closer look into Parkinson's Disease GWAS genes: Targeted next-generation sequencing approach
Institution name	McGill University
Expected presentation date	Aug 2020

Additional Data

Portions	Figure 1: Clinical symptoms and time course of Parkinson's disease progression Page 898 (3)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Uladzislau Rudakou

Milestones in Parkinson's disease—Clinical and pathologic features

Author: Glenda Halliday, Andrew Lees, Matthew Stern
Publication: Movement Disorders
Publisher: John Wiley and Sons
Date: May 27, 2011

Copyright © 2011 Movement Disorder Society

Order Completed

Thank you for your order.

This Agreement between Mr. Uladzislau Rudakou ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4832010952203
License date	May 18, 2020

Printable Details

Licensed Content

Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Movement Disorders
Licensed Content Title	Milestones in Parkinson's disease—Clinical and pathologic features
Licensed Content Author	Glenda Halliday, Andrew Lees, Matthew Stern
Licensed Content Date	May 27, 2011
Licensed Content Volume	26
Licensed Content Issue	6
Licensed Content Pages	7

Order Details

Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No

About Your Work

Title	A closer look into Parkinson's Disease GWAS genes: Targeted next-generation sequencing approach
Institution name	McGill University
Expected presentation date	Aug 2020

Additional Data

Portions	Fig.1 Stylized representation of the Braak staging for Parkinson's disease showing the initiation sites in the medulla oblongata and olfactory bulb through to the later infiltration of Lewy pathology into the cortical regions Page 2

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Uladzislau Rudakou

Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies

Author:
Mike A Nalls,Cornelis Blauwendraat,Costanza L Vallerga,Karl Heilbron,Sara Bandres-Ciga,Diana Chang,Manuela Tan,Demis A Kia,Alastair J Noyce,Angli Xue,Jose Bras,Emily Young,Rainer von Coelln,Javier Simón-Sánchez,Claudia Schulte,Manu Sharma et al.

Publication: The Lancet Neurology

Publisher: Elsevier

Date: December 2019

© 2019 Elsevier Ltd. All rights reserved.

Order Completed

Thank you for your order.

This Agreement between Mr. Uladzislau Rudakou ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4832011331218
License date	May 18, 2020

Printable Details

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Uladzislau Rudakou

Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders

Author:
Brian J. O'Roak,Laura Vives,Wenqing Fu,Jarrett D. Egertson,Ian B. Stanaway,Ian G. Phelps,Gemma Carvill,Akash Kumar,Choli Lee,Katy Ankenman,Jeff Munson,Joseph B. Hiatt,Emily H. Turner,Roie Levy,Diana R. O'Day,Niklas Krumm,Bradley P. Coe,Beth K. Martin,Elhanan Borenstein,Deborah A. Nickerson,Heather C. Mefford,Dan Doherty,Joshua M. Akey,Raphael Bernier,Evan E. Eichler,Jay Shendure

Publication: Science

Publisher: The American Association for the Advancement of Science

Date: Dec 21, 2012

Copyright © 2012, Copyright © 2012, American Association for the Advancement of Science

Order Completed

Thank you for your order.

This Agreement between Mr. Uladzislau Rudakou ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4833750304591	Printable Details
License date	May 21, 2020	

Ethics certificate

MUHC REB – NEUPSY / CÉR CUSM – NEUPSY
3801, rue University, # 686
Montréal (Québec) H3A 2B4
Tel. 514.398.1046
reb.neuro@mcgill.ca
www.leneuro.ca

Annual renewal submission

Submit date: **2019-12-04 11:55**
Project's REB approbation date: **2016-12-01**
Project number: **2017-2944**
Form status: **Approved**

Submitted by: **Zaharieva, Vessela**
Nagano identifier: **RBD-PD**
Form: **F9 - 50014**

Administration

1. **MUHC REB Panel & Co-chair(s):**
   Neurosciences-Psychiatry (NEUPSY)
   Co-chairs: Judith Marcoux, Brigitte Pâquet

2. **REB Decision:**
   Approved - REB delegated review

3. **Comments on the decision:**
   Please note that this Annual Renewal Form was submitted prior to the REB approval expiry.

4. **Renewal Period Granted:**
   Until 2020-12-01

5. **Date of the REB final decision & signature**
   2019-12-04
   Signature

Ms. Brigitte Paquet, LL.B.
Brigitte Pâquet, LL.B.
Co-Chair, MUHC REB (NEUPSY Panel)

6. FWA 00000840 - FWA 00004545

7. **Local REB number**
IRB00010120

A. General information

1. **Indicate the full title of the research study**

The genetics of REM sleep behavior disorder, Parkinson's disease and other synucleinopathies

2. **If relevant, indicate the full study title in French**

3. **Indicate the name of the Principal Investigator in our institution (MUHC)**

Gan-Or, Ziv

From which department is the principal investigator?

Neurology

4. **Are there local co-investigators & collaborators involved in this project?**

Yes

List all the local co-investigators & collaborators involved in the research study

Rouleau, Guy

Department of co-investigators & collaborators

Neurology

Fon, Edward A

Department of co-investigators & collaborators

Neurology

Postuma, Ronald

Department of co-investigators & collaborators

Neurology

B. Project development

1. **Study start date:**

2016-12-19

2. **Expected ending date of the study:**

☐ Determined date
☑ Undetermined date

3. **Date of recruitment of the first participant?**

☑ 1st enrollment date is...
☐ No participant enrolled

1st participant enrollment date:
2016-12-21

4. **Add a brief statement on the study status**

The project is still on-going

5. **Information about the participants at this institution, since the beginning of the project**

Number of participants who have been recruited
2647
Number of participants who have not yet completed the study (still in progress)
0
Number of participants who've completed the study
2647
Number of participants who were recruited to the study, but who were then excluded or withdrawn:
0
Number of participants who dropped out (voluntary withdrawal):
0
Number of participants who died during the study
0

6. **Information about the participants at this institution (MUHC) since the previous REB approval**

Number of participants who have been recruited
566
Number of participants who have not yet completed the study (still in progress)
0
Number of participants who've completed the study
566
Number of participants who were recruited to the study, but who were then excluded or withdrawn:
0
Number of participants who dropped out (voluntary withdrawal):
0
Number of participants who died during the study
0

7. **Since the previous REB approval (annual renewal or initial approval):**

Were there any changes to the protocol (or to the databank management framework) ?

No

Specify the current version/date:

V3 2018-06-21

Date approved by the REB:

2018-06-21

Were there any changes to the information and consent form?

No

Specify the current version/date:

V1

REB approval date:

2016-03-14

Were there any reportable adverse events at this site (or, for multi-center projects, at an institution under the jurisdiction of our REB) that should be reported to the REB under section 5.2.1 of " SOP-REB- 404001 " ?

https://muhc.ca/cae/page/standard-operating-procedures-sops

No

Has there has been any new information likely to affect the ethics of the project or influence the decision of a participant as to their continued participation in the project ?

No

Were there any deviations / major violations protocol (life -threatening or not meeting the inclusion / exclusion criteria)?

No

Was there a temporary interruption of the project?

No

Have the project results been submitted for publication, presented or published?

No

Has the REB been notified of a conflict of interest - (apparent , potential or actual), of one or more members of the research team - that was not known when it was last approved project?

No

Do you want to bring any other info to the REB's attention?

No

8. **Is there a data safety monitoring committee analyzing data on the safety and efficacy of the treatment?**

No

C. Signature

1. **Comments**

All samples come from the RouBank that continues to increase throughout the study period. We need as many as possible samples and therefore we have not a target for number of participants

2. **I confirm that all information is complete & accurate.**

First & last name of person who completed the submission

Vessela Zaharieva