LUND UNIVERSITY

The etiology of Parkinson's disease. Genetic and non-genetic risk factors for a multifactorial disease

Brolin, Kajsa

2023

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Brolin, K. (2023). *The etiology of Parkinson's disease. Genetic and non-genetic risk factors for a multifactorial disease*. [Doctoral Thesis (compilation), Department of Experimental Medical Science]. Lund University, Faculty of Medicine.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The etiology of Parkinson's disease

Genetic and non-genetic risk factors for a multifactorial disease

KAJSA BROLIN
DEPT. OF EXPERIMENTAL MEDICAL SCIENCE | FACULTY OF MEDICINE | LUND UNIVERSITY

The etiology of Parkinson's disease

The etiology of Parkinson's disease

Genetic and non-genetic risk factors for a multifactorial disease

Kajsa Brolin

LUND UNIVERSITY

DOCTORAL DISSERTATION

By due permission of the Faculty of Medicine at Lund University, Sweden.
To be defended at Segerfalksalen, Biomedical Centre (BMC), Lund, Sweden.
March 27, 2023, at 9.00.

Faculty opponent

Professor Beate Ritz, M.D., Ph.D.
University of California, Los Angeles, USA

Organization
LUND UNIVERSITY
Faculty of Medicine
Dept of Experimental Medical Science
Translational Neurogenetics Unit

Author: Kajsa Brolin

Document name
DOCTORAL DISSERTATION

Date of issue 2023-03-27

Sponsoring organization

Title and subtitle
The etiology of Parkinson's disease - Genetic and non-genetic risk factors for a multifactorial disease

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder affecting over six million people worldwide. It is characterized by the progressive loss of dopaminergic neurons in the substantia nigra, resulting in motor symptoms such as bradykinesia, rigidity, and tremor, and non-motor symptoms such as REM sleep behavior disorder (RBD). However, both the phenotypic features and the etiology of PD are heterogenous, suggesting that the disease does not exist as a single entity. It is known that the etiology of PD is complex, involving both genetic and non-genetic factors (e.g., environment and lifestyle). The genetic contributors to PD exist across a continuum, ranging from causal and highly penetrant genetic variants (monogenic PD) to common variants with small effect sizes associated with an increase disease risk (idiopathic PD). Known monogenic loci have been reported to explain a fraction of the observed familial aggregation of PD and known common risk variants explain only up to 36% of the heritable PD risk. Thus, a considerable large part of the genetic component in PD etiology remains unidentified. Additionally, the global burden of PD is increasing and considering that a person's genetic make-up is largely stable, unknown non-genetic factors likely contribute substantially to PD development.

The overall aim of this thesis was to extend the knowledge of genetic and non-genetic factors in PD etiology. Firstly, the prevalence of known pathogenic mutations in the dominant PD genes *LRRK2* and *SNCA* was investigated in Sweden, showing that these mutations rarely contribute to PD in Sweden. Subsequently, the gene *RIC3,* which had been reported to cause autosomal dominant PD in the Indian population, was investigated in cohorts of European, Latin American, and East Asian ancestry where no association was observed. Next, the contribution of common genetic and non-genetic factors to PD was investigated in Sweden. Multiple factors could be confirmed to be associated with PD, including prior pesticide exposure and the use of the Swedish moist tobacco snus. A new and potential population-specific genetic risk variant was observed in the *PLPP4* locus. This finding was further investigated in Swedish, Norwegian and mixed European cohorts, concluding that a potential population-specific effect of *PLPP4* cannot be ruled out. Additionally, the genetic architecture of common variants in RBD was examined, indicating a potential difference in the genetic background of PD, RBD, and other synucleinopathies. In summary, this thesis contributes to the understanding of PD etiology and can help to pave the way for development of predictive, preventative, and new therapeutical approaches in PD.

Key words Parkinson's disease, genetics, GWAS, risk factors, epidemiology, REM sleep behavior disorder

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

ISSN and key title: 1652-8220

ISBN: 978-91-8021-380-6

Recipient's notes

Number of pages 113

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Kajsa Brolin

Date 2023-02-15

The etiology of Parkinson's disease

Genetic and non-genetic risk factors for a multifactorial disease

Kajsa Brolin

LUND UNIVERSITY

Cover photo by Kajsa Brolin using the assistance of DALL-E 2 and with modifications by Matilde Negrini

Copyright pp 1-113 Kajsa Brolin

Paper 1 © The Authors. Published by Elsevier

Paper 2 © The Authors. Published by Elsevier

Paper 3 © The Authors. Published by IOS Press

Paper 4 © The Authors. (Manuscript unpublished)

Paper 5 © The Authors. Published by Nature Communications

Faculty of Medicine
Department of Experimental Medical Science

ISBN 978-91-8021-380-6
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2023

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

‘You don't have to be brilliant. It's enough to become progressively less stupid.’

- Marshall Rosenberg

Table of Contents

Original papers and manuscripts **11**
Papers outside of this thesis 13
Abstract **15**
Lay summary **16**
Populärvetenskaplig sammanfattning **18**
Abbreviations **20**
Introduction **23**
Neurodegenerative disorders 23
Parkinson's disease 23
The heterogeneity of Parkinson's disease 25
Prodromal Parkinson's disease 27
Treatments in Parkinson's disease 28
Epidemiology 29
Parkinson's disease genetics 32
Monogenic Parkinson's disease 33
Idiopathic PD 39
Genetic variants associated with PD expressivity 40
Non-genetic risk factors in Parkinson's disease 41
Sex and age as risk factors for Parkinson's disease 41
Lifestyle and dietary factors 42
Environmental factors 44
Comorbidities, infections, and drugs in PD risk 46
Gene-environment interaction 47
Aims of the thesis **48**
Methodological considerations **49**
Ethical considerations and open science 49
Cohorts 50
Multipark's biobank sample collection (MPBC) 50
Other Swedish cohorts 51

The International Parkinson's Disease Genomics Consortium 53
Accelerating Medicines Partnership - Parkinson's Disease..... 55
The Global Parkinson's Genetics Program 56
Epidemiological analyses 56
Genetic analyses 57
Genotyping 57
Quality control (QC) 57
Imputation 59
Power calculations for GWAS 59
Genetic association analyses 60
Polygenic risk scores (PRS) 61
Haplotype analyses..... 62
Cumulative effect of risk variants 62
Other GWAS down-stream analyses..... 63
Summary of results and discussion 64
Paper I | Low prevalence of known pathogenic mutations in dominant PD genes: A Swedish multicenter study..... 64
Paper II | RIC3 variants are not associated with Parkinson's disease in large European, Latin American, or East Asian cohorts 67
Paper III | Insights on genetic and environmental factors in Parkinson's disease from a regional Swedish case-control cohort 70
Paper IV | Investigating the potential association of *PLPP4* with Parkinson's disease in Scandinavian and European cohorts 77
Paper V | Genome-wide association study of REM sleep behavior disorder identifies polygenic risk and brain expression effects..... 81
Concluding remarks and future perspectives 87
References 91
Acknowledgements 111
Appendix (Papers I-V) 113

Original papers and manuscripts

The thesis is based on the following published articles and manuscripts:

I. Low prevalence of known pathogenic mutations in dominant PD genes: A Swedish multicenter study.

   Puschmann A, Jiménez-Ferrer I, Lundblad-Andersson E, Mårtensson E, Hansson O, Odin P, Widner H, **Brolin K**, Mzezewa R, Kristensen J, Soller M, Rödström EY, Ross OA, Toft M, Breedveld GJ, Bonifati V, Brodin L, Zettergren A, Sydow O, Linder J, Wirdefeldt K, Svenningsson P, Nissbrandt H, Belin AC, Forsgren L, Swanberg M.

   Parkinsonism & Related Disorders. 66:158-165, 2019.

II. *RIC3* variants are not associated with Parkinson's disease in large European, Latin American, or East Asian cohorts

   Brolin K, Bandres-Ciga S, Leonard H, Makarious MB, Blauwendraat C, Mata IF, Foo JN, Pihlstrøm L, Swanberg M, Gan-Or Z, Tan MMX; International Parkinson's Disease Genomics Consortium.

   Neurobiology of Aging. 109:264-268, 2021.

III. Insights on Genetic and Environmental Factors in Parkinson's Disease from a Regional Swedish Case-Control Cohort.

   Brolin K, Bandres-Ciga S, Blauwendraat C, Widner H, Odin P, Hansson O, Puschmann A, Swanberg M.

   Journal of Parkinson's Disease. 12(1):153-171, 2022.

IV. Investigating the potential association of *PLPP4* with Parkinson's disease in Scandinavian and European cohorts

Brolin K, Ran C, Bäckström D, Svenningsson P, Forsgren L, Global Parkinson's Genetics Program (GP2) Consortium, Tan MM, Pihlström L, Hansson O, Carmine Belin A, Swanberg M.

Manuscript

V. Genome-wide association study of REM sleep behavior disorder identifies polygenic risk and brain expression effects.

Krohn L, Heilbron K, Blauwendraat C, Reynolds RH, Yu E, Senkevich K, Rudakou U, Estiar MA, Gustavsson EK, **Brolin K**, Ruskey JA, Freeman K, Asayesh F, Chia R, Arnulf I, Hu MTM, Montplaisir JY, Gagnon JF, Desautels A, Dauvilliers Y, Gigli GL, Valente M, Janes F, Bernardini A, Högl B, Stefani A, Ibrahim A, Šonka K, Kemlink D, Oertel W, Janzen A, Plazzi G, Biscarini F, Antelmi E, Figorilli M, Puligheddu M, Mollenhauer B, Trenkwalder C, Sixel-Döring F, Cochen De Cock V, Monaca CC, Heidbreder A, Ferini-Strambi L, Dijkstra F, Viaene M, Abril B, Boeve BF; 23andMe Research Team, Scholz SW, Ryten M, Bandres-Ciga S, Noyce A, Cannon P, Pihlstrøm L, Nalls MA, Singleton AB, Rouleau GA, Postuma RB, Gan-Or Z.

Nature Communications, 5;13(1):7496, 2022.

Papers outside of this thesis

- The Parkinson's Disease Genome-Wide Association Study Locus Browser.

  Grenn FP, Kim JJ, Makarious MB, Iwaki H, Illarionova A, **Brolin K**, Kluss JH, Schumacher-Schuh AF, Leonard H, Faghri F, Billingsley K, Krohn L, Hall A, Diez-Fairen M, Periñán MT, Sandor C, Webber C, Fiske BK, Gibbs JR, Nalls M, Singleton A, Bandres-Ciga S, Reed X, Blauwendraat C, on behalf of the International Parkinson's Disease Genomics Consortium (IPDGC).

  Movement Disorders. 35(11):2056-67, 2020.

- Effect Modification between Genes and Environment and Parkinson's Disease Risk. *Review*.

  Periñán MT*, **Brolin K*,** Bandres-Ciga S, Blauwendraat C, Klein C, Gan-Or Z, Singleton A, Gomez-Garre P, Swanberg M, Mir P, Noyce A.

  Annals of Neurology. 92(5):715-724, 2022.

  *Shared first authorship.

- The IPDGC/GP2 Hackathon - an open science event for training in data science, genomics, and collaboration using Parkinson's disease data.

  Leonard H, Murtadha R, Martinez-Carrasco A, Muller-Nedebock A, Gil-Martinez AL, Illarionova A, Moore A, Bustos B, Jadhav B, Huxford B, Storm C, Towns C, Vitale D, Chetty D, Yu E, Jama A, Grenn F, Salazar G, Rateau G, Iwaki H, Elsayed I, Foote I, Jansen van Rensburg Z, Kim J, Yuan J, Lake J, **Brolin K**, Senkevich K, Wu L, Tan MMX, Periñán MT, Makarious M, Ta M, Nikita Simone Pillay, Oswaldo Lorenzo Betancor, Paula Reyes-Pérez, Jerez P, Saini P, al-Ouran R, Sivakumar R, Real R, Reynolds R, Hu R, Abrahams S, Rao S, Antar T, Leal T, Iankova V, Scotton W, Song Y, Singleton A, Nalls M, Dey S, Bandrés-Ciga S, Blauwendraat C, Noyce A.

  Accepted in NPJ Parkinson's Disease, 2023.

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder affecting over six million people worldwide. It is characterized by the progressive loss of dopaminergic neurons in the substantia nigra, resulting in motor symptoms such as bradykinesia, rigidity, and tremor, and non-motor symptoms such as REM sleep behavior disorder (RBD). However, both the phenotypic features and the etiology of PD are heterogenous, suggesting that the disease does not exist as a single entity. It is known that the etiology of PD is complex, involving both genetic and non-genetic factors (e.g., environment and lifestyle). The genetic contributors to PD exist across a continuum, ranging from causal and highly penetrant genetic variants (monogenic PD) to common variants with small effect sizes associated with an increase disease risk (idiopathic PD). Known monogenic loci have been reported to explain a fraction of the observed familial aggregation of PD and known common risk variants explain only up to 36% of the heritable PD risk. Thus, a considerable large part of the genetic component in PD etiology remains unidentified. Additionally, the global burden of PD is increasing and considering that a person's genetic make-up is largely stable, unknown non-genetic factors likely contribute substantially to PD development.

The overall aim of this thesis was to extend the knowledge of genetic and non-genetic factors in PD etiology. Firstly, the prevalence of known pathogenic mutations in the dominant PD genes *LRRK2* and *SNCA* was investigated in Sweden, showing that these mutations rarely contribute to PD in Sweden. Subsequently, the gene *RIC3,* which had been reported to cause autosomal dominant PD in the Indian population, was investigated in cohorts of European, Latin American, and East Asian ancestry where no association was observed. Next, the contribution of common genetic and non-genetic factors to PD was investigated in Sweden. Multiple factors could be confirmed to be associated with PD, including prior pesticide exposure and the use of the Swedish moist tobacco snus. A new and potential population-specific genetic risk variant was observed in the *PLPP4* locus. This finding was further investigated in Swedish, Norwegian and mixed European cohorts, concluding that a potential population-specific effect of *PLPP4* cannot be ruled out. Additionally, the genetic architecture of common variants in RBD was examined, indicating a potential difference in the genetic background of PD, RBD, and other synucleinopathies. In summary, this thesis contributes to the understanding of PD etiology and can help to pave the way for development of predictive, preventative, and new therapeutical approaches in PD.

Lay summary

Parkinson's disease is a neurodegenerative disorder affecting over six million people worldwide. In the disease, neurons in the midbrain that produce the signal molecule dopamine die. Dopamine is crucial for regulating movements and the lower levels of dopamine in Parkinson's disease results in symptoms such as slowness of movement, stiffness, and shaking while at rest. There are also less visible symptoms that occur in the disease, such as cognitive impairment and sleep problems, including the rapid eye movement sleep behavior disorder, RBD. These less visible symptoms can occur years before the classical symptoms that affects movements and the symptoms in Parkinson's disease vary substantially between affected individuals. It is not fully understood why some individuals develop Parkinson's disease as the etiology -the cause(s) of the disease - is complex and involves a combination of genetic and non-genetic factors, such as environmental and lifestyle factors.

A strong genetic component is seen in a fraction of all individuals with Parkinson's disease. These individuals have changes (mutations) in certain genes that can be inherited and cause the disease. However, for the majority of patients, it is instead several common variations in certain genes that have contributed to the disease. Our knowledge regarding genes behind Parkinson's disease is still low, meaning that a large part of the genetic contribution to disease development is unknown. Globally, cases of Parkinson's disease have increased in recent years, and are expected to continue to increase substantially in the coming years. This increase is partly due to a growing elderly population in the world since age is the number one risk factor for the disease. Additionally, exposure to detrimental environmental and lifestyle factors, such as pesticides, may contribute to the increased numbers. Understanding the factors connected to an increased risk of Parkinson's disease is therefore of great importance for developing preventative measures and new interventions.

In this thesis, I have investigated genetic and non-genetic factors that contribute to Parkinson's disease etiology. *In the first part*, I investigated mutations in two genes, *SNCA* and *LRRK2*, that are known to cause Parkinson's disease, using samples from individuals with the disease from Sweden. The results showed that mutations in these genes are a rare cause of Parkinson's disease in Sweden and that other factors likely underlie the Parkinson's disease risk in the Swedish population. *In the second part*, I investigated the link between Parkinson's disease and a potential disease-causing gene, named *RIC3*, which previously had been reported to cause

Parkinson's disease in an Indian population. However, I did not observe a link between this gene and Parkinson's disease in a study population with individuals of European, Latin American, and East Asian ancestry.

In the third part, I investigated the contribution of common genetic and non-genetic factors to Parkinson's disease in Sweden. I confirmed a link between several factors and Parkinson's disease, such as an increased risk of the disease in individuals who had been exposed to pesticides and a decreased risk in those using Swedish moist tobacco, also known as snus. Additionally, a genetic risk variant in the gene *PLPP4* was found to be associated to the risk of Parkinson's disease which had not been reported to be linked to the disease before. *In the fourth part*, I further investigated the *PLPP4* gene in more individuals from Sweden, Norway, and other countries in Europe, and noted that the connection between the risk gene variant and Parkinson's disease may actually be specific to the Swedish population.

In the fifth part, I investigated the potential genetic overlap between Parkinson's disease and the sleep disorder RBD as some individuals that have Parkinson's disease are also diagnosed with RBD, which may occur before or after the Parkinson's disease diagnosis. I observed that genetics partly determine if an individual develops RBD and Parkinson's disease and plays a role in whether an individual is diagnosed with RBD pre- or post-Parkinson's disease diagnosis.

In conclusion, this thesis contributes to the understanding of Parkinson's disease etiology and can help to pave the way for the development of future predictive, preventative, and new therapeutical approaches in the disease.

Populärvetenskaplig sammanfattning

Parkinsons sjukdom är en neurodegenerativ sjukdom som drabbar över sex miljoner människor världen över. Sjukdomen påverkar förmågan att styra kroppens rörelser då celler som tillverkar signalsubstansen dopamin i mellanhjärnan gradvis dör. Detta leder till lägre nivåer av dopamin och då hjärnan använder dopamin för att kontrollera rörelser uppkommer typiska symptom så som långsamma rörelser, stelhet och skakningar vid vila. Även andra symptom är vanliga i sjukdomen och inkluderar sämre kognitiva funktioner och sömnstörningar så som REM-sömnstörningar (rapid eye movement sleep behavior disorder, RBD). Dessa symptom kan uppkomma flera år innan de typiska symptomen i sjukdomen. Generellt är det en stor variation i symptomen mellan olika individer med Parkinson sjukdom. En stor variation ses också för orsakerna till att Parkinsons sjukdom uppkommer, vilket kallas etiologi. Orsakerna till Parkinsons sjukdom är ännu inte helt fastställda men det är känt att etiologin är komplex och involverar en kombination av genetiska och icke-genetiska faktorer, så som miljöfaktorer och livsstil.

En stark genetisk risk orsakar sjukdomen hos en bråkdel av alla individer där förändringar (mutationer) i specifika gener leder till sjukdomen. För majoriteten finns dock ingen stark genetisk förklaring utan istället bidrar flera vanliga genetiska variationer till en ökad sjukdomsrisk. Kunskapen kring genetik i Parkinsons sjukdom är fortfarande relativt låg, vilket innebär att en stor del av den genetiska risken bakom sjukdomsutveckling är okänd.

Den globala förekomsten av Parkinsons sjukdom har ökat de senaste åren och tros fortsätta öka kraftigt kommande år. Ökningen beror delvis på en åldrande population i världen då stigande ålder är den främsta riskfaktorn för Parkinsons sjukdom. Dock bidrar även miljöfaktorer och livsstil till den ökade förekomsten, där exempelvis exponering för bekämpningsmedel är kopplat till en ökad sjukdomsrisk. Ökad förståelse för faktorer som är kopplade till en ökad risk för Parkinsons sjukdom är av stor betydelse för att kunna vidta förebyggande åtgärder och utveckla nya behandlingar.

I denna avhandling har jag studerat genetiska och icke-genetiska faktorer som bidrar till etiologin i Parkinsons sjukdom. *I den första delen* undersökte jag förekomsten av mutationer i två gener, *SNCA* och *LRRK2*, vilka är kända att orsaka Parkinsons sjukdom, hos individer med sjukdomen från Sverige. Resultatet visade att

mutationer i dessa gener är ovanliga och endast förklarar en bråkdel av den totala ärftligheten av Parkinsons sjukdom i Sverige. *I den andra delen* undersökte jag kopplingen mellan Parkinsons sjukdom och genen *RIC3*, vilken tidigare har rapporterats kunna orsaka sjukdomen i en indisk familj. Dock observerade jag ingen koppling till Parkinsons sjukdom o populationer med europeiskt, latinamerikanskt, eller östasiatiskt ursprung.

I den tredje delen undersökte jag vanliga genetiska och icke-genetiska riskfaktorer för Parkinsons sjukdom i Sverige. Jag bekräftade en koppling mellan flertalet faktorer och sjukdomen, såsom en ökad sjukdomsrisk hos individer som exponerats för pesticider och en lägre sjukdomsrisk hos individer som snusade. Dessutom identifierades en riskvariant i genen *PLPP4* vilken var kopplad till sjukdomsrisk. Denna riskvariant har inte rapporterats vara kopplad till Parkinsons sjukdom tidigare och kan vara specifik för den svenska populationen. *I den fjärde delen* undersökte jag detta vidare hos fler individer från Sverige, Norge och andra europeiska länder och såg att kopplingen troligen är populationsspecifik.

I den femte delen undersökte jag potentiella gemensamma genetiska riskfaktorer för Parkinsons sjukdom och sömnstörningen RBD. En individ kan utveckla RBD före eller efter diagnosen Parkinsons sjukdom, eller så har individen Parkinsons sjukdom utan RBD. Resultatet från studien visade på att dessa grupper genetiskt skiljer sig åt och att detta kan avgöra om RBD förekommer samtidigt med PD.

Sammantaget bidrar denna avhandling till förståelsen kring etiologin i Parkinsons sjukdom, vilket är av stor betydelse för framtida utveckling av diagnostisering, förebyggande åtgärder och nya behandlingar i sjukdomen.

Abbreviations

AAD	Age at diagnosis
AAI	Age at inclusion
AAO	Age at onset
AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
AMP-PD	Accelerating Medicines Partnership Parkinson’s disease
BMI	Body mass index
CEU	Utah residents with Northern and Western European ancestry
CISI-PD	Clinical Impression of Severity Index - Parkinson’s Disease
COMT	Catechol-O-methyltransferase
DAGs	Directed acyclic graphs
DBS	Deep brain stimulation
DDC	Dopa decarboxylase
DLB	Dementia with Lewy bodies
EOPD	Early onset Parkinson's disease
EQ-5D-3L	EuroQol five-dimension - 3-level
eQTL	Expression quantitative trait loci
FDA	The US Food and Drug Administration
GBD	The Global Burden of Disease consortium
GD	Gaucher’s disease
GDPR	The General Data Protection Regulation
GEoPD	Genetic Epidemiology of Parkinson's disease
GP2	The Global Parkinson’s Genetics Program
GRCh	Genome Reference Consortium Human Build
GRS	Genetic risk score
GWAS	Genome-wide association study
H&Y	Hoehn and Yahr
HRC	Haplotype Reference Consortium
HWE	Hardy–Weinberg equilibrium
IBD	Inflammatory bowel diseases
IPDGC	The International Parkinson’s Disease Genomics Consortium
LARGE-PD	The Latin American Research Consortium on the Genetics of Parkinson’s Disease
LD	Linkage disequilibrium
LRRK2	Leucine-rich repeat kinase 2
MAF	Minor allele frequency
MAO-B	Monoamine oxidase type B
MDS	International Parkinson and Movement Disorder Society
MJFF	The Michael J. Fox Foundation
MPBC	Multipark's biobank sample collection

MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MR	Mendelian randomization
MRI	Magnetic resonance imaging
MSA	Multiple system atrophy
nAChRs	Nicotinic acetylcholine receptors
NDD	Neurodegenerative disorders
NIH	National Institutes of Health
OR	Odds ratio
PARKIN	Parkin RBR E3 ubiquitin protein
PARLU	The PARkinson LUnd study
PCA	Principal component analysis
PCs	Principal component
PD	Parkinson's disease
PDQ-8	Parkinson’s Disease Questionnaire - 8
PINK1	PTEN-induced putative kinase 1
PLPP	Phospholipase phosphatase
PLPP4	Phospholipase phosphatase 4
PRS	Polygenic risk score
QC	Quality control
QQ	Quantile-Quantile
RBD	REM sleep behavior disorder
REM	Rapid eye movement
SD	Standard deviation
SDI	Sociodemographic index
SKAT	Sequence kernel association test
SKAT-O	Optimal sequence kernel association test
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SSD	Schizophrenia spectrum disorders
T2D	Type 2 diabetes
T2T	Telomer-to-Telomer
TSI	Toscani in Italia
TTO	Time trade-off
VAS	Visual analogue scale
WGG	Whole-genome genotyping
WGS	Whole-genome sequencing
YOPD	Young-onset Parkinson's disease

Introduction

In this chapter, I will provide a general overview of Parkinson's disease (PD) and PD etiology. I will discuss the heterogeneity on the disease and its prodromal phase, highlighting the PD-associated sleep disorder rapid eye movement (REM) sleep behavior disorder (RBD). Finally, I will discuss genetic and non-genetic factors associated with PD and shortly address their potential interactions.

Neurodegenerative disorders

Neurodegenerative disorders (NDD) are a heterogenous group of disorders characterized by progressive damage and degeneration of the cells and structures of the central or peripheral nervous systems. NDD are often considered as separate entities, targeting different brain regions with distinct pathology and symptoms (1). However, numerous common traits and patterns are seen in the different disorders, where one example are pathologic lesions composed of aggregated and misfolded proteins. Synucleinopathies are a spectrum of NDD characterized by aggregated α-synuclein (α-syn) protein in selectively vulnerable populations of neurons and glia, known as Lewy bodies and Lewy neurites. Synucleinopathies includes PD, dementia with Lewy bodies (DLB), and other rarer disorders such as multiple system atrophy (MSA) (2). Neurological disorders, which includes NDD, are the leading cause of disability and the second leading cause of death worldwide. The prevalence of the major disabling neurological disorders increases with age and considering growing and aging populations, the burden of these disorders is expected to increase substantially (3).

Parkinson's disease

The first published description of PD was made by James Parkinson in his essay on the shaking palsy in the year 1817, describing a few individuals with the disease having symptoms of e.g., "involuntary tremulous motion, with lessened voluntary power" and "propensity to bend the trunk forwards" (4). As of today, PD is the second most common NDD after Alzheimer's disease (AD) and 6.1 million

individuals worldwide had the disease in 2016. It has been reported to be the fastest growing neurological disorder in prevalence, disability, and deaths and over the past generation, the global burden of PD has more than doubled (5).

PD is called a NDD mainly due to the characteristic feature of a degeneration of dopaminergic neurons in the basal ganglia structure substantia nigra pars compacta (SNpc), located in the midbrain. Together with the intracellular inclusions of aggregated α-syn, these two major neuropathologies are characterizing PD when observed together at post-mortem pathological examination (6). A definitive diagnosis of PD can only be established in post-mortem brains and PD is therefore a clinical diagnosis based on motor symptoms, being bradykinesia (slowness of movement) plus rigidity and/or resting tremor, often together with a beneficial response to dopaminergic therapy (7, 8).

The dopaminergic neuronal loss within the SNpc is believed to be the main cause of the motor symptoms observed in the disease. However, a dramatic loss of these neurons has been reported to occur already before the onset of motor symptoms in the disease (7, 9). Besides the two major neuropathological characteristics of PD, the pathophysiology of the disease includes a complex interplay of additional features such as mitochondrial dysfunction, abnormal lysosomal and vesicle transport, and neuroinflammation. In addition, neuronal loss in the disease may also occur in other brain regions besides the SNpc such as the amygdala and the hypothalamus. The loss of nigrostriatal dopaminergic cells causes a deficiency in dopamine, resulting in an imbalance of the complex system of direct and indirect pathways trough the basal ganglia (7). Simplified, in a normal situation the activation of the nigrostriatal projection by dopamine leads to opposite but synergistic effects with an increased responsiveness of the direct pathway (through D1 receptors) and a decreased responsiveness of the indirect pathway (through D2 receptors). This leads to a decreased inhibitory outflow (disinhibition) from the basal ganglia and a subsequent increased excitability of upper motor neurons, facilitating movements. In PD, the reduction of dopamine results in an abnormally high inhibitory outflow from the basal ganglia and a lower excitability of upper motor neurons which is reflected in the symptoms occurring in the disease, such as bradykinesia (10).

Less visible but common non-motor symptoms also occur in the disease. These include for example cognitive impairment, olfactory dysfunction, orthostatic hypotension, depression, and sleep disorders such as RBD. These non-motor symptoms can occur years before the onset of the classical motor symptoms in the so-called prodromal phase of the disease (Figure 1) (9). PD frequently manifest with highly variable symptoms. However, recognizable clinical subtypes can still be observed among groups of individuals where some symptoms coincide. These subtypes can for example be a mild-motor-predominant subtype where individuals more commonly have a younger age at diagnosis (AAD), a tremor-dominant subtype, or a rigidity-dominant subtype (11-15).

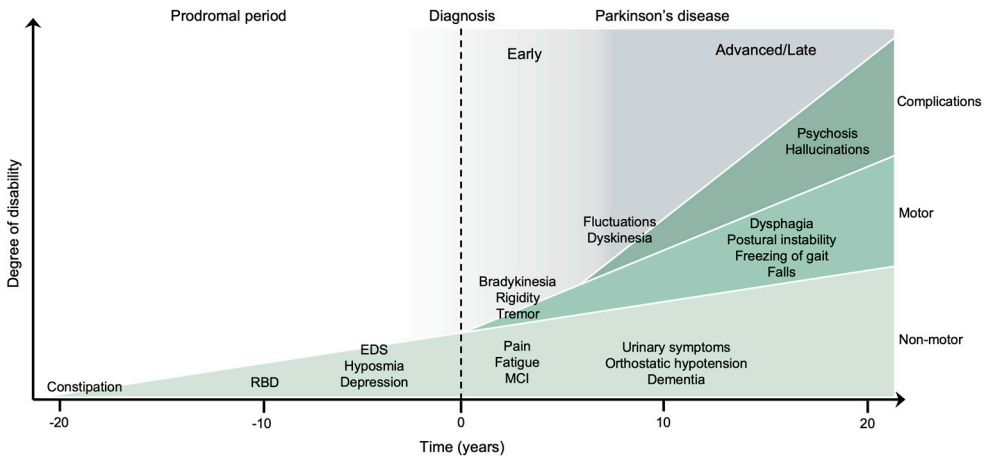

Figure 1. Clinical symptoms and time course of Parkinson's disease progression. The diagnosis of PD occurs at the onset of motor symptons (time 0 years). The disease is preceded by a prodromal phase which can occur decades prior to diagnosis and includes non-motor symptoms such as constipation and RBD. Additional motor and non-motor symptoms develops years after diagnosis as the disease progresses, leading to disability. Complications of long-term dopaminergic treatment such as fluctuations and dyskinesia also contributes to disability. EDS: Excessive daytime sleepiness, MCI:mild cognitive impairment, RBD: REM sleep behaviour disorder. Adapted from *Kalia et al.* (2015) with permission from Elsevier (license number: 5461980440265)

The heterogeneity of Parkinson's disease

'...every person has their own unique Parkinson's disease.'
- Bloem et al. 2021 (7)

PD does not appear to exist as a single entity. Instead, it can be argued that multiple subtypes or sub-diseases are combined in the term PD, resulting in a similar but heterogenous clinical syndrome. Supporting this is the fact that the symptoms seen in the disease are highly variable, the age at onset (AAO) varies, as well as the pattern of disease progression (7, 16). In 2008, a review entitled "There is no Parkinson's disease" was even published, arguing for the difficulties of having a single name for such a heterogeneous syndrome (17). A new conceptual model, named "the PD mountain range model" was proposed by Farrow *et al.* in 2022 with the aim to capture the multiple phenotypes that collectively form PD (Figure 2) (18). In the model, each phenotype that can be observed within the spectrum which is PD is represented by a mountain within a mountain range (with the number of mountains limited for simplicity). An individual's genetic PD risk is then represented by the position in the valley (i.e., "basecamp") and marks where the individual starts climbing and indicate that the route(s) that can be taken are limited depending on your genetic predisposition. Non-genetic factors (such as environmental factors), and age interact with the individual's genetic factors to alter

aspects of the disease, including if, and/or when an individual will develop the disease, which is indicated in the model as at the time an individual begins climbing. The topology of the valley floor represents the variation in an individual's risk, combining factors such as genetics, sex, and age. The topology of the mountain represents extrinsic and intrinsic factors that influence how quickly an individual climbs the route (being the rate of progression) and thus the type and severity of symptoms. The existence of multiple routes to each mountain peak represents that the individual diseases that together comprise PD are heterogenous in and of themselves. The routes are not independent, but they are both merging and diverging, representing that an individual can take a different route depending on its specific combination of intrinsic and extrinsic factors. Additionally, each route has different "checkpoints", representing various biomarkers that provide a snapshot of the disease and that potentially can be used to track disease development within individual patients. However, these can change over time and course of the disease and can potentially be influenced by various factors (18). The model treats PD as a group of diseases with different but overlapping etiologies and is a relatively accurate model since it takes the complexity of PD into consideration.

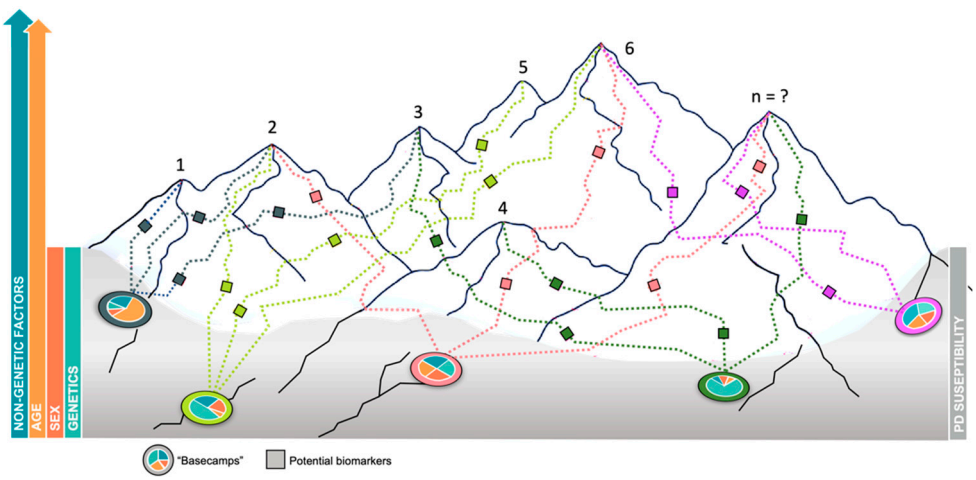

Figure 2: Conceptual model assimilating the different diseases within Parkinson's disease (PD) to mountains within a mountain range. The topology of the valley floor represents the suseptibility to PD based on interactions between non-genetic factors, age, sex and genetics. An individuals risk is represented by the "basecamps" in the valley, indicating that different factors can play a larger or smaller role for an individuall's PD suseptibility. Exposure to non-genetic factors and increasing age interacts with an individual's suseptibility, affecting when (age at onset) and how fast they climb (disease progression). The topology of the mountain affects how quickly each individual climbs the route and each mountain represents a phenotype of the disease (likely there are more than indicated here). The small boxes along the routes of ascent represent potential biomarker. However, these biomarkers likely change over the course of the disease.

Prodromal Parkinson's disease

The existence of stages or phases of PD prior to the PD diagnosis have been described, starting with a risk phase in which genetic and non-genetic factors modulate the risk of developing the disease in susceptible individuals. The subsequent preclinical phase is characterized by the initiation of progressive neurodegenerative pathology before any clinical symptoms or signs are evident. This is followed by the prodromal phase which is defined by the emergence of symptoms due to an underlying dopaminergic neuronal dysfunction and neurodegeneration (19, 20). These prodromal symptoms of PD can be difficult to recognize as associated with a neurodegeneration and a subsequent PD diagnosis. In other NDD, the prodromal phase can be identified by abnormalities similar to the characteristic symptoms later in the disease, such as that mild cognitive impairment is the primary prodromal marker of AD in which individuals are affected with cognitive symptoms and dementia (21). However, for PD, the heterogeneity observed in the disease starts already in the prodromal phase where the onset and progression of symptoms can differ tremendously between individuals. The earliest symptoms of PD have reported to occur on an average 10 years prior to a PD diagnosis but can start as early as 20 years before the onset of motor symptoms (22). Common symptoms that are seen in the prodromal phase includes e.g., constipation, hyposmia (the reduced ability to smell), depression, and RBD (7). In 2012, the International Parkinson and Movement Disorder Society (MDS) Task Force on the definition of PD was created to clarify challenges to our current definition of PD (20) and three years later, the MDS presented their criteria for prodromal PD (23). Due to ethical reasons with the lack of any clear neuroprotective or disease-modifying therapy in PD, the criteria were developed for research purpose only and not as diagnostic criteria.

However, identifying individuals in the prodromal phase of PD is of importance in the selection of appropriate individuals for inclusion in clinical trials of experimental disease-modifying therapeutics since these hopefully would delay or even prevent PD when administered at an early stage (7). The clinical trials conducted that have focused on putative disease-modifying interventions in patients with "early" PD have failed, despite beneficial preclinical results. The reasons for the failures are several, including ineffective drugs or an insufficient understanding of the biological mechanisms in the disease. However, another reason behind the failures could be that the definitions of early PD are inadequate and describes relatively advanced PD (20). Most prodromal symptoms are relatively non-specific for PD and a large majority of individuals having these symptoms will never develop the disease. A noticeable exception is idiopathic or isolated RBD (iRBD), which is the prodromal symptom with the highest risk of subsequent phenoconversion to PD (24).

Rapid-eye movement (REM) sleep behavior disorder (RBD)

RBD is characterized by dream enactment and an excess in muscle tones and/or phasic muscle twitching during REM sleep (25). It can be categorized into iRBD and symptomatic (or secondary RBD, sRBD) forms. iRBD is defined as having RBD without other significant clinical neurological signs whereas sRBD is mainly associated with synucleinopathies (25). Observational studies have suggested that most patients with iRBD eventually will develop a NDD, most commonly an α-synucleinopathy. A multi-center study from 2019 investigated the disease conversion among 1,280 individuals with iRBD and saw an overall conversation rate from iRBD to a NDD of 6.3% per year, and that 73.5% of the individuals had converted after the 12-year follow-up (24). Among PD patients, 30-60% are estimated to have RBD. This can either be that they have had iRBD prior to the PD diagnosis, or that they develop sRBD after the diagnosis. RBD is also associated with a more rapid progression of non-motor symptoms in PD (24). In addition, RBD has been reported to be a strong predictor for the development of dementia in PD, where individuals with PD and concomitant RBD showed significantly poorer performance on standardized tests measuring cognitive functions compared to patients with PD without RBD (26).

Treatments in Parkinson's disease

Due to the heterogeneity of PD, it would be an ideal disease for precision medicine in which various treatments could be tailored to match the priority, need, and/or genetic and molecular setup of each individual (7). However, available therapies for PD are rather uniform and only treat symptoms of the disease. A major goal of PD research is therefore the development of disease-modifying treatments that can slow down or even stop the underlying neurodegenerative process and thereby the progression of the disease.

The pharmacological treatments available for PD are primarily dopamine based to enhance the intracerebral dopamine concentrations which primarily improve motor symptoms. These include the dopamine precursor levodopa (given together with a dopa decarboxylase [DDC] inhibitor to prevent peripheral decarboxylation of levodopa to dopamine), dopamine agonists, monoamine oxidase type B (MAO-B) inhibitors, and catechol-*O*-methyltransferase (COMT) inhibitors. MAO-B and COMT inhibitors prevent the degradation of dopamine and thereby prolongs its effect (27). As the disease progresses, medication tolerance and loss of efficacy of levodopa treatment occur and therefore, more frequent, and higher levodopa doses over time are required. Following several years of dopaminergic pharmacotherapy, fluctuations in the treatment response typically develops (9).

For individuals with PD who have complications such as "off periods" or dyskinesias, and when oral treatments are not giving an adequate therapeutic effect,

more advance treatment options are available such as deep brain stimulation (DBS), magnetic resonance imaging (MRI)-guided focused ultrasound, and therapy with levodopa-carbidopa enteral suspension. “Off periods” is the worsening of symptoms when a drug dose wears off, whereas dyskinesia is the occurrence of involuntary muscle movements which commonly can occur when the concentration of levodopa is at its maximum (i.e., peak-dose dyskinesia), or less commonly, at the beginning or the end of a levodopa dose (or both) (9).

The surgical treatment DBS is a well-established advanced treatment for motor symptoms in PD. In DBS, wires are surgically placed unilateral or bilateral in the brain regions of the subthalamic nucleus or the globus pallidus interna. The wires are then attached to a battery in the chest, similar to a pacemaker battery. DBS can be used in individuals with PD experiencing problems with medication-resistant tremor, dyskinesia, or complicated off periods. The MRI-guided focused ultrasound uses ultrasound beams to burn a target, being the thalamus, and has shown to have a beneficial effect on tremor. The procedure is done unilaterally to minimize the risk of adverse events such as worsening of speech and balance. However, both of these methods are difficult and invasive surgeries, and are only able to treat a subset of the motor symptoms in PD. Levodopa-carbidopa enteral suspension is a levodopa and carbidopa (DDC inhibitor) gel which is administered continuously via a pump through a percutaneous endoscopic transgastric jejunostomy, resulting in a more stable levodopa concentration compared to oral administration (27).

Abnormalities in other neurotransmitters than dopamine also contribute to symptoms in PD and other pharmacological therapies therefore include drugs that target acetylcholine, serotonin, norepinephrine, glutamate and/or adenosine. The symptomatic treatments for non-motor symptoms in PD are similar to treatments for these symptoms in the general population, where for example, selective serotonin reuptake inhibitors may be useful for treating depression (27).

New and more effective treatments are needed in PD that can halt disease progression or potentially regenerate or replace the dopaminergic neurons. Stem-cell based therapies for PD, primarily aiming to restore lost motor functions by replacing the lost dopaminergic neurons are under development, showing great potential in early PD clinical trials (28).

Epidemiology

Epidemiology refers to i) the study of the distribution of disease in human populations and ii) the factors determining that distribution. In this section, the focus will lie on the first of these parts of the definition, while the second part will be discussed in more details in the section of non-genetic risk factors in PD.

The frequency of a disease in a population can be measured as incidence, which refers to the number of individuals diagnosed with a disease in a specified period,

or as prevalence, being the proportion of persons currently diagnosed with the disease at a given time, regardless of when they were diagnosed. The disease prevalence is nonetheless influenced by the incidence since it is affected by both the number of newly diagnosed but also by the survival of previously diagnosed individuals. Considering a stable diagnostic efficiency, the incidence more directly reflects the effect of risk factors for a disease, as compared to the prevalence.

PD is an age-related disorder, and it is commonly stated that PD affects approximately 1% of individuals over the age of 60 years. In 2015, PD was reported by the Global Burden of Disease consortium (GBD) to be the fastest growing neurological disorder in prevalence, disability, and deaths (3). Interestingly, the increase in prevalence was not solely due to a growing elderly population in the world. Standardizing for age, a 21.7% increase between 1990 and 2016 was observed for the prevalence rate of PD (5). The same study also observed that the global prevalence of PD had more than double between 1990 and 2016, with 2.5 million individuals affected in 1990, as compared to 6.1 million in 2016. This was partly due to a growing elderly population and longer disease duration, but also likely due to other factors such as non-genetic factors (5). In a systematic review and meta-analysis from 2014, the estimated worldwide prevalence of PD was approximately 315 per 100,000 individuals from the age of 40 years. The prevalence of the disease increased steadily with age with the highest reported prevalence of 1,903 per 100,000 individuals at the age of 80 years or older (29).

The disease prevalence also differs between countries having a high socio-demographic index (SDI) compared to countries having a low SDI, with a higher prevalence in high SDI countries (Figure 3). SDI is a measurement of a countries development status that uses several variables including income per capita, education, and fertility to provide a better estimate compared to binary descriptions such as developed and developing countries (30). The higher prevalence could be explained by a better ascertainment of PD in high SDI countries through e.g., better health-care access and disease recognition, along with a longer life expectancy. However, the higher PD prevalence in high SDI countries could also be linked to higher exposure to environmental factors tied to industrialization in these countries which are associated with PD risk. In fact, the country with the largest increase in age-adjusted prevalence rates between 1990 and 2016 was China, a country which underwent a rapid industrial growth between these years (5). Applying the same estimate for the increase in the prevalence of PD as the one observed between 1990 and 2016 would result in almost 13 million affected individuals by 2040, meaning that the future burden of PD will increase substantially (5, 31).

High-quality estimates of PD incidence are typically more difficult to accomplish compared to prevalence estimates, but they are of great importance for several reasons, including the understanding of disease risk and prevention, and anticipating the healthcare need and capacity (32). The worldwide incidence of PD in 2019, measured as the age-standardized rate, was estimated to approximately 13.4 per

100,000 individuals which was an increase by almost 160% since 1990 (33). The largest increased was observed in the United States of America (USA) and Norway with an estimated annual percentage change of 2.9 respectively 2.1 (33). As for the prevalence, the incidence of PD clearly also increases with age. For example, a recently published study investigated the incidence of PD in the year of 2012 across five epidemiological cohorts in North America of individuals aged 45 years and older and estimated that the age and sex adjusted incidence of PD ranged from 47 to 77 per 100,000 among individuals aged 45 and older, and 108 to 212 per 100,000 among individuals from 65 years and above (32).

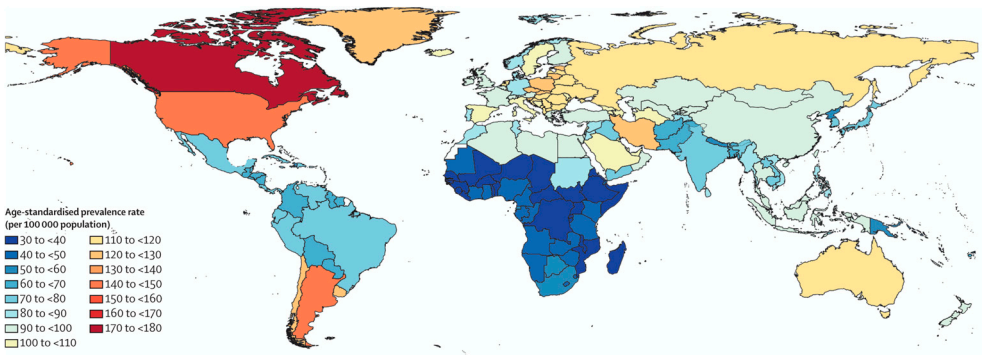

Figure 3: Age-standardised prevalence of Parkinson's disease per 100 000 population by location for both sexes, for year 2016. Adapted from the Global Burden of Disease (GBD) 2016 Parkinson's Disease (PD) Collaborators systematic review on the global, regional, and national burden of PD (5) under the terms of the Creative Commons CC-BY licence.

However, it is important to note that the disease does not exclusively affect older individuals and it is stated that approximately 3-5% of PD patients have an early onset PD (EOPD). EOPD refers to individuals with an AAO younger than 50 years (34), (although this age varies in different publications). A genetic etiology is more common among individuals with EOPD as compared to later onset PD (35-37). EOPD can be further subdivided into young-onset PD (YOPD), which is arbitrarily defined as PD onset between 21 and 40 years of age, and juvenile parkinsonism, which is when onset is before the age of 21 years (34).

Although PD affects both sexes, the disease appear to affect men more frequently and the male sex is often stated to be one of the main risk factors for the disease. To what extent although differ between studies. In the GBD study, the prevalence was reported to be 1.4 times higher in men than in women (5). A meta-analysis from 2014 observed that when stratified by age, a statistically significant difference in prevalence between the sexes was only seen for the age group 50-59 years with a prevalence of 134 per 100,000 among men vs 41 per 100,000 for women. Even if a slight male dominance was observed for the majority of the other age groups, there was no significant difference between the sexes (29). A more recent meta-analysis

also highlighted the fact that the sex differences in PD prevalence might not be as pronounced as previously thought. An overall male/female prevalence ratio of 1.18 was observed (38) as compared to previous reports of a ratio of 1.5 (39) or even 2 (40). The known associations between PD and the different sexes will be further discussed under the section of non-genetic risk factors in PD.

Adding on to the complexity of PD is the disease etiology, where numerous factors are involved, and research is still needed to understand these factors and to untangle their connections. However, when discussing PD etiology, it is known that three main aspects are highly relevant: genetics, non-genetic factors (e.g., environment and lifestyle), and their interactions (7).

Parkinson's disease genetics

'The longest-standing question in genetics is to understand how genetic variation contributes to phenotypic variation'
(Boyle et al 2017) (41)

Our understanding of human genetics owes much to the work of Gregor Mendel (1822-1884) whose studies of inheritance patterns provided a foundation for understanding monogenic diseases, also called Mendelian diseases. Pathogenic mutations following Mendelian inheritance can follow a dominant pattern, where the effect of a mutation on one allele is sufficient to cause disease. Alternative, it follows a recessive pattern, where disease occurs only in individuals with two mutant alleles (the same - homozygous, or different - compound-heterozygous), most commonly with a high penetrance. Penetrance is defined as the probability that individuals carrying a disease-associated genotype will exhibit a disease-associated phenotype. Penetrance is not to be confused with expressivity, which refers to the degree of variation of the phenotype in those individuals carrying the same disease-associated genotype (42). However, most diseases are complex, where associations can be observed between the disease and common variants such as single nucleotide polymorphisms/variants (SNPs/SNVs), spread across most of the genome, including near genes without an obvious connection to the disease (41).

Historically, the etiology of PD was considered to be sporadic and before the 1990s, heredity was not believed have a significant role in PD (43). However, we now know that genetics is highly involved in both PD etiology and progression. A family history of PD (defined as having a first degree relative with the disease) is seen in approximately 10-15% of PD patients (44). The genetic contributors to PD exist across a continuum, ranging from causal and highly penetrant genetic variants to common variants with small effect sizes that individually only increase the disease risk slightly (Figure 4). Although simplified, the genetics in PD etiology is therefore often divided into two categories: monogenic and idiopathic (or sporadic) PD.

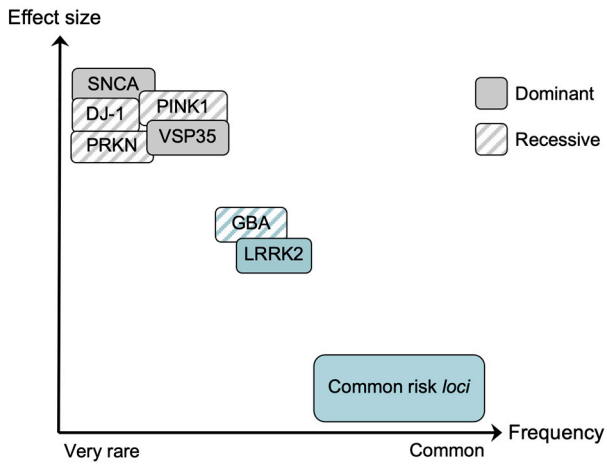

Figure 4. Schematic overview of effect sizes and frequencies of genetic variants in known Parkinson's disease (PD) genes. *GBA* is concidered a recessive gene since homozygous mutations in *GBA* cause Gaucher's disease whereas heterozygous mutations are associated with increased PD risk.

Monogenic Parkinson's disease

Monogenic PD is commonly stated to affect 10-15% of all individuals with PD. However, this is likely an overestimation and the most common mutations associated with PD are though to explain 3-5% of PD cases in most populations (45). The gradient between monogenic and idiopathic PD makes it difficult to estimate the accurate fraction of PD patients with monogenic PD (45). Adding on to the complexity of PD etiology, the disease will still not manifest (referred to as incomplete penetrance) for some individuals carrying disease causing mutations. Also, the expressivity can vary, i.e., the AAO, clinical presentation, and progression, even among carriers of the same mutation and within the same family (46). Seven genes - *SNCA, LRRK2, VPS35, PRKN, PINK1, GBA,* and *DJ* - have been convincingly associated with typical PD. However, rare variants in more than 20 genes have been reported to cause PD but their relevance is debated and should be considered putative until replication and/or validation studies can confirm their connection to PD (46).

SNCA

The first discovery of disease-causing mutations in PD was reported in 1997 when a missense variant (resulting in a different amino acid being encoded) in the gene *SNCA* was identified in a large Italian kindred and three unrelated Greek families (47). *SNCA* encodes for α-syn and mutations in the gene are associated with autosomal dominant PD. Around the same time as the identification of *SNCA* mutations in PD etiology, α-syn was found to be the major constituent of Lewy bodies, collectively identifying *SNCA* and α-syn as key factors in PD etiology and

pathology (48). Subsequently, triplications of the *SNCA* locus were reported in a kindred with familial PD (49-52), demonstrating that both *SNCA* mutations and copy number variants (CNV; duplications, triplications) can lead to PD (53, 54). The individuals with PD carrying four copies of *SNCA* rather than the normal two, were also found to have a corresponding doubling of α-syn mRNA in postmortem brain tissue (55). It has been demonstrated that the gene dosage of α-syn has an impact on disease initiation and severity of progression (54). This is also reflected in the AAO, where triplication carriers have an earlier AAO at an average of 34.5 years (± 7.4) (56). Individuals having duplications have been reported to be phenotypically similar to individuals with idiopathic PD, although with an average AAO of 47 years (± 10.5) (56-58). *SNCA* CNV are rare and have been reported in approximately 60 families to date (56, 59). An interesting study in the UK Biobank, which is a large-scale population prospective study that includes around 500,000 individuals, identified six individuals carrying *SNCA* duplications (59). However, none of these individuals were reported to have PD, parkinsonism, or dementia. It cannot be ruled out that these individuals could develop PD at a later stage, although the average age of recruitment for the six carriers was 61 years (range 49-69) (59). These individuals are interesting candidates for follow-up studies and more thorough investigation of *SNCA*, i.e., general *SNCA* biology, and how these duplication carriers potentially escape the development of PD.

Pathogenic missense variants in *SNCA* are even more rare than CNV in the general population. Five mutations in the gene (A30P, E46K, G51D, A53E and A53T) have been reported as disease-causing mutations following an autosomal dominant pattern. Carriers most commonly have a more rapid disease progression and earlier AAO compared to idiopathic PD (59). Despite the apparent genetic and pathologic link between *SNCA* and PD, the cellular function of α-syn is still not fully understood. The protein is predominantly and ubiquitously expressed in the brain, but it has also been detected in red blood cells and other tissues throughout the body. The presynaptic location of α-syn strongly suggest a regulatory function associated with the synapse and its interactions with lipid surfaces is believed to be a key mechanism for mediating its cellular functions (60). Multiple animal models of PD are based on e.g., overexpression or intracerebral injection of α-syn, resembling the pathological hallmarks of PD with neuronal loss and parkinsonian phenotype, strengthening the pathological role of α-syn in PD (61).

LRRK2

Mutations in the gene *LRRK2*, encoding the leucine-rich repeat kinase 2, are the most frequent known cause of PD. *LRRK2* mutations act in an autosomal dominant matter and a minimum of six disease-causing mutations have been described (N1437H, R1441C/G/H, Y1699C, I2020T, and G2019S). The G2019S mutation accounts for almost 90% of the *LRRK2* mutations and is reported to occur in 1% of PD patients with idiopathic PD and 4% of patients with familial PD (62). Frequency

studies of the G2019S mutation have revealed population specificity with wide variations, ranging from being rare in Asian countries with a carrier frequency among PD patients of around 0.1% to as high as 14-19% among Ashkenazi Jews and 36-42% in familial North-African Arab PD patients (63-66). The frequency also varies even within Europe, with higher frequencies in southern European countries such as Italy, Spain, and Portugal compared to countries in northern Europe (63). The penetrance estimates of the *LRRK2* G2019S mutation also varies widely between populations and is difficult to study due to the occasional occurrence of the mutation in asymptomatic carriers. The penetrance of the mutation at age 80 (in non-Ashkenazi Jewish carriers) has been reported to be approximately 43%, however with a wide range from 26% to 66% (67).

The mean AAO among *LRRK2* mutation carriers is approximately 58 years and the cumulative risk of PD among carriers increase substantially with age. For *LRRK2* non-Gly2019Ser mutations combined, the cumulative PD risk was 40% at 59 years and 84% at 79 years. Among *LRRK2* G2019S carriers, the risk was slightly lower with 28% at 59 years and 74% at 79 years (62). In general, the PD phenotype among *LRRK2* carriers resembles that in idiopathic PD. However, *LRRK2* G2019S carriers have a lower risk of cognitive impairment and hyposmia compared to patients with presumably idiopathic PD (62).

The LRRK2 protein is a kinase and GTPase consisting of multiple protein-protein interaction domains and mutations in the gene tend to cluster within these catalytic domains (68). For example, the G2019S mutation is in the kinase domain and the mutation result in a higher kinase activity (68, 69). The different *LRRK2* mutations result in different defects in the activation mechanism of the protein although the common output appears to be increased kinase activity and reduced GTPase activity, resulting in increased LRRK2 signaling (70). The increased signaling is thought to induce neurodegeneration by interfering various molecular processes, including the endolysosomal pathway and vesicular trafficking (68). Interestingly, the LRRK2 activity has also been found to be enhanced in postmortem brain tissue from PD patients with an apparent idiopathic PD (71). LRRK2 has emerged as a promising target for potential disease-modifying treatments in PD, targeting the kinase activity of the protein. Clinical trials with small-molecule *LRRK2* kinase inhibitors are ongoing, were at least one had reached phase III by the end of 2022 (70, 72).

VPS35

In 2011, the D620N (Asp620Asn) mutation in the gene *VPS35* (encoding vacuolar protein sorting 35 retromer complex component), was separately identified to cause PD in an Australian kindred (73) and in a Swiss kindred (74). The heterozygous mutation was found in all affected family members and in a few unaffected, suggesting an autosomal dominant inheritance with incomplete penetrance. Since 2011, the mutation has been confirmed in independent studies to cause autosomal

dominant PD (75, 76). Other identified missense variants have been reported to be associated with PD, but their pathological role still remain uncertain (77). Mutations in the *VSP35* gene appear to be a rare cause of PD. However, PD patients carrying the mutation have been reported to be phenotypically similar to individuals with apparent idiopathic PD which could mean that a number of patients with *VPS35* mutations are being overlooked. The VPS35 protein is a component of the retromer cargo-recognition complex which is critical for endosome-trans-golgi trafficking and membrane-protein recycling, which supports the hypothesis that vesicle formation and trafficking is involved in the pathophysiology of PD (77).

Parkin

Autosomal recessive homozygous or compound heterozygous loss-of-function mutations in PD have been identified and replicated in three genes, *PRKN* (also referred to as *PARK2* or *Parkin*), *PINK-1*, and *DJ-1*. Mutations in these genes are rare but appear to be responsible for a substantial proportion of YOPD (78). Already in 1998, mutations in *Parkin* were reported to cause autosomal recessive juvenile parkinsonism (79). In a study of EOPD (here defined as an AAO below 45 years), 77% of those with juvenile PD carried at least one mutation in *Parkin*. Among the individuals with an AAO between 20-30 years, 26% carried a mutation whereas 2-7% for age group 30-45 years where carriers (80). In a systematic review, 958 *Parkin* mutation carriers were identified having a median AAO of 31 years where around a quarter had an AAO over 40 years (81). More than 135 disease-causing variants have been reported in *Parkin* where homozygous mutations are most common, most likely followed by compound-heterozygous individuals carrying at least one pathogenic variant (81). Aside from the early AAO, PD patients with *Parkin* mutations are often characterized by a good response to dopamine replacement therapy and a low risk for non-motor symptoms such as cognitive decline (82). Reasonably, it has been suggested that heterozygous pathogenic *Parkin* mutations increase the risk of PD. However, these mutations have shown to be common in the general population and that they do not appear to increase PD risk (83). *Parkin* encodes the Parkin RBR E3 ubiquitin protein ligase which plays an important role in mitochondrial quality control and turnover. Along with other mechanisms, the Parkin protein ligase has been shown to be recruited from the cytosol to the outer mitochondrial membrane in impaired mitochondria to mediate mitophagy (the selective autophagic removal of the damaged organelle) (84, 85).

PINK1

PINK1, encoding the PTEN-induced putative kinase 1, has been shown to regulate the translocation of Parkin and hence mitophagy when accumulated at the outer mitochondrial membrane (85). Mutations in *PINK1* were first identified in three European consanguineous EOPD families where all affected family members were homozygous carriers (86). Mutations in *PINK1* are now recognized as the second most common recessive gene associated with monogenic PD. The frequency of

PINK1 mutations have been reported to be 1-9% among PD patients where a monogenic cause is suspected (such as for patient groups with a young AAO and positive family history) in different ethnic groups (45). In a systematic MDSgene review from 2018 evaluating 46 *PINK1* studies in PD, 62 different disease-causing variants were reported in the gene, the most common being the missense mutation Leu347Pro (81). Among the carriers, 83% were homozygous and 17% were compound heterozygous. The majority, 62%, had an EOPD and 15% even had juvenile PD. *PINK1* mutations were also associated with late onset PD in 22% of the patients, though the median AAO was 32 years among carriers (81). It has also been postulated that heterozygous mutations in *PINK1* increase the risk of PD (87). However, a large subsequent study did support these findings but did not exclude that heterozygous variants in *PINK1* have a minor effect on PD risk (88).

DJ-1

The third gene associated with autosomal recessive PD is *DJ-1* (also known as Parkinson disease protein 7, PARK2). Both pathogenic point mutations and structural mutations (Q163L, L166P, exon 1–5 deletion, and g.168-185dup) in the gene was first identified in two consanguineous families in genetically isolated communities in Italy and the Netherlands (89). Mutations in *DJ-1* are less common in EOPD compared to *Parkin* and *PINK1*. As for the other two recessive genes, the AAO is low among carriers and the median AAO for *DJ-1* mutation carriers has been reported to be 27 years. Hence, a majority of carriers have an EOPD and only 4% were reported in the MDSgene review to have had a late AAO (81). In the same study they identified 20 disease-causing variants where 93% of the patients were homozygous carriers, and 7% compound-heterozygous (81). The DJ-1 protein acts as a sensor and protector from oxidative stress. Therefore, it is highly expressed in cells with higher levels of reactive oxygen species, being cells with high energy demand. It therefore likely act in similar pathways as *PINK1* and *Parkin* and the absence of *DJ-1* has been seen to alter mitochondrial morphology (90, 91).

GBA

It could be stated that one of the most important genes in PD is *GBA* (also called *GBA1*) since between 5-20% of PD patients in different populations carry variants in the gene (92). *GBA* is located on chromosome 1 (1q21) and encodes glucocerebrosidase, a lysosomal enzyme involved in the metabolism of glucosylceramide. Recessive mutations in the gene cause Gaucher's disease (GD), a lysosomal storage disorder with three clinical types: nonneuropathic (type I), acute neuropathic (type II), and chronic neuropathic (type III) (93). Depending on the type of GD a mutation is causing, the mutations can be categorized as mild or severe. Mild mutations are those causing GD type I whereas type II and III are caused by severe mutations (94). GD can manifest early in childhood with a combination of symptoms such as severe neurological manifestations, with brainstem involvement (i.e., spasticity and hypotonia) but it can also remain undiagnosed until adulthood

when the phenotype is mild. Among the many phenotypes associated with the disease are patients with parkinsonian symptoms (95). Today, we know that approximately 300 *GBA* mutations have been described in GD, and many of them have also been reported to be associated with PD risk (92). Heterozygous carriers of the severe type of *GBA* variants have a higher risk of PD and an earlier AAO compared to milder mutations carriers (92), and a faster cognitive decline (96). However, most carrier will not develop PD since the penetrance of heterozygous *GBA* variants is reduced and ranges between 8% to 30% in different populations, increasing with age (97-99).

Other genes

Known monogenic loci have been reported to only explain a fraction of the observed familial aggregation of PD (44). In addition to the previously described genes, several additional genes have been reported as putative disease-causing genes in PD. A study evaluated the probability of these genes to be disease-causing in PD, taking into consideration the number of families reported and if the finding has been replicated (46). The three genes *ATP13A2*, *FBXO7*, and *PLA2G6* were then reported with very high confidence to be disease-causing. Four additional genes (*POLG*, *DNAJC6*, *SYNJ1,* and *VSP13C*) were reported to be disease-causing with high confidence, while a numerous genes were reported with low confidence (46). Even more genes have been nominated as putative disease-causing genes, including *RIC3.*

A missense variant (P57T) in the gene *RIC3* was identified in a large Indian family showing autosomal dominant inheritance (100). In addition, another rare heterozygous missense *RIC3* variant, V168L, was found in the same study through targeted screening in an independent Indian PD case-cohort study (100). The *RIC3* gene encodes for a member of the resistance to inhibitors of cholinesterase 3-like family. and functions as a chaperone for nicotinic acetylcholine receptors (nAChRs), specifically the alpha-7 subunit of homomeric nicotinic acetylcholine receptors (CHRNA7). nAChRs are widely expressed in the central nervous system and regulates processes crucial for network operations and influence physiological functions such as cognitive functions (101). Stimulation of nAChRs in the basal ganglia has been suggested to result in functional alterations at a cellular level that may include the control of locomotor activity and protection against nigrostriatal degeneration (102). The function hence supports a potential involvement in PD pathology and Sudhaman *et al.* showed that the identified mutations in *RIC3* reduced the level of CHRNA7 in mutant cell lines (100). However, the association with PD have not been replicated. When investigating the variants in French-Canadian and French cohorts (103) and a Chinese cohort (104), no association with PD was found. This was further investigated in this thesis by investigating *RIC3* in large European, Latin American, and East Asian cohorts in **Paper II**.

Idiopathic PD

At this point in the chapter, the complexity of PD has hopefully come across. Adding on to the complexity is the vague and not fully correct terms idiopathic, or sporadic, PD. The implication that PD would arise with an unknown cause (as idiopathic suggest), or spontaneously (as sporadic suggest) is inaccurate. However, for simplicity, these terms will be used to describe the disease when no monogenic cause is known.

A large proportion of individuals with PD have a genetic susceptibility for PD through common genetic risk factors. Genetic risk factors for a disease are commonly identified using genome-wide association studies (GWASs). A GWAS is a hypothesis-free method to investigate whether genetic variants are associated with a particular trait or phenotype. The "GWAS era" is commonly stated to have started in 2005 with a small study of age-related macular degeneration in which an intronic and common variant in the complement factor H gene (*CFH*) was identified to be associated with the disease (105). The first GWAS in PD was done in 2009, consisting of 5,074 cases and 8,551 controls of European origin in which two genetic risk loci for PD was identified in *SNCA* and in *MAPT* (microtubule-associated protein tau) (106). Prior to the GWAS area, genetic associations to a disease were often studied using a candidate gene approach. Using this approach, variants in *SNCA*, *LRRK2*, *MAPT*, and *GBA* had already before the first PD GWAS been reported to be important PD susceptibility factors (107-115).

The largest PD GWAS meta-analysis conducted as of today was done in 2019 and included data from more than 37,000 patients, 17,000 proxy cases (individuals with a parent with PD) and 1.4 million controls (116). Approximately 7.8 million SNPs were analyzed, and 90 independent genome-wide significant risk variants were identified as associated with PD risk across 78 genomic loci. Using polygenic risk scores (PRS), it was estimated that these 90 variants explained 16–36% of the heritable PD risk (depending on prevalence), indicating that a considerable large part of the genetic component in PD remains unidentified. Several of the identified loci are near genes associated with monogenic PD, such as *SNCA*, *LRRK2*, *GBA*, and *VPS13C* (116). These genetic regions can be called pleomorphic risk loci, being risk loci at which both rare and common genetic variants contribute to a specific disease or trait (117). Some of the identified risk loci are also close to genes that are associated with other diseases, including the genes *MAPT*, *GRN*, and *NEK1* which are associated with frontotemporal dementia and amyotrophic lateral sclerosis (ALS) (118), strengthening the theory that PD is not a single entity.

Genetic variants associated with PD expressivity

Along with the identification of factors contributing to PD risk, the research on factors contributing to altered expressivity (modifiers) is of great interest. Despite the higher incidence of PD with increasing age, the AAO is variable. An earlier AAO is an essential contributor to the overall disease burden of PD since the prevalence is affected by both the incidence and the disease duration. Hence, if the AAO could be postponed, the prevalence and burden of PD would decrease, and if postponed long enough, some individuals might never even develop PD. The burden of PD, both on an individual and societal level, could also be affected in a positive matter if the progression would be haltered. Therefore, it is of great interest to identify genetic and non-genetic modifiers of both AAO and PD progression. In this section, the focus will be on giving a short description on what is known regarding genetic modifiers of PD expressivity.

There is an overlap between the common variant associated with PD risk and AAO, meaning that they both contributes to increased PD risk but also to a younger AAO. A large PD AAO GWAS identified two loci associated with PD AAO, *SNCA* and *TMEM175/GAK* (119). As previously described, rare variants in *SNCA* are associated with monogenic PD whereas common variants are associated with idiopathic PD. The *TMEM175/GAK* locus was also associated with PD risk in the latest GWAS meta-analysis (116). A subsequent GWAS AAO meta-analysis could confirm the association between the *SNCA* and PD AAO, but the *TMEM175/GAK* locus did not reach genome-wide significance (120). However, the study identified another independent locus, *BST1*, as associated with PD AAO. The loci were associated with rather small effects on AAO. e.g., the *SNCA* locus was associated with an average delay of AAO of 0.57-0.72 years (119, 120). Genetic modifiers associated with PD progression (cognitive decline and/or progression of motor symptoms) have been reported to be variants in *GBA*, *SLC44A1*, *APOE*, and *RIMS2* (121-123). The variant in *APOE* driving the association was an *APOE4* tagging variant specifically associated with greater cognitive decline in PD patients. Interestingly, the *APOE4* allele is overrepresented in patients with diseases commonly related to a loss of cognitive function, i.e., DLB and AD (122).

As described previously, the phenotypic features of PD varies and subtypes have been described based on common patterns of symptoms (12-15). Suggestive associations between genetic variants and the different subtypes have been made, including an association between a variant at the *STK32B* locus and the subtype tremor-dominant PD (124). Another variant at the locus has previously been reported to be associated with essential tremor, suggesting a potential overlap between the genetic risk of tremor-dominant PD and essential tremor (125)

Non-genetic risk factors in Parkinson's disease

Non-genetic risk factors also play a vital role in PD etiology. Even if there is still a lot to learn regarding the genetic architecture of PD, the study of genetics has the advantage that the genome can be deciphered, and a person's genetic make-up is largely stable. In contrast, non-genetic risk factors such as the environment are for the most part constantly changing and even if an exposure is constant, the level of exposure is varying. The sum of all potential factors in the environment that modulate the risk of a disease can be referred to as the "environmentome". A similar word for factors that we are exposed to is the "exposome". Here, I will instead use the term non-genetic risk factors to include all factors not directly related to genetics, such as exposures, environment, lifestyle, and comorbidities. The contribution of non-genetic factors in PD etiology is more pronounced in idiopathic PD but it likely play an important role also in monogenic PD, influencing both penetrance and expressivity (126). Epidemiological studies rarely distinguish monogenic PD from idiopathic PD since they often rely on clinical diagnostic criteria. Therefore, this section will be focused on non-genetic risk factors for PD without further specification, if not stated otherwise. Numerous non-genetic risk factors have been proposed for PD where many were identified in observational retrospective case-control studies. These types of studies are prone to various forms of bias where observed associations could be a consequence of e.g., *reverse causation* (when the association is affected by disease-related changes in the level of exposure), *recall bias* (caused by false accuracy of the reports of past exposures), or *selection bias* (when there is an imbalance between the study groups affecting the outcome), making causal interference challenging. Although prospective cohort studies, where the assessment of an exposure occur before the onset of PD, often are a better type of study design, they may also be subject to bias such as reverse causality due to the prodromal phase of the disease (127). Nevertheless, several factors have been reported to be associated with PD even if the causality of them can be debated.

Sex and age as risk factors for Parkinson's disease

The two factors that often are described as the top risk factors for PD are advanced age and the male sex. As previously described, both the PD incidence and prevalence increase with age, and more men than women get the disease. It could be argued that both age and sex are genetically determined factors and that they therefore not should be included as non-genetic risk factors in PD. However, it is unclear whether the observed difference between the sexes is due to genetic or non-genetic factors, or a combination of the two. In fact, it has been suggested that the observed sex differences in PD is determined by genetic, environmental, and hormonal influences on the nigrostriatal dopaminergic system (128, 129). Additionally, sex-specific differences in the phenotype and progression of PD have

been observed between the sexes. For example, women often present with a more tremor-dominant PD and experience more symptoms such as anxiety and depression compared to men (129). To investigate if common genetic variants contribute to PD etiology on a sex-specific manner, a sex stratified GWAS meta-analysis was done, indicating that common genetic variation on the autosomes do not explain the observed sex differences (130). Moreover, no strong associations between variants on the Y-chromosome and PD has been observed (131). Interestingly though, an X-chromosome-wide association analysis identified two significant loci associated with PD at *GPM6B* and a high gene density region including *RPL10, ATP6A1, FAM50A*, and *PLXNA3*, indicating that the X-chromosome contributes to PD risk in both men and women (132). Estimating the effect of these loci on PD risk in men and women could be a difficult task considering the X-chromosome inactivation occurring in women during embryonic development. X-chromosome inactivation results in an activation ratio of both sex chromosomes of approximately 1:1 which means that an X-linked mutation causing complications in men can be compensated by the other X-chromosome in 50% of the cells in women (133). Although simplified, this would mean that heterozygous women carrying one risk allele on the X-chromosome are at lower risk compared to male carriers, whereas women being homozygous carriers would be at a similar risk. Among other factors, women and men have been reported to be differently exposed to various non-genetic PD risk factors, such as smoking. Despite inconclusive reports, a protective effect of estrogen in PD has also been suggested (134). Overall, the sex-related differences observed in PD further highlight the complexity of the disease and its etiology.

Lifestyle and dietary factors

An inverse association and potential protective effect of smoking on PD risk has been observed in numerous studied, which is surprising considering the increased risk of chronic diseases such as heart and lung diseases among smokers (135-137). The estimated risk reductions are large, ranging from approximately 30% to up to 60%. However, the effect estimates of smoking on PD risk is substantially smaller when analyzed using mendelian randomization (MR), reporting approximately a 5% lower PD risk among ever-smokers (138). MR can be used as a method to investigate the effect of an exposure on a specific outcome, using genetic variants associated with the exposure as instrumental variables, minimizing potential bias from reverse causality. Individuals who tend to smoke might do so due to higher dopamine levels, resulting in a stronger reward mechanism from smoking. This would also mean that a larger reduction in dopamine levels is needed before PD symptoms occur in these individuals (7). It has also been seen that individuals with PD were able to quit smoking more easily compared to controls, and individuals with a greater difficulty quitting smoking were less likely to develop PD (139). Moreover, passive smoke exposure has not associated with PD risk (140). These findings support that there might be a decreased responsiveness to nicotine during

the prodromal phase of PD. Additionally, an inverse association between other tobacco products such as the Swedish moist tobacco snus and PD risk has also been reported (141). This finding was also replicated in **Paper III** of this thesis. It cannot be ruled out that there is a protective effect of these products on PD risk, and it has been seen that nicotine have a neuroprotective effect in parkinsonian animal models by modulation of dopamine release by acting at nAChRs on dopaminergic nerve terminals (142, 143). Whether this also occur in humans is not known and clinical trials with transdermal nicotine patches have presented contradictory results when it comes to improving symptoms in PD (144-146). If the inverse association is truly causal, then the tendency to smoke less, which is seen on a global level, could partly explain the observed and expected increase in PD incidence (147). If this is the case, the highest increase will be observed for the groups where the largest decline in smoking prevalence is seen, being women in high SDI countries (148). Even if the relationship between smoking and PD has been known since the 1960s, more research is still needed to fully elucidate the role of nicotine products in PD (149, 150). The observed association between smokeless tobacco products and PD at least reduces the number of chemicals that could drive the association.

Other factors that can be related to the reward system has also been observed to be inversely associated with PD, including alcohol and coffee consumption. These observations have mainly been made in retrospective case-control studies and are therefore at high risk of bias from reverse causality. Alcohol has been proposed to be inversely associated with PD risk in multiple studies, indicating a potential protective effect of alcohol consumption (151, 152). A recent meta-analysis also supported an inverse, dose-dependent association and saw an overrepresentation of never drinkers among individuals with PD. However, the study included both prospective and retrospective studies and a high study heterogeneity was observed. The authors therefore conclude that in the absence of a known neuroprotective effect of alcohol, there might be reasons to doubt a true biological effect in PD (153). Interestingly, an MR study supported the opposite, that a higher daily alcohol intake instead increased the risk of PD (138). Hence, the direction of the association between alcohol and PD remains uncertain. The possible relationship between PD risk and caffeinated beverages, such as coffee, was reported already in the 1980s when caffeine was identified to act as an antagonist of adenosine A2A receptors and that this was coupled with antiparkinsonian effects in animal models (154, 155). Coffee also been reported in numerous epidemiological studies to be inversely associated with PD (127, 136, 156). However, as for smoking and alcohol, it is not clear if the association is due to reverse causation and a link between coffee and PD risk has not been confirmed in MR studies (157, 158).

Numerous additional lifestyle related factors have been reported to modulate PD risk. For example, intake of dairy products (159), and having a higher educational level (160) have been reported to be associated with increased risk whereas moderate to high levels of physical exercise (particularly among men) has been

associated with a decreased risk in a dose-response matter (161). As for to the other described factors, the causality of the association can be debated, and intervention studies are needed to show if these truly are PD risk factors. It is likely that individuals with prodromal PD have subtle symptoms keeping them from exercising but promoting physical activity at any stage could be beneficial since it has been shown to have a positive impact on symptoms and quality of life in individuals with PD and the effects could be positive already at a prodromal stage (162).

A tendency for a lower body mass index (BMI) is often seen in individuals with PD in case-control studies (163). However, in a meta-analysis of prospective studies, no support of an association between BMI and PD risk was observed (164). Interestingly, a MR study saw that variants influencing BMI also had effects on PD risk where a higher BMI lead to a lower risk of the disease. A lifetime elevation in BMI of 5 kg/m^2 was associated with a lower PD risk with an OR of 0.82 (165). For a reference, a "normal" BMI is said to be in the range of 18.5-25 kg/m^2. The authors conclude that BMI is a potential modifiable risk factor for PD but that the negative health impacts of raising BMI, such as the increased risk of type 2 diabetes (T2D), need to be taken into consideration (165). Highlighting the importance of not drawing conclusions regarding the association between PD and BMI based on current BMI from case-control studies is the fact that a combination of factors may produce weight loss in PD. These include factors linked to reduced energy intake such as dysphagia, gastrointestinal dysfunction and factors linked to increased energy expenditure such as enhanced glucose metabolism and dyskinesia from levodopa treatment (166). Taken together, multiple evidence points to an important role of lifestyle and dietary factors in PD. However, the causal effect of these factors needs to be determined before potential preventative actions are carried out.

Environmental factors

Environmental factor showing strong evidence to be linked to PD risk are toxins such as pesticides. The link between toxins and PD was observed the first time in 1983 when four individuals developed parkinsonian symptoms following intravenous injections of "a new synthetic heroin" contaminated with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). The toxic MPTP metabolite MPP+ causes cell death of dopaminergic neurons in the substantia nigra by accumulation in mitochondria and subsequently inhibiting complex I of the electron transport chain, thereby interfering with mitochondrial respiration (167). Today, MPTP is a commonly used toxin in animal models to induce a parkinsonian phenotype (168). Since the discovery of MPTP, several epidemiological studies have observed an association between an increased PD risk and pesticides as well as factors related to pesticide exposure. These factors could function as surrogate measures for pesticide exposure and include for example farming, and well-water drinking (136, 156, 169, 170). Pesticides are chemicals that kill or harm pests and depending on the type of

pest, they can be categorized into e.g., insecticides, herbicides, fungicides, but also include garden chemicals and household disinfectants. They can also be categorized based on their chemical composition and nature of active ingredients, which is a more complex classification system. However, modern pesticides are commonly organic chemicals and includes e.g., organochlorines, organophosphates, carbamates, and pyrethroids (171).

Specific pesticides that have been associated with an increased PD risk includes e.g., rotenone (insecticide and piscicide), maneb (fungicide), and paraquat (herbicide) (172, 173). Rotenone act with a similar mechanism of action as MPTP by directly inhibiting mitochondrial complex I (174) and causes selectively nigrostriatal dopaminergic degeneration in PD animal models (175). Despite its structural similarity with MPTP, rotenone is commonly stated to be a safe and natural alternative to synthetic pesticides since it is derived from the roots of certain tropical and subtropical plant species and is used as an insecticide and piscicide (kills of fish) (175). In 2022, rotenone was the only piscicide allowed for essential use according to the European Union (EU) directive Biocidal Products Directive and it is stated to be the only effective piscicide approved by European environmental authorities to eradicate fish species in freshwater (176).

Exposure to a combination of maneb and paraquat in PD mice models has been seen to result in several neurodegenerative events, including increased substantia nigra pathology (177) and a progressive reduction in dopamine metabolites and dopamine turnover (178). It has also been shown that exposure to the pesticides increase the risk of PD in humans (156, 172). In addition to the increased PD risk of paraquat exposure, it is reported to be an extremely hazardous substance and was therefore banned in the EU in 2007 (179). However, paraquat is still one of the most widely used herbicides in certain states of USA as of today, and both paraquat and maneb are used on crops such as potatoes (172). By the end of 2022, the Michael J. Fox Foundation (MJFF) for PD research announced that the Environmental Protection Agency (EPA) in USA had agreed to reconsider the evidence of paraquat being linked to PD and other neurological disorders, which could lead to a ban of paraquat (180). Like paraquat, maneb is banned in EU (since 2009) but appear to be widely used in other parts of the world.

Considering that factors that could act as surrogate measures for pesticide exposure also have been associated with PD, it is likely that additional pesticides also are associated with PD risk. In fact, a not yet peer-reviewed study investigating 300 pesticides identified that long-term exposure to 53 different active ingredients in pesticides were associated with PD (181). Subsequently, it was reported that residential or workplace proximity to higher amounts of 10 out of these 53 pesticides was associated with a faster PD progression (182). Among these 10 pesticides, the broad-spectrum pesticide copper sulfate, and the herbicide 2-methyl-4-chlorophenoxyacetic acid (MCPA) was associated with all three measured progression endpoints, being cognitive, depressive, and motor symptom events.

These findings contribute to important insights into the biological mechanisms of pesticides in PD etiology but also provide a guideline for future agricultural policy.

Exposure to other potential toxic chemicals such as solvents (most notably trichloroethylene) and various metals, including manganese, copper, mercury, lead, iron, zinc, aluminum, and amalgam has been reported to be associated with an increased risk of PD (183, 184). However, additional replication studies are needed to validate these associations. The same goes for air pollution, which also has been suggested to be an emerging risk factor for PD (185). Considering the general worldwide exposure to air pollution from traffic and other sources, and the toxicological effect it generally has on human health (186), it would be an important modifiable factor not only for PD risk.

Comorbidities, infections, and drugs in PD risk

An increased PD risk has been observed among individuals with complex neuropsychiatric disorders such as bipolar disorder (187), obsessive-compulsive disorder (188), and schizophrenia spectrum disorders (SSD) (189). The causality is debated, and it has for example been lifted that the increased PD among individuals with SSD could be due to increased vulnerability of the dopamine system or misdiagnosis of drug-induced parkinsonism (189, 190). There are also report of an increased risk of PD among individuals with inflammatory bowel diseases (IBD), such as Crohn's disease (191) and T2D (192). The association between IBD and PD could be due to partly shared genetic architecture since *LRRK2* is a common susceptibility locus for both diseases (193). Increasing evidence is also pointing towards a shared biology between T2D and PD. For example, insulin has been suggested to have neuroprotective roles and can influence pathways related to PD pathogenesis, such as pathways involved in mitochondrial dysfunction and neuroinflammation (192, 194). In fact, the repurposing of drugs used to treat T2D, such as the GLP-1 analogue, is tested in clinical trials for a potential disease-modifying effect in PD with potential beneficial effects (195).

Associations between PD risk and other factors related to health, such as serum urate levels, certain drug use ,infections, appendectomy, and head trauma has also been reported. For example, higher serum urate levels have been reported to be associated with a decreased PD risk, at least in men (156). Various drugs such as the non-steroidal anti-inflammatory drug ibuprofen, thiazolidinediones, calcium channel blockers, and statins have also been reported to decrease the risk of PD (196-199). Hormone replacement therapy, and beta blockers have on the other hand been reported to potentially be associated with an increased risk (156, 200).

Certain infections may also be potential risk factors for PD and several infectious agents have been suggested to increase the risk such as influenza, viral hepatitis, and *Helicobacter pylori* (201). Recent publications have seen an increased risk of

PD up to 10 years after influenza infection (202, 203) and the question regarding the long-term effects of COVID-19 infection has been lifted following the recent pandemic. However, whether it will be associated with subsequent NDD remain to be seen. In addition to the potential link between IBD and PD, there is also a discussion regarding the gut microbiome in PD risk and individuals who have gone through an appendectomy have been reported to have a lower PD risk (204). A history of head trauma has on the other hand been reported to be linked to an increased risk of developing PD (156).

In summary, a substantial number of non-genetic factors have been reported to be associated with PD risk. However, several need replication and high quality observational and interventional studies are required to try to elucidate if any of the factors are causal in PD.

Gene-environment interaction

Interactions between genetic and non-genetic factors are thought to explain a part of the missing link in PD etiology. Interactions have been shown in focused analyses, showing synergistic effects between specific genetic and non-genetic factors. For example, a less pronounced inverse association of smoking has been seen in carriers of the minor alleles in variants in the genes *RXRA* (rs4240705) and *SLC17A6* (rs1900586) (205) and a higher PD risk was observed among patients with T2D having the common *SNCA* rs356221-AT/TT genotype (206).

An inverse association between PD and three amino acids (V11, H13, and H33) encoded by *HLA-DRB1* has been seen (207) and subsequently, it was shown that the inverse association between smoking and PD only was present in patients without the alleles encoding the protective amino acids (208). An increased PD risk and a synergistic effect between a genetic variant in another gene in the major histocompatibility complex class II locus, *HLA-DRA*, and pesticide exposure has also been seen (209). Other studies have reported increased PD risk and synergistic effects between pesticide exposure and genetic variants in genes such as *BCHE (210)*, *PPARGC1α* (211), and *ABCB1 (212)*. Large-scale replication studies are of interest to understand these potential interactions in PD risk.

Studies have also aimed at combining PD PRS and information on non-genetic factors to study interactions and identify individuals at higher risk of PD (213, 214). Using this approach, interactions between PRS and diabetes (type not specified) in PD was identified where diabetes was seen to be a stronger risk factors in individuals with a lower PD PRS and potentially having protective (likely through the treatments) effects in those at higher genetic risk (214). Additionally, it is plausible that genetic and non-genetic factors interact through epigenetic mechanisms where e.g., environmental exposures may cause alterations in gene expression, making it a research area that needs to be further investigated in PD etiology (215).

Aims of the thesis

The overall aim of this thesis was to contribute to an extended knowledge of genetic and non-genetic factors in PD etiology.

The specific key research objectives addressed in this thesis were as followed:

- To investigate the prevalence of autosomal dominant mutations in Sweden (**Paper I**)
- To investigate the associations between unconfirmed genes suggested to be linked to PD risk **(Paper II and IV**)
- To investigating genetic and environmental factors in PD in Sweden (**Paper III and IV**)
- To study the genetics and potential mechanisms of the prodromal PD clinical symptom RBD (**Paper V**)

Methodological considerations

This section aims to provide an overview of the key methods and cohorts used in the papers in this thesis. For additional details, please refer to the methods section of the respective paper (see appendices)

Ethical considerations and open science

When doing research on humans (or on information or biological samples from humans), special ethics concerns must be considered since the data contains sensitive health and non-health related information. Therefore, it is necessary to adopt privacy safeguards when analyzing this type of data for research. When the General Data Protection Regulation (GDPR) entered into force in 2016 with the goal of harmonizing data protection across the EU, pseudonymized data was defined as identifiable (personal) data (with the use of additional information), and genetic data was put in the catalog of special categories of data as sensitive data (216). Processing special categories of data has generally been prohibited but a specific article of GDPR permits processing this type of data when it is necessary e.g., scientific, or statistical purposes. This can be done without the explicit consent from the study participants if this is permitted under EU or Member State law, and if adequate safeguards are in place (216, 217). However, the Member States are allowed to set further limitations on processing genetic data for research purposes which could hamper cross-border processing, and a personal reflection is that the different countries interpret this slightly different. My experience is that interpretation of GDPR is difficult and as a researcher you must acquaint oneself with the legal aspects of working with human data. This can lead to that cross-border collaborations are hampered, and it is of important that researchers have access to expertise regarding the processing of human data in a correct manner. Collaborations and sharing of data are though necessary to further understand mechanisms contributing to disease etiology and pathology (217, 218). A benefit of GWAS meta-analyses and other downstream analyses is the usage of GWAS summary statistics. Data is then only utilized on a group-level, facilitating cross-country collaborations. However, an important aspect in these studies is to try to harmonize the study design, as well as the data collection and analyses. Even if sharing data might be difficult in human research, an effort should always be made

to utilize and enable new discoveries using already existing data since this can accelerate the research while being time-, cost-, and labor-saving (219). Additionally, it is of importance to make the code for research analyses publicly available to foster open science, research reproducibility, and replicability.

Cohorts

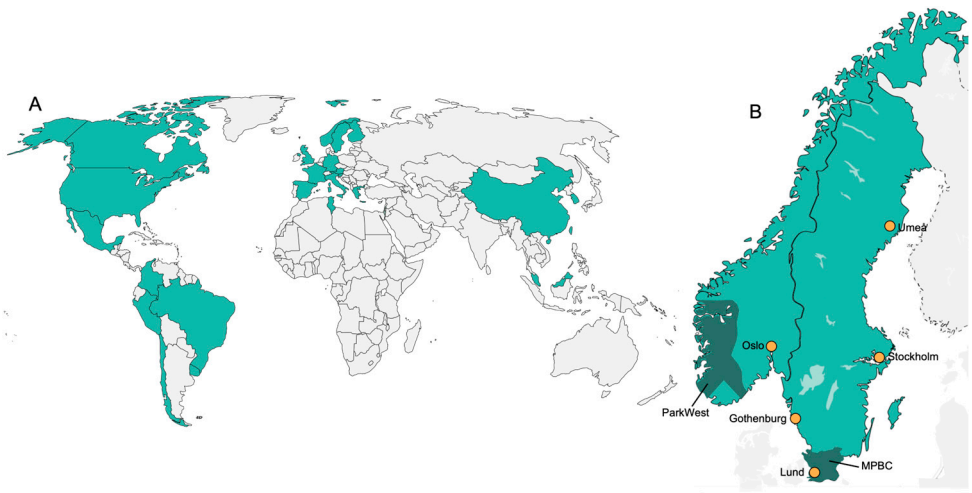

Figure 5. Geographical representation of the countries and cohorts analyzed in this thesis. A) World map with the countries wherefrom study samples originated highlighted in green. B) Map over Sweden and Norway with the cities (orange) and regions (dark green) highligthed. Modified from Wikimedia Commons, Author: ArnoldPlaton.

Multipark's biobank sample collection (MPBC)

The PD case-control study Multipark's biobank sample collection (MPBC) has been a central part in this thesis work and is included in **Paper I, III**, and **IV**. MPBC is a cross-sectional case-control study with individuals specifically from the southernmost province of Sweden, Scania. For PD patient inclusion, individuals primarily diagnosed with PD according to the International Classification of Disease, tenth revision *(ICD-10-SE)* code G20.9, were assessed for eligibility. The patients were included at Neurology clinics in nine cities in Scania. For inclusion, the patients needed to visit one of the neurology clinics, resulting in a potential loss of more disabled PD patients. Additionally, patients living at nursing homes were not invited to participate in the study. In total, 2,119 PD patients were invited to participate between November 2014 and July 2017, whereof 1,011 were included (48.3% inclusion rate). It is estimated that the number of PD patients in the region

are approximately 2,000, indicating that most patients in the region were invited to participate. For each patient, population-based controls matched by date of birth, sex, and residential area were randomly selected from the Swedish Population Register and invited between March 2015 and April 2018. A total of 1,001 individuals consented and completed data collection (18.5% inclusion rate). Out of the completed controls, 953 were unique matchings, resulting in a matched control for 953 PD patients (94.3%).

All study participants gave blood samples and filled out a questionnaire covering basic demographic data, lifestyle, habits, family history of PD, comorbid diseases, environmental exposures, medications, health status, and perceived motor- and non-motor symptoms. Additional data for the PD rating tools Hoehn and Yahr (H&Y), Clinical Impression of Severity Index - Parkinson's Disease (CISI-PD), and Parkinson's Disease Questionnaire - 8 (PDQ-8), as well as information on AAD was retrieved from the Swedish Parkinson's registry.

DNA extraction was done from whole blood at the central biobanking facility (Biobank Syd: BD4BD47) and at LGC Biosearch Technologies (UK, GEN-9300-120). LGC Biosearch Technologies (Germany) also performed genotyping and subsequent "Basic BioIT" with technical quality control (QC) and genotype clustering. The no-call GenCall threshold was set to the standard score cutoff for Infinium data of 0.15. Matched case-control samples were genotyped using the Infinium Global Screening Array-24 v.2.0 with the Multi-Disease drop in panel v.2.0 (GSAMD-24v2.0) containing 712,189 variants annotated to the Genome Reference Consortium Human Build 37 (GRCh37). Genetic imputation was done using the University of Michigan server (220) with the Haplotype Reference Consortium (HRC) r1.1 2016 European reference panel (221). A further elaboration on genotyping and imputation can be found under the section of genetic analyses.

Other Swedish cohorts

Samples from multiple other Swedish cohorts have been studied in **Paper I** and **IV** in this thesis, either as entire cohorts or parts of them. For simplicity, several of these cohorts have been named after the city where the sample inclusion occurred.

Umeå

The study was approved by the ethics committee of Umeå University and all participants gave written informed consents. Individuals with a possible PD diagnosis (ICD-10 G20, G21 and G23) were identified through patient registries in the Västerbotten county in Sweden and a PD diagnosis was validated through neurological medical records by a specialist in neurology and movement disorders. For most PD patients, samples had been donated to the biobank of the neurology clinic at the University Hospital of Umeå. The remaining PD patients were

identified in the Northern Sweden Health and Disease Cohort (NSHDC), containing samples from over 135,000 unique persons in Västerbotten. For each patient, controls were matched by age, sex, and residential area. In **Paper I**, genotype data for variants of interest for all investigated cohorts was obtained using TaqMan SNP genotyping assays whereas for **Paper IV**, whole genome genotype data had been obtained from deCODE Genetics/Amgen.

Stockholm

Samples from several different cohort were obtained from the area around Stockholm through collaborators at the Karolinska Institute (KI). Ethical approval for all studies was retrieved prior to participants inclusion from the Swedish ethical review authority and all participants gave their informed consent. Patients diagnosed with PD were recruited from the Neurology clinic at the Karolinska University Hospital. Controls were healthy elderly individuals obtained from the longitudinal study The Swedish National Study on Aging and Care in Kungsholmen (SNAC-K), spouses of PD patients, individuals visiting the Neurology clinic, and healthy blood donors. Genotype data for variants of interest was obtained using TaqMan SNP genotyping in both **Paper I** and **Paper IV**.

Gothenburg

Samples from individuals with PD from the area of Gothenburg had been recruited between 2000 and 2012 and were included in **Paper I**. Blood samples and clinical data had been collected during study visits to research nurses.

PARLU

The PARkinson LUnd study (PARLU) was analyzed in **Paper I**. PARLU included individuals with PD on the grounds that they lived in a certain geographical area (part of the counties Blekinge and Skåne) with or without a family history of PD. Blood samples were collected along with a standardized questionnaires on possible exposure to environmental risk factors, non-motor symptoms and their family background. Controls included the patient's spouses, caregivers, or friends and only controls of the same age groups as the patients were included in the study.

SweGen and ACpop

The SweGen dataset is a reference cohort of 1,000 individuals representing the Swedish population, utilized in **Paper III** and **IV**. All samples were sequenced, and allele frequency data was generated by Science for Life Laboratory (222). The ACpop variant frequency dataset is also a reference cohort containing whole-genome variant frequencies for 300 individuals selected from the NSHDC (223).

The International Parkinson's Disease Genomics Consortium

The International Parkinson's Disease Genomics Consortium (IPDGC) was founded in 2009 by a small group of investigators aiming to put together a research consortium focused on the genetics of PD and based on collaborative research (224). The IPDGC was borne out of the realization that a highly collaborative approach was required to deliver both high quality and high quantity genetic results in PD. By 2020, the IPDGC had grown significantly, including over 100 researchers (excluding students) from all around the world and the research focus had expanded to also include clinical and functional investigations in PD and related disorders such as DLB, progressive supranuclear palsy, and MSA. The IPDGC has been a major contributor to the majority of GWAS in PD, including the latest GWAS meta-analysis (116). The ongoing and future aims of the consortium is to expand the known genetic architecture in PD, including in non-European ancestry populations through initiatives in Southeast Asia and China, and across Africa along with collaborations centered in India, the Luxembourg-German-Indian Alliance on Neurodegenerative diseases and Therapeutics (LUX-GIANT) and the Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD) (224). Following a IPDGC meeting in London 2019, I have been very involved in the IPDGC trainee network with over 130 students collaborating on projects and creating resources for education and training. Through the IPDGC, collaborative research projects have been conducted during my PhD, which is reflected in **Paper II**, **IV**, and **V**. Since the consortium includes multiple different investigators, it also functions as a hub for multiple data collections and cohorts whereof some have been investigated in the projects included in this thesis. A brief description of these cohorts is found below.

Oslo Parkinson's Disease Study

The Oslo Parkinson's Disease Study (Oslo) was approved by the Regional Committee for Medical Research Ethics in South-East Norway, and all participants gave informed written consent. Most patients were recruited at the Oslo University Hospital in a tertiary-care unit for movement disorders. A large proportion of these patients were evaluated for advanced treatment options, including DBS. Consequently, the patient group is characterized by an overrepresentation of patients with EOPD, having more severe motor fluctuations, a good levodopa- response, and better cognitive function relative to the average PD cohort (225). The remaining fraction of PD patients were recruited at the Drammen Hospital. Controls were recruited among spouses of patients, volunteers from Rotary clubs in the Oslo area, and outpatients in primary care without any neurological disease. A total of 479 PD patients and 464 controls were included and genotyped the Illumina Infinium OmniExpress array and subsequently imputed at the University of Michigan server (220) with the HRC r1.1 2016 reference population (221).

ParkWest

The Norwegian ParkWest study is a prospective longitudinal multicenter cohort study of patients with PD in Western and Southern Norway. The study was approved by the Regional Committee for Medical Research Ethics, University of Bergen. The study intended to recruit all residents in the study area of Sogn and Fjordane, Hordaland, Rogaland, and Aust-Agder with incident PD between 1 November 2004 and 31 August 2006. A total of 265 individuals fulfilled the clinical research criteria of PD at their latest clinical visit. A comprehensive description of the patient inclusion and incidence of PD in the study area has been published in 2009 by Alvez *et al.* (226). As for the Oslo cohort, all samples had been genotyped using the Illumina Infinium OmniExpress array and imputed with the HRC r1.1 2016.

Latin American Research consortium on the Genetics of PD (LARGE-PD)

The LARGE-PD was formed in 2009 and is an ongoing consortium involving 35 institutions in 12 countries across Latin America. In **Paper II**, a total of 1,481 study participants from Uruguay, Mexico, Peru, Chile, Brazil, and Colombia were analyzed which had been recruited from 2007 to 2015 and genotyped using the Multi-Ethnic Genotyping Array (MEGA) from Illumina (227). The PD patients had been diagnosed by a local movement disorder specialist using the UK PD Society Brain Bank clinical diagnostic criteria. Individuals who did not exhibit neurological symptoms were included as controls. All study participants gave written informed consent according to their respective country's requirements. Imputation was performed using the TOPMed Imputation Server using a reference panel of 125,568 haplotypes from diverse samples. A more detailed description of the cohort was published 2020 by Loesch BA *et al.* (227). The consortium is still active, and at least 1,800 PD patients and 2,000 controls are included as of today.

Cohort of East Asian ancestry

Individuals with a PD diagnosis (based on the UK Parkinson's Disease Society Brain Bank criteria) and matched controls (ethnically and regionally) were recruited from six regions across East Asia during January 1, 2016, to December 31, 2018, (228). The study was approved by the local review board and written informed consents were obtained from all participants. Genotyping data was available for 6,724 PD patients and 24,851 controls after genotyping with the GSAMD-24v2.0, imputation using the multiethnic 1000 Genomes Project phase 3 reference panel, and pre- and post-imputation QC (228).

RBD cohorts

Two cohorts were used in **Paper V** for the RBD GWAS meta-analysis, an iRBD cohort with 1,061 patients and 8,386 controls, and a cohort of PD patients (1,782) with probable RBD (pRBD) and controls (N=131,250) which was genotyped and analyzed by 23andMe, Inc. The patient group in the first cohort was collected by

the International RBD Study Group and were genotyped and analyzed at McGill University and was composed of individuals of predominantly European ancestry. Genotype data for controls were obtained from four different sites, McGill University (N = 871), the HYPERGENES Project (229) (*N* = 557), the Wellcome Trust Case Control Consortium (230) (N = 5,516), and European control samples genotyped in the Laboratory of Neurogenetics, National Institute on Aging, National institutes of Health (NIH) (N=1,442). All individuals were genotyped using the Illumina OmniExpress GWAS chip and imputed using the HRC r1.1 2016 reference panel. In the second cohort, pRBD was identified using the RBD Single-Question Screen (RBD1Q). The RBD1Q has a sensitivity of almost 94% and specificity of 87% in PD. Individuals without PD but with pRBD were excluded since some of them potentially had iRBD. All study participants gave their signed informed consents, and the study was approved by the institutional review boards.

23andMe

Although being a commercial company, the personal genetics company 23andMe Inc., collaborates closely with academic researchers to study foremost genetics but also non-genetic factors in PD. All individuals included gave informed consent and subsequently sent in saliva for genotyping and answered surveys online according to the 23andMe human subject protocol, which was approved by the Ethical & Independent Review Services (231). The PD diagnosis among the participants in 23andMe is therefore self-reported but to try reducing the number of individuals to falsely state a diagnosis, their physician's name, phone number, and institution needed to be provided (232). The company provided a substantial part of the samples in the GWAS meta-analyses from Nalls *et al.* 2019 and in **Paper V** of this thesis. More information on the company and their research can be found at their website https://research.23andme.com/.

Accelerating Medicines Partnership - Parkinson's Disease

The Accelerating Medicines Partnership Parkinson’s disease (AMP-PD) was launched 2018 with the ambitious aim to find a cure for PD (233). It was formed as a partnership between the NIH, the US Food and Drug Administration (FDA), five pharmaceutical and life science companies (GSK, Pfizer, Sanofi, Bristol-Myers Squibb, Verily) and the non-profit organization MJFF with focus on discovering new therapeutic targets and develop biomarkers to help validate existing therapeutic targets for PD. An important part of the AMP-PD is to share data and analyses with the biomedical community and therefore, they provide a harmonized dataset from the eight cohorts BioFIND, the Harvard Biomarkers Study (HBS), LBD, The LRRK2 Cohort Consortium (LCC), The Parkinson's Disease Biomarkers Program (PDBP), the Parkinson’s Progression Markers Initiative (PPMI), STEADY-PD3 and the Study of URate Elevation in Parkinson’s Disease, phase 3 (SURE-PD3) to

be used by researchers. Including other forms of data, the cohort contains whole genome sequencing data for over 10,000 individuals, including over 3,000 PD patients which can be accessed and analyzed on the cloud-native platform Terra (https://terra.bio/). The AMP-PD data has been used in **Paper II** and **IV** (manuscript) in this thesis.

The Global Parkinson's Genetics Program

The Global Parkinson's Genetics Program (GP2) is a supported resource project of the Aligning Science Across Parkinson's initiative aimed towards creating a worldwide collaborative effect to accelerate the research and identification of genetic contributors in PD (218). The principles of GP2 includes to increase the diversification both regarding researchers and participants, to increase transparency and reproducibility, and to provide training opportunities. To ensure an efficient structure, several working groups centered on specific aims have been created, including the complex disease genetics working group. This working group has the long-term aim to genotype 150,000 individuals from all around the world (with a specific focus on underrepresented populations). The project is still at a rather early state and as of today, the complex disease data consists of 14,902 genotyped participants (8,190 PD patients and 6,712 individuals without PD). Like the AMP-PD data, the GP2 data is available for researchers on Terra. Individuals of European ancestry from this cohort were analyzed in **Paper IV** (manuscript).

Epidemiological analyses

Non-genetic factors associated with PD was evaluated in a Swedish case-control study in **Paper III** of this thesis. Due to the cross-sectional case-control study design, variables related to past exposures were analyzed, excluding variables related to exposures at time or near time of inclusion. Case-control studies have multiple advantages, including the possibility to study the association between exposures and disease and to study multiple potential risk factors at once. However, these types of studies are prone to recall bias, being the risk that individuals in the case group will recall and report an exposure more frequently, even if the true exposure was the same as in the control group (234). Case-control studies cannot establish causation but can be used to investigate associations between exposure and outcome. It is important to try to reduce confounding bias which can occur when a variable is related to both the exposure and the outcome. To test for associations between potential risk factors and PD, both unadjusted and multivariate logistic regression analyses were used. To report an exposure as 'independent' in a multivariable regression, other related variables are often included as covariates in the model. To consider each variable in relation to both the exposure and outcome

of interest, directed acyclic graphs (DAGs) were made for all exposures of interest to help identify confounding variables to adjust for in the multivariate analyses. It is not fully known whether this strategy led to a more accurate effect estimate. However, it is a tool to increase transparency on how the confounding variables were identified since this often is unclear in observational studies (235). The pathways with confounders having the least amount of missing data were prioritized and complete-case analyses were run to estimate the odds ratio (OR) and 95% confidence intervals (CI). All epidemiologic data in **paper III**, including demographic characteristics, was analyzed using R version 4.0.0 (236).

Genetic analyses

Genotyping

Genetic analyses and data that have been utilized in this thesis includes selective genotyping for candidate variant association analyses, and whole-genome genotyping (WGG) for GWAS and down-stream analyses, mapped to either GRCh37 or GRCh38. Whole-genome sequencing (WGS) data has also been accessed and analyzed through AMP-PD, although only looking at SNV. For the candidate variant association analyses, variants of interest were genotyped in patients and controls and evaluated. WGG and GWAS are explorative methods where typically 300,000–1,000,000 SNV are genotyped on an array. These SNVs should have a genome-wide coverage and tag adjacent SNVs through patterns of linkage disequilibrium (LD), giving a representation of the entire genome. Since these patterns of LD can differ between populations, a genotyping array that provides the best genome-wide coverage for the studied population should ideally be selected (237). WGS offers an even more unbiased approach for association analyses compared to WGG but is considerably more costly.

Quality control (QC)

An important step when working with whole-genome genotype data is QC since false-negative or false-positive associations can occur if the data has not been cleaned properly, which also affects downstream analyses. Various factors such as batch effect, population stratification, and sample relatedness can confound genetic association analyses and hence, these need to be taken into consideration. In this section, common QC steps for addressing this are discussed. Focus will not be put on the various threshold used in the papers since these vary slightly between populations depending on sample size. For details, please refer to the methods section of the respective paper (see appendices).

The free, open-source whole-genome association analysis toolset PLINK is commonly used for analyses of genotype data and has been the main software used in the projects in this thesis, in particularly version 1.9 (238, 239). An initial QC step includes to investigate the missingness of variants and samples. Variants with low genotype calls, are here excluded, as well as samples from individuals with low genotype calls. Discrepancies between an individual's reported sex and their genetic sex is checked since deviations from this might indicate sample mix-ups. Genetic sex is based on X-chromosome heterozygosity/homozygosity rates where men should have an X-chromosome homozygosity estimate approximately over 0.75 and women less than 0.25.

The samples' heterozygosity rates are evaluated (excluding the sex chromosomes) to identify and exclude those with an excessive or reduced proportion of heterozygote genotypes, since this can indicate sample contamination or inbreeding. Samples from closely related individuals, commonly defined as sharing more than 12.5% of alleles, are also excluded. A major source of confounding in genetic studies is population stratification, where the differences in genotype frequency between cases and controls is observed due to differences in the population's ancestry rather than a difference due to the investigated disease (240). Hence, to reduce the potential population stratification effect during QC, individuals of divergent ancestry are removed. Principal component analysis (PCA) is a common method to identify and exclude ancestry outliers. For this, the publicly available HapMap phase 3 data set is commonly used (241). The genotype data is then pruned and merged with the HapMap phase 3 data set to identify samples with divergent ancestry. Commonly, samples are determined as having European ancestry if they cluster around the combined population mean of the HapMap populations Utah residents with Northern and Western European ancestry (CEU) and Toscani in Italia (TSI) and outliers are removed. The use of PCA in genetic population studies have recently been criticized as being biased depending on e.g., the sample size of the reference population (242). Novel methods of estimating ancestry using machine learning with larger and more divergent reference populations, such as GenoTools are under development. For example, in **Paper V**, the supervised learning model support vector machine was used to classify haplotypes into one of 31 reference population available through 23andMe. Until these methods and reference populations become more easily accessible, careful evaluating the dataset is of great importance during QC to minimize confounding factors.

Additional variant level QC include to exclude variants where the genotype rate differs between patients and controls, where genotypes that are missing by haplotype, or variants that deviate from Hardy–Weinberg equilibrium (HWE) since these are common indicators of genotyping error. Variants with a low MAF are also removed since they are more prone to genotyping errors and the power to detect associations for these variants is often lacking. Several publications are available explaining the QC steps in further details (240, 243).

Imputation

Genotype imputation is a key method of genetic association studies. It predicts and imputes genotypes that have not been directly genotyped in the study samples and is used to e.g., increase power and facilitate meta-analyses (244). Considering that the human genome contains around 85 million SNVs (245), the genotype arrays are far from covering all variants but are still considered sufficient to cover a substantial amount of variation and are attractive due to the lower cost than of WGS. Additionally, imputation can be used to improve the genome coverage and guide fine-mapping, normally using the autosomal QC:ed genotype data as input data. Genotype imputation uses the LD structure among variants (haplotypes) in an external reference panel/cohort having genotype information on a much denser set of variants. Various algorithms are available for phasing (assigning alleles to the parental and maternal chromosomes) each individual at the typed variants and different software utilize these algorithms to perform the imputation. Commonly used software for phasing includes Beagle, Eagle, and Shapeit, and software for imputation includes Beagle, Impute and Minimac. All software provides a general accurate imputation for common variants but still differ slightly in accuracy, memory usage, and run time (246). Eagle v2.4 and Minimac4 has been used for phasing and imputation in several of the projects of this thesis, having a good imputation accuracy for low frequency variants and a low memory usage. The imputation accuracy also increases with the number of haplotypes available in the reference panel, especially for rare and low frequency variants (MAF<5%) (220).

Importantly, a post-imputation QC step is also required, excluding variants with low MAF (depending on the sample size) and a low imputation accuracy. The accuracy is commonly calculated as the squared Pearson correlation between allele dosages and best-guess genotypes and is referred to as Rsq (247). Varying thresholds are used for exclusion but commonly, variants with Rsq<0.3 are excluded. The commonly used reference panel for samples of European origin, HRC, contains data from >32,000 individuals, creating a panel of 4,976 human haplotypes at 39 million variants and provides a generally good imputation accuracy even for variants of low frequency (0.1–0.5% or less) (221). Larger reference panels such as TOPMed impute variants at even lower frequency of approximately 0.01%, providing power to identify associations at a low MAF, if the study cohort has sufficient power to detect them (248). More diverse and population-specific reference panels have also been developed for improving imputation for non-European samples (249, 250).

Power calculations for GWAS

Preferably, genetic power calculations should be performed prior to cohort inclusion to avoid performing underpowered GWAS studies. The power indicates the probability of detecting an effect of a particular size and GWAS often require large

sample sizes to identify genome-wide significant associations. The calculations often take sample size into the calculation along with disease prevalence, estimated magnitude of effect, allele frequency ,and threshold for significance. Multiple tools are available for these calculations, including the online Genetic Association Study Power Calculator and The Genetic Power Calculator (251).

Genetic association analyses

Candidate gene association analyses

Candidate gene studies focus on a selected gene (or genes) which are known, or have been suggested, to be related to the disease through e.g., experimental models. Candidate gene studies are relatively cheap and quick to perform and have been used to identify genes associated with PD prior to the first PD GWAS. The variants of interest are genotyped in a case-control study, and tests i.e., chi-square test can be used to determine if a certain allele or genotype occurs more often in patients than in controls. Genotyping can be done with e.g., TaqMan qPCR (as in **Paper I and IV**) using TaqMan Genotyping Assays. Alternatively, the specific genotype data can also be extracted from already existing WGG datasets. As for all studies, replication studies are needed in other cohorts to validate the potential findings.

Genome-wide association analysis

To identify associations of genotypes with phenotypes on a genome-wide level, GWAS can be used. Typically, the input data in a GWAS is QC:ed imputed genotype data. Both linear and logistic regression models can be used to test for association depending on whether the phenotype is continuous (e.g., AAO) or binary (e.g., the presence or absence of a disease). Since millions of genetic variants are tested for an association with a phenotype, a stringent threshold for significance is needed to avoid false positive results and is commonly set to $p < 5 \times 10^{-8}$ ('genome-wide significance'). Covariates such as age, sex, and principal components (PCs, representing ancestry) are included to avoid confounding effects of demographic factors and account for potential population stratification. Multiple tools can be used for running GWAS, including PLINK and RVTESTS, both of which have been used in **Paper III** and **V** in this thesis (238, 252). A MAF threshold of 1-5% is commonly seen in GWAS due to the lack of power to detect rare variants. The output data is plotted in a classic Manhattan plot with the SNVs genomic coordinate on the x-axis and the transformed p-value on the y-axis.

A quantile-quantile (QQ) plot is created to graphically represent the deviation of the observed p-values from the null hypothesis where the observed values should in general correspond to the expected values, with the exception from variants associated with the trait of interest (Figure 6A).

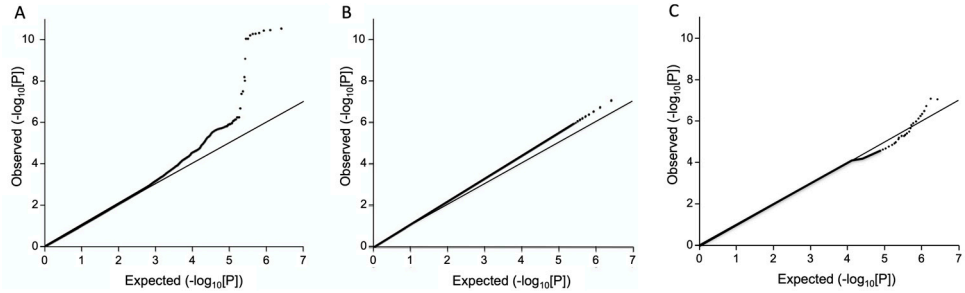

Figure 6. Quantile-quantile (QQ) plot of simulated data from a genome-wide association study (GWAS). A) QQ plot where the observed values generally correspond to the expected values, with the exception from variants associated with the investigated trait. B) QQ plot showing an early separation of the observed from the expected, suggesting e.g., population stratification. C) QQ plot with a section of deflated values which could indicate a too low MAF threshold in the GWAS. A and B adapted from Ehret *et al.* (2010) with permission from Springer Nature (license number: 5478670327871).

The QQ-plot gives an indication if there is a problem of e.g., population stratification in the cohort. Population stratification is likely an issue if an early separation occurs between the expected and the observed value since it is indicating a systematic difference between cases and controls in allele frequencies (Figure 6B) (253). Deflated p-values (deviating toward the expected p-values) can indicate that the MAF threshold used is too low for the sample size in the GWAS (Figure 6C).

Additionally, GWAS meta-analyses is an important tool, enabling pooling of studies using GWAS summary statistics instead of individual-level data, improving power to study complex diseases. Tools such as METAL can be used to efficiently run GWAS meta-analysis (254).

Polygenic risk scores (PRS)

A polygenic risk score (PRS), or cumulative genetic risk score (GRS), is an estimate of an individual's genetic liability for a disease/trait , calculated from their genotype profile and relevant GWAS data (255). PRS has been used in **Paper III** and **Paper V** in this thesis. The PRS score is calculated as the sum of risk alleles that an individual is carrying, weighted by the risk allele effect size estimate from a GWAS on the phenotype using software such as PRSice-2 (256). For PD, effect size estimates for the 90 risk variants from the GWAS meta-analysis from Nalls *et al.* are commonly used when testing the PRS for PD in another cohort (116). Including more variants in the PRS calculation than those that reached genome-wide significance could provide a greater predicative power, however, there is a risk of 'overfitting'. Overfitting is when the PRS model has been over-optimized to the investigated population and the validity of the model should therefore always be tested in an additional cohort (255). Even if the PRS for PD can be used to predict an individual's genetic risk for a disease, the models are used today for research purpose and can be utilized to e.g., stratify patients (46).

Haplotype analyses

A haplotype is in its most general sense referring to a set of variants along a chromosome that are in LD, and therefore tend to be inherited together. Haplotype-based analysis may help improve the detection of causal genetic variants as, unlike SNV-based analysis, it is possible to assign variants to a strand and combine information from multiple variants to identify rarer causal variants (257). An example of a haplotype associated with PD risk is a common *MAPT* extended haplotype. Two *MAPT* haplotypes are present in Caucasians, H1 and H2, where the H1 haplotype is associated with an increased risk of PD (258, 259). The use of haplotype analyses has decreased since the accuracy of genotype imputation has increased and more rare variants can be studied directly in larger cohorts. However, haplotype analyses can still be used to e.g., investigate differences between the size of haplotype blocks at PD risk loci, where differences have been seen between PD patients of different countries. Haplotypes of more admixed European populations were reported to be smaller compared than less admixed European populations, highlighting the importance of studying diverse populations to accelerate the mapping of PD risk loci (260). Various software can be used for analyzing haplotypes, including Haploview (261, 262), PLINK (238), or the use of BEAGLE for phasing and R packages such as haplo.stats to analyze associations.

Cumulative effect of risk variants

Since rare variants cannot be included in GWAS due to power issues, other methods have been developed to run association tests for rare variants in a cumulative way, such as region- or gene-based multivariant tests (263). For this, rare variant test software such as RVTESTS can be used (252). Multiple tests exist, being more or less suitable depending on the cohort and whether the outcome is linear or logistic. Examples include burden tests, the non-burden sequence kernel association test (SKAT), and the optimal SKAT (SKAT-O). Burden tests collapse rare variants into genetic scores and are powerful when many of the variants are associated with the outcome and the effect is in the same direction. The SKAT instead tests the variance of the genetic effects and is powerful when a large fraction of the variants in a region are not associated with the outcome or if the effects of the associated variants are in different directions. SKAT-O combines burden and variance component test and is more robust if there are variants with effects in different directions (263).

Expression quantitative trait loci (eQTL) analysis, colocalization, and expression specificity measures

A common GWAS downstream analysis is the expression quantitative trait loci (eQTL) analysis to try to understand the function of the variants associated with the disease of interest. The identified variants could exert their effects by regulating expression levels of local or distant genes and the eQTL analysis aims to associate

the genetic variants with variation of gene expression levels (264). Publicly available gene expression datasets are commonly used such as the bulk-tissue RNA-sequencing of 53 human tissues from the Genotype-Tissue Expression consortium (GTEx) (265). The probability that a risk locus and another variable, i.e., an eQTL share a single causal variant can be also evaluated using tools such as coloc (266). Publicly available eQTL data can be downloaded from sources such as eQTLGen (267) and analyzed together with GWAS data to try to identify a single shared causal variant and hence being considered as colocalized. Cell-type and tissue specificity measures can be done also utilizing data from e.g., GTEx. Specificity values can be generated as a value between 0 and 1, representing the proportion of a gene's total expression attributable to one cell type or tissue. The values 0 represents that a gene is not expressed whereas 1 means that a gene only is expressed in that cell type or tissue. Online platforms also exist to give quick estimate some of these downstream analyses including the Functional Mapping and Annotation of GWAS (FUMA, https://fuma.ctglab.nl/) (268) or the Human Protein Atlas (proteinatlas.org) (269)

Other GWAS down-stream analyses

Multiple additional GWAS down-stream analyses can be done, some of which were included in **Paper V** of this thesis, being pathway analysis, and investigation of heritability and genetic correlation, which will be described briefly. Additional detail and code can be found in the paper.

Pathway analysis

Pathway analysis can be performed on genes identified through a GWAS using e.g., gene-set enrichment analysis in which the enrichment of a provided set of genes, is examined using predetermined lists of genes involved in various functional pathways. This method can help identify biological pathways that are enriched in a gene list more than what is expected by chance (270). Online tools such as the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) (271) or software such as RVTESTS can be used to run the analysis (252). The level to which the genes are over-expressed in the investigated biological pathways is represented by an enrichment score, giving an indication of the potential role of various biological pathways in disease pathogenesis.

Heritability and genetic correlation

Heritability of a disease and the genetic correlation, being the shared heritability across traits, can be calculated in patients using linkage-disequilibrium (LD) score regression. Tools available for this include the database and web interface LDHub (272). This was used in **Paper V** where traits to test were selected based on previous associations to synucleinopathies, e.g., smoking and the genetic correlation with RBD was evaluated.

Summary of results and discussion

Even if the knowledge of PD etiology is increasing, a tremendous amount of work is needed to fully elucidate the mechanisms behind the disease. The papers of this thesis work are small, although important, pieces of the puzzle that explains PD etiology. In **Paper I**, the prevalence of known pathogenic mutations in the dominant PD genes *LRRK2* and *SNCA* is investigated, highlighting that these mutations rarely contribute to PD in Sweden. Known monogenic loci have been reported to explain a fraction of the observed familial aggregation of PD and several new potential disease-causing genes in PD have been reported. In **Paper II**, we investigate one of those genes, *RIC3*, but cannot confirm an association with PD in the investigated cohorts of European, Latin American, and East Asian ancestry. We further set out to investigate the contribution of common genetic and non-genetic factors to PD in Sweden in **Paper III**, where we could identify a new and potential population-specific genetic risk variant for PD. This finding was further investigated in **Paper IV**, concluding that a population-specific effect of this risk variant cannot be ruled out at this point. In **Paper V**, the genetic architecture of common variants in the PD-associated sleep disorder RBD was examined, indicating a potential difference in the genetic background of PD, RBD, and other synucleinopathies such as DLB.

Paper I | Low prevalence of known pathogenic mutations in dominant PD genes: A Swedish multicenter study

Heterozygous mutations in *SNCA* and *LRRK2*, as well as CNVs in *SNCA* are known to cause PD with autosomal dominant inheritance (81). Mutations and CNVs in *SNCA* are rare whereas *LRRK2* G2019S is the most common variant associated with PD. However, the frequency of *LRRK2* G2019S varies substantially between different populations, even within Europe (63). At the time of the study design of **Paper I**, few reports on systematic screening of pathogenic variants for PD (in more than one gene) in larger sample collections had been published, making it hard to estimate the relative frequency of known pathogenic mutations for autosomal dominant PD. In this study, we therefore aimed to establish the frequency of known

pathogenic mutations in *SNCA* and *LRRK2*, including CNVs in *SNCA*, in Swedish PD patients.

Seven mutations in *SNCA* (A30P and A53T) and *LRRK2* (N1437H, R1441H, Y1699C, G2019S, and I2020T), along with *SNCA* CNVs (duplications and triplications) were analyzed in a total of 2,206 PD patients from four different regions in Sweden (Figure 7).

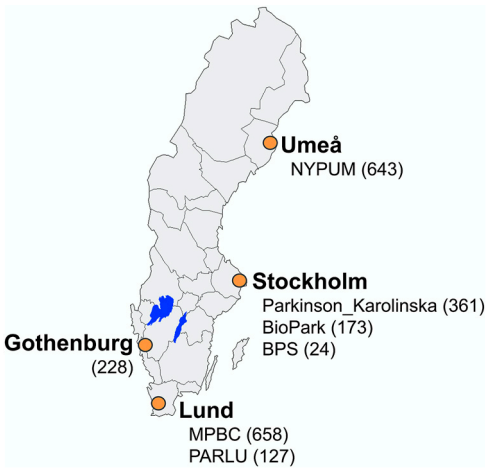

Figure 7. Map over Sweden showing the locations and names of the contributing cohorts. The number of DNA samples from PD patients analyzed within this study is written after the cohort name. BioPark: Biomarkers in Parkinsonian Syndromes, MPBC: Multipark's biobank sample collection, NYPUM: Ny Parkinson i Umeå, PARLU: PARkinson LUnd study

The study cohort included over 10% of the expected number of PD patients in Sweden. The age at disease onset or the AAD was reported for the study participants and was at an average 60.7 and 64.4 years, respectively. Among the patients, 21.6% had at least one first- or second-degree relative with PD, where a first-degree relative was reported for 12.1%. *LRRK2* G2019S was the only mutation present in the patient group and was identified in 12 patients (0.54%) (Table 1), whereof four had been reported previously (273). The mutation was also investigated in a control cohort of 942 individuals in which one G2019S carrier was identified (carrier frequency of 0.11%). This finding goes in line with the reports of the incomplete and varying penetrance of the G2019S variant (67, 274). Additionally, one patient carried a *SNCA* duplication and had previously been included in the PARLU study since the patient belongs to a large kindred with *SNCA* multiplications, the Swedish Lister family (275). Only five out of the 13 identified patients reported a positive family history of PD. This stress the importance of not solely selecting PD patients based on a positive family history when investigating a potential monogenic disease etiology since this could exclude many carriers. This in particularly true for genes with reduced penetrance, such as *LRRK2*, where relatives might be carriers without showing any symptoms associated with PD.

Using information from the Genome Aggregation Database (gnomAD), the carrier frequency of G2019S in all populations available through gnomAD was estimated

to be 0.098%, with the highest in frequency (1.63%) among the Ashkenazi Jewish population. When investigating the frequency of all known, or suspected, disease-causing mutations in dominant PD genes, *LRRK2* G2019S was by far the most common, representing 90.4% of the mutations.

Table 1: Characteristics of PD patients with mutations in dominant PD genes

Individual	Mutation	Sex	AAO	AAI	Family history	Brady-kinesia	Rigidity	Tremor	RBD symptoms	Cognitive dysfunction	Ortho-statism
#01	*LRRK2* G2019S	M	53*	59	Yes	Yes	Yes	No	No	No	No
#02	*LRRK2* G2019S	F	50*	75	No	Yes	Yes	No	Yes	Yes	Yes
#03	*LRRK2* G2019S	F	45*	49	Yes	Yes	Yes	No	No	No	No
#04	*LRRK2* G2019S	M	59*	63	Yes	Yes	Yes	No	N.A.	No	No
#05	*LRRK2* G2019S	F	56*	63	No	Yes	Yes	Yes	N.A.	No	No
#06	*LRRK2* G2019S	M	64*	66	No	N.A.	Yes	Yes	No	No	No
#07	*LRRK2* G2019S	M	75	79	No	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
#08	*LRRK2* G2019S	F	58	74	No	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
#09	*LRRK2* G2019S	M	47	51	No	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
#10	*LRRK2* G2019S	M	53	58	Yes	N.A.	Yes	Yes	N.A.	N.A.	N.A.
#11	*LRRK2* G2019S	M	54	56	No	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
#12	*LRRK2* G2019S	F	48	60	No	Yes	Yes	Yes	No	No	Yes
#13	*SNCA* dup.	F	52	54	Yes	Yes	Yes	No	Yes	Yes	Yes

This table summarizes the clinical data on the 13 patients carrying one of the known pathogenic mutations tested and has been modified from Paper I. AAI, Age at inclusion; AAO, Age at onset; F, Female; M, Male; N.A., not available. Family history refers to having at least one first- or second-degree relative with PD. *Age at diagnosis

Hence, the *LRRK2* G2019S appear to be the most common mutation in dominant PD genes in Sweden but it can only explain a small fraction of the disease etiology in the population, and similarly only a minor proportion of the observed familial aggregation of PD in the cohort. Not all dominant pathogenic variants in PD were analyzed in this study, including the relatively common variants R1441H and R1441C in *LRRK2,* and D620N in *VPS35*, which is a limitation. However, these mutations have, to our knowledge, not been documented in Sweden or in neighboring countries (273, 276, 277). The findings show that the investigated *SNCA* and *LRRK2* mutations are very rare events among the Swedish PD population, and rather few carriers seem to have a positive PD family history. These are important findings which potentially could have an impact for the decisions on clinical genetic testing.

Paper II | RIC3 variants are not associated with Parkinson's disease in large European, Latin American, or East Asian cohorts

Multiple genes have been nominated as putative disease-causing candidates in PD, including the gene *RIC3*. A missense variant (P57T) in *RIC3* was previously identified in a large Indian family showing autosomal dominant inheritance. The variant was present in all nine affected individuals and absent in five unaffected family members. Strengthening the hypothesis of *RIC3* being a risk gene for PD was the finding of another rare heterozygous missense *RIC3* variant, V168L, in a patient with YOPD using an independent cohort of 220 unrelated Indian PD patients and 186 controls (100). In this study, we investigated the association between *RIC3* and PD in several large cohorts of different ethnicities: European (IPDGC and AMP-PD), Latin American (LARGE-PD), and East Asian.

A total of 926 variants in *RIC3* were identified in the genotype data from IPDGC (n=143) and sequencing data from AMP-PD (n=773). None of the variants was associated with PD risk after Bonferroni correction (Figure 8A, B). No significant association was observed when the cumulative effect of more rare variants (MAF≤3%) was investigated in the datasets. Similar results were observed when investigating common variants (MAF>1%) using summary statistics from LARGE-PD and from an East Asian GWAS, were variants in *RIC3* were not associated with PD (Figure 8C, D). Neither of the two variants P57T or V168L identified in the original study were present in any of the analyzed cohorts.

We did not find evidence supporting the hypothesis that *RIC3* is associated with PD risk in individuals of European, Latin American, or East Asian ancestry. This is in line with previous studies, including large PD GWAS meta-analyses (116, 227, 228) and targeted analyses of *RIC3* in smaller French-Canadian, French, and Han Chinese cohorts (103, 104) which also have not found evidence supporting the pathogenicity of *RIC3*.

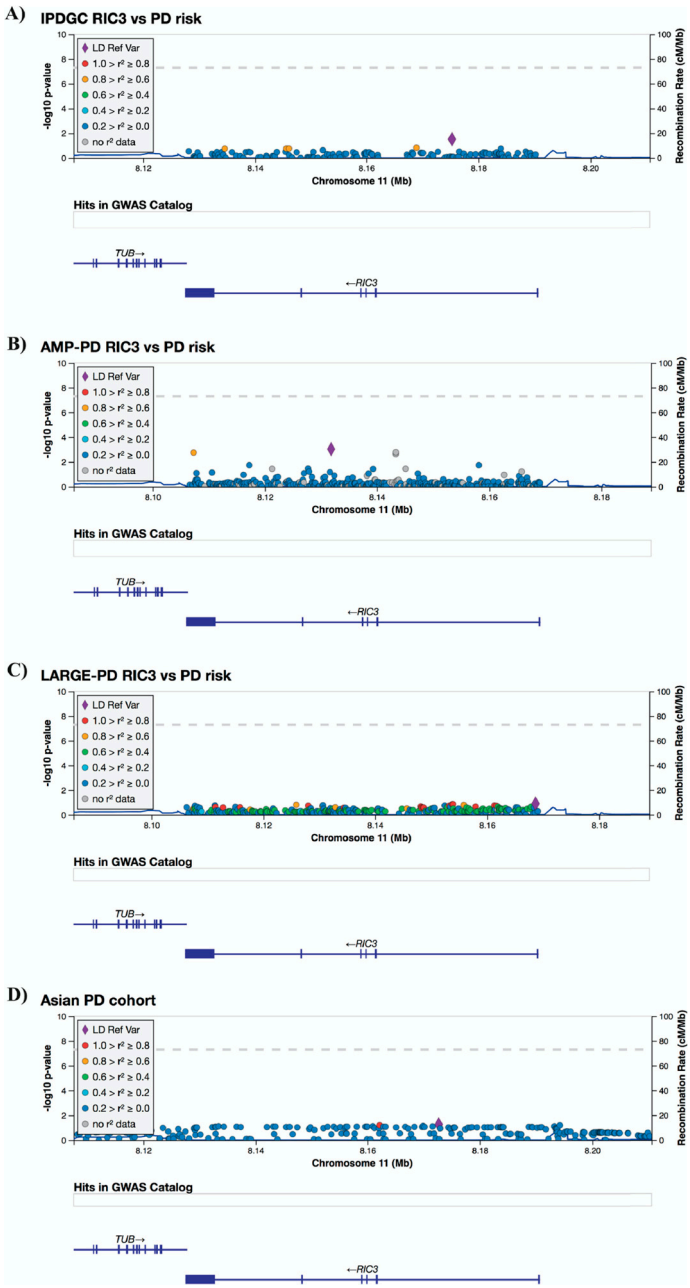

Figure 8. LocusZoom plots of the RIC3 region. Showing the p-values and recombination rate of the variants analyzed in the (A) IPDGC genotyping data (GRCh37), (B) AMP-PD WGS data (GRCh38), (C) LARGE-PD GWAS summary statistic data (GRCh38), and (D) East Asian GWAS summary statistic data (GRCh37). The variant with the lowest p-value from the logistic regression analyses is annotated with a purple diamond in each plot. AMP-PD: Accelerating Medicines Partnership in Parkinson's Disease, GWAS: Genome-Wide Association Study, IPDGC: International Parkinson's Disease Genomics Consortium, LARGE-PD: Latin American Research consortium on the Genetics of PD, RIC3: resistance to inhibitors of cholinesterase, WGS: whole genome sequencing.

However, it cannot be ruled out that rare variants in *RIC3* are linked to PD in specific families or populations since the two variants P57T or V168L were not identified in any of the datasets. Importantly, the original variants were identified in an Indian population and the analyses of rare variants in this study could only be conducted in the European cohorts. It is possible that rare variants in *RIC3* have a population-specific effect on PD risk. Although being rare, the variants reported in the original study have a higher allele frequency in the South Asian population in gnomAD (278) with a MAF of 0.05% for P57T and 0.08% for V168L, as compared to not being present in populations of European, Latino, and East Asian ancestry. This supports that the variants potentially are population-specific but do not answer the question of whether they are associated with PD risk. Further studies or rare variants in *RIC3* in the South Asian populations are needed to evaluate the suggested role of the gene in PD etiology.

Paper III | Insights on genetic and environmental factors in Parkinson's disease from a regional Swedish case-control cohort

Characteristics of the MPBC cohort

The regional Swedish case-control study MPBC displayed typical PD demographics and characteristics, as summarized in Table 2 and Table 3. MPBC is a matched case-control study, making it a highly appropriate cohort to study both genetic and non-genetic factors in PD. The case-control matching was seen successful, with 36% women in both groups and with an average year of birth of 1944 for both groups. This resulted in an average age at inclusion at 71 years for the patient group and 72 years for the controls due to control matching occurring after patient inclusion. Also, similar distributions for Swedish ancestry (defined as self-reporting that both parents were born in Sweden), highest completed education and marital status were seen in the two groups.

The patients rated their health-related quality-of-life lower than the controls on the EuroQol five-dimension-3-level (EQ-5D-3L) instrument with a calculated mean index of 0.8 on the time trade-off (TTO) scale and 69 on the visual analogue scale (VAS) compared to 0.9 and 82 for the controls (Figure 9A, B). As a reference, the general Swedish population has an estimated mean TTO index of 0.9 and a VAS index of 80 (Figure 9C) (279). The most common self-reported motor-symptom was muscle stiffness, occurring among 72.7% vs 10.0% of the patients and controls, respectively, whereas the most reported non-motor symptom was nocturia (71.6% vs. 58.4%). Approximately 92% of the patients in the cohort reported using levodopa therapy.

Figure 9. Visualization of the calculated mean index on the visual analogue scale (VAS). A) Multipark's biobank sample collection (MPBC) PD patient group, B) MPBC control group, and C) general Swedish population.

We further evaluated the PD-associated health status and AAD among male and female PD patients (Table 2) were the median AAD was identical for both sexes of 67 years. However, a wide range of 29–84 years for men and 35–89 years for women was observed, indicating that MPBC represents the heterogeneity in age seen in PD.

Health status among the patients at inclusion was evaluated using the H&Y scale, CISI-PD, and PDQ-8. Both sexes had a median of 2.0 on the H&Y, corresponding to bilateral involvement without impairment of balance (Table 3). Similar total

scores for both sexes were also reported on the CISI-PD scale (median 5.0, range 0–24) whereas women had a slightly higher median PDQ-8 total score as compared to men (7.0 vs. 6.0 out of 32). The overall low scoring on the different scales indicated that the patients in general were in a rather early stage of their disease at inclusion. This could be confirmed when evaluating the time since PD diagnosis at inclusion. The median time since diagnosis varied from study inclusion the same year as diagnosis to up to 36 years for men and 33 years for women, with a median time of 4 years for both sexes.

Table 2: Characteristics of participants (individuals with Parkinson's disease (PD) and individuals without PD [controls]) in the MPBC cohort

	Patients (N=929)	Controls (N=935)
Sex		
Men	599 (64.5%)	598 (64.0%)
Women	330 (35.5%)	337 (36.0%)
Birth year		
Mean (SD)	1944 (8)	1944 (8)
Range	1920–1979	1920–1980
Age at inclusion		
Mean (SD)	71 (8)	72 (8)
Range	37–96	38–97
Swedish ancestry		
No	112 (12.1%)	128 (13.9%)
Yes	810 (87.9%)	793 (86.1%)
Highest completed education		
< Primary and lower secondary education	8 (0.9%)	4 (0.4%)
Primary and lower secondary education	241 (26.1%)	232 (25.0%)
Upper secondary education	403 (43.7%)	384 (41.3%)
University	271 (29.4%)	309 (33.3%)
BMI at inclusion (kg/m^2)		
Mean (SD)	25.8 (4.4)	26.0 (3.9)
Range	15.2–48.8	15.9–42.3
Marital status		
Single	57 (6.2%)	62 (6.7%)
Married/Civil partnership	719 (77.7%)	713 (76.6%)
Divorced/Separated	71 (7.7%)	60 (6.4%)
Widow/Widower	78 (8.4%)	96 (10.3%)
TTO Index		
Mean (SD)	0.8 (0.1)	0.9 (0.1)
Range	0.4–1.0	0.5–1.0
VAS Index		
Mean (SD)	68.9 (16.8)	82.0 (9.7)
Range	17.2–88.9	20.9–88.9

Table 3: Characteristics of individuals with Parkinson's disease in the MPBC cohort

	Men (N=599)	Women (N=330)
Age at diagnosis		
Median (Q1, Q3)	67 (60, 72)	67 (59, 73)
Range	29–84	35–89
Disease length at inclusion		
Median (Q1, Q3)	4 (2, 8)	4 (2, 7)
Range	0–36	0–33
Hoehn & Yahr		
Median (Q1, Q3)	2.0 (1.5, 3.0)	2.0 (1.5, 3.0)
Range	0.0–5.0	0.0–5.0
CISI-PD		
Median (Q1, Q3)	5.0 (3.0, 8.0)	5.0 (2.0, 7.0)
Range	0.0–21.0	0.0–20.0
PDQ-8		
Median (Q1, Q3)	6.0 (2.0, 10.0)	7.0 (3.0, 11.0)
Range	0.0–25.0	0.0–21.0

Data retrieved from the Swedish Parkinson Registry. CISI-PD: Clinical Impression of Severity Index – Parkinson's Disease, summary score; PDQ-8: Parkinson's Disease Questionnaire – 8, summary score. Age at diagnosis and disease length at inclusion was available for 84% of men and 88% of the women. Data for the disease ranking scale were available as followed: H&Y – Men(M):92%, Women(W):89%; CISI-PD – M:86%, W:91%; PDQ-8 – M: 6%, W: 91%.

Epidemiological characteristics of PD in Sweden

For investigating potential risk factors for PD in MPBC, we selected those where exposure more likely could have occurred prior to the PD diagnosis to reduce bias from reverse causation. As an example, we observed a strong inverse association between PD and alcohol consumption during the past year prior to study inclusion but we do believe that this greatly was affected by inverse causality.

Nevertheless, multiple previously reported risk factor for PD could be confirmed in MPBC (Figure 10), including an association between pesticide exposure and PD with an OR of 2.26. The number of participants exposed to pesticides was overrepresented among those who had reported farming and well-water consumption, but no association was observed between these variables and PD. No information was available for the type of pesticides used and further investigations are needed to obtain this information. A positive family history was overrepresented among the PD patients (20% vs. 11%) and having a relative diagnosed with PD doubled the risk of having PD. We could observe a statistically significant association between PD and a history of head trauma at an OR of1.30, and it has previously been suggested that the risk of PD increases after different forms of head trauma (280, 281). However, among the participants in MPBC who has reported head trauma, no association was observed between loss of consciousness and PD. A slightly higher OR of 1.05 was seen for a higher BMI at the age of 20 years. However, since weight loss frequently is observed among PD patients (163), this variable might be even more prone to recall bias. Additionally, the result is

contradictory to the reports of a potential protective effect of a higher BMI on PD risk (165). An inverse association between PD and ever having smoked was initially observed (ever- vs. never-smoking OR=0.82) but the association was not statistically significant after adjustment for confounders. This could be due to the common use of the Swedish moist tobacco snus among smokers. Likely due to the harm reduction, it is more common to transition from smoking to snus than the other way around. Interestingly, a statistically significant inverse association between snus and PD was observed also after adjusting for confounders, such as ever-smoking (OR=0.53). An inverse association has previously been reported in non-smoking, snus-using Swedish men who had 60% lower risk of PD compared to men who had never used snus (141). Using snus is more common than smoking tobacco among men in Sweden, 18% vs 7% for the year 2019. Among women, 5% used snus and 7% smoke daily in 2019. The prevalence of smoking is decreasing in Sweden, where 6% of both men and women smoked daily in 2021 whereas snus use increased to 20% for men and 6% for women, making smoking and snus equally common among women (282).

A trend of lower OR for PD with increasing amount of coffee drinking at all investigated age groups was observed, where drinking >5 cups of coffee per day was inversely associated with PD. Previous studies have also reported a relationship between increased coffee consumption and decreased risk of developing PD (127, 136, 156). However, as described in the introduction, it is not clear if this association (or any of the observed associations) is due to reverse causation (157, 158). Other variables such as physical activity was only analyzed to describe potential differences between the groups at inclusion due to the high risk of reverse causation where we could observe that the patient group engaged in less physical activity compared to the control group.

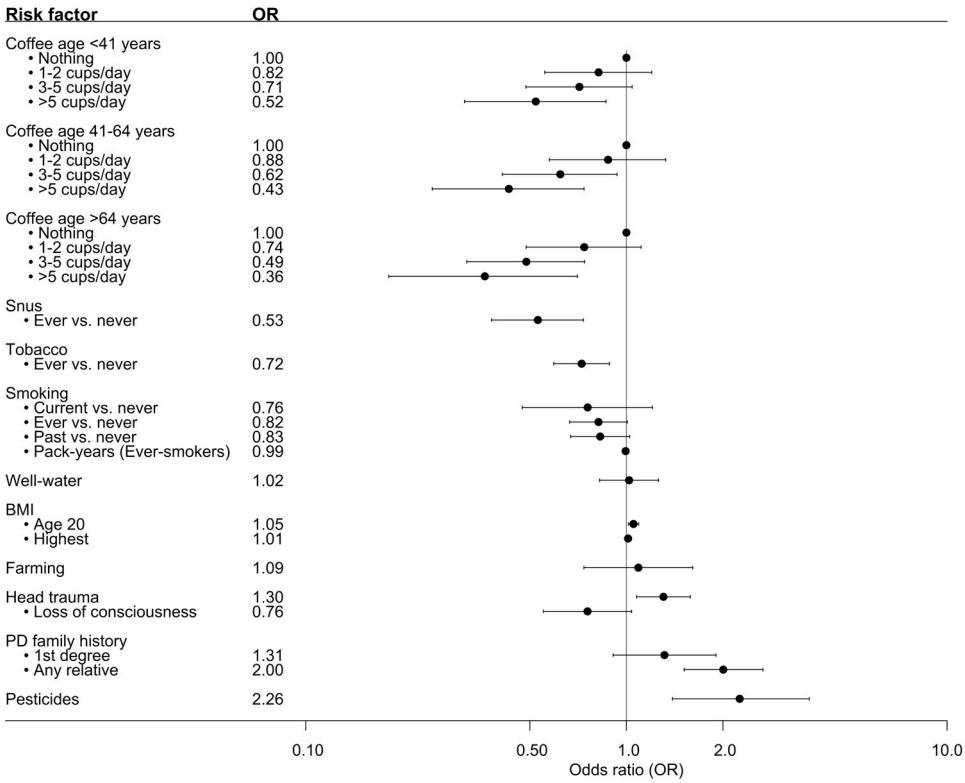

Figure 10: Forest plot over the associations between risk factors and PD in MPBC showing the adjusted odds ratio (OR) and 95% CI.

GWAS of PD risk nominated a novel genome-wide significant variant in the PLPP4 locus

To evaluate genetic risk variants in PD in Sweden, we performed the first GWAS of PD composed solely of PD patients from Sweden. Although the relatively small sample size is a limitation in this study, issues with population stratification were expected to be lower due to region-specific study recruitment and the matched case-control design. Four variants in the *SNCA* locus were observed at p-values near genome-wide significance (Figure 11) whereof one had been reported in the latest GWAS meta-analysis as one of the 90 variants associated with PD risk (116). Another 23 of the 90 risk loci were replicated at an uncorrected p-value ($p<0.05$). This study had insufficient statistical power to detect variants with low MAF or small effect size, and the majority of the 90 risk variants in the GWAS meta-analysis had an OR of 0.8–1.2. However, a novel genome-wide significant association signal in the *PLPP4* locus (rs12771445) was identified at an OR of 0.64. The significant variant was an intron variant in the gene Phospholipid Phosphatase 4 (PLPP4) and was imputed with Rsq=0.99 and MAF=0.31. The estimated MAF

was 0.27 in the patient group and 0.36 in the control group, as compared to 0.32 in the reference population SweGen (222). We also identified a haplotype in the same region being significantly associated with PD, following the same pattern with being more common in the control group (OR=0.69). Furthermore, we investigated the cumulative effect of less common variants (MAF<5%) in *PLPP4* on PD risk. Due to the few coding variants in the data, we adapted the non-burden SKAT and observed a significant association with PD risk for the genotyped variants. A potential joint effect of variants in *PLPP4* on PD risk was observed also in the imputed data but did not pass multiple testing correction.

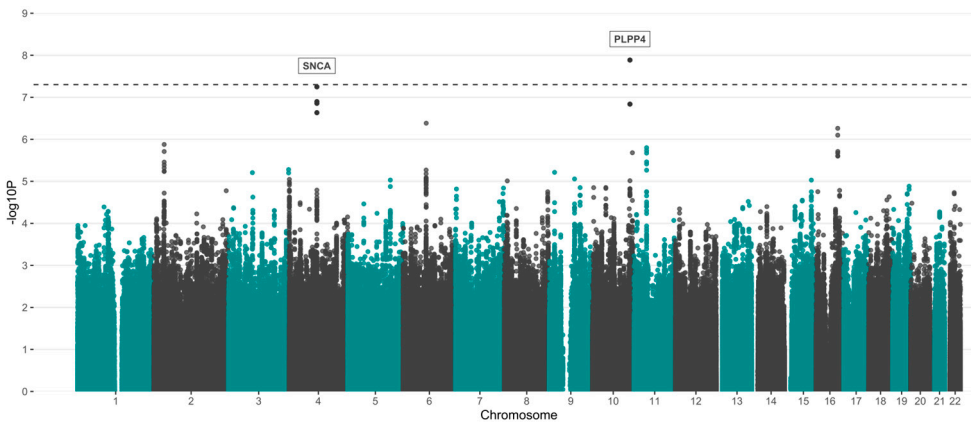

Figure 11: Manhattan plot showing the result from PD GWA analysis. A total of 5,445,841 SNPs (MAF>5%) were tested for 929 PD patients vs 935 controls. The y-axis represents the negative log (two-sided *p*-values) for association of variants with PD and the x-axis represent the genomic position on genome build GRCh37. The horizontal dashed line indicates the genome-wide significance level (*p*=5E-08)

Genetic PD risk profile

The effect estimates of the 90 risk variants from Nalls *et al.* GWAS meta-analysis (116) was used to calculate PRS (here named GRS) in the cohort, revealing an increase in OR by 1.8 by each standard deviation (SD) increase from the reference mean value (Figure 12A).We also observed that study participants in the highest GRS quartile were estimated to be 4.7 times more likely to have PD compared to the participants in the lowest quartile. There was also an inverse association between the GRS and AAD, where one SD increase in GRS was associated with approximately one-year earlier AAD (Figure 12B, C). The lack of an association with PD in this Swedish cohort for multiple of the 90 risk variants could have an impact on the fit of the model. However, the results are in line with previous results in other larger cohorts (116, 119, 260). Although the model is functional enough to compare risk for disease status at a population level, caution should be taken to not evaluate an individual's risk based on GRS due to the complexity of the disease etiology.

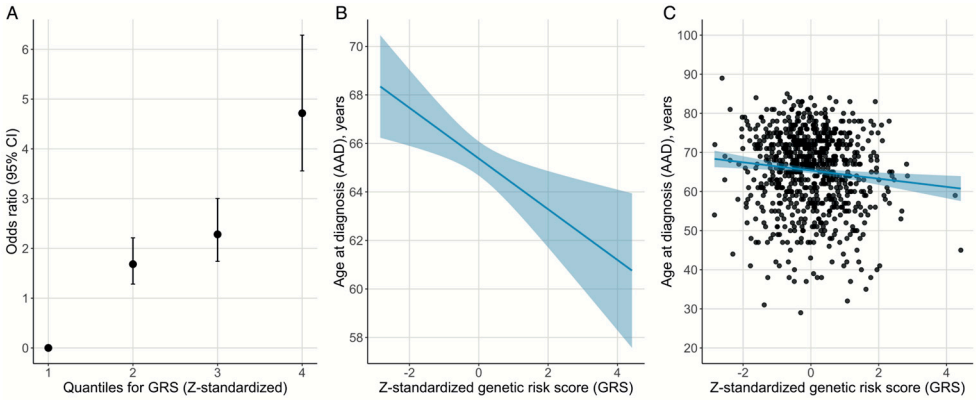

Figure 12: Genetic risk score (GRS) quartiles vs. disease status (A) and age at diagnosis (AAD) (B and C). A) Odds ratio of PD status per risk quartile of the Z-standardized GRS. B) Regression line for the association between the Z-standardized GRS and AAD. The line represents the parameter estimate and the shading the 95% CI of the regression model. The model was adjusted for sex, PD family history, and PC1-5. C) AAD and Z-standardized GRS for each study participant with the regression model in plot B fitted to the plot.

In summary, this work provides a comprehensive description of a PD case-control study from southern Sweden and contributes to the knowledge of environmental and genetic risk factors in PD in the Swedish population. We nominate a novel PD risk variant at the *PLPP4* locus which has not been reported to be associated with PD previously. It could be that this common variant is a tagging a population-specific rare variant as the Swedish population contain a substantial number of genetic variants that are not represented in other European populations (222). Hence, subsequent studies are needed to validate whether *PLPP4* is associated with PD within the Swedish, and other, populations.

Paper IV | Investigating the potential association of *PLPP4* with Parkinson's disease in Scandinavian and European cohorts

The results described in **Paper III** indicate a role of *PLPP4* in PD risk and even if the study had sufficient statistical power to detect risk variants of that effect size, replication of associations discovered by GWASs is essential to rule out false associations and improve effect estimates (283). Based on previous results, we hypothesized that the association is population- specific to the Swedish population. In **Paper IV (manuscript)**, we therefore aimed to further investigate the potential association between variants in *PLPP4* and PD risk in Swedish, Norwegian, and mixed European cohorts.

PLPP4 (also named *PPAPDC1A*) encodes for a phospholipase phosphatase (PLPP), which belongs to a superfamily of integral membrane glycoproteins with broad substrate specificity. These glycoproteins have multiple functions, including catalyzing the dephosphorylation of several bioactive lipid mediators such as phosphatidic acid (PA), lysophosphatidic acid and sphingosine 1-phosphate, to attenuate cell activation (284). A potential link between these pathways and PD has previously been suggested where e.g., *DGKQ*, encoding a diacylglycerol kinase regulating PA content, has been reported as a potential PD risk factor (285, 286).

The Swedish PD GWAS top hits at the *PLPP4* locus, rs12771445 and rs11596921 (in LD) were analyzed for an association to PD. An association between the variant rs11596921 and PD was confirmed in the previously investigated MPBC cohort by genotyping, showing similar result as the imputed GWAS genotypes with an OR=0.70 (Table 4). No genotyping result could be obtained for rs12771445 due to genotyping difficulties. An association was also observed in a cohort from Oslo, Norway with OR=0.78, indicating a lower risk of having PD among individuals carrying the minor allele of the variant in these populations. However, no association was observed in the in two Swedish case-cohorts from the regions of Umeå (UMU) or Stockholm (KI), nor in the meta-analysis of the Norwegian Oslo cohort and the three Swedish case-control study cohorts (MPBC, UMU, and KI). However, a trend of a higher MAF in the control group was observed for all cohorts except for the UMU cohort. We also observed that the allele frequency of the two variants varies between two different Swedish reference populations (SweGen and ACpop), with a lower MAF in the cohort from the more northern part of Sweden (ACpop). Also, the MAF for the variants were substantially higher among patients in the Norwegian cohort ParkWest as compared to the patients in the other Norwegian Oslo cohort and in the Swedish MPBC cohort. These results indicate that the frequency of the variants varies within the populations of Sweden and Norway (287). Large genetic differences have been observed within Sweden, particularly between the northern and southern parts of the country, and a similar

genetic structure in terms of inter-country differences has been reported in Norway (287, 288). To fully elucidate the potential association between *PLPP4* and PD in Sweden and Norway, larger and potential regional case-control studies are necessary. We additionally show through PCAs, combining the cohorts with data from the 1000 Genome Project (1KGP), that the Swedish MPBC cohort is genetically different from other mixed European cohorts in AMP-PD, which also could explain the observed differences in variant frequency between the cohorts.

Table 4: Overview of the analyzed cohorts, divided by patients and controls, and the minor allele frequencies (MAF) of rs11596921 and rs12771445

	MPBC	**KI**	**UMU**	**Oslo**	**AMP-PD**	**GP2**	**ParkWest**	**SweGen**	**ACpop**
Ancestry	Swedish	Swedish	Swedish	Norwegian	European	European	Norwegian	Swedish	Swedish
Type of cohort	Case-control	Case-Control	Case-Control	Case-Control	Case-Control	Case-Control	Patients only	Controls only	Controls only
Patients									
N	896	494	862	479	2454	4101	147	NA	NA
AAD (mean, SD)	65.5 (10.0)	59.4 (10.7)	63.3 (10.7)	55.5 (12.1)	60.5 (10.5)	58.3 (11.4)	67.1* (9.3)	NA	NA
Age (mean, SD)	71.0 (8.3)	66.9 (10.2)	77.9 (9.6)	65.3 (9.3)	64.3 (9.4)	63.0 (11.5)	67.3 (9.3)	NA	NA
Women (%)	35.5	39.6**	39.8	35.9	36.3	37.0	36.7	NA	NA
rs11596921 $MAF_{(C)}$	0.285	0.304	0.317	0.292	0.324	0.338	0.371	NA	NA
rs12771445 $MAF_{(T)}$	0.273	NA	NA	0.285	0.308	0.322	0.378	NA	NA
Controls									
N	897	661	1256	464	3118	2311	NA	1000	300
Age (mean, SD)	72.3 (8.3)	74.9 *(7.3)	78.6 (9.4)	61.8 (11.0)	69.8 (13.0)	68.0 (8.8)	NA	65.2***	N.A.†
Women (%)	36.0	53.8**	42.2	42.5	51.3	53.4	NA	49.4	50.0
rs11596921 $MAF_{(C)}$	0.363	0.310	0.309	0.358	0.330	0.341	NA	0.331	0.310
rs12771445 $MAF_{(T)}$	0.359	NA	NA	0.346	0.312	0.325	NA	0.321	0.297
rs11596921									
OR	0.70	0.97	1.04	0.78	0.98	0.97	NA	NA	NA
95% CI	0.61-0.81	0.81-1.16	0.91-1.19	0.63-0.96	0.90-1.07	0.89-1.06	NA	NA	NA
p-value	**1.34E-06**	0.737	0.548	**2.08E-02**	0.621	0.515	NA	NA	NA
rs12771445									
OR	0.67	NA	NA	0.79	0.98	0.98	NA	NA	NA
95% CI	0.58-0.77	NA	NA	0.64-0.97	0.90-1.07	0.90-1.07	NA	NA	NA
p-value	**2.29E-08**	NA	NA	**2.54E-02**	0.635	0.682	NA	NA	NA

AMP-PD: Age at diagnosis was available for 1517 patients (63.3%). GP2: Age was available for 2055 controls (88.9%) and 3286 patients (80.1%). Age at diagnosis was available for 1301 patients (31.72%). No Swedish samples in the data release. MPBC: Age at diagnosis available for 792 patients (85.3%). KI: Age was available for 320 patients (64.8%) and 308 controls (46.6%). Age at diagnosis was available for 320 patients (64.7%). *Age at onset. **Information on sex was available for 78.6% of the cohort: 65.0% of the PD patients and 88.8% of the controls. ***Median age. †To be considered for inclusion in ACpop, an individual had to have reached an age of at least 80 years. AMP-PD: Accelerating Medicines Partnership in Parkinson's Disease, GP2: Global Parkinson's Genetics Program, KI: Karolinska Institutet, MPBC: Multipark's biobank sample collection, UMU: Umeå cohort.

Using open access resources (i.e., the Human Protein Atlas), we observed that *PLPP4* RNA is tissue enriched in the brain, however with low region specificity. Protein expression was identified to be cytoplasmic and membranous expression in the central nervous system and RNA single cell type specificity showed enrichment in oligodendrocyte precursor cells. The *PLPP4* gene was also identified to be part of a cluster with 344 other genes involved in nervous system development. An eQTL analysis revealed that rs12771445, representing the *PLPP4* locus, is a significant eQTL for *PLPP4* and the neighboring transcript *C10orf85* (an RNA gene affiliated with the long-noncoding RNA [lncRNA] class) specifically in the cerebellum and in the cerebellar hemisphere (Table 5). The result showed that the major allele (C) of rs12771445, which was more common in the patient group in MPBC, increased the expression of *PLPP4* and *C10orf85* in the tissues. Similar results were observed for all tested variants, including rs11596921.

PLPPs are involved in the generation of the lipid-signaling molecule diacylglycerol (DAG) (289) and PD patients have been reported to have significantly increased levels of DAG in the frontal cortex compared to controls. It has also been seen that changes in the levels of DAG occur early in PD and are amplified as the disease progresses (290). Damage both in the deep nuclei and white matter of the cerebellum is seen in PD and the cerebellum plays a role in motor symptoms, particularly tremor, but also in nonmotor symptoms such as cognitive deficits (291, 292). An investigation of DAG levels in other brain regions such as the cerebellum in PD would therefore be of interest.

Our eQTL analysis also showed that rs12771445 is a tissue-specific cis-eQTL for the neighboring transcript of *PLPP4*, *C10orf85*. *C10orf85* has not been reported to be associated with PD but lncRNA have important roles in many biological processes and significant changes in expression of 87 different lncRNAs have been identified in the substantia nigra of PD patients (293, 294).

Table 5: eQTL lookup of the *PLPP4* variant rs12771445 from GTEx in brain tissues

Gene symbol	P-value	FDR	Tissue
PLPP4	7.205E-09	5.417E-44	Cerebellar Hemisphere
C10orf85	2.061E-06	1.762E-44	Cerebellar Hemisphere
C10orf85	2.789E-06	1.633E-52	Cerebellum
PLPP4	2.527E-05	1.850E-50	Cerebellum

eQTL = expression quantitative trait locus; GTEx = Genotype-Tissue Expression Project

Cis-regulation of gene expression has been proposed to be a major mechanism behind a large proportion of genetic associations to PD. A study testing 12 risk genes identified by a PD GWAS, allele-specific expression in brain tissue samples was observed for 9 genes (295). Another study investigating all 90 PD risk variants from

the latest PD GWAS meta-analysis observed that 76 (84%) of the risk variants were identified as eQTLs. Hence, variants that alter gene expression could play an important role in neurological disorders, where altered gene expression in different tissues could affect disease risk (296)

Our findings show that the MPBC cohort is genetically different from other cohorts of mixed European ancestries and the intracountry variations within the Swedish and Norwegian populations may contribute to the observed variances in allele frequencies between various Scandinavian cohorts (222, 287). As a result, it is not possible to rule out a potential population-specific influence of *PLPP4* on PD risk in southern Sweden and the Oslo region of Norway, and further research into the *PLPP4* gene in PD is necessary.

Paper V | Genome-wide association study of REM sleep behavior disorder identifies polygenic risk and brain expression effects

It is estimated that over 70% of patients with the sleep disorder RBD will convert within 10–15 years (on average) to PD or other forms of synucleinopathies. Additionally, RBD is linked to more malignant forms of synucleinopathies (24, 297). The genetics of RBD had previously only been investigated through candidate gene analyses and not through hypothesis-free methods. We therefore performed GWASs and a GWAS meta-analysis with a total of 2,843 RBD patients and 139,636 controls to better understand the genetics of RBD and early synucleinopathies.

Genetic risk variants for RBD

Two novel RBD-associated loci were identified in *SCARB2* and *INPP5F* and already known RBD associations could be confirmed near *SNCA* (231), *TMEM175* (298), and *GBA* (299, 300) (Figure 13 and Table 6). These loci have all been implicated in PD, however, the identified variants in *SNCA* and *SCARB2* associated with RBD are not in LD with the top PD-associated variants at these loci. Therefore, they were considered as independent, suggesting that RBD-associated synucleinopathy only have a partial overlapping genetic background with PD. The common variants identified in the GWAS were estimated to explain approximately 12.3% of the heritable iRBD risk, indicating that a large part of the genetic component of RBD remains unidentified.

We further examined if an RBD specific PRS can distinctly identify RBD as opposed to PD without RBD, by using an RBD PRS profile with the variants identified in this meta-analysis. The PRS could differentiate between iRBD cases and controls were individuals in the top quartile for the RBD PRS were 2.9 times more likely to have RBD. The performance was similar for PD patients with pRBD (PD+pRBD), where individuals with PD+pRBD in the top PRS quartile were 2.4 times more likely to have PD+pRBD compared to individuals in the lowest quartile. The RBD PRS did not perform as well for PD patients without RBD (PD-pRBD), and when comparing PD+pRBD to PD–pRBD, the RBD PRS was not a strong predictor. It is likely that this RBD PRS do not capture the genetic difference between PD with and without RBD and further investigations of this difference are therefore needed.

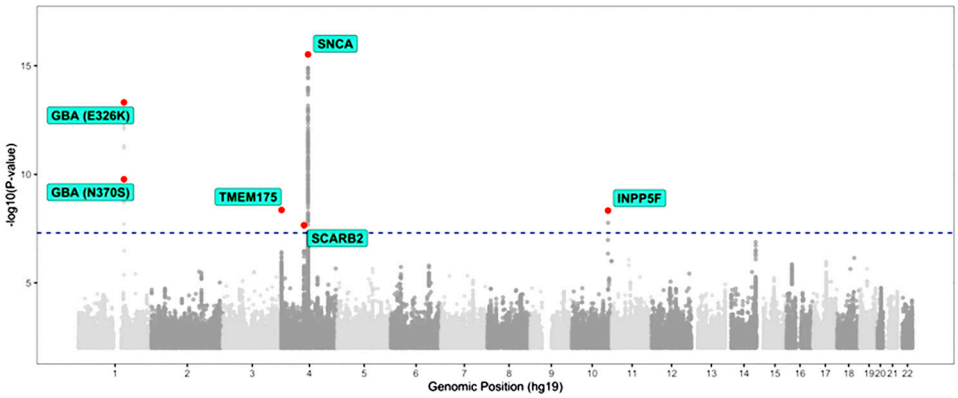

Figure 13: The Manhattan plot highlights the 6 genome-wide association study (GWAS)-nominated loci after the REM sleep bahaviour disorder (RBD) meta-analysis. GWAS was performed as repeated logistic regression across the genome, adjusted for age, sex, and principal components. Each point represents the log adjusted *p*-value at each genomic site. A locus was considered significant if the two-sided *p*-value was less than the corrected GWAS-significant *p*-value threshold of 5E-08, visualized in this plot with the dashed line. The points in red show the top variant at that locus, as well as any secondary independent associations.

Table 6: Independent RBD risk loci nominated by GWAS meta-analysis

Position (GRCh37)	rsID	Closest gene	EA	RA	EAF	OR	95% CI	p-value	Het I2 (%)
4:90757272	rs3756059	*SNCA*	A	G	0.5	1.26	1.19–1.33	3.02E-16	94.4
1:155205378	rs12752133	*GBA*	T	C	0.01	2.09	1.73–2.54	4.87E-14	0
1:155205634	rs76763715	*GBA*	C	T	0.004	2.84	2.06–3.92	1.68E-10	0
4:951947	rs34311866	*TMEM175*	C	T	0.19	1.22	1.14–1.31	4.41E-09	0
10:121536327	rs117896735	*INPP5F/BAG3*	A	G	0.02	1.80	1.48–2.19	4.70E-09	0
4:77132634	rs7697073	*SCARB2*	T	C	0.34	1.18	1.11–1.25	2.21E-08	0

EA=Effect allele; EAF=Effect allele frequency; Het=Heterogeneity; OR=Odds ratio

Colocalization analyses demonstrate tissue and cell-specific differential effects of RBD-associated variants

Colocalization analyses were done to investigate whether risk variants for RBD also were associated with gene expression in the brain or the blood. Strong evidence for colocalization in the *SNCA* locus with *SNCA*antisense-1 (*SNCA-AS1*) expression in the brain was observed. Variants in the region surrounding *SNCA-AS1* showed an inverse relationship between RBD risk and *SNCA-AS1* expression, indicating that an increased *SNCA-AS1* expression is associated with reduced RBD risk.

Evidence of colocalization in the *SNCA* locus with *MMRN1* expression in the blood was also observed. *MMRN1* encodes the Multimerin1 protein, a factor V/Va-binding protein found in platelets and in the endothelium of blood vessels. Since both the *SNCA-AS1* and *MMRN1* colocalizations were observed at the same RBD risk locus but in different tissues, a tissue-specific regulation of the expression was

hypothesized. We therefore explored the tissue- and cell-type-specific patterns of *SNCA-AS* and *MMRN1* expression and observed that *SNCA-AS1* expression was predominantly brain-specific, while *MMRN* expression was most specific to thyroid, adipose and lung tissues and least specific to brain tissues. At a cellular level, *SNCA-AS* demonstrated neuronal specificity, while *MMRN1* expression was specific to microglia in data of cell types derived from human middle temporal gyrus.

Differential eQTL effects in different brain regions could explain the independent associations of SNCA in RBD and PD

SNCA and *SCARB2* are GWAS-nominated risk loci for both PD and RBD, but the associations are driven by independent variants. We therefore hypothesized that the different variants at these loci may be associated with different patterns of expression of their respective genes in different brain regions. Variants in *SNCA* have previously been associated with *SNCA-AS1* expression (301). Therefore, the effect of the genetic variants on the expression of *SCARB2*, *SNCA* and *SNCA-AS1* was examined (using eQTL data from GTEx). The colocalization analysis indicated that the RBD locus is driven by *SNCA-AS1* expression since the RBD *SNCA* variant (rs3756059) correlated strongly with decreased *SNCA-AS1* expression in the cerebellum, cerebellar hemisphere, frontal cortex, and anterior cingulate cortex. The PD *SNCA* variant (rs356182) was only significantly associated with decreased expression in the anterior cingulate cortex (Figure 14). The RBD risk variant at the *SCARB2* locus was most strongly associated with increased expression in the cortex while the PD risk variant at the same locus was most strongly associated with increased expression in the substantia nigra. The differential association for PD and RBD at the *SNCA* locus may provide a mechanistic hypothesis for gene expression-dependent regional vulnerability of different brain areas.

Pathway analysis reveals potential role for the autophagy-lysosomal pathway in RBD pathogenesis

Pathway enrichment analysis was performed using the GWAS-nominated genes for cellular components and biological processes to examine potential pathway enrichments. The nominated cellular components included whole membrane, lysosome, and vacuole whereas the biological processes included positive regulation of receptor recycling, vacuole organization, and endocytosis. These results suggest an involvement of the autophagy lysosomal pathway which is a key mechanism for clearing α-syn (302, 303). This pathway has a major role also in PD and DLB and involves the enzyme encoded by *GBA,* glucocerebrosidase. In fact, four of the five nominated RBD-associated GWAS genes in this study directly engage with glucocerebrosidase function, highlighting its potential role in RBD pathology.

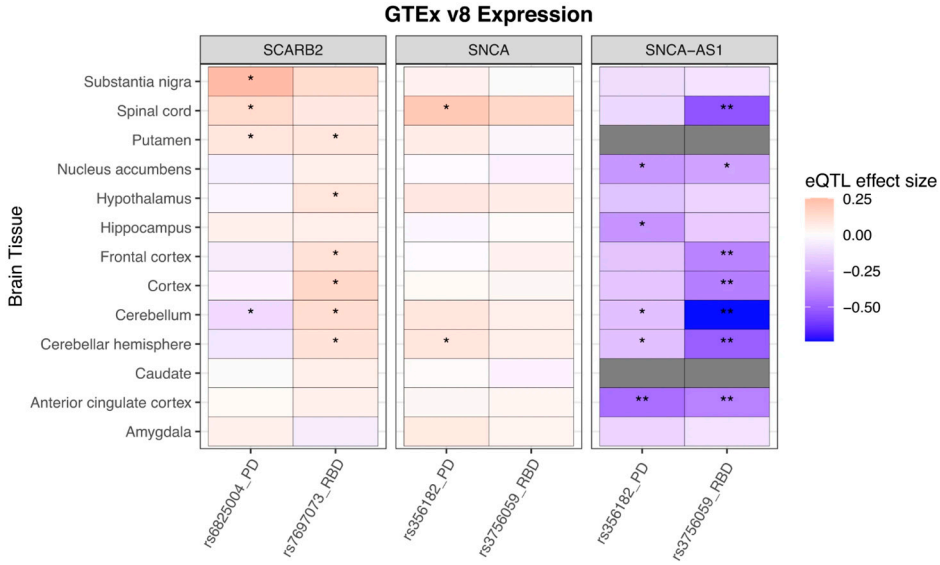

Figure 14: eQTL data from GTEx version 8 for RBD and PD top variants in differing loci. The effect sizes represent the slope of linear regression on normalized gene expression data versus the genotype status using single-tissue eQTL analysis, performed by the GTEx consortium. Nominal associations are indicated with * ($p < 0.05$) while FDR-corrected significant associations are indicated with **. Dark gray indicates missing data. GTEx: Genotype-Tissue Expression Consortium, v8: version 8, RBD:REM sleep behavior disorder, PD:Parkinson's disease, eQTL: expression quantitative trait loci.

Genetic differences between PD, iRBD, and PD+pRBD

Given the observed differential associations in the *SNCA* and *SCARB2* loci in PD and RBD, we wanted to further examine how PD GWAS loci behave in RBD. A list of nominated synucleinopathy variants was composed based on GWAS results in PD (116), DLB (301), PD AAO (119), and RBD and the effects of the variants were compared between the PD variants vs the variants of this RBD GWASs (iRBD, PD+pRBD, and the meta-analysis). Additionally, the potential genetic overlap between PD and DLB was evaluated. Figure 15 shows the similarities and differences in effect size and direction of the variants between the different diseases. A difference was observed at some key PD loci (Figure 15a) for iRBD, including *SNCA*, *CYLD*, and *FYN* where the direction of effect in iRBD was opposite of that seen in PD (however without corrected significance in iRBD). In difference, all the PD+pRBD variants showed the same direction of effect with PD (Figure 15b), as did all DLB nominally significant variants (Figure 15d). However, important PD loci, including *SNCA* 3' and *LRRK2* were not significant for PD+pRBD in our study, despite sufficient power. DLB also genetically deviates from PD as variants in *MAPT*, *LRRK2*, and *SNCA* 3' not were statistically significant in a previous DLB GWAS (301). When running similar analyses using PD AAO summary statistics, most variants increasing the risk of any of the synucleinopathies were associated with an earlier PD AAO. However interestingly, the PD *SCARB2* variant was nominally associated with an earlier AAO of PD but a decreased risk of iRBD.

Additionally, the *SNCA* variant rs356182 was associated with an increased risk and earlier AAO of PD but had a potential protective effect in iRBD.

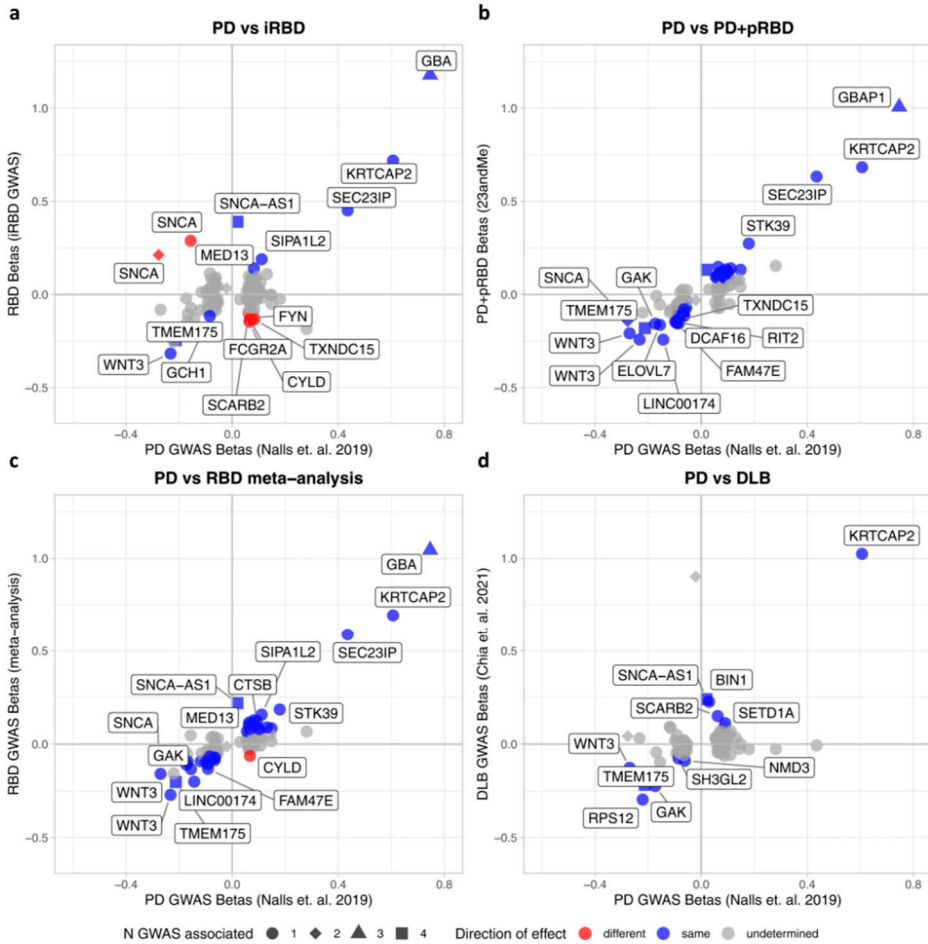

Figure 15: Beta-beta plots comparing the significance and direction of variants from synucleinopathy genome-wide association study (GWAS) summary statistics to the latest PD GWAS. PD GWAS summary statistics was compared to (a) iRBD, (b) PD+RBD, (c) RBD meta-analysis, and (d) previously published DLB summary statistics. Colored points indicate variants with the same (blue) or opposite (red) direction of effect in both studies, with a nominally significant p-value ($p < 0.05$) in their respective GWAS. Gray points indicate those with undetermined direction (non-significant). The shapes of the points indicate the number of synucleinopathy GWAS where the locus reaches GWAS significance (counting PD, PD age at onset, DLB, and this RBD meta-analysis). Gene names indicate the closest gene to the represented variant. PD: Parkinson's disease, RBD: REM sleep behavior disorder, GWAS: genome-wide association study, pRBD:probable RBD, DLB: dementia with Lewy bodies.

LD-score regression was used to examine the genetic correlation between RBD and other diseases (i.e., T2D and AD) and exposures (i.e., smoking). However, most of the results were only observed with nominal significance and did not pass multiple

corrections. Nevertheless, the results reflect that there is a potential difference between iRBD and PD+pRBD. Although iRBD and PD+pRBD were positively correlated at a nominal significance, PD+pRBD was seen to be strongly correlated with PD whereas iRBD was not. In addition, iRBD was potentially genetically correlated with T2D, whereas PD+pRBD correlated with AD. Since these results were not observed after multiple corrections, no conclusions could be made. Additionally, it has previously been reported that those with T2D are at increased risk for AD and that the two diseases share genetic risk architecture (304, 305). We also observed that all RBD cohorts showed genetic correlation, although again at a nominate level, with variables such as less smoking, similar to what has been seen in PD.

Despite the inclusion of PD patients with pRBD in this study, known PD risk loci such as *LRRK2* and *MAPT* were not observed in the GWAS results, despite sufficient power to detect these variants at genome-wide significance (if they had comparable effects for RBD as in the PD GWAS results). Similar goes for DLB GWAS loci such as *BIN1* and *APOE*. The results of this study cannot rule out associations between these loci and RBD, but it is evident that they do not play an important role in iRBD or RBD-related PD. This further support that RBD is a distinct disease subtype, both clinically and genetically. The results of this study support the heterogenous nature of PD, showing that the subtypes/diseases likely include different genetic subgroups, where some are associated with variants in *LRRK2*, *MAPT*, and *APOE*, while the subgroup defined by having iRBD prior to PD onset (or DLB), is not (Figure 16). The results therefore suggest that the genetic background of RBD, PD and DLB only partially overlap, and larger RBD studies is needed to further elucidate the genetic architecture of RBD and the link to PD.

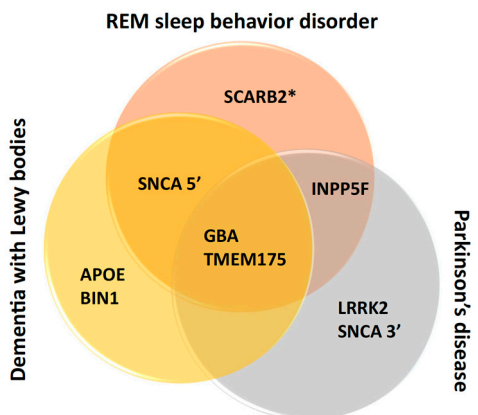

Figure 16: Key GWAS-significant loci across three synucleinopathies. Only *GBA* and *TMEM175* are shared between all three, both of which play a role in the autophagy-lysosomal pathway. *SNCA* plays a role in PD, DLB, and RBD risk, yet the strongest risk locus for PD is at the 3' end of the gene while RBD and DLB share a risk locus at the 5' end. Similarly, *SCARB2* is a risk factor for PD as well as RBD, however, the RBD locus is independent of the variant identified for PD risk (as indicated by the asterisk in the figure)

Concluding remarks and future perspectives

This thesis further highlights the complexity of PD etiology. I truly believe that worldwide collaborations are necessary to accelerate the identification and understanding of factors contributing to PD pathology and etiology. My work further emphasizes that factors contributing to PD can vary between populations, even when viewed on a European scale. However, the work in this thesis also highlights the importance of studying PD in different population to understand how the disease varies across populations. Our findings for apparent monogenic PD revealed that:

- *The prevalence of known pathogenic mutations in the dominant PD genes LRRK2 and SNCA rarely contribute to PD in Sweden.*
- *The suggested autosomal dominant PD gene RIC3 is not associated with PD in cohorts of European, Latin American, or East Asian ancestry.*

These findings indicate the need of studying monogenic PD at a regional level in different populations. Taking Sweden as an example, the genomes of Swedes have been reported to contain many genetic variants that appear to not be present in other European cohorts and substantial genetic differences have been reported even within the country (222, 288). This has an impact also on the genetics contributing to PD. For the rather common *LRRK2* G2019S, the frequency has been reported to vary between countries in Europe, with a higher frequency of more admixed populations in southern Europe as compared to countries in northern Europe, such as Sweden, strengthening our report (63). Differences between Sweden and other European countries has also been reported for the T369M variants in *GBA* where it has been associated with an increased risk for PD in a meta-analysis, but the variant does not appear to be a risk factor for PD in Sweden (306, 307).

Our findings for apparent idiopathic PD further support the need of studying PD in different countries, but also points out the relevance of conducting studies at a local/regional level within a country, since we observed that:

- *Both non-genetic and genetic risk factors for PD can be population-specific, as we observed a new potential genetic risk factor for PD in the PLPP4 locus which appear to be region-specific.*

- *The Swedish moist tobacco snus was associated with a decreased risk of having PD in a PD case-control study from southern Sweden.*

Much of my work involves analyzes of PD patients on a group level, despite the heterogeneity in PD and the importance of not viewing PD as a single entity. One of the difficulties is to define these subtypes and to obtain sufficient sample sizes to conduct e.g., GWASs. Multiple difficulties exist when trying to define subtypes of PD, including the fact that motor subtypes can shift during the disease progression (14). Until we have the knowledge to differentiate subtypes of PD more clearly, it is necessary to study PD at a group level, including all potential subtypes of the disease.

Furthermore, to understand PD etiology, it is important to try to study subtypes of prodromal PD. In this thesis, we studied the genetics of RBD, as this is a well-defined characteristic and more than 70% of individuals with RBD will convert to an overt synucleinopathy, such as PD (24). In this thesis work, we observed that:

- *There are potential separate genetic backgrounds to RBD, PD, and other synucleinopathies such as DLB, supporting that RBD is a distinct subtype, both genetically and clinically.*
- *Apparent idiopathic PD have genetic subtypes, where those without iRBD prior to PD onset are associated with variants in e.g., LRRK2 whereas subgroups defined as having iRBD prior to PD onset are not.*

These findings support the need of studying the prodromal phase of PD to further understand PD etiology and the biological pathways involved. Individuals at high risk of PD, such as those with iRBD, would be of high interest to include in prospective studies evaluating non-genetic risk factors associated with phenoconversion to PD. This to understand what differentiates the individuals converting to PD or other synucleinopathies from the ones that do not. In fact, a recent study of 319 RBD patients investigated this, and observed a phenoconversion to a NDD of 46% after a mean follow-up of 5.8 years, whereof around 50% developed PD (308). They observed that individuals with older age and those who used nitrate derivates (drugs for short-term symptomatic relief of angina pectoris) where more likely to develop an NDD. Interestingly though, prior exposure to pesticides, a well-known risk factor for PD, was associated with a lower NDD phenoconversion risk. However, among the converters, the use of nonoccupational herbicides was significantly associated with the development of PD rather than other NDD (308). A question arises whether this study had sufficient statistical power to observe true associations and the findings further support the need of larger prospective studies to better understand non-genetic factors contributing to PD etiology. In fact, a study protocol for an Italian multicenter longitudinal study (FARPRESTO) was recently published with the primary aim to stratify the risk of

phenoconversion in a cohort of incident and prevalent individuals with iRBD which might provide us with important knowledge in the future (309).

A highly important issue that has not been highlighted in this thesis is the need of diversification in PD research. The understanding of PD has to this point largely been centered on research in individuals of European ancestry. Therefore, we do not know if the current knowledge about PD etiology is generalizable to all populations. Efforts are therefore needed, and underway (e.g., the GP2), to diversify studies on genetics in PD, particularly in underrepresented populations from around the world (218, 310). Similar efforts are needed to study non-genetic risk factors in PD, to identify clear, and preferably modifiable risk factors for PD. To my knowledge, initiatives such as the MDS Epidemiology Study Group and the Genetic Epidemiology of PD (GEoPD) consortium have been made to investigate epidemiology and genetic epidemiology in PD (311). Additionally, the GP2 aim to serve as a hub for investigators to form collaborations to also explore the aspects of non-genetic risk factors in PD, including epidemiology and environmental exposures (218). However, considering the present and future rise in the global burden of PD, this research needs to be further prioritized.

In 2022, the World Health Organization launched a technical brief, outlining the global burden, treatment gaps and crucial areas for actions in PD. This was followed by a publication identifying six main action steps with the focus to address the global disparities in PD, where one of those steps was "prevention and risk reduction" with suggestions of measures to take (312). For example, it emphasizes the importance of protective measures when using pesticides associated with PD, the need of screens for toxic effects prior to releases of new pesticides to the market and replacing hazardous pesticides with safer alternatives (312). Much is needed to be done at a societal level to reduce environmental toxins linked to increased PD risk, such as pesticides and air pollutants to try to reduce the increased global burden of the disease. On an individual level, measures such as increased physical activity and coffee consumption could be easy actions to take to potentially reduce the risk. However, the causal effect of other potential risk factors in PD needs to be determined to develop interventions and to take preventative actions.

Studying genetics can provide us with important knowledge on disease mechanisms and potentially identify novel drug targets. Recently, the Telomer-to-Telomer (T2T) Consortium presented the complete sequence of a human genome (T2T-CHM13), containing 3.055 billion base pairs, completing the sequence for the 8% of the genome that is missing from the GRCh38 build due to past technological limitations (313). It will truly be interesting to see the outcome of this novel information on future studies of genetics in PD where genetic variations across the entire human genome can be analyzed for the first time. Strengthening the importance of studying genetics for drug discovery was a recent publication reporting that 66% of the FDA-approved drugs in 2021 was supported by genetic evidence, i.e., that genes coding for the drug targets, or for proteins that interacted with the target previously had

been associated with the drug indication, or a closely related phenotype (314). Additionally, despite not being within the scope of this thesis, I believe that gene-environment interactions also play a major role in PD etiology and future studies using hypothesis-free approaches are crucial for our further understanding of the mechanisms underlying PD etiology and pathology.

In conclusion. studies of non-genetic and genetic factors, including their interactions, contributing to PD etiology are crucial to understand the mechanisms underlying the disease and to identify and develop predictive, preventative, and new therapeutical approaches.

References

1. Focus on neurodegenerative disease. *Nature Neuroscience*. 2018;21(10):1293.
2. Jellinger KA. Synucleinopathies. Editors: Kompoliti K, Metman LV. *Encyclopedia of Movement Disorders*. Oxford: Academic Press; 2010. p. 203-207.
3. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*. 2019;18:459-480.
4. James Parkinson, Member of the Royal College of Surgeons. An Essay on the Shaking Palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 2002;14(2):223-236.
5. GBD 2016 Parkinson's Diseaase Collaborators. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*. 2018;17(11):939-953.
6. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, et al. Parkinson disease. *Nature Reviews Disease Primers*. 2017;3(1):17013.
7. Bloem BR, Okun MS, Klein C. Parkinson's disease. *The Lancet*. 2021;397(10291):2284-2303.
8. Marsili L, Rizzo G, Colosimo C. Diagnostic Criteria for Parkinson's Disease: From James Parkinson to the Concept of Prodromal Disease. *Front Neurol*. 2018;9:156.
9. Kalia LV, Lang AE. Parkinson's disease. *The Lancet*. 2015;386(9996):896-912.
10. McGregor MM, Nelson AB. Circuit Mechanisms of Parkinson's Disease. *Neuron*. 2019;101(6):1042-1056.
11. De Pablo-Fernández E, Lees AJ, Holton JL, Warner TT. Prognosis and Neuropathologic Correlation of Clinical Subtypes of Parkinson Disease. *JAMA Neurology*. 2019;76(4):470-479.
12. von Coelln R, Shulman LM. Clinical subtypes and genetic heterogeneity: of lumping and splitting in Parkinson disease. *Current Opinion in Neurology*. 2016;29(6):727-734.
13. Thenganatt MA, Jankovic J. Parkinson disease subtypes. *JAMA Neurology*. 2014;71(4):499-504.
14. Eisinger RS, Hess CW, Martinez-Ramirez D, Almeida L, Foote KD, Okun MS, et al. Motor subtype changes in early Parkinson's disease. *Parkinsonism & related disorders*. 2017;43:67-72.
15. Marras C, Lang A. Parkinson's disease subtypes: lost in translation? *Journal of Neurology, Neurosurgery & Psychiatry*. 2013;84(4):409-415.
16. Titova N, Padmakumar C, Lewis SJG, Chaudhuri KR. Parkinson's: a syndrome rather than a disease? *Journal of Neural Transmission.* 2017;124(8):907-914.
17. Weiner WJ. There Is No Parkinson Disease. *Archives of Neurology.* 2008;65(6):705-708.

18. Farrow SL, Cooper AA, O'Sullivan JM. Redefining the hypotheses driving Parkinson's diseases research. *NPJ Parkinson's disease.* 2022;8(1):45.
19. Berg D, Borghammer P, Fereshtehnejad S-M, Heinzel S, Horsager J, Schaeffer E, et al. Prodromal Parkinson disease subtypes — key to understanding heterogeneity. *Nature Reviews Neurology.* 2021;17(6):349-361.
20. Stern MB, Lang A, Poewe W. Toward a redefinition of Parkinson's disease. *Movement Disorders*. 2012;27(1):54-60.
21. Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. *Nature Reviews Disease Primers.* 2015;1(1):15056.
22. Fereshtehnejad SM, Yao C, Pelletier A, Montplaisir JY, Gagnon JF, Postuma RB. Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. *Brain*. 2019;142(7):2051-2067.
23. Berg D, Postuma RB, Adler CH, Bloem BR, Chan P, Dubois B, et al. MDS research criteria for prodromal Parkinson's disease. *Movement Disorders.* 2015;30(12):1600-1611.
24. Postuma RB, Iranzo A, Hu M, Högl B, Boeve BF, Manni R, et al. Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*. 2019;142(3):744-759.
25. Dauvilliers Y, Schenck CH, Postuma RB, Iranzo A, Luppi P-H, Plazzi G, et al. REM sleep behaviour disorder. *Nature Reviews Disease Primers.* 2018;4(1):19.
26. Vendette M, Gagnon JF, Décary A, Massicotte-Marquez J, Postuma RB, Doyon J, et al. REM sleep behavior disorder predicts cognitive impairment in Parkinson disease without dementia. *Neurology*. 2007;69(19):1843-1849.
27. Armstrong MJ, Okun MS. Diagnosis and Treatment of Parkinson Disease: A Review. *JAMA*. 2020;323(6):548-560.
28. Barbuti PA, Barker RA, Brundin P, Przedborski S, Papa SM, Kalia LV, et al. Recent Advances in the Development of Stem-Cell-Derived Dopaminergic Neuronal Transplant Therapies for Parkinson's Disease. *Movement Disorders*. 2021;36(8):1772-1780.
29. Pringsheim T, Jette N, Frolkis A, Steeves TDL. The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders.* 2014;29(13):1583-1590.
30. GBD 2016 Mortality Collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1084-1150.
31. Dorsey ER, Bloem BR. The Parkinson Pandemic—A Call to Action. *JAMA Neurology*. 2018;75(1):9-10.
32. Willis AW, Roberts E, Beck JC, Fiske B, Ross W, Savica R, et al. Incidence of Parkinson disease in North America. *NPJ Parkinson's disease.* 2022;8(1):170.
33. Ou Z, Pan J, Tang S, Duan D, Yu D, Nong H, et al. Global Trends in the Incidence, Prevalence, and Years Lived With Disability of Parkinson's Disease in 204 Countries/Territories From 1990 to 2019. *Frontiers in Public Health.* 2021;9.
34. Mehanna R, Smilowska K, Fleisher J, Post B, Hatano T, Pimentel Piemonte ME, et al. Age Cutoff for Early-Onset Parkinson's Disease: Recommendations from the International Parkinson and Movement Disorder Society Task Force on Early Onset Parkinson's Disease. *Movement Disorders Clinical Practice.* 2022;9(7):869-878.

35. Riboldi GM, Frattini E, Monfrini E, Frucht SJ, Di Fonzo A. A Practical Approach to Early-Onset Parkinsonism. *Journal of Parkinson's disease.* 2022;12(1):1-26.
36. Golbe LI. Young-onset Parkinson's disease: a clinical review. *Neurology*. 1991;41(2 (Pt 1)):168-173.
37. Quinn N, Critchley P, Marsden CD. Young onset Parkinson's disease. *Movement Disorders.* 1987;2(2):73-91.
38. Zirra A, Rao SC, Bestwick J, Rajalingam R, Marras C, Blauwendraat C, et al. Gender Differences in the Prevalence of Parkinson's Disease. *Movement Disorders Clinical Practice*. 2023;10:86-93.
39. Wooten GF, Currie LJ, Bovbjerg VE, Lee JK, Patrie J. Are men at greater risk for Parkinson's disease than women? *Journal of Neurology, Neurosurgery & Psychiatry*. 2004;75:637-639.
40. Baldereschi M, Di Carlo A, Rocca WA, Vanni P, Maggi S, Perissinotto E, et al. Parkinson's disease and parkinsonism in a longitudinal study: two-fold higher incidence in men. ILSA Working Group. Italian Longitudinal Study on Aging. *Neurology*. 2000;55(9):1358-1363.
41. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017;169(7):1177-1186.
42. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human Genetics*. 2013;132(10):1077-1130.
43. Duvoisin RC. Is Parkinson's disease acquired or inherited? *Canadian Journal of Neurological Sciences.* 1984;11(1 Suppl):151-155.
44. Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiology of disease*. 2020;137:104782.
45. Klein C, Westenberger A. Genetics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine.* 2012;2(1):a008888.
46. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *The Lancet Neurology*. 2020;19(2):170-178.
47. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science.* 1997;276(5321):2045-2047.
48. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. *Nature*. 1997;388(6645):839-840.
49. Spellman GG. Report of familial cases of parkinsonism. Evidence of a dominant trait in a patient's family. *JAMA*. 1962;179:372-374.
50. Waters CH, Miller CA. Autosomal dominant Lewy body parkinsonism in a four-generation family. *Annals of Neurology.* 1994;35(1):59-64.
51. Muenter MD, Forno LS, Hornykiewicz O, Kish SJ, Maraganore DM, Caselli RJ, et al. Hereditary form of parkinsonism--dementia. *Annals of Neurology.* 1998;43(6):768-781.
52. Farrer M, Gwinn-Hardy K, Muenter M, DeVrieze FW, Crook R, Perez-Tur J, et al. A chromosome 4p haplotype segregating with Parkinson's disease and postural tremor. *Human Molecular Genetics.* 1999;8(1):81-85.

53. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003;302(5646):841.
54. Devine MJ, Gwinn K, Singleton A, Hardy J. Parkinson's disease and α-synuclein expression. *Movement Disorders*. 2011;26(12):2160-2168.
55. Miller DW, Hague SM, Clarimon J, Baptista M, Gwinn-Hardy K, Cookson MR, et al. Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. *Neurology*. 2004;62(10):1835-1838.
56. Book A, Guella I, Candido T, Brice A, Hattori N, Jeon B, et al. A Meta-Analysis of α-Synuclein Multiplication in Familial Parkinsonism. *Frontiers in Neurology.* 2018;9:1021.
57. Ibáñez P, Bonnet AM, Débarges B, Lohmann E, Tison F, Pollak P, et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet*. 2004;364(9440):1169-1171.
58. Chartier-Harlin MC, Kachergus J, Roumier C, Mouroux V, Douay X, Lincoln S, et al. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet*. 2004;364(9440):1167-1169.
59. Blauwendraat C, Makarious MB, Leonard HL, Bandres-Ciga S, Iwaki H, Nalls MA, et al. A population scale analysis of rare SNCA variation in the UK Biobank. *Neurobiology of disease.* 2021;148:105182.
60. Burré J. The Synaptic Function of α-Synuclein. *Journal of Parkinson's disease.* 2015;5(4):699-713.
61. Visanji NP, Brotchie JM, Kalia LV, Koprich JB, Tandon A, Watts JC, et al. α-Synuclein-Based Animal Models of Parkinson's Disease: Challenges and Opportunities in a New Era. *Trends in Neurosciences.* 2016;39(11):750-762.
62. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *The Lancet Neurology.* 2008;7(7):583-590.
63. Correia Guedes L, Ferreira JJ, Rosa MM, Coelho M, Bonifati V, Sampaio C. Worldwide frequency of G2019S LRRK2 mutation in Parkinson's disease: A systematic review. *Parkinsonism & related disorders.* 2010;16(4):237-242.
64. Hulihan MM, Ishihara-Paul L, Kachergus J, Warren L, Amouri R, Elango R, et al. LRRK2 Gly2019Ser penetrance in Arab-Berber patients from Tunisia: a case-control genetic study. *The Lancet Neurology.* 2008;7(7):591-594.
65. Orr-Urtreger A, Shifrin C, Rozovski U, Rosner S, Bercovich D, Gurevich T, et al. The LRRK2 G2019S mutation in Ashkenazi Jews with Parkinson disease: is there a gender effect? *Neurology.* 2007;69(16):1595-1602.
66. Ozelius LJ, Senthil G, Saunders-Pullman R, Ohmann E, Deligtisch A, Tagliati M, et al. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *New England Journal of Medicine.* 2006;354(4):424-425.
67. Lee AJ, Wang Y, Alcalay RN, Mejia-Santana H, Saunders-Pullman R, Bressman S, et al. Penetrance estimate of LRRK2 p.G2019S mutation in individuals of non-Ashkenazi Jewish ancestry. *Movement Disorders.* 2017;32(10):1432-1438.
68. Jeong GR, Lee BD. Pathological Functions of LRRK2 in Parkinson's Disease. *Cells*. 2020;9(12).
69. West AB, Moore DJ, Biskup S, Bugayenko A, Smith WW, Ross CA, et al. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *PNAS.* 2005;102(46):16842-16847.

70. Wojewska DN, Kortholt A. LRRK2 Targeting Strategies as Potential Treatment of Parkinson's Disease. *Biomolecules*. 2021;11(8).
71. Di Maio R, Hoffman EK, Rocha EM, Keeney MT, Sanders LH, De Miranda BR, et al. LRRK2 activation in idiopathic Parkinson's disease. *Science Translational Medicine.* 2018;10(451).
72. Kingwell K. LRRK2-targeted Parkinson disease drug advances into phase III. *Nature Reviews Drug Discovery.* 2023;22(1):3-5.
73. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, Offman MN, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *American Journal of Human Genetics*. 2011;89(1):168-175.
74. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 mutations in Parkinson disease. *American Journal of Human Genetics.* 2011;89(1):162-167.
75. Sharma M, Ioannidis JPA, Aasly JO, Annesi G, Brice A, Bertram L, et al. A multi-centre clinico-genetic analysis of the VPS35 gene in Parkinson disease indicates reduced penetrance for disease-associated variants. *Journal of Medical Genetics*. 2012;49(11):721.
76. Klein C, Hattori N, Marras C. MDSGene: Closing Data Gaps in Genotype-Phenotype Correlations of Monogenic Parkinson's Disease. *Journal of Parkinson's disease.* 2018;8(s1):25-30.
77. Sassone J, Reale C, Dati G, Regoni M, Pellecchia MT, Garavaglia B. The Role of VPS35 in the Pathobiology of Parkinson's Disease. Cellular and Molecular *Neurobiology*. 2021;41(2):199-227.
78. Kilarski LL, Pearson JP, Newsway V, Majounie E, Knipe MDW, Misbahuddin A, et al. Systematic Review and UK-Based Study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. *Movement Disorders.* 2012;27(12):1522-1529.
79. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*. 1998;392(6676):605-608.
80. Lücking CB, Dürr A, Bonifati V, Vaughan J, De Michele G, Gasser T, et al. Association between Early-Onset Parkinson's Disease and Mutations in the Parkin Gene. *New England Journal of Medicine.* 2000;342(21):1560-1567.
81. Kasten M, Hartmann C, Hampf J, Schaake S, Westenberger A, Vollstedt E-J, et al. Genotype-Phenotype Relations for the Parkinson's Disease Genes Parkin, PINK1, DJ1: MDSGene Systematic Review. *Movement Disorders*. 2018;33(5):730-741.
82. Puschmann A. Monogenic Parkinson's disease and parkinsonism: Clinical phenotypes and frequencies of known mutations. *Parkinsonism & related disorders.* 2013;19(4):407-415.
83. Zhu W, Huang X, Yoon E, Bandres-Ciga S, Blauwendraat C, Billingsley KJ, et al. Heterozygous PRKN mutations are common but do not increase the risk of Parkinson's disease. *Brain*. 2022;145(6):2077-2091.
84. Narendra D, Tanaka A, Suen DF, Youle RJ. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *Journal of Cell Biology.* 2008;183(5):795-803.

85. Quinn PMJ, Moreira PI, Ambrósio AF, Alves CH. PINK1/PARKIN signalling in neurodegeneration and neuroinflammation. *Acta Neuropathologica Communications.* 2020;8(1):189.
86. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science.* 2004;304(5674):1158-1160.
87. Puschmann A, Fiesel FC, Caulfield TR, Hudec R, Ando M, Truban D, et al. Heterozygous PINK1 p.G411S increases risk of Parkinson's disease via a dominant-negative mechanism. *Brain*. 2017;140(1):98-117.
88. Krohn L, Grenn FP, Makarious MB, Kim JJ, Bandres-Ciga S, Roosen DA, et al. Comprehensive assessment of PINK1 variants in Parkinson's disease. *Neurobiology of aging.* 2020;91:168.e1-.e5.
89. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science.* 2003;299(5604):256-259.
90. Borsche M, Pereira SL, Klein C, Grünewald A. Mitochondria and Parkinson's Disease: Clinical, Molecular, and Translational Aspects. *Journal of Parkinson's disease.* 2021;11(1):45-60.
91. Repici M, Giorgini F. DJ-1 in Parkinson's Disease: Clinical Insights and Therapeutic Perspectives. *Journal of Clinical Medicine.* 2019;8(9): 1377.
92. Gan-Or Z, Amshalom I, Kilarski LL, Bar-Shira A, Gana-Weisz M, Mirelman A, et al. Differential effects of severe vs mild GBA mutations on Parkinson disease. *Neurology*. 2015;84(9):880-887.
93. Riboldi GM, Di Fonzo AB. GBA, Gaucher Disease, and Parkinson's Disease: From Genetic to Clinic to New Therapeutic Approaches. *Cells*. 2019;8(4):364.
94. Beutler E, Gelbart T, Scott CR. Hematologically important mutations: Gaucher disease. *Blood Cells, Molecules, and Diseases.* 2005;35(3):355-364.
95. Rosenbloom B, Balwani M, Bronstein JM, Kolodny E, Sathe S, Gwosdow AR, et al. The incidence of Parkinsonism in patients with type 1 Gaucher disease: data from the ICGG Gaucher Registry. *Blood Cells, Molecules, and Diseases*. 2011;46(1):95-102.
96. Liu G, Boot B, Locascio JJ, Jansen IE, Winder-Rhodes S, Eberly S, et al. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Annals of Neurology.* 2016;80(5):674-685.
97. Anheim M, Elbaz A, Lesage S, Durr A, Condroyer C, Viallet F, et al. Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. *Neurology*. 2012;78(6):417-420.
98. Rana HQ, Balwani M, Bier L, Alcalay RN. Age-specific Parkinson disease risk in GBA mutation carriers: information for genetic counseling. *Genetics in Medicine.* 2013;15(2):146-149.
99. Balestrino R, Tunesi S, Tesei S, Lopiano L, Zecchinelli AL, Goldwurm S. Penetrance of Glucocerebrosidase (GBA) Mutations in Parkinson's Disease: A Kin Cohort Study. *Movement Disorders.* 2020;35(11):2111-2114.
100. Sudhaman S, Muthane UB, Behari M, Govindappa ST, Juyal RC, Thelma BK. Evidence of mutations in RIC3 acetylcholine receptor chaperone as a novel cause of autosomal-dominant Parkinson's disease with non-motor phenotypes. *Journal of Medical Genetics*. 2016;53(8):559-566.

101. Gotti C, Clementi F. Neuronal nicotinic receptors: from structure to pathology. *Progress in Neurobiology*. 2004;74(6):363-396.
102. Quik M, Kulak JM. Nicotine and Nicotinic Receptors; Relevance to Parkinson's Disease. *NeuroToxicology*. 2002;23(4):581-594.
103. Ross JP, Dupré N, Dauvilliers Y, Strong S, Dionne-Laporte A, Dion PA, et al. RIC3 variants are not associated with Parkinson's disease in French-Canadians and French. *Neurobiology of aging.* 2017;53:194.e9-.e11.
104. He D, Hu P, Deng X, Song Z, Yuan L, Yuan X, et al. Genetic analysis of the RIC3 gene in Han Chinese patients with Parkinson's disease. *Neuroscience Letters.* 2017;653:351-354.
105. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science.* 2005;308(5720):385-389.
106. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genetics.* 2009;41(12):1308-1312.
107. Lill CM. Genetics of Parkinson's disease. *Molecular and Cellular Probes.* 2016;30(6):386-396.
108. Krüger R, Vieira-Saecker AM, Kuhn W, Berg D, Müller T, Kühnl N, et al. Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype. *Annals of Neurology.* 1999;45(5):611-617.
109. Biskup S, Mueller JC, Sharma M, Lichtner P, Zimprich A, Berg D, et al. Common variants of LRRK2 are not associated with sporadic Parkinson's disease. *Annals of Neurology.* 2005;58(6):905-908.
110. Skipper L, Li Y, Bonnard C, Pavanni R, Yih Y, Chua E, et al. Comprehensive evaluation of common genetic variation within LRRK2 reveals evidence for association with sporadic Parkinson's disease. *Human Molecular Genetics*. 2005;14(23):3549-3556.
111. Paisán-Ruíz C, Lang AE, Kawarai T, Sato C, Salehi-Rad S, Fisman GK, et al. LRRK2 gene in Parkinson disease: mutation analysis and case control association study. *Neurology.* 2005;65(5):696-700.
112. Golbe LI, Lazzarini AM, Spychala JR, Johnson WG, Stenroos ES, Mark MH, et al. The tau A0 allele in Parkinson's disease. *Movement Disorders.* 2001;16(3):442-447.
113. Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, et al. Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. *JAMA*. 2001;286(18):2245-2250.
114. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS genetics.* 2012;8(3):e1002548.
115. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *New England Journal of Medicine*. 2009;361(17):1651-1661.
116. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology.* 2019;18(12):1091-1102.

117. Singleton A, Hardy J. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. *Human Molecular Genetics.* 2011;20(R2):R158-162.
118. Nguyen HP, Van Broeckhoven C, van der Zee J. ALS Genes in the Genomic Era and their Implications for FTD. *Trends in Genetics.* 2018;34(6):404-423.
119. Blauwendraat C, Heilbron K, Vallerga CL, Bandres-Ciga S, von Coelln R, Pihlstrom L, et al. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and alpha-synuclein mechanisms. *Movement Disorders.* 2019;34(6):866-875.
120. Grover S, Kumar Sreelatha AA, Pihlstrom L, Domenighetti C, Schulte C, Sugier P-E, et al. Genome-wide Association and Meta-analysis of Age at Onset in Parkinson Disease. Evidence From the COURAGE-PD Consortium. *Neurology*. 2022;99(7):e698-e710.
121. Liu G, Peng J, Liao Z, Locascio JJ, Corvol JC, Zhu F, et al. Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease. *Nature genetics.* 2021;53(6):787-793.
122. Tan MMX, Lawton MA, Jabbari E, Reynolds RH, Iwaki H, Blauwendraat C, et al. Genome-Wide Association Studies of Cognitive and Motor Progression in Parkinson's Disease. *Movement Disorders.* 2021;36(2):424-433.
123. Iwaki H, Blauwendraat C, Leonard HL, Kim JJ, Liu G, Maple-Grødem J, et al. Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. *Movement Disorders.* 2019;34(12):1839-1850.
124. Alfradique-Dunham I, Al-Ouran R, von Coelln R, Blauwendraat C, Hill E, Luo L, et al. Genome-Wide Association Study Meta-Analysis for Parkinson Disease Motor Subtypes. *Neurology Genetics.* 2021;7(2):e557.
125. Müller SH, Girard SL, Hopfner F, Merner ND, Bourassa CV, Lorenz D, et al. Genome-wide association study in essential tremor identifies three new loci. *Brain*. 2016;139(Pt 12):3163-3169.
126. Periñán MT, Brolin K, Bandres-Ciga S, Blauwendraat C, Klein C, Gan-Or Z, et al. Effect modification between genes and environment, and Parkinson's disease risk. *Annals of Neurology.* 2022;92(5):715–724.
127. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *The Lancet Neurology.* 2006;5(6):525-535.
128. Gillies GE, Pienaar IS, Vohra S, Qamhawi Z. Sex differences in Parkinson's disease. *Frontiers in Neuroendocrinology.* 2014;35(3):370-384.
129. Jurado-Coronel JC, Cabezas R, Ávila Rodríguez MF, Echeverria V, García-Segura LM, Barreto GE. Sex differences in Parkinson's disease: Features on clinical symptoms, treatment outcome, sexual hormones and genetics. *Frontiers in Neuroendocrinology.* 2018;50:18-30.
130. Blauwendraat C, Iwaki H, Makarious MB, Bandres-Ciga S, Leonard HL, Grenn FP, et al. Investigation of Autosomal Genetic Sex Differences in Parkinson's Disease. *Annals of Neurology.* 2021;90(1):35-42.
131. Grenn FP, Makarious MB, Bandres-Ciga S, Iwaki H, Singleton AB, Nalls MA, et al. Analysis of Y chromosome haplogroups in Parkinson's disease. *Brain Communications*. 2022;4(6).
132. Le Guen Y, Napolioni V, Belloy ME, Yu E, Krohn L, Ruskey JA, et al. Common X-Chromosome Variants Are Associated with Parkinson Disease Risk. *Annals of Neurology*. 2021;90(1):22-34.

133. Pereira G, Dória S. X-chromosome inactivation: implications in human disease. *Journal of Genetics.* 2021;100.
134. Subramanian I, Mathur S, Oosterbaan A, Flanagan R, Keener AM, Moro E. Unmet Needs of Women Living with Parkinson's Disease: Gaps and Controversies. *Movement Disorders.* 2022;37(3):444-455.
135. Li X, Li W, Liu G, Shen X, Tang Y. Association between cigarette smoking and Parkinson's disease: A meta-analysis. *Archives of Gerontology and Geriatrics.* 2015;61(3):510-516.
136. Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism & related disorders*. 2016;23:1-9.
137. Mappin-Kasirer B, Pan H, Lewington S, Kizza J, Gray R, Clarke R, et al. Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors. *Neurology*. 2020;94(20):e2132-e8.
138. Heilbron K, Jensen MP, Bandres-Ciga S, Fontanillas P, Blauwendraat C, Nalls MA, et al. Unhealthy Behaviours and Risk of Parkinson's Disease: A Mendelian Randomisation Study. *Journal of Parkinson's disease.* 2021;11(4):1981-1993.
139. Ritz B, Lee PC, Lassen CF, Arah OA. Parkinson disease and smoking revisited: ease of quitting is an early sign of the disease. *Neurology*. 2014;83(16):1396-1402.
140. Gatto NM, Deapen D, Bordelon Y, Marshall S, Bernstein L, Ritz B. Passive smoking and Parkinson's disease in California Teachers. *Parkinsonism & related disorders.* 2017;45:44-49.
141. Yang F, Pedersen NL, Ye W, Liu Z, Norberg M, Forsgren L, et al. Moist smokeless tobacco (Snus) use and risk of Parkinson's disease. *International Journal of Epidemiology.* 2017;46(3):872-880.
142. Quik M, Perez XA, Bordia T. Nicotine as a potential neuroprotective agent for Parkinson's disease. *Movement Disorders.* 2012;27(8):947-957.
143. Quik M, Wonnacott S. α6β2* and α4β2* nicotinic acetylcholine receptors as drug targets for Parkinson's disease. *Pharmacological reviews.* 2011;63(4):938-966.
144. Lieberman A, Lockhart TE, Olson MC, Smith Hussain VA, Frames CW, Sadreddin A, et al. Nicotine Bitartrate Reduces Falls and Freezing of Gait in Parkinson Disease: A Reanalysis. *Frontiers in Neurology.* 2019;10:424.
145. Vieregge A, Sieberer M, Jacobs H, Hagenah JM, Vieregge P. Transdermal nicotine in PD: a randomized, double-blind, placebo-controlled study. *Neurology*. 2001;57(6):1032-1035.
146. Watts RL, Jankovic J, Waters C, Rajput A, Boroojerdi B, Rao J. Randomized, blind, controlled trial of transdermal rotigotine in early Parkinson disease. *Neurology*. 2007;68(4):272-276.
147. Rossi A, Berger K, Chen H, Leslie D, Mailman RB, Huang X. Projection of the prevalence of Parkinson's disease in the coming decades: Revisited. *Movement Disorders.* 2018;33(1):156-159.
148. Dai X, Gakidou E, Lopez AD. Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. *Tobacco Control.* 2022;31(2):129.
149. Nefzger MD, Quadfasel FA, Karl VC. A retrospective study of smoking in Parkinson's disease. *American Journal of Epidemiology.* 1968;88(2):149-158.

150. Kahn HA. The Dorn study of smoking and mortality among U.S. veterans: report on eight and one-half years of observation. *National Cancer Institute monograph,*. 1966;19:1-125.
151. Zhang D, Jiang H, Xie J. Alcohol intake and risk of Parkinson's disease: A meta-analysis of observational studies. *Movement Disorders*. 2014;29(6):819-822.
152. Shao C, Wang X, Wang P, Tang H, He J, Wu N. Parkinson's Disease Risk and Alcohol Intake: A Systematic Review and Dose-Response Meta-Analysis of Prospective Studies. *Frontiers in Nutrition,* 2021;8:709846.
153. Mitchell E, Chohan H, Bestwick JP, Noyce AJ. Alcohol and Parkinson's Disease: A Systematic Review and Meta-Analysis. *Journal of Parkinson's disease.* 2022;12(8):2369-2381.
154. Schwarzschild MA. Caffeine in Parkinson disease: better for cruise control than snooze patrol? *Neurology*. 2012;79(7):616-618.
155. Fuxe K, Ungerstedt U. Action of caffeine and theophyllamine on supersensitive dopamine receptors: considerable enhancement of receptor response to treatment with DOPA and dopamine receptor agonists. *Medical Biology.* 1974;52(1):48-54.
156. Chen Y, Sun X, Lin Y, Zhang Z, Gao Y, Wu IXY. Non-Genetic Risk Factors for Parkinson's Disease: An Overview of 46 Systematic Reviews. *Journal of Parkinson's disease.* 2021;11(3):919-935.
157. Domenighetti C, Sugier PE, Sreelatha AAK, Schulte C, Grover S, Mohamed O, et al. Mendelian Randomisation Study of Smoking, Alcohol, and Coffee Drinking in Relation to Parkinson's Disease. *Journal of Parkinson's disease.* 2022;12(1):267-282.
158. Larsson SC, Woolf B, Gill D. Plasma Caffeine Levels and Risk of Alzheimer’s Disease and Parkinson’s Disease: Mendelian Randomization Study. *Nutrients*. 2022;14(9):1697.
159. Jiang W, Ju C, Jiang H, Zhang D. Dairy foods intake and risk of Parkinson's disease: a dose-response meta-analysis of prospective cohort studies. *European Journal of Epidemiology.* 2014;29(9):613-619.
160. Fardell C, Torén K, Schiöler L, Nissbrandt H, Åberg M. High IQ in Early Adulthood Is Associated with Parkinson’s Disease. *Journal of Parkinson's disease.* 2020;Preprint:1-8.
161. Fang X, Han D, Cheng Q, Zhang P, Zhao C, Min J, et al. Association of Levels of Physical Activity With Risk of Parkinson Disease: A Systematic Review and Meta-analysis. *JAMA Network Open.* 2018;1(5):e182421.
162. Gamborg M, Hvid LG, Dalgas U, Langeskov-Christensen M. Parkinson's disease and intensive exercise therapy - An updated systematic review and meta-analysis. *Acta Neurologica Scandinavica.* 2022;145(5):504-528.
163. van der Marck MA, Dicke HC, Uc EY, Kentin ZH, Borm GF, Bloem BR, et al. Body mass index in Parkinson's disease: a meta-analysis. *Parkinsonism & related disorders.* 2012;18(3):263-267.
164. Wang YL, Wang YT, Li JF, Zhang YZ, Yin HL, Han B. Body Mass Index and Risk of Parkinson's Disease: A Dose-Response Meta-Analysis of Prospective Studies. *PLoS One.* 2015;10(6):e0131778.
165. Noyce AJ, Kia DA, Hemani G, Nicolas A, Price TR, De Pablo-Fernandez E, et al. Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. *PLoS Medicine.* 2017;14(6):e1002314.

166. Kashihara K. Weight loss in Parkinson's disease. *Journal of Neurology.* 2006;253 Suppl 7:VII38-41.
167. Przedborski S, Vila M. MPTP: a review of its mechanisms of neurotoxicity. *Clinical Neuroscience Research*. 2001;1(6):407-418.
168. Schober A. Classic toxin-induced animal models of Parkinson's disease: 6-OHDA and MPTP. *Cell and Tissue Research.* 2004;318(1):215-224.
169. Breckenridge CB, Berry C, Chang ET, Sielken RL, Jr., Mandel JS. Association between Parkinson's Disease and Cigarette Smoking, Rural Living, Well-Water Consumption, Farming and Pesticide Use: Systematic Review and Meta-Analysis. *PLoS One.* 2016;11(4):e0151841.
170. Gatto NM, Cockburn M, Bronstein J, Manthripragada AD, Ritz B. Well-water consumption and Parkinson's disease in rural California. *Environmental Health Perspectives.* 2009;117(12):1912-1918.
171. Hassaan MA, El Nemr A. Pesticides pollution: Classifications, human health impact, extraction and treatment techniques. *The Egyptian Journal of Aquatic Research.* 2020;46(3):207-220.
172. Costello S, Cockburn M, Bronstein J, Zhang X, Ritz B. Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California. *American Journal of Epidemiology.* 2009;169(8):919-926.
173. Tanner CM, Kamel F, Ross GW, Hoppin JA, Goldman SM, Korell M, et al. Rotenone, paraquat, and Parkinson's disease. *Environmental Health Perspectives.* 2011;119(6):866-872.
174. Sherer TB, Richardson JR, Testa CM, Seo BB, Panov AV, Yagi T, et al. Mechanism of toxicity of pesticides acting at complex I: relevance to environmental etiologies of Parkinson's disease. *Journal of Neurochemistry.* 2007;100(6):1469-1479.
175. Betarbet R, Sherer TB, MacKenzie G, Garcia-Osuna M, Panov AV, Greenamyre JT. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature Neuroscience.* 2000;3(12):1301-1306.
176. Sandodden R, Aune S, Bardal H, Adolfsen P, Nøst T. Rotenone application and degradation following eradication of invasive roach *Rutilus rutilus* in three Norwegian lakes. *Management of Biological Invasions.* 2022;13(1):233-245.
177. Norris EH, Uryu K, Leight S, Giasson BI, Trojanowski JQ, Lee VM. Pesticide exposure exacerbates alpha-synucleinopathy in an A53T transgenic mouse model. *The American Journal of Pathology,.* 2007;170(2):658-666.
178. Thiruchelvam M, McCormack A, Richfield EK, Baggs RB, Tank AW, Di Monte DA, et al. Age-related irreversible progressive nigrostriatal dopaminergic neurotoxicity in the paraquat and maneb model of the Parkinson's disease phenotype. *The European Journal of Neuroscience.* 2003;18(3):589-600.
179. Kervégant M, Merigot L, Glaizal M, Schmitt C, Tichadou L, de Haro L. Paraquat poisonings in France during the European ban: experience of the Poison Control Center in Marseille. J*ournal of Medical Toxicology.* 2013;9(2):144-147.
180. Teichholtz H. Breaking News: EPA to Reconsider Herbicide Paraquat Michael J Fox Foundation2022 [updated December 23 2022; cited January 30 2023]. Available from: https://www.michaeljfox.org/news/breaking-news-epa-reconsider-herbicide-paraquat

181. Paul KC, Krolewski RC, Moreno EL, Blank J, Holton K, Ahfeldt T, et al. Coupling comprehensive pesticide-wide association study to iPSC dopaminergic screening identifies and classifies Parkinson-relevant pesticides. *bioRxiv.* 2022:2022.02.06.479305.
182. Li S, Ritz B, Gong Y, Cockburn M, Folle AD, Del Rosario I, et al. Proximity to residential and workplace pesticides application and the risk of progression of Parkinson's diseases in Central California. *Science of The Total Environment.* 2023;864:160851.
183. Cheng P, Yu J, Huang W, Bai S, Zhu X, Qi Z, et al. Dietary intake of iron, zinc, copper, and risk of Parkinson's disease: a meta-analysis. *Neurol Sci.* 2015;36(12):2269-2275.
184. Ascherio A, Chen H, Weisskopf MG, O'Reilly E, McCullough ML, Calle EE, et al. Pesticide exposure and risk for Parkinson's disease. *Annals of Neurology*. 2006;60(2):197-203.
185. Murata H, Barnhill LM, Bronstein JM. Air Pollution and the Risk of Parkinson's Disease: A Review. *Movement Disorders.* 2022;37(5):894-904.
186. Babatola SS. Global burden of diseases attributable to air pollution. *Journal of public health in Africa*. 2018;9(3):813.
187. Faustino PR, Duarte GS, Chendo I, Castro Caldas A, Reimão S, Fernandes RM, et al. Risk of Developing Parkinson Disease in Bipolar Disorder: A Systematic Review and Meta-analysis. *JAMA Neurology.* 2020;77(2):192-198.
188. Liou YJ, Bai YM, Chen MH, Cheng CM, Chen TJ, Tsai SJ. Obsessive-Compulsive Disorder and Risk of Parkinson Disease: A Nationwide Longitudinal Study. *Psychosomatic Medicine.* 2022;84(9):1096-1102.
189. Kuusimäki T, Al-Abdulrasul H, Kurki S, Hietala J, Hartikainen S, Koponen M, et al. Increased Risk of Parkinson's Disease in Patients With Schizophrenia Spectrum Disorders. *Movement Disorders.* 2021;36(6):1353-1361.
190. Bandres-Ciga S, Blauwendraat C, Singleton AB. Assessment of Genetic Association Between Parkinson Disease and Bipolar Disorder. *JAMA Neurology.* 2020;77(8):1034-1035.
191. Weimers P, Halfvarson J, Sachs MC, Saunders-Pullman R, Ludvigsson JF, Peter I, et al. Inflammatory Bowel Disease and Parkinson's Disease: A Nationwide Swedish Cohort Study. *Inflammatory Bowel Diseases.* 2019;25(1):111-123.
192. Chohan H, Senkevich K, Patel RK, Bestwick JP, Jacobs BM, Bandres Ciga S, et al. Type 2 Diabetes as a Determinant of Parkinson's Disease Risk and Progression. *Movement Disorders.* 2021;36(6):1420-1429.
193. Herrick MK, Tansey MG. Is LRRK2 the missing link between inflammatory bowel disease and Parkinson’s disease? *NPJ Parkinson's disease.* 2021;7(1):26.
194. Athauda D, Foltynie T. Insulin resistance and Parkinson's disease: A new target for disease modification? *Progress in Neurobiology.* 2016;145-146:98-120.
195. Vijiaratnam N, Girges C, Auld G, Chau M, Maclagan K, King A, et al. Exenatide once weekly over 2 years as a potential disease-modifying treatment for Parkinson's disease: protocol for a multicentre, randomised, double blind, parallel group, placebo controlled, phase 3 trial: The 'Exenatide-PD3' study. *BMJ Open.* 2021;11(5):e047993.
196. Bai S, Song Y, Huang X, Peng L, Jia J, Liu Y, et al. Statin Use and the Risk of Parkinson's Disease: An Updated Meta-Analysis. *PLoS One.* 2016;11(3):e0152564.

197. Zhu Y, Pu J, Chen Y, Zhang B. Decreased risk of Parkinson's disease in diabetic patients with thiazolidinediones therapy: An exploratory meta-analysis. *PLoS One.* 2019;14(10):e0224236.
198. Gao X, Chen H, Schwarzschild MA, Ascherio A. Use of ibuprofen and risk of Parkinson disease. *Neurology.* 2011;76(10):863-869.
199. Lang Y, Gong D, Fan Y. Calcium channel blocker use and risk of Parkinson's disease: a meta-analysis. *Pharmacoepidemiology and drug safety.* 2015;24(6):559-566.
200. Noyce AJ, Bestwick JP, Silveira-Moriyama L, Hawkes CH, Giovannoni G, Lees AJ, et al. Meta-analysis of early nonmotor features and risk factors for Parkinson disease. *Annals of Neurology.* 2012;72(6):893-901.
201. Smeyne RJ, Noyce AJ, Byrne M, Savica R, Marras C. Infection and Risk of Parkinson's Disease. *Journal of Parkinson's disease.* 2021;11(1):31-43.
202. Levine KS, Leonard HL, Blauwendraat C, Iwaki H, Johnson N, Bandres-Ciga S, et al. Virus exposure and neurodegenerative disease risk across national biobanks. *Neuron*. 2023: S0896-6273(22)01147-3.
203. Cocoros NM, Svensson E, Szépligeti SK, Vestergaard SV, Szentkúti P, Thomsen RW, et al. Long-term Risk of Parkinson Disease Following Influenza and Other Infections. *JAMA Neurology.* 2021;78(12):1461-1470.
204. Liu B, Fang F, Ye W, Wirdefeldt K. Appendectomy, Tonsillectomy and Parkinson's Disease Risk: A Swedish Register-Based Study. *Frontiers in Neurology.* 2020;11:510.
205. Lee PC, Ahmed I, Loriot MA, Mulot C, Paul KC, Bronstein JM, et al. Smoking and Parkinson disease: Evidence for gene-by-smoking interactions. *Neurology*. 2018;90(7):e583-e592.
206. Liu Y, Bai H, Gen S, Zhang H, Wang S, Hua L, et al. Interaction between SNCA gene polymorphisms and T2DM with Parkinson's disease. *Acta Neurologica Scandinavica.* 2020;142(5):443-448.
207. Yu E, Ambati A, Andersen MS, Krohn L, Estiar MA, Saini P, et al. Fine mapping of the HLA locus in Parkinson's disease in Europeans. *NPJ Parkinson's disease.* 2021;7(1):84.
208. Domenighetti C, Douillard V, Sugier P-E, Sreelatha AAK, Schulte C, Grover S, et al. The Interaction between HLA-DRB1 and Smoking in Parkinson's Disease Revisited. *Movement Disorders.* 2022;37(9):1929-1937.
209. Kannarkat GT, Cook DA, Lee JK, Chang J, Chung J, Sandy E, et al. Common Genetic Variant Association with Altered HLA Expression, Synergy with Pyrethroid Exposure, and Risk for Parkinson's Disease: An Observational and Case-Control Study. *NPJ Parkinson's disease.* 2015;1:15002.
210. Rösler TW, Salama M, Shalash AS, Khedr EM, El-Tantawy A, Fawi G, et al. K-variant BCHE and pesticide exposure: Gene-environment interactions in a case–control study of Parkinson's disease in Egypt. *Scientific reports.* 2018;8(1):16525.
211. Paul KC, Sinsheimer JS, Cockburn M, Bronstein JM, Bordelon Y, Ritz B. NFE2L2, PPARGC1α, and pesticides and Parkinson's disease risk and progression. *Mechanisms of Ageing and Development.* 2018;173:1-8.
212. Narayan S, Sinsheimer JS, Paul KC, Liew Z, Cockburn M, Bronstein JM, et al. Genetic variability in ABCB1, occupational pesticide exposure, and Parkinson's disease. *Environmental Research.* 2015;143(Pt A):98-106.

213. Chairta PP, Hadjisavvas A, Georgiou AN, Loizidou MA, Yiangou K, Demetriou CA, et al. Prediction of Parkinson's Disease Risk Based on Genetic Profile and Established Risk Factors. *Genes.* 2021;12(8).
214. Jacobs BM, Belete D, Bestwick J, Blauwendraat C, Bandres-Ciga S, Heilbron K, et al. Parkinson's disease determinants, prediction and gene–environment interactions in the UK Biobank. *Journal of Neurology, Neurosurgery & Psychiatry.* 2020;91(10):1046.
215. Angelopoulou E, Paudel YN, Papageorgiou SG, Piperi C. Environmental Impact on the Epigenetic Mechanisms Underlying Parkinson's Disease Pathogenesis: A Narrative Review. *Brain Sciences.* 2022;12(2).
216. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics*. 2018;26(2):149-156.
217. Knoppers BM. Framework for responsible sharing of genomic and health-related data. The *HUGO Journal.* 2014;8(1):3.
218. Program TGPsG. GP2: The Global Parkinson's Genetics Program. *Movement Disorders.* 2021;36(4):842-851.
219. Sielemann K, Hafner A, Pucker B. The reuse of public datasets in the life sciences: potential risks and rewards. *PeerJ.* 2020;8:e9954.
220. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nature genetics.* 2016;48(10):1284-1287.
221. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics.* 2016;48(10):1279-1283.
222. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. European journal of human genetics. *European Journal of Human Genetics: EJHG*. 2017;25(11):1253-1260.
223. Rentoft M, Svensson D, Sjödin A, Olason PI, Sjöström O, Nylander C, et al. A geographically matched control population efficiently limits the number of candidate disease-causing variants in an unbiased whole-genome analysis. *PLoS One.* 2019;14(3):e0213350.
224. Ten Years of the International Parkinson Disease Genomics Consortium: Progress and Next Steps. *Journal of Parkinson's disease.* 2020;10(1):19-30.
225. Pihlstrøm L, Morset KR, Grimstad E, Vitelli V, Toft M. A cumulative genetic risk score predicts progression in Parkinson's disease. *Movement Disorders.* 2016;31(4):487-490.
226. Alves G, Müller B, Herlofson K, HogenEsch I, Telstad W, Aarsland D, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *Journal of Neurology, Neurosurgery & Psychiatry.* 2009;80(8):851.
227. Loesch DP, Horimoto A, Heilbron K, Sarihan EI, Inca-Martinez M, Mason E, et al. Characterizing the Genetic Architecture of Parkinson's Disease in Latinos. *Annals of Neurology.* 2021(90):353-365.
228. Foo JN, Chew EGY, Chung SJ, Peng R, Blauwendraat C, Nalls MA, et al. Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. *JAMA Neurology.* 2020;77(6):746-754.

229. Salvi E, Kutalik Z, Glorioso N, Benaglio P, Frau F, Kuznetsova T, et al. Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension*. 2012;59(2):248-255.
230. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-678.
231. Krohn L, Wu RYJ, Heilbron K, Ruskey JA, Laurent SB, Blauwendraat C, et al. Fine-Mapping of SNCA in Rapid Eye Movement Sleep Behavior Disorder and Overt Synucleinopathies. *Annals of Neurology.* 2020;87(4):584-598.
232. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS genetics.* 2011;7(6):e1002141.
233. The Lancet N. Accelerating research for Parkinson's disease. *The Lancet Neurology.* 2018;17(4):289.
234. Tenny S, Kerndt CC, Hoffman MR. Case Control Studies. StatPearls. Treasure Island (FL): StatPearls Publishing. Copyright © 2022, *StatPearls Publishing LLC.;* 2022.
235. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International Journal of Epidemiology.* 2020;50(2):620-632.
236. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: *R Foundation for Statistical Computing;* 2020.
237. Verlouw JAM, Clemens E, de Vries JH, Zolk O, Verkerk AJMH, am Zehnhoff-Dinnesen A, et al. A comparison of genotyping arrays. *European Journal of Human Genetics: EJHG. 2021*;29(11):1611-1624.
238. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics.* 2007;81(3):559-575.
239. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4:7.
240. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature protocols.* 2010;5(9):1564-1573.
241. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The International HapMap Project. *Nature*. 2003;426(6968):789-796.
242. Elhaik E. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific reports.* 2022;12(1):14683.
243. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research.* 2018;27(2):e1608-e.
244. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics.* 2010;11(7):499-511.
245. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.

246. De Marino A, Mahmoud AA, Bose M, Bircan KO, Terpolovsky A, Bamunusinghe V, et al. A comparative analysis of current phasing and imputation software. *PLoS One*. 2022;17(10):e0260177.
247. Sun Q, Yang Y, Rosen JD, Jiang MZ, Chen J, Liu W, et al. MagicalRsq: Machine-learning-based genotype imputation quality calibration. *American Journal of Human Genetics. 2022*;109(11):1986-1997.
248. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299.
249. O'Connell J, Yun T, Moreno M, Li H, Litterman N, Kolesnikov A, et al. A population-specific reference panel for improved genotype imputation in African Americans. *Communications Biology.* 2021;4(1):1269.
250. Wall JD, Stawiski EW, Ratan A, Kim HL, Kim C, Gupta R, et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019;576(7785):106-111.
251. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics.* 2003;19(1):149-150.
252. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics.* 2016;32(9):1423-1426.
253. Ehret GB. Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Current Hypertension Reports.* 2010;12(1):17-25.
254. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
255. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols.* 2020;15(9):2759-2772.
256. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience.* 2019;8(7).
257. Yunusbaev U, Valeev A, Yunusbaeva M, Kwon HW, Mägi R, Metspalu M, et al. Reconstructing recent population history while mapping rare variants using haplotypes. *Scientific reports.* 2019;9(1):5849.
258. Farrer M, Skipper L, Berg M, Bisceglio G, Hanson M, Hardy J, et al. The Tau H1 Haplotype is associated with Parkinson's disease in the Norwegian population. *Neuroscience Letters.* 2002;322(2):83-86.
259. Skipper L, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M, et al. Linkage Disequilibrium and Association of MAPT H1 in Parkinson Disease. *The American Journal of Human Genetics.* 2004;75(4):669-677.
260. Bandres-Ciga S, Ahmed S, Sabir MS, Blauwendraat C, Adarmes-Gómez AD, Bernal-Bernal I, et al. The Genetic Architecture of Parkinson Disease in Spain: Characterizing Population-Specific Risk, Differential Haplotype Structures, and Providing Etiologic Insight. *Movement Disorders.* 2019;34(12):1851-1863.
261. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science.* 2002;296(5576):2225-2229.

262. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263-265.
263. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *American Journal of Human Genetics.* 2014;95(1):5-23.
264. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2013;368(1620):20120362.
265. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348(6235):648-660.
266. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics.* 2014;10(5):e1004383.
267. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature genetics*. 2021;53(9):1300-1310.
268. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications.* 2017;8(1):1826.
269. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
270. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature protocols.* 2019;14(2):482-517.
271. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic acids research.* 2017;45(W1):W130-W137.
272. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2017;33(2):272-279.
273. Carmine Belin A, Westerlund M, Sydow O, Lundströmer K, Håkansson A, Nissbrandt H, et al. Leucine-rich repeat kinase 2 (LRRK2) mutations in a Swedish Parkinson cohort and a healthy nonagenarian. *Movement Disorders.* 2006;21(10):1731-1734.
274. Marder K, Wang Y, Alcalay RN, Mejia-Santana H, Tang MX, Lee A, et al. Age-specific penetrance of LRRK2 G2019S in the Michael J. Fox Ashkenazi Jewish LRRK2 Consortium. *Neurology*. 2015;85(1):89-95.
275. Puschmann A, Wszolek ZK, Farrer M, Gustafson L, Widner H, Nilsson C. Alpha-synuclein multiplications with parkinsonism, dementia or progressive myoclonus? *Parkinsonism & related disorders.* 2009;15(5):390-392.
276. Aasly JO, Toft M, Fernandez-Mata I, Kachergus J, Hulihan M, White LR, et al. Clinical features of LRRK2-associated Parkinson's disease in central Norway. *Annals of Neurology.* 2005;57(5):762-765.
277. Turski P, Chaberska I, Szukało P, Pyska P, Milanowski Ł, Szlufik S, et al. Review of the epidemiology and variability of LRRK2 non-p.Gly2019Ser pathogenic mutations in Parkinson's disease. *Frontiers in Neuroscience.* 2022;16.

278. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.
279. Burström K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin L, et al. Swedish experience-based value sets for EQ-5D health states. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2014;23(2):431-442.
280. Jafari S, Etminan M, Aminzadeh F, Samii A. Head injury and risk of Parkinson disease: a systematic review and meta-analysis. *Movement Disorders.* 2013;28(9):1222-1229.
281. Marras C, Hincapié CA, Kristman VL, Cancelliere C, Soklaridis S, Li A, et al. Systematic review of the risk of Parkinson's disease after mild traumatic brain injury: results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation.* 2014;95(3 Suppl):S238-244.
282. The Public Health Agency of Sweden. Tobacco. [cited January 20 2023]. Available from: https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/living-conditions-and-lifestyle/andtg/tobacco/
283. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. *Statistical Science.* 2009;24(4):561-573.
284. Zhang X, Zhang L, Lin B, Chai X, Li R, Liao Y, et al. Phospholipid Phosphatase 4 promotes proliferation and tumorigenesis, and activates Ca(2+)-permeable Cationic Channel in lung carcinoma cells. *Molecular Cancer.* 2017;16(1):147.
285. Simón-Sánchez J, Kilarski LL, Nalls MA, Martinez M, Schulte C, Holmans P, et al. Cooperative genome-wide analysis shows increased homozygosity in early onset Parkinson's disease. *PLoS One.* 2012;7(3):e28787.
286. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics.* 2014;46(9):989-993.
287. Mattingsdal M, Ebenesersdóttir SS, Moore KHS, Andreassen OA, Hansen TF, Werge T, et al. The genetic structure of Norway. *European journal of human genetics: EJHG.* 2021;29(11):1710-1718.
288. Humphreys K, Grankvist A, Leu M, Hall P, Liu J, Ripatti S, et al. The genetic structure of the Swedish population. *PLoS One.* 2011;6(8):e22547-e.
289. Sciorra VA, Morris AJ. Roles for lipid phosphate phosphatases in regulation of cellular signaling. *Biochimica Biophysica Acta.* 2002;1582(1-3):45-51.
290. Wood PL, Tippireddy S, Feriante J, Woltjer RL. Augmented frontal cortex diacylglycerol levels in Parkinson's disease and Lewy Body Disease. *PLoS One.* 2018;13(3):e0191815.
291. Riou A, Houvenaghel JF, Dondaine T, Drapier S, Sauleau P, Drapier D, et al. Functional Role of the Cerebellum in Parkinson Disease: A PET Study. *Neurology*. 2021;96(23):e2874-e2884.
292. Zhong Y, Liu H, Liu G, Zhao L, Dai C, Liang Y, et al. A review on pathology, mechanism, and therapy for cerebellum and tremor in Parkinson’s disease. *NPJ Parkinson's disease.* 2022;8(1):82.
293. Lyu Y, Bai L, Qin C. Long noncoding RNAs in neurodevelopment and Parkinson's disease. *Animal Models and Experimental Medicine.* 2019;2(4):239-251.

294. Ni Y, Huang H, Chen Y, Cao M, Zhou H, Zhang Y. Investigation of Long Non-coding RNA Expression Profiles in the Substantia Nigra of Parkinson's Disease. *Cellular and Molecular Neurobiology.* 2017;37(2):329-338.
295. Langmyhr M, Henriksen SP, Cappelletti C, van de Berg WDJ, Pihlstrøm L, Toft M. Allele-specific expression of Parkinson's disease susceptibility genes in human brain. *Scientific reports.* 2021;11(1):504.
296. Farrow SL, Schierding W, Gokuladhas S, Golovina E, Fadason T, Cooper AA, et al. Establishing gene regulatory networks from Parkinson's disease risk loci. *Brain*. 2022;145(7):2422-2435.
297. Postuma RB, Gagnon JF, Vendette M, Fantini ML, Massicotte-Marquez J, Montplaisir J. Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology*. 2009;72(15):1296-1300.
298. Krohn L, Öztürk TN, Vanderperre B, Ouled Amar Bencheikh B, Ruskey JA, Laurent SB, et al. Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies. *Annals of Neurology.* 2020;87(1):139-153.
299. Gan-Or Z, Mirelman A, Postuma RB, Arnulf I, Bar-Shira A, Dauvilliers Y, et al. GBA mutations are associated with Rapid Eye Movement Sleep Behavior Disorder. *Annals of Clinical and Translational Neurology.* 2015;2(9):941-945.
300. Krohn L, Ruskey JA, Rudakou U, Leveille E, Asayesh F, Hu MTM, et al. GBA variants in REM sleep behavior disorder: A multicenter study. *Neurology*. 2020;95(8):e1008-e16.
301. Chia R, Sabir MS, Bandres-Ciga S, Saez-Atienzar S, Reynolds RH, Gustavsson E, et al. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nature genetics.* 2021;53(3):294-303.
302. Gan-Or Z, Dion PA, Rouleau GA. Genetic perspective on the role of the autophagy-lysosome pathway in Parkinson disease. *Autophagy*. 2015;11(9):1443-1457.
303. Martini-Stoica H, Xu Y, Ballabio A, Zheng H. The Autophagy-Lysosomal Pathway in Neurodegeneration: A TFEB Perspective. *Trends in Neurosciences.* 2016;39(4):221-234.
304. Chatterjee S, Mudher A. Alzheimer's Disease and Type 2 Diabetes: A Critical Assessment of the Shared Pathological Traits. *Frontiers in Neuroscience.* 2018;12:383.
305. Zhu Z, Lin Y, Li X, Driver JA, Liang L. Shared genetic architecture between metabolic traits and Alzheimer's disease: a large-scale genome-wide cross-trait analysis. *Human Genetics.* 2019;138(3):271-285.
306. Ran C, Brodin L, Gellhaar S, Westerlund M, Fardell C, Nissbrandt H, et al. Glucocerebrosidase variant T369M is not a risk factor for Parkinson's disease in Sweden. *Neuroscience Letters.* 2022;784:136767.
307. Mallett V, Ross JP, Alcalay RN, Ambalavanan A, Sidransky E, Dion PA, et al. GBA p.T369M substitution in Parkinson disease: Polymorphism or association? A meta-analysis. *Neurology Genetics*. 2016;2(5):e104.
308. Zhang H, Iranzo A, Högl B, Arnulf I, Ferini-Strambi L, Manni R, et al. Risk Factors for Phenoconversion in Rapid Eye Movement Sleep Behavior Disorder. *Annals of Neurology.* 2022;91(3):404-416.

309. Puligheddu M, Figorilli M, Antelmi E, Arnaldi D, Casaglia E, d'Aloja E, et al. Predictive risk factors of phenoconversion in idiopathic REM sleep behavior disorder: the Italian study "FARPRESTO". *Neurological Sciences.* 2022;43(12):6919-6928.
310. Schumacher-Schuh AF, Bieger A, Okunoye O, Mok KY, Lim S-Y, Bardien S, et al. Underrepresented Populations in Parkinson's Genetics Research: Current Landscape and Future Directions. *Movement Disorders.* 2022;37(8):1593-1604.
311. Farrer MJ, Bardien S, Hattori N, Lesage S, Ross OA, Mellick GD, et al. Editorial: Celebrating the Diversity of Genetic Research to Dissect the Pathogenesis of Parkinson's Disease. *Frontiers in Neurology.* 2021;12.
312. Schiess N, Cataldi R, Okun MS, Fothergill-Misbah N, Dorsey ER, Bloem BR, et al. Six Action Steps to Address Global Disparities in Parkinson Disease: A World Health Organization Priority. *JAMA Neurology.* 2022;79(9):929-936.
313. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze Andrey V, Mikheenko A, et al. The complete sequence of a human genome. *Science.* 2022;376(6588):44-53.
314. Ochoa D, Karim M, Ghoussaini M, Hulcoop DG, McDonagh EM, Dunham I. Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nature Reviews. Drug Discovery.* 2022;21(8):551.

Acknowledgements

I would like to express my sincere gratitude to all of you that, in different ways, contributed to the work in my PhD thesis. I am forever grateful for the knowledge, advice, and particularly the support that I have received from all of you during this time. I apologize if I miss to acknowledge you all by name, but please know that I highly appreciate every single one of you.

First and foremost, I would like to acknowledge and thank all individuals contributing to science in different ways, donating your time and samples to research. Without your contribution, the research process would be significantly hampered (and this thesis would never have happened). I hope that we in return can contribute with new knowledge which at some point leads to the development of disease-modifying treatments.

Secondly, and as important for making this thesis happened, I would like to thank my main supervisor, *Maria Swanberg*. Thank you for all your feedback and for always listening to me, supporting me, and encouraging me to form my thesis work after what I have wanted.

I would also like to thank my co-supervisor, *Andreas Puschmann*. Thank you for your valuable advice and feedback on my work, in particularly when it comes to the clinical aspects.

To be honest, I don't know how this thesis work would have turned out if it wasn't for all the fantastic people, I have gotten to know through the IPDGC and GP2 consortiums. Research is a competitive field, but these consortiums truly shows that research can be done, and should be done, collaboratively. I would like to thank every single one of you; *Sara*, *Manuela*, *Lynne*, *Mary*, *Hampton*, *Teresá*, *Pilar*, and so many more, for your warm support and for becoming my colleagues and friends during these years (which was especially appreciated during the corona pandemic).

To all present and former members of the Translational Neurogenetics Unit; *Filip*, *Lautaro*, *Olivia*, *Itzia*, *Alfredo*, *Michael*, and all students that have been with us, my biggest thank you for being the best colleagues anyone could have wished for. Also, I would like to thank everyone who I have been lucky to interact with during these years, including everyone in MDR, the fantastic people working in various research groups at BMC, the Neurology research nurses, the people involved in the SBW2021, and so many more. Thank you for all the fika breaks, discussions,

travels, for supporting me in different ways during these years, and for making BMC a great place to work: *Megg*, *Malin, Martino, Jana, Emma N., Rosalia, Sabine, Laura, Lluis, Caroline, Alexander, Eduardo, Elna, Matilde, Carolina, Andreas, Robert, Emma P., Gunilla, Christin, Martina, Sorina, Sigurd, Esther, Deborah, Suze, Joel, and many more*. Also, a big thank you to everyone else at A10 for all the great times and nice conversations we had during these years.

To *my family and friends*: Thank you for always supporting me in every imaginable situation, I am forever grateful to have you in my life. *Viggo*, thank you for being my best colleague when working from home, and for forcing me to take needed breaks during my work and thesis writing. *Erik*, thank you for being my biggest supporter, and for being who you are.

Lastly, I would like to thank all funding agencies that have supported the work in this thesis. The work in this thesis was financially supported by Promobilia Foundation, the Royal Physiographic Society of Lund, the Foundation in the Memory of Sigurd and Elsa Golje, MultiPark, NEURO Sweden, the Åke Wiberg Foundation, the Åhlén Foundation, and the Swedish Research Council.

Appendix (Papers I-V)

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Experimental Medical Science

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2023:41
ISBN 978-91-8021-380-6
ISSN 1652-8220