UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2023-103
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 29/09/2023 in Esch-sur-Alzette, Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN BIOLOGIE

by

Sinthuja PEIRIS PACHCHEK
Born on 31 March 1986 in Pesalai (Sri Lanka)

DECIPHERING THE GENETIC ARCHITECTURE OF PARKINSON'S DISEASE IN THE LUXEMBOURGISH POPULATION

Dissertation defence committee

Prof Dr Rejko Krüger, dissertation supervisor
Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, Luxembourg

Prof Dr Michel Mittelbronn, Chairman
Luxembourg Center of Neuropathology, National Center of Pathology, Luxembourg

Dr Patrick May, Vice Chairman
Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, Luxembourg

Dr Manu Sharma
Institute for Clinical Epidemiology and Applied Biometry, Germany

Dr Kathrin Brockmann
University of Tübingen, Hertie Institute for Clinical Brain Research HiH, German Centre for Neurodegenerative Diseases DZNE, Germany

Affidavit

I hereby confirm that the Ph.D. thesis entitled "*Deciphering the genetic architecture of Parkinson's disease in the Luxembourgish population*" has been written independently and without any other sources than cited.

Belval, Luxembourg
Sinthuja Pachchek Peiris
29/08/2023

Acknowledgements

I thank God ***Jesus*** for the incredible journey I had as a student at the ***Luxembourg Centre for Systems Biomedicine*** (LCSB) during the years 2019-2023. These years of academic research were the most rewarding in terms of education. I cherish the experiences and knowledge gained during this period.

I would like to thank the ***Fonds National de la Recherche*** (FNR) for funding this work through the ***National Center for Excellence in Research on Parkinson's Disease*** program (NCER-PD) and the ***doctoral training unit*** (PARK-QC). I would like to thank all the participants who actively took part in the ***Luxembourg Parkinson's disease study***, as their invaluable contributions made this research possible.

Foremost, I wish to express my sincere gratitude to ***Prof. Dr. Rejko Krüger*** and ***Dr. Patrick May***, my thesis supervisors, for offering me the invaluable opportunity to join their group and for their continuous support which allowed me to become fully involved in the field of genomic studies and in-depth genetics research. Their mentoring, innovative ideas and invaluable advice and support along the journey have played a crucial role in making me the confident scientist I am today.

Additionally, I would like to extend my gratitude to ***Prof. Dr. Michel Mittelbronn***, ***Dr. Manu Sharma*** and ***Dr. Kathrin Brockmann***, who kindly agreed to be part of the esteemed jury members for this thesis. Their involvement in the evaluation process was a great honor, and I am deeply grateful for the valuable time and expertise they devoted to reviewing my work.

Thanks to all the LCSB and Bio-Core team members, including past and present students and postdoctoral researchers, especially the Genome Analysis Groupe, ***Dr. Dheeraj Reddy Bobbili*** and ***Nuria Terrado Ortuno,*** for cultivating a friendly and highly productive environment. My special thanks go to ***Dr. Zied Landoulsi***. Throughout our journey together, we have faced many challenges and achieved many goals. I am deeply grateful to him for his invaluable advice, his sharing of knowledge, his patience and his constant moral support over the past four years.

I sincerely thank the neurologists, ***Dr. Lukas Pavelka*** and ***Dr. Olena Tsurkalenko***, and the clinical team from Luxembourg Institute of Health, for their extraordinary contribution to exploring the patient cohort and for the insightful discussions we had together.

I would like to express my deepest gratitude to my thesis partners, ***Axel***, ***Vera***, and ***Laure***. In the face of academic rigors and personal struggles, their constant support, shared experiences and our shared funny moments kept me optimistic and motivated.

I would like to express my gratitude to all those who supported me with this project and contributed to it, particularly the project managers ***Dr. Tainá Marques***, ***Dr. Clarissa Gomes, Dr. Nancy Ramia*** and ***Dr. Bianca Dragomir***.

I would like to thank all those who helped me to write, correct, and layout this manuscript.

I would like to thank my friends who supported me, with a special mention to ***Dr. Kundavai Natchiar***, whose encouragement and motivation inspired me to begin this doctoral journey.

I would like to thank my family and friends for giving me joy and happiness during the final phase of writing the thesis in Pesalai Sri Lanka, particularly the ***Babylona*** team.

I am deeply grateful to my family, especially my parents ***Thomas Selvartnam*** and ***Carmel Sinnarani***, for their constant love and support. I would also like to express my deep gratitude to my sister and brother, ***Brighty*** and ***Jérémy***, my sweet sister-in-law ***Prya***, and my nieces and nephews ***Kandus, ShareShare, Tshushi, Chico***, ***Anjali***, ***Sahana*** and ***Shalou.***

Lastly, but most importantly, I want to offer my heartfelt and immense thanks to my husband ***Anton Merino*****.** From the beginning, he was my greatest support, constantly encouraging me to persevere in the pursuit of my Ph.D. goal and never once doubting my ability to achieve it. He was invaluable in holding me up when things got very difficult. I am truly blessed to have him by my side. I thank him for being who he is and for forgiving me for all the evenings and weekends I had to dedicate to my studies instead of spending them with him and our daughter, ***Kayalvili***. I hope that when our little girl grows up, she will realize how much I love her and how hard I have worked for her.

*to my husband **Anton Merino***

உன் அன்பும், ஆதரவும் இல்லாமல் என்னால் இந்தப் படிப்பை முடித்திருக்க முடியாது

"without your love and support, I could never have completed this thesis"

Table of Contents

List of Figures

Introduction

Chapter 4

List of Tables

Introduction

Chapter 4

List of abbreviations

aa	Amino acid
AAA	Age-at-assessment
AAO	Age-at-onset
ACMG	American College of Medical Genetics and Genomics
AD	Alzheimer's disease
ALP	Autophagy-lysosome pathway
ALS	Amyotrophic lateral sclerosis
Annovar	Annotate variation tool
AP	Atypical parkinsonism
ATP	Adenosine triphosphate
BMI	Body Mass Index
bp	Base pair
CADD	Combined annotation dependent depletion
CBD	Corticobasal degeneration
cDNA	Complementary DNA
CI	Confidence interval
CNS	Central nervous system
CNVs	Copy number variations
COMT	Catechol-O-Methyl transferase
DBS	Deep brain stimulation
dbSNP	Single nucleotide polymorphism database
ddNTPs	Dideoxy-nucleoside triphosphates
DGV	Database of genomic variants
DHS	DNase I hypersensitive sites
DLB	Dementia with Lewy bodies
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
eQTL	Expressed Quantitative Trait Loci
ER	Endoplasmic Reticulum
ERT	Enzyme replacement therapy
FTD	Frontotemporal disorders
GATK	Genomic analysis toolkit
GBA1	Glucosylceramidase beta 1
$GBA1_{PVs}$	GBA1 pathogenic variants

GBAP1	Glucosylceramidase beta 1 pseudogene
GCase	Glucocerebrosidase
GD	Gaucher's disease
GlcCer	Glucosylceramide
GlcSph	Glucosylsphingosine
gnomAD	Genome aggregation database
gnomAD-SV	Genome aggregation database structural variants
GP2	Global Parkinson's Genetics program
GSLs	Glycosphingolipids
GTPase	Guanosine triphosphate binding enzyme
GWAs	Genome-wide association studies
HGMD	Human gene mutation database
HLA-DQ1b	Human leukocyte antigen DQ1b
IGV	Integrative Genomics Viewer
Indels	Small insertions and deletions
iPSCs	Induced pluripotent stem cells
kb	Kilobase
L-DOPA	Levodopa
LBs	Lewy bodies
LEDD	Levodopa-Equivalent Daily Dosage
LID	Levodopa-Induced Dyskinesia
Lns	Lewy neurites
LoF	Loss-of-function
LRS	Long-read sequencing
MAF	Minor allele frequency
MAO-B	Monoamine oxidase B
MAPT	Microtubule-associated tau protein
MDS	Movement disorder society
MISTIC	Missense deleteriousness predictor
MJFF	Michael J. Fox foundation's
MLPA	Multiplex ligation dependent probe amplification
MPP+	1-methyl-4-phenylpyridinium
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
mRNA	Messenger ribonucleic acid
MSA	Multiple system atrophy
mvPPT	Pathogenicity Prediction Tool for missense variants
NBIA	Neurodegenerations with brain iron accumulation

NCBI	National Center for Biotechnology Information
NCER-PD	National Center for Excellence in Research on Parkinson's Disease program
NFE	Non-Finnish European
NGS	Next-generation sequencing
ONT	Oxford nanopore technologies
OR	Odds Ratio
PacBio	Pacific biosciences platforms
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD	Parkinson's disease
PRS	Polygenic risk score
PSP	Progressive supranuclear palsy
RBD	REM Sleep behaviour disorders
REM	Rapid eye movement
REVEL	Rare Exome Variant Ensemble Learner
RNA	Ribonucleic acids
ROS	Reactive oxygen species
SD	Standard deviation
SMCs	Small molecule chaperones
SMRT	Single-molecule real-time sequencing
SNP	Single nucleotide polymorphism
SNpc	Substantia nigra pars compacta
SNVs	Single nucleotide variants
SRT	Substance reduction therapy
SVs	Structural variants
TFBS	Transcription factor binding sites
TIM	Triose-phosphate isomerase
TRS	Tandem repeat sequences
UPDRS	Unified Parkinson’s disease rating scale
UPR	Unfolded Protein Response
UPS	Ubiquitin proteasome system
UTR	Untranslated region
VCF	Variant call format
VUS	Variants of unknown significance
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

List of publications

This thesis is based on work that has been published or is under review by the respective journal or is ready for submission.

Manuscripts used in this cumulative thesis have all been submitted with the candidate as the first or equally contributing first author.

All projects presented in the thesis have been carried out with several collaborators and were supported by the National Center for Excellence in Research on Parkinson's Disease consortium (NCER-PD).

Publications resulting from this work

- **Pachchek, S.**, Landoulsi, Z., Pavelka, L. Schulte, C., Buena-Atienza, E., Gross, C., Hauser, A.-K., Bobbili, D.R., Casadei, N., May, P., Krüger, R.. Accurate long-read sequencing identified GBA1 as major risk factor in the Luxembourgish Parkinson's study. *npj Parkinsons Dis*. 9, 156 (2023). https://doi.org/10.1038/s41531-023-00595-w

- Landoulsi, Z.[*], **Pachchek, S.**[*], Bobbili, D.R., Pavelka, L., May, P., Krüger, R., August 2023. Genetic landscape of Parkinson's disease in Luxembourg. *Frontiers in Aging Neuroscience*. (Manuscript under review)

- Hassanin, E.[*], Landoulsi, Z.[*], **Pachchek, S.**[*], Krawitz, P., Krüger, R., Maj, C., May, P., Bobbili D.R., August 2023. Parkinson's risk: the interplay between polygenic background and *GBA1* pathogenic variants. (Manuscript in preparation)

  *Equal contribution

Additional publications

- Smajić, S., Prada-Medina, C.A., Landoulsi, Z., Ghelfi, J., Delcambre, S., Dietrich, C., Jarazo, J., Henck, J., Balachandran, S., **Pachchek, S**., Morris, C.M., Antony, P., Timmermann, B., Sauer, S., Pereira, S.L., Schwamborn, J.C., May, P., Grünewald, A., Spielmann, M., 2022. Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. Brain J. Neurol. 145, 964–978. https://doi.org/10.1093/brain/awab446

- Pavelka, L., Rauschenberger, A., Landoulsi, Z., **Pachchek, S.**, May, P., Glaab, E., Krüger, R., NCER-PD Consortium, 2022. Age at onset as stratifier in idiopathic Parkinson's disease - effect of ageing and polygenic risk score on clinical phenotypes. NPJ Park. Dis. 8, 102. https://doi.org/10.1038/s41531-022-00342-7

- Pavelka, L., Rauschenberger, A., Landoulsi, Z., **Pachchek, S.**, Marques, T., Gomes, C.P.C., Glaab, E., May, P., Krüger, R., NCER-PD Consortium, 2022. Body-First Subtype of Parkinson's Disease with Probable REM-Sleep Behavior Disorder Is Associated with Non-Motor Dominant Phenotype. J. Park. Dis. 12, 2561–2573. https://doi.org/10.3233/JPD-223511

Abstract

Dissecting the genetic complexity of parkinsonism in the Luxembourg Parkinson's Study

Parkinson's Disease (PD) is the fastest-growing neurodegenerative disorder (Bloem et al., 2021; Dorsey et al., 2018). It is characterized by motor symptoms such as postural instability, rest tremor, bradykinesia and rigidity (Kalia and Lang, 2015). Additionally, PD exhibits non-motor symptoms including cognitive impairment, neuropsychiatric features like hallucinations, depression, and anxiety, as well as sleep disturbances, hyposmia and autonomic dysfunction (Khoo et al., 2013).

While the exact causes of PD remain incompletely understood, research has identified genetic risk factors as important contributing factors in its development. Although PD appears typically as sporadic, approximately 30% of cases can be linked to genetic factors involving monogenic forms (Billingsley et al., 2018). Although motor symptoms can be effectively addressed by pharmacological therapies at least in early disease stages, there is currently no treatment that may interfere with the chronic progressive neurodegeneration. Therefore, there is a strong motivation among clinicians and researchers to perform early diagnoses and identify genetic variants that could lead to future causative and thereby neuroprotective therapies. However, PD is a complex disorder influenced by a combination of environmental and genetic factors. Only a limited subset of genes have been conclusively associated with typical PD, as their causative role in Mendelian forms of PD has been consistently replicated in multiple studies involving large populations of PD patients (Bandres-Ciga et al., 2020). These rare Mendelian forms follow distinct inheritance patterns and have a notable impact on disease development (e.g., *SNCA*, *LRRK2*, *VPS35, PRKN*, *PINK1* and *PARK7*). There are also low-frequency variants with significant effects, such as the *GBA1* and *LRRK2* genes. These variants are not as rare as monogenic mutations, but their frequency remains relatively low in the general population. Then, there are the common variants that exert minor effects and highlight genetic variants that are prevalent in the population but individually contribute only a modest risk toward PD (Manolio et al., 2009).

Although not as extensively studied as single nucleotide variants, small copy number variants have received considerable interest due to their potential pathogenic implications (La Cognata et al., 2017; Pankratz et al., 2011; Toft and Ross, 2010).

As part of the Luxembourg Parkinson's study, which includes both healthy individuals and patients with PD and atypical parkinsonism, we aim to investigate the genetic background of PD-causal genes (*LRRK2*, *SNCA*, *VPS35*, *PRKN*, *PARK7*, *PINK1* and *ATP13A2*) and other known PD-associated genes, looking for rare Single-Nucleotide Variants (SNVs), Copy Number Variation (CNVs), and

estimating the effect of multiple common SNVs in predicting an individual's probability of developing PD, using Polygenic Risk Scores (PRS).

The Luxembourg Parkinson's study, a large monocentric longitudinal cohort, included 1791 participants, 911 of whom were diagnosed cases and 880 neurologically healthy controls. The mean age of the control group was 64.7 ± 12.1 years in 2023. The mean age at onset for PD patients was 62.4 ± 11.7 years.

We identified 12.1% of PD patients and 5% of healthy controls carried *GBA1* variants. Additionally, four *GBA1* variants were discovered in patients with progressive supranuclear palsy and dementia with Lewy bodies. We identified different categories of pathogenic *GBA1* variants, including those with severe, mild and risk-associated effects. We then studied the relationships between genotypes and phenotypes, to better understand the impact of each type of variant and how they contribute to disease severity. We discovered a total of 60 rare SNVs within seven PD-causal genes. Notably, nine of these variants were found to be disease-causing in *LRRK2*, *PINK1*, and *PRKN* genes. Additionally, we identified eleven rare CNVs in the *PRKN* gene, encompassing seven duplications and four deletions.

We showed that the PRSs were significantly associated with PD and highlighted the important role of polygenic background plays in modulating PD risk in carriers of pathogenic *GBA1* variants.

Moreover, in an explorative study, where we are looking for loss of function variants, we identified 134 rare variants and eight rare copy number variations in PD-related genes, that could potentially contribute to PD, but these results are not statistically reliable and further analysis is required. The failure to discover novel genetic variants with whole-genome sequencing data can be attributed to the limited power to detect rare variants with small effects, as well as the relatively small sample size of the study.

We carried out an in-depth genetic analysis of the participants in the Luxembourg Parkinson's study. Our findings should help future research to unravel the complex genetic landscape of PD. This knowledge will make it possible to classify participants according to their genetic profile, improving the effectiveness of future precision medicine approaches. These targeted therapies can then be tailored to attack specific molecular targets, paving the way for a new era of personalized treatment strategies.

Aims and Objectives

The primary objective of this thesis was to conduct a genetic analysis on participants of the Luxembourg Parkinson's study, which includes patients with typical PD and atypical parkinsonism. The initial investigation involved examining genetic risk variants in both patients and healthy controls in 1791 participants.

This analysis aimed to identify and characterize genetic variants associated with PD in this specific population. The next objective was to assess the prevalence of PD-related genetic mutations in the Luxembourg population. By analyzing the genetic data from patients and healthy controls, the study aimed to determine the frequency and distribution of specific risk variants in this cohort. To better understand the impact of each type of variant and how they affect disease severity, we identified three different categories of pathogenic *GBA1* variants (severe, mild and risk). We also characterized genotype-phenotype correlations.

Multiple genetic data sources were used to establish a comprehensive overview of the genetic landscape of PD in Luxembourg. These included whole genome sequencing data, which facilitated the analysis of rare and novel variants, target-*GBA1* Pacific Biosciences data, which explored the involvement of the *GBA1* gene in PD susceptibility, genotyping data, which allowed the identification of common risk variants and Sanger sequencing, which allowed us to validate the presence of the variants in individuals.

By focusing on these aims and objectives, this thesis hoped to improve our knowledge of the genetic foundation of PD and provide important information to guide translational researchers, and clinicians to develop future precision medicine approaches for patients with parkinsonism in Luxembourg. The discovery of *GBA1* variants as an important genetic risk factor for Parkinson's disease in the Luxembourg Parkinson's disease study provides new therapeutic opportunities. The current clinical trial aims to attenuate cognitive decline in Parkinson's patients with *GBA1* mutations using prasinezumab, an antibody meticulously designed to hinder the transmission of α-synuclein aggregates between neurons.

Introduction

I.1. Genetics, an overview

Genetics is the science that studies genes and their inheritance. The genome refers to the complete set of genetic material found within the cells of an individual. In humans, there are a total of 23 pairs of chromosomes, with one pair inherited from the mother and the other from the father. Among these, 22 pairs are known as autosomes and are numbered from 1 to 22 based on their decreasing size. The remaining pair is composed of sex chromosomes, which determine an individual's biological sex (XX for females and XY for males).

I.1.1. DNA

Chromosomes consist of DeoxyriboNucleic Acid (DNA) molecules, which are composed of four distinct nucleotides. DNA carries genetic information and is found in all living organisms, spanning bacteria, plants, and animals. In 1953, Watson and Crick unveiled the double helix structure of DNA. This structure is formed by four nucleotide bases: Adenine (A), Thymine (T), Guanine (G), and Cytosine (C), linked together by a sugar (deoxyribose) and a triphosphate. These bases work in pairs, A and G known as purines combining with pyrimidine bases T and C respectively, to ensure the consistency of the genetic code and the structural stability of the double helix. To minimize space within the cell, this double helix associates with organizing proteins called histones, coiling around them to form a highly condensed structure known as a chromosome (Alberts et al., 2002) (Figure 1). The human genome contains nearly 3 billion base pairs organized into functional units called genes. It is estimated that the human genome contains between 25,000 and 30,000 genes. Genes consist of two main components: coding regions called exons and non-coding regions called introns. The exons undergo a process of transcription, resulting in the production of RiboNucleic Acids (RNA), which are subsequently translated into proteins.

Apart from genes, the genome also encompasses various types of non-coding regions. These non-coding regions serve different purposes, such as acting as structural elements like centromeres or playing a role in regulating gene expressions, such as Transcription Factor Binding Sites (TFBS), or intergenic regions existing between genes. However, the precise functions of these non-coding regions are still largely unknown and continue to be an area of active research.

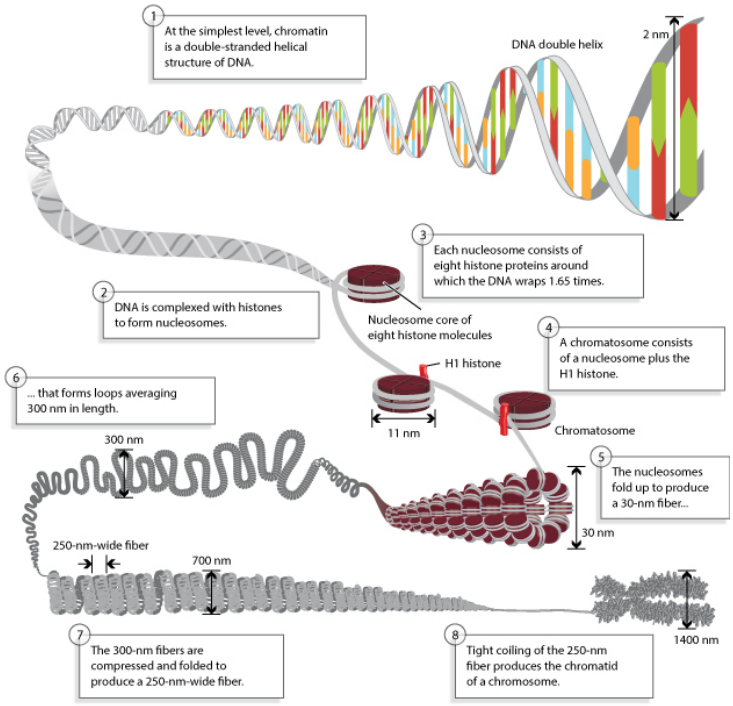


Figure 1 : Chromosome structure with several levels of organization

I.1.2. Genomic variability

Genetic variants refer to alterations in the DNA sequence which, when manifested in a germ cell, result in an unexpected, heritable modification of a DNA fragment. The influence of these variants depends on their precise localization and impact on protein function (Marian, 2020). Various mechanisms can trigger genetic alterations. These changes can result from errors in DNA replication. Furthermore, certain substances, including chemicals, radiation, and some viruses, have the power to damage DNA directly or block DNA replication and repair. Transposons, mobile genetic elements, tend to disrupt genes or regulatory elements when inserted into new genomic regions. Another factor comes into play during meiosis, where the crossover mechanism initiates genetic recombination, resulting in the creation of new gene combinations. Radiation, chemicals, pollution and even dietary factors are examples of environmental elements that can influence the rate of genetic change.
Genetic changes are a fundamental aspect of evolution, as they promote genetic diversity within populations. While some alterations can have negative consequences for a species, leading to disease, others can confer advantages in specific environments or facilitate improvement in a species' adaptability over time (Eichler, 2019; Marian, 2020).

I.1.2.1. Substitution

Among humans, substitutions are the most prevalent form of genetic variants, accounting for over 90% of all genetic variants. Substitutions occur when a single nucleotide is altered within the DNA sequence. When a substitution is found in less than 1% of the population under study, it is referred to as a Single Nucleotide Variant (SNV). Conversely, if the substitution is relatively common, present in more than 1% of the population, it is called a Single Nucleotide Polymorphism (SNP). SNPs are extensively utilized in constructing genetic maps, such as the well-known Single Nucleotide Polymorphism database (dbSNP) and serve as genetic markers for individual identification.

SNV can be categorized into nonsynonymous, synonymous, splice, regulatory and intronic variants, based on their localization into the DNA sequence and impact on protein production (Marian, 2020).

Nonsynonymous variants occur when a nucleotide change leads to the substitution of one amino acid with another within the protein sequence. As a result, the function of the protein may be altered to varying degrees, depending on the specific amino acid change. A subset of non-synonymous variants are nonsense variants, which result in a premature stop codon in the protein-coding sequence and induce the expression of aberrant transcripts resulting in truncated proteins that are non-functional or prone to degradation (Marian, 2020).

Synonymous variants do not alter the amino acid sequence within the protein and are denoted as silent variants. Nevertheless, these synonymous SNVs could potentially influence factors such as transcriptional and translational efficiency, splicing, and Messenger RiboNucleic Acid (mRNA) stability (Marian, 2020).

Splice variants occur at exon-intron boundaries defined by splicing consensus sequences. Introns generally begin with GT sequences and end with AG sequences, called splice donor and splice acceptor sites respectively. The motifs are recognized by the splicing machinery, composed of protein-RNA complexes, which excise the introns and ligate the subsequent exons. Genetic variants located at consensus splice acceptor (5′) or donor (3′) sites can disrupt the correct splicing of the primary transcript, resulting in exon skipping, intron retention or a reading frame mismatch in the final mRNA production (Pan et al., 2008).

Regulatory variants are located in genomic regulatory regions, including positions in the promoter, enhancer and 3′ regulatory regions. SNVs found in 5′ regulatory regions can impact mRNA transcription by altering the binding affinity of the transcription factors involved. Similarly, SNVs located in amplifying regions can affect the expression of many genes. Similarly, variants located in 3′ untranslated regions have the potential to influence transcript stability and the interaction of transcripts with microRNAs (Marian, 2020).

Intronic variants, the most common SNVs**,** are located in deep intronic regions (excluding conventional splice sites) and intergenic regions. In general, intronic variants are not expected to have any biological impact, unless they create novel splice sites, leading to the incorporation of a pseudo-exon or causing a frameshift in the protein. These variants may also influence the non-coding RNAs transcribed from the respective genomic regions (Marian, 2020).

I.1.2.2. Insertion and Deletions

Insertions and deletions, commonly known as indels, represent the insertion or deletion of nucleotide fragments that affect less than 50 nucleotides (Mullaney et al., 2010). When Indels occur within the coding sequence, they cause a frameshift. Consequently, incorrect DNA reads are typically generated, resulting in the production of non-functional proteins. On average, the human genome is estimated to contain between 192 and 280 frameshifts (1000 Genomes Project Consortium et al., 2010). However, the majority of the indels are located in the intergenic regions, and prediction of their functionality is quite challenging (Marian, 2020).

I.1.2.3. Structural variants

Structural Variants (SVs) encompass significant chromosomal rearrangements, such as Copy Number Variations (CNVs), inversions, Tandem Repeat Sequences (TRS), and chromosomal

rearrangements. These SVs typically exceed 1 kb in size and involve various combinations of DNA gains, losses, or rearrangements (Alkan et al., 2011).

CNVs are a specific subtype of SVs mostly characterized by deletions and duplications that cover at least 50 bp (Carvalho and Lupski, 2016). CNVs have the potential to impact gene expression levels and contribute to the development of various pathologies (Jakobsson et al., 2008; Redon et al., 2006). For instance, in the context of neurodegenerative diseases, it was discovered that an early stage of dystonia-parkinsonism disease was associated with the insertion of a retrotransposon in an intron of the *TAF1* gene (Bragg et al., 2017). Notably, CNVs alone affect more than 12% of the human genome.

TRS represents the repetition of a specific pattern within a sequence. They are alternatively known as satellites, minisatellites, or microsatellites, depending on their size. Microsatellites consist of motifs of 1 to 5 bp, repeated 2 to 50 times, with a total size of less than 300 bp. Minisatellites, on the other hand, comprise motifs of 15 to 100 bp, repeated 15 to 50 times, resulting in a total size of 1 to 5 kb. Satellites are large motifs repeated consecutively and predominantly participate in cellular mechanisms like meiosis.
The repetition of these sequences can influence the proper functioning of a genetic unit. In neurological diseases, for instance, extensions of ATTCC repeats in the *ATXN10* gene have been frequently associated with PD (Schüle et al., 2017). Additionally, CAG expansions in Huntington's disease have been extensively studied (McColgan and Tabrizi, 2018).

Inversion refers to the process of modifying several nucleotides by exchanging the original sequence with an identical sequence in reverse order.

Chromosomal rearrangements encompass significant changes that occur at a macro level, involving deletions, duplications, insertions, inversions, substitutions, and translocations. These alterations impact the structure and organization of chromosomes. One example of a disorder resulting from a chromosomal rearrangement is Trisomy 21, where there is an additional copy of chromosome 21, leading to Down syndrome. This disorder illustrates how changes in the number of chromosomes can have significant implications for an individual's health and development.

An individual genome contains over 4 million SNVs, 100,000 Indels and hundreds of SVs. These variants differ depending on ethnic origin and geographic distribution, illustrating the significant genetic variety of humans (1000 Genomes Project Consortium et al., 2015).

I.2. Genetic testing techniques

I.2.1. Sanger sequencing

The Sanger sequencing technique was developed by Frederick Sanger and his colleagues in 1977 (Sanger et al., 1977). It serves to identify the arrangement of nucleotide bases in a DNA fragment with a size smaller than 1,000 bp. This method relies on the detection of fluorophore-labeled dideoxyNucleotide TriPhoshates (ddNTPs) that are randomly incorporated into the DNA fragment during replication by a DNA polymerase. Sanger sequencing has an impressive accuracy rate of 99.99% for detecting nucleotide bases, making it widely recognized as the "gold standard" for DNA sequence validation. However, due to its cost and time-consuming, the Sanger sequencing method was not feasible for sequencing the whole genome.

I.2.2. Microarray

Genotyping, which involves the analysis of multiple SNVs using SNP arrays, has emerged as a practical and cost-effective method for generating high-quality genotyping data for a large number of individuals (Das et al., 2016; LaFramboise, 2009). Array genotyping technology not only provides medically relevant information concerning neurological diseases such as Alzheimer's Disease (AD), Parkinson's disease (PD), Dementia with Lewy Body (DLB), FrontoTemporal Dementia (FTD), Amyotrophic Lateral Sclerosis (ALS), Progressive Supranuclear Palsy (PSP), CorticoBasal Degeneration (CBD), and Multiple System Atrophy (MSA), but it also offers non-medical data including ethnicity and individual traits (Blauwendraat et al., 2017). This advanced array offers researchers and clinicians an enhanced platform for studying neurological disorders and gaining deeper insights into the genetic landscape associated with these diseases.

The NeuroChip, an updated version of NeuroX (Nalls et al., 2015), is an example of a genome-wide genotyping array (Infinium HumanCore-24 v1.0). It comprises 306,670 tagging variants and 179,467 variants associated with neurodegenerative diseases. Recently, NeuroChip released its latest update, named Neuro Booster array, featuring an extensive variant backbone of 1.9 million variants and over 95,000 custom contents.

The genotyping array method has been extensively employed in Genome-Wide Association studies (GWAS) for many years (Hirschhorn and Daly, 2005). The GWAS are employed to determine the association between distinct genetic changes and specific traits, such as diseases. This technique involves studying the genomes of many unrelated individuals, with or without a disease, to search for statistically significant variations in SNP frequency that can be used to predict the presence of a disease. A GWAS study should include more than 1000 patients and controls and more than 300,000 markers to allow for statistical comparison of allele frequencies between cases and controls (Balding, 2006). The GWAS catalog (https://www.ebi.ac.uk/gwas/) contains all the GWAS that have been

performed. Most SNPs do not directly affect gene products because they are found in non-coding regions of the genome. Instead, it is believed that they have a role in controlling gene expression. The analysis of such variants in the context of gene expressions measured in cells or tissues has spawned a big field in human genetic study known as Expressed Quantitative Trait Loci (eQTL) (Nica and Dermitzakis, 2013).

Polygenic Risk Score analysis (PRS) is calculated from the summary statistical data of the GWAS (Torkamani et al., 2018). It is a statistical tool that quantifies an individual's genetic predisposition to a specific trait or disease based on the cumulative effect of multiple SNPs. The first disease studied with the PRS method was schizophrenia (Lesage and Brice, 2009).

I.2.3. Next-Generation Sequencing

Next-Generation Sequencing (NGS) is a powerful technique used to determine the precise order of nucleotides in whole genomes or targeted regions of DNA or RNA. In comparison to Sanger sequencing, which can generate sequences of up to 1,000 bp with a per-base accuracy of 99.9%, the NGS technique employs parallel amplification and shorter read-length sequencing, resulting in an average raw error rate of 1% to 1.5%. Despite this error rate, NGS technology has proven to be superior in terms of cost and time efficiency. Several well-known brands dominate the NGS technology landscape for genome sequencing. These include Illumina's HiSeq2000, Roche's 454, Thermo Fisher's SOLiD and IonTorrent, Oxford Nanopore Technologies (ONT), and Pacific Biosciences platforms (PacBio). Each platform has its own set of advantages and disadvantages in terms of pricing, reliability, time required for analysis, and overall cost-effectiveness (Table 1).

Table 1 : Comparison of read length between NGS platforms

	Roche454	**HiSeq2000**	**Solid**	**IonTorrent**	**Oxford nanopore**	**PacBio**
Company	Roche	Illumina	Thermo Fisher	Thermo Fisher	Oxford Nanopore Technologies	Pacific bioscience
Maximum read length	700 bp	2 X 100 bp	75 + 35 bp	200 bp	30 kbp	50 kbp

While longer-read technologies offer advantages in terms of resolving repetitive and ambiguous regions, their cost remains a significant barrier, particularly for large-scale genome projects. It is expected that costs will continue to decline, making this long-read technique more accessible and beneficial for a wider range of genomic research activities.

I.2.3.1. Next-Generation Sequencing workflow

Regardless of the specific instrument technology employed, a typical NGS experiment follows a similar workflow consisting of three main steps: library preparation, library amplification, and sequencing. In this process, the starting material can be DNA or RNA, which can be transcribed into complementary DNA (cDNA) if required and used for library preparation. In the library preparation

step, the DNA or cDNA sample undergoes fragmentation into short double-stranded fragments ranging from 100 to 800 bp in size. This fragmentation can be achieved through methods such as sonication, nebulization, or enzymatic digestion. The resulting DNA fragments are then ligated to appropriate adaptor sequences specific to the sequencing technology, forming a fragment library. In contrast to unidirectional sequencing, paired-end libraries are used to sequence DNA fragments from both ends. Subsequently, during the library amplification step, each unique DNA molecule in the library is attached to a solid surface, and the Polymerase Chain Reaction (PCR) amplification is performed. This amplification process generates multiple identical clones of the DNA fragment, increasing the detectable signal for each target during sequencing. Most sequencing devices employ optical detection methods to track the incorporation of nucleotides during DNA synthesis. This allows for the generation of short sequencing reads, which are then aligned or mapped to the reference human genome. Following alignment, a variant calling process is performed to identify genetic variants. To assess the potential biological impact of these detected alterations, further annotation and analysis of these variants are necessary. This can be achieved by utilizing publicly available tools and databases, which provide additional information and context to interpret the significance of the observed genetic variants.

I.2.3.2. Short read sequencing

Whole-Exome Sequencing (WES) is a highly reliable method for studying genetic variants specifically within the coding regions of the genome, whereas Whole-Genome Sequencing (WGS) offers a comprehensive analysis of the entire genome. WES, with an average coverage depth of 100X, remains a more cost-effective option compared to WGS, which typically has an average coverage depth of 30X, and is priced at around 400€ per sample (Goh and Choi, 2012). Furthermore, WES data per patient is approximately six times smaller than WGS data, resulting in faster processing times and reducing the financial strain on data storage resources (Clayton-Smith et al., 2011). However, it's important to note that WES does have certain limitations when compared to WGS. Since WES focuses on coding regions, some protein-coding areas of the genome may not be fully covered due to incomplete annotation. Additionally, WES excludes potentially functional non-coding elements such as untranslated regions, enhancers, and long non-coding RNAs. Furthermore, WES has limited capability in identifying structural variants like CNVs, translocations, and inversions (Goh and Choi, 2012). Despite these limitations, WES remains a valuable and cost-effective approach for investigating coding variants, while WGS offers a more comprehensive but relatively more expensive analysis of the entire genome.

I.2.3.3. Long-read sequencing

Long-Read Sequencing (LRS), also known as third-generation sequencing, is a DNA sequencing technology that surpasses conventional short-read sequencing methods by enabling the identification

of nucleotide sequences in lengthy DNA fragments, ranging from 10,000 to 100,000 bp. Moreover, LRS provides valuable data on DNA methylation, which offers insights into epigenetic modifications (Logsdon et al., 2020; Miller et al., 2021). LRS enhances the accuracy of *de novo* assembly, mapping, transcript isoform identification, and SVs discovery, whereas these regions in short-read sequencing present major difficulties (Chaisson et al., 2019; Eichler, 2019; Hiatt et al., 2021). Currently, two dominant producers of long-read sequencing technologies are PacBio with their Single-Molecule Real-Time sequencing (SMRT) platform, and ONT (Amarasinghe et al., 2020). Both PacBio and ONT have developed platforms for real-time sequencing of nucleic acids (DNA and RNA) that offer faster sequencing capabilities compared to traditional short-read technologies. Since their commercial release in 2011 for SMRT and 2014 for nanopore sequencing, both technologies have gained popularity and have been employed in a wide range of applications (Amarasinghe et al., 2020).

Nanopore sequencing (MinION, GridION, and PromethION) works by detecting changes in ionic current when single-stranded DNA fragments pass through biological nanopores, tiny protein-based pores encapsulated within membranes (Jain et al., 2016; Rang et al., 2018). The different DNA sequences cause distinct levels of resistance as they traverse these pores, enabling the precise determination of the nucleotide sequence. One limitation of nanopore sequencing is the challenge of incorporating very high molecular weight DNA into the pore, which can impact throughput (Jain et al., 2018).

On the other hand, SMRT sequencing (RSII, Sequel, and Sequel II) relies on detecting variations in fluorescence levels during the replication of a target DNA sequence using modified nucleotides (Merker et al., 2018; Roberts et al., 2013). This process occurs in a series of wells and is constrained by the quality and lifespan of the DNA polymerase used. SMRT sequencing allows for sequencing library inserts ranging from 250 bp to 50 kb, while ONT achieves insert sizes between 10 to 30 kb. SMRT and ONT technologies demonstrate raw base-call error rates of less than 1% (Wenger et al., 2019) and 5% (Jain et al., 2018), respectively.

I.2.4. Multiplex ligation dependent probe amplification

Multiplex Ligation dependent Probe Amplification (MLPA) is a technique used to identify abnormal variations in the copy numbers of specific genomic regions (Stuppia et al., 2012). It can also be employed to detect abnormal DNA methylation patterns. MLPA involves a multiplex PCR assay using up to 40 probes, each designed to target a specific DNA sequence of interest. Each probe consists of a target-specific sequence and a universal primer sequence, enabling simultaneous amplification of all the probes by PCR (Stuppia et al., 2012). Capillary electrophoresis is then used to separate the amplified products based on their lengths. The amount of target sequence present in the sample is directly proportional to the amount of amplification product generated by MLPA. To

identify copy number abnormalities, the heights of the fluorescence peaks derived from the PCR amplification products are compared to control DNA samples with known copy numbers. Homozygous deletions can be identified by the absence of peaks specific to the target gene. Heterozygous CNVs, including deletions or duplications, result in variable peak heights, making interpretation challenging. The presence of varying PCR reaction efficiencies further complicates the interpretation process. MLPA probes, which can recognize sequences ranging from 60 to 80 nucleotides, enable the detection of single exon duplications and deletions (Stuppia et al., 2012). Currently, MLPA analysis remains the gold standard for the investigation of CNVs and SVs, providing a reliable and widely accepted method for such analyses.

I.3. Parkinson's disease

I.3.1. Pathophysiology and epidemiology

PD is the second most prevalent neurodegenerative disorder, after AD. PD affects approximately 2-3% of individuals over the age of 65 (Tysnes and Storstein, 2017). Currently, an estimated 6 million people worldwide are affected by PD and this number is expected to double in 2040 due to an aging population and increased longevity (Dorsey et al., 2018). PD is a complex multifactorial disorder influenced by both genetic and environmental factors. Monogenetic PD is rare, accounting for 30% of familial and 3 to 5% of sporadic cases, the inheritance pattern follows the classical Mendelian pattern (Kumar et al., 2011). However, the vast majority of PD patients are sporadic and the etiology is believed to be multifactorial with environmental influences (Lesage and Brice, 2009). The prevalence of PD is observed to be higher in men than in women, possibly attributed to increased estrogenic activity that provides a protective effect (Haaxma et al., 2007; Pringsheim et al., 2014).

The pathologic hallmark of PD is the degeneration of dopaminergic neurons, with dopamine deficiency, in the *Substantia Nigra pars compacta* (*SNpc*) (Greenfield and Bosanquet, 1953). In addition to this degeneration, abnormal α-synuclein protein aggregates are present in neuronal cell bodies and neurites called Lewy Bodies (LBs) or Lewy Neurites (LNs), respectively (Dickson, 2018). The presence of these accumulations can be detected using anti-synuclein antibodies and are observed not only in PD but also in other diseases such as MSA, Neurodegenerations with Brain Iron Accumulation (NBIA), AD, and even in aged healthy individuals (Hardy et al., 2009; Stefanis, 2012). However, each disease exhibits distinct accumulation patterns and subcellular distributions (Stefanis, 2012). Synucleinopathies are a group of neurodegenerative disorders characterized by the accumulation of abnormal α-synuclein protein aggregates in the brain. Examples include MSA and DLB (Galpern and Lang, 2006). In contrast, different neurodegenerative conditions involve the accumulation of distinct proteins predominant. Tauopathies, for example, involve the irregular aggregation of the tau protein in the brain. AD, PSP, FTD and CBD are examples of tauopathies.

PD signs and symptoms were first described by James Parkinson in 1817 and initially referred to as "Shaking Palsy" (Parkinson, 2002). The discovery of the Lewy Bodies by Friedrich Lewy in 1912 led to the identification of the Substantia Nigra as the primary cerebral structure affected in PD in 1919 (Goedert et al., 2013).

The clinical hallmark of PD is the presence of postural instability, muscular rigidity and bradykinesia (slowness of movement) (Tolosa et al., 2006). To diagnose PD, the Movement Disorder Society (MDS) recommends the presence of motor symptoms such as tremors, rigidity, bradykinesia, postural instability, and abnormal walking, along with other features like hyposmia, dyskinesia due to levodopa (L-DOPA), or a significant response to dopamine replacement therapy (Postuma et al., 2015). It is now well-established that most PD patients experience various non-motor symptoms, including autonomic dysfunction (e.g., constipation, orthostatic hypotension, urogenital dysfunction), neuropsychiatric dysfunction (e.g., apathy, anxiety, psychosis, dementia), sleep behavior disorders (e.g., sleep fragmentation, insomnia, Rapid Eye Movement (REM) Sleep Behavior Disorder (RBD)), and sensory symptoms (e.g., abnormal sensations, pain) (Braak et al., 2003; Poewe, 2008). Some of these non-motor symptoms may precede the onset of motor symptoms by several years, making them potentially useful for early-stage PD diagnosis (Mahlknecht et al., 2015) (Figure 2). PSP, CBS, and MSA are among the neurodegenerative diseases that can present resembling PD symptoms. Although each exhibits a distinct syndrome from PD (Bhidayasiri et al., 2019), they usually do not respond effectively to PD treatments such as L-DOPA administration (Lamb et al., 2016; Owolabi, 2013).

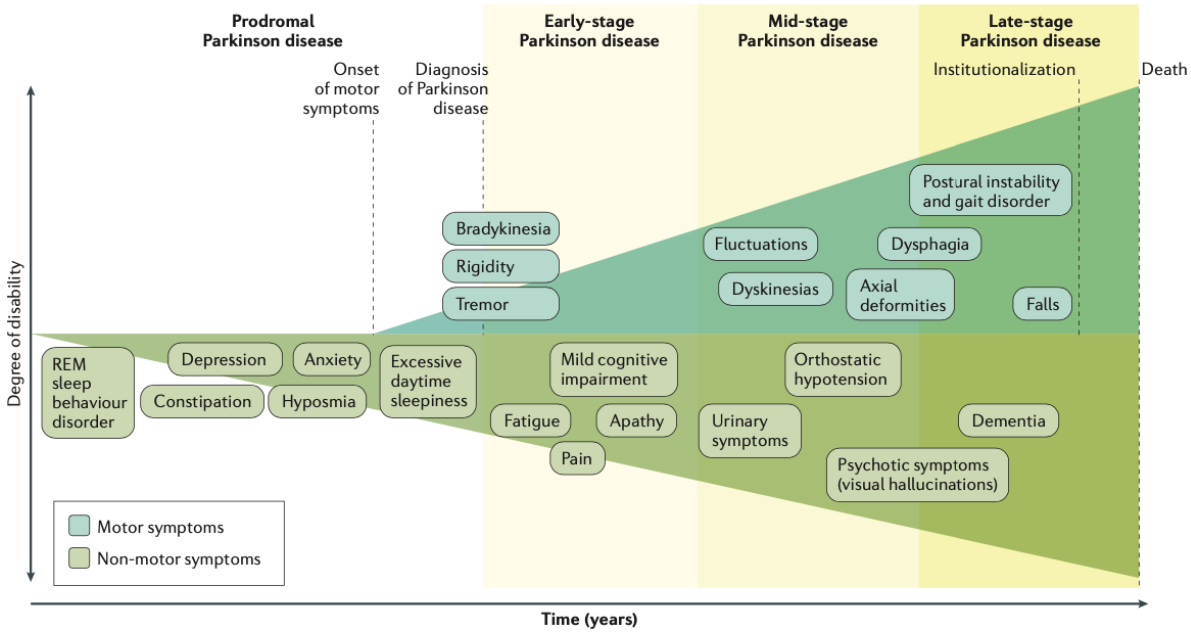


Figure 2 : Clinical symptoms of Parkinson's disease development
The onset of PD is preceded by a prodromal phase marked by non-motor symptoms like insomnia, constipation, and hyposmia. People with PD experience both motor and non-motor symptoms throughout the disease. (Source: Copyright 2017 Nature Reviews Disease Primers from Werner Poewe et al, Macmillan Publishers Limited, Springer Nature)

I.3.2. Genetics architecture of Parkinson's disease

Initially, PD was largely considered a non-genetic disease. However, the identification of PD cases within the same family led to investigations into the potential genetic basis of the disease. PD is now categorized into two main types: familial and sporadic, based on the presence or absence of affected family members, particularly first-degree relatives (Kenborg et al., 2015). Around 15% of all PD cases are classified as familial, with 5-10% of patients carrying variants associated with monogenic forms of the disease (Lesage and Brice, 2009).

I.3.2.1. Genetic causes of Parkinson's disease

To date, 13 loci and 20 genes have been identified as associated with autosomal dominant or recessive forms of PD or atypical parkinsonism (Del Rey et al., 2018). The designation "PARK" is used as a prefix for genes implicated in familial monogenic forms of PD, following the chronological order of their discovery (Coppedè, 2012) (Figure ***3***). It is important to note that certain PARK loci, such as PARK1 and PARK4, refer to the same gene *SNCA*, while some earlier PARK loci, like PARK5, are no longer considered disease-causing (Table 2). Current recommendations suggest using gene names instead of numbered loci (Marras et al., 2016). However, only a limited subset of genes, including *SNCA*, *LRRK2*, *VPS35*, *PRKN*, *PINK1*, and PARK7 (*DJ-1)*, have been positively linked to typical PD (Figure 4), as their associations have been replicated in numerous studies involving large populations of PD patients(Bandres-Ciga et al., 2020).

Additionally, genetic variants in *GBA1* and the Microtubule Associated Protein Tau (*MAPT*) are established risk factors for PD without the Mendelian segregation (Pastor et al., 2000; Sidransky et al., 2009). Other genes, such as *UCHL1*, *HTRA2*, *GIGYF2*, *EIF4G1*, *SMPD1*, *DNAJC13*, *CHCHD2*, *TMEM230*, *RIC3*, *LRP10*, *NUS1*, and *ARSA*, have been identified as potential disease-causing candidates, although their functional confirmation and replication validation are still required (Deng et al., 2016; Fernández-Santiago and Sharma, 2022; Guo et al., 2018; Makarious et al., 2019; Quadri et al., 2018). Furthermore, genes such as *POLG*, *ATP13A2*, *FBXO7*, *PLA2G6*, and *SYNJ1* have been associated with atypical Parkinsonian syndromes (Fernández-Santiago and Sharma, 2022).

These rare Mendelian forms follow distinct inheritance patterns and have a notable impact on disease development (e.g., *SNCA*, *LRRK2*, *VPS35, PRKN*, *PINK1* and *PARK7*). *GBA1* and *LRRK2* genes are low-frequency variants with significant effects. These variants are not as rare as monogenic mutations, but their frequency remains relatively low in the general population. Then, there are the common variants that exert minor effects and highlight genetic variants that are prevalent in the population but individually contribute only a modest risk toward PD (Manolio et al., 2009).

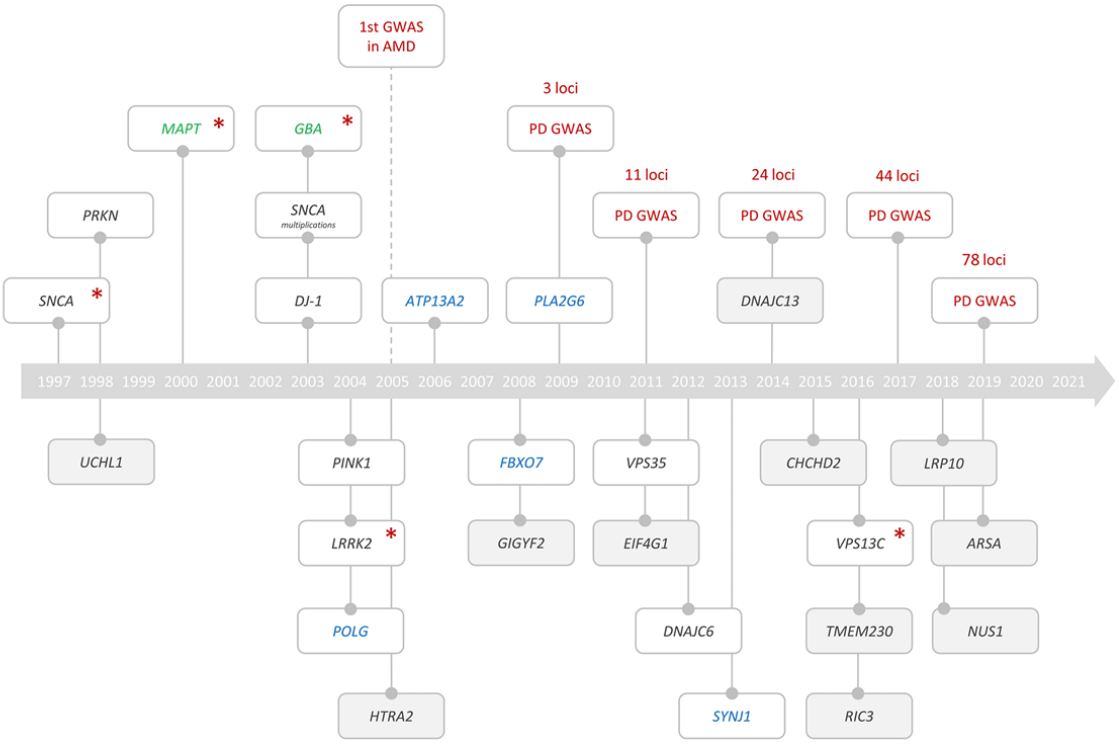


Figure 3 : Genetic research into Parkinson's disease over the past 25 years
New loci linked to the disease are in red. Mendelian forms of PD in black or atypical parkinsonism in blue. Genes nominated by GWAS are marked with red asterisks. PD risk genes without Mendelian segregation in green. (Source: Copyright 2022 from Rubén Fernández-Santiago, Manu Sharma, Ageing Research Reviews, Published by Elsevier B.V. Licensed under the Creative Commons Attribution CC BY license (https://creativecommons.org/licenses/by/4.0/))

Table 2 : Parkinson's disease related locus and genes

Locus	Gene	Name	Inheritance	Lead to PD	Protein function
PARK1/PARK4 (4q21-22)	SNCA	α-synuclein	AD	Mendelian forms	Synaptic protein. Major component of LB
PARK2 (6q25.2-27)	PRKN	parkin	AR	Mendelian forms	E3 ubiquitin ligase involved in protein degradation
PARK3 (2q13)	Unknown	Parkinson Disease 3	AD		Unknown
PARK5 (4p13)	UCHL1	Ubiquitin C-Terminal Hydrolase L1	AD		Processing of ubiquitin precursors
PARK6 (1p35-36)	PINK1	PTEN-induced kinase 1	AR	Mendelian forms	Mitochondrial kinase involved in mitochondrial quality control
PARK7 (1p36)	PARK7 (DJ-1)	Protein deglycase Daisuke-Junko-1	AR	Mendelian forms	Redox-sensitive chaperone with an anti-oxidative stress function
PARK8 (12q12)	LRRK2	Leucine-rich repeat kinase 2	AD	Mendelian forms	Multiple functions by several protein domains
PARK9 (1p36)	ATP13A2	ATPase cation transporting 13A2	AR	Mendelian forms	Maintenance of lysosomal and mitochondrial cation homeostasis
PARK10 (1p32)	Unknown	Parkinson Disease 10	AD		Unknown
PARK11 (2q36-37)	GIGYF2	GRB10 interacting GYF protein 2	AD		May be involved in the regulation of tyrosine kinase receptor signaling
PARK12 (Xq21-25)	Unknown	Parkinson Disease 12	X-linked		Unknown
PARK13 (2q12)	HTRA2	High Temperature Requirement Protein A2	AD		Serine protease involved in caspase-dependent apoptosis
PARK14 (22q13.1)	PLA2G6	phospholipase A2 group VI	AR		Lipase involved in phospholipid metabolism
PARK15 (22q12-13)	FBXO7	F-Box Protein 7	AR		E3 ubiquitin ligase involved in protein degradation
PARK16 (1q32)	PARK16	Parkinson Disease 16	Unknown		Unknown
PARK17 (16q11.2)	VPS35	Vacuolar protein sorting-associated protein 35	AD	Mendelian forms	Mitochondria-peroxisomes and endosome-trans-Golgi trafficking
PARK18 (3q27.1)	EIF4G1	Eukaryotic Translation Initiation Factor 4 Gamma 1	AD		Recruitment of mRNA to the ribosome
PARK19 (1p31.3)	DNAJC6	DnaJ Heat Shock Protein Family (Hsp40) Member C6	AR		Regulation of molecular chaperone activity
PARK20 (21q22.11)	SYNJ1	Synaptojanin 1	AR		Regulation of synaptic vesicle dynamics
PARK21 (3q22.1)	DNAJC13	DnaJ Heat Shock Protein Family (Hsp40) Member C13	AD		Required For Receptor-Mediated Endocytosis
PARK22 (7p11.2)	CHCHD2	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 2	AD		Negative regulator of mitochondria-mediated apoptosis
PARK23 (15q22.2)	VPS13C	Vacuolar Protein Sorting 13C	AR		Delivery of damaged mitochondria cargo to lysosomes

AD, autosomal dominant; AR, autosomal recessive.

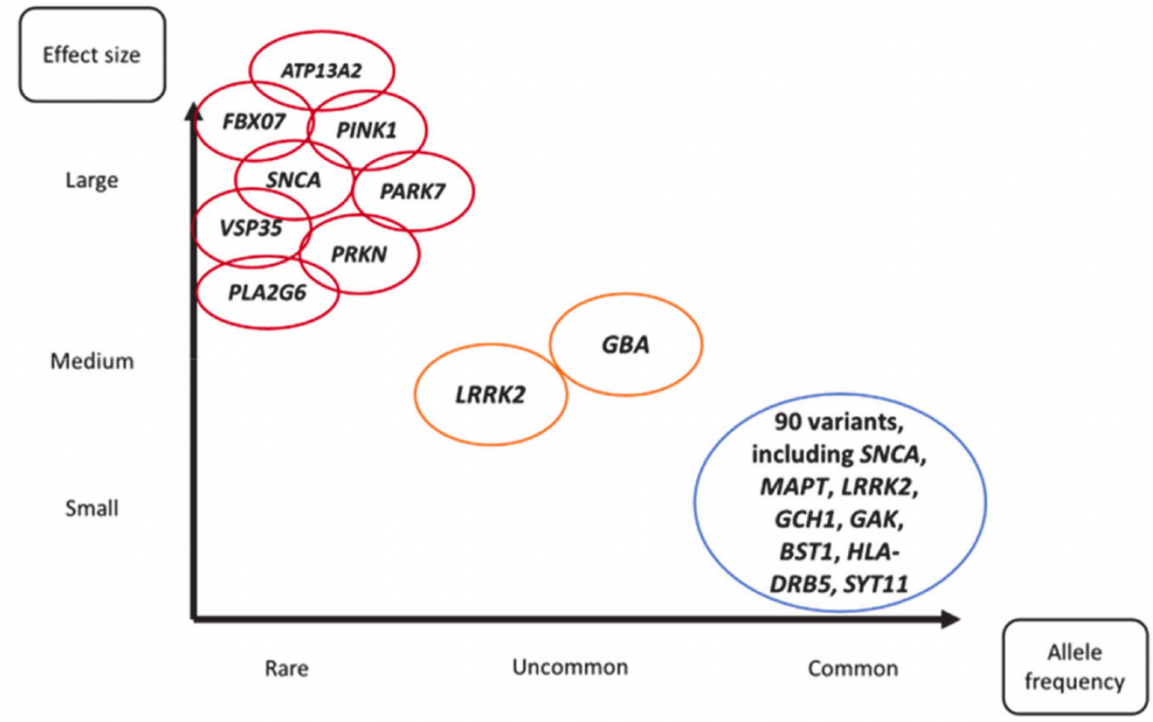


Figure 4 : The genetic architecture of Parkinson's disease
Correlation between genetic variants frequency and penetrance in established PD-associated genes. Monogenic traits, which are rare variants, typically have high penetrance, while polygenic traits, which are more common variants, typically have lower penetrance. (Source: Copyright 2021 Day and Mullin. Licensed under the Creative Commons Attribution CC BY license (https://creativecommons.org/licenses/by/4.0/))

PARK1/4: α-synuclein (SNCA)

The *SNCA* gene encoded the α-synuclein protein, which is primarily expressed in neural tissue and plays a crucial role in the formation of Lewy bodies. This protein tends to aggregate, forming oligomers and fibrils. Variants in the codon sequence (p. A30G, p.A30P, p.E46K, p. H50Q, p. G51D, p.A53T, p.A53V, p.A53E, p.T72M, and p.E83Q), as well as duplications and triplications of the entire *SNCA* gene, have been associated with PD (Guo et al., 2021; Houlden and Singleton, 2012; Krüger et al., 1998; Magistrelli et al., 2021; Siddiqui et al., 2016; Xia et al., 2022). The precise mechanism by which *SNCA* variants and CNVs lead to PD is not yet fully understood (Guo et al., 2021). However, these variants seem to impact the multimerization of the α-synuclein protein and alter synaptic regulation and maturation, resulting in reduced dopamine levels in the striatum, α-synuclein accumulation, and tau deposition (Poulopoulos et al., 2012). Under certain circumstances, α-synuclein can form inclusions in oligodendroglial cells, similar to those found in MSA (Asi et al., 2014). Further research suggests that the *SNCA* protein may also be involved in processes such as vesicle endocytosis and dissolution, mitochondrial function, and interaction with lysosomal membranes (Benskey et al., 2016). Individuals with *SNCA* variants often exhibit dementia and hallucinations, resembling the symptoms seen in DLB (Houlden and Singleton, 2012).

PARK2: Parkin (PRKN)

The *PRKN* gene encodes the Parkin protein, a cytosolic E3 ubiquitin ligase that plays a significant role in the Ubiquitin Proteasome System (UPS), which targets abnormal proteins for degradation (Chan and Chan, 2011). Furthermore, *PRKN* is essential for the autophagy process involved in clearing damaged mitochondria (Fiesel et al., 2015). Various mutations, including point mutations, frameshift mutations, deletions, and exon duplications, have been identified in the *PRKN* gene (Abbas et al., 1999). These mutations are commonly found in patients with early-onset PD before the age of 31 (Kasten et al., 2018). The mutations result in a loss of normal E3 ligase activity, leading to the accumulation of substrates such as synphilin-1, synaptotagmin XI, and Hsp70 (Shimura et al., 2000). Individuals with *PRKN* variants typically develop focal dystonia, psychosis, bradykinesia, tremor and experience a slower disease progression (Goldman and Postuma, 2014; Lücking et al., 2000). Generally, these patients do not exhibit Lewy body pathology.

PARK6: PTEN-induced kinase 1 (PINK1)

The *PINK1* gene is responsible for encoding the PTEN-induced putative kinase-1 protein. This protein is ubiquitously expressed and plays a crucial role in the *PINK1* and *PRKN* protein complex, which is involved in the degradation of damaged mitochondria (Geisler et al., 2010; Matsuda et al., 2010; Valente et al., 2004). Mutations in *PINK1* lead to the accumulation of α-synuclein in LBs, but interestingly, the interaction between *PINK1* and α-synuclein can reduce the neurotoxicity associated

with Lewy bodies by inducing autophagy in damaged neurons (Valente et al., 2004). Pathogenic variants in *PINK1* have been identified as a cause of early-onset PD (onset before the age of 32) and variant carriers are susceptible to dopa-induced dyskinesia (Brooks et al., 2009; Kasten et al., 2018).

PARK7: DJ-1

The DJ-1 protein is ubiquitously expressed and tends to form dimers. Mutations in the *PARK7* gene disrupt the formation of the dimers (Dawson and Dawson, 2003). Although the exact function of *DJ-1* is still being investigated, there is evidence to suggest that it plays a role in oxidative stress and mitochondrial function (Raninga et al., 2017). Patients with *PARK7* mutations typically develop the disease at an early age, before the age of 30 (Klein and Westenberger, 2012). Individuals with *PARK7* mutations exhibit abnormal characteristics such as tremors and bradykinesia and are highly responsive to L-DOPA treatment (Kilarski et al., 2012).

PARK8: Leucine-rich repeat kinase 2 (LRRK2)

The *LRRK2* gene is responsible for encoding the leucine-rich repeat kinase 2 protein, also known as dardarin. This gene produces a large multidomain protein that includes a roc (ras of complex protein) domain containing GTPase and a kinase domain (Gilsbach and Kortholt, 2014; Li et al., 2014). Mutations in different domains of *LRRK2* have been reported to cause various pathogenic effects (Zimprich et al., 2004). Certain pathogenic mutations, such as p.G2019S, p.R1441C, and p.I2020T, have been shown to increase *LRRK2* kinase activity and autophosphorylation (Gloeckner et al., 2006; West et al., 2005), while others like p.R1441H and p.Y1699C have been found to decrease GTPase activity (Liao et al., 2014; Nguyen and Moore, 2017). *LRRK2* mutation carriers typically exhibit symptoms between the ages of 50 and 60, with tremors and dystonia being predominant in these patients (Healy et al., 2008). The prevalence of *LRRK2* mutations in PD patients of European ancestry is around 1-7%, with the most common mutation (p.G2019S) being found in 20% of Ashkenazi Jewish patients (Lesage et al., 2006; Ozelius et al., 2006). *LRRK2* is associated with several pathways, including the endosomal, lysosomal, autophagic, protein synthesis, and cytoskeletal arrangement pathways (Beilina et al., 2014; Biskup et al., 2006; Martin et al., 2014; Meixner et al., 2011). Particularly, *LRRK2* plays an important role in the autophagy-lysosomal pathway (ALP). Lysosomal homeostasis and endosomal trafficking are regulated by the phosphorylation of Rab GTPases by *LRRK2*. The *LRRK2* mutant impairs chaperone-mediated autophagy, which causes α-synuclein to bind and oligomerize on lysosomal membranes (Pang et al., 2022).

PARK17: vacuolar protein sorting associated protein 35 (VPS35)

The *VPS35* gene encodes the vacuolar protein sorting 35 ortholog protein, which serves as a core subunit of the retromer complex along with *VPS26*, *VPS29*, and *SNX1* (Williams et al., 2017). The retromer complex plays a role in retrograde protein transport from endosomes to the trans-Golgi

network (Mukadam and Seaman, 2015). Individuals carrying the *VPS35* mutation typically exhibit a late onset of symptoms and are susceptible to L-DOPA treatment. Genetic mutations in this gene lead to reduced autophagic flux and decreased lysosomal mass, leading to the accumulation of α-synuclein (Hanss et al., 2021). In addition, the neurons presented dysfunctional mitochondria characterized by reduced membrane potential, impaired mitochondrial respiration and increased generation of reactive oxygen species. This mitochondrial impairment was correlated with a deficiency in mitophagy-mediated mitochondrial quality monitoring (Hanss et al., 2021).

***GBA1* a genetic risk factor**

The Glucosylceramidase Beta 1 gene (*GBA1*) is responsible for encoding the lysosomal enzyme known as Glucocerebrosidase (GCase). This enzyme catalyzes the catabolism of GlucosylCeramide (GlcCer) and GlucosylSphingosine (GlcSph) (Sidransky and Lopez, 2012). Homozygous mutations in *GBA1* result in Gaucher's Disease (GD), a rare genetic disorder (Pastores and Hughes, 1993). On the other hand, heterozygous mutations in *GBA1* have been identified as significant risk factors for PD. *GBA1* mutations occur more frequently than other genes linked to familial PD, such as *LRRK2*, *SNCA*, and *PARK2*, and are present in 5–15% of PD cases, making them the most significant genetic risk factor for PD (Smith and Schapira, 2022). *GBA1* variants represent a genetic risk factor, that increases the risk of developing PD by 5 to 30, depending on age, ethnic origin and the mutations considered in the analysis (Lesage et al., 2011; Migdalska-Richards and Schapira, 2016; Sidransky et al., 2009). There are approximately 300 known pathogenic mutations in the *GBA1* gene, which can lead to loss or gain of function (Smith and Schapira, 2022). The most common pathogenic mutations observed in *GBA1* are p.N409S and p.L483P (Neumann et al., 2009). The penetrance and frequency of *GBA1* mutations vary among different ethnicities (Mitsui et al., 2009). A more detailed description of the *GBA1* gene will be provided in a subsequent paragraph (I.3.5).

I.3.2.2. Sporadic Parkinson's disease

Extensive research has been conducted to investigate genetic variants associated with an increased risk of PD in the general population, aiming to establish a disease risk profile. The common disease-common variant hypothesis has served as the foundation for identifying risk variants in common diseases like PD. In this regard, case-control GWAS, focusing on common genetic variants, has significantly advanced our understanding of the genetic basis of sporadic PD. GWAS studies are effective in identifying genetic loci that contribute significantly to the traits of interest and can provide insights into the underlying biological processes. A GWAS study typically involves more than 1,000 patients and controls, as well as over 300,000 markers, enabling a statistical comparison of allele frequencies between cases and controls (Balding, 2006). The GWAS catalog (https://www.ebi.ac.uk/gwas/) serves as a comprehensive repository of all conducted GWAS studies. Although some criticisms have been raised regarding the use of GWAS, such as the small effect sizes

and lack of therapeutic relevance for certain risk loci, they still offer valuable insights into the molecular pathways underlying disease development. Additionally, due to population stratification, risk variants with odds ratios less than 1.5 may represent false positives. Furthermore, most risk variants identified by GWAS are located in non-coding regions with unclear functions, but they may serve as proxies for the causative variables (Visscher et al., 2017).

The first GWAS study on PD, conducted in 2005 with a small cohort, failed to identify any risk loci (Maraganore et al., 2005). However, subsequent independent GWAS studies with larger sample sizes revealed significant associations of PD with *LRRK2*, *SNCA*, *PARK16*, *GAK/DGKQ*, and HLA region, and established the importance of *SNCA* and *MAPT* in PD risk (Hamza et al., 2010; Pankratz et al., 2009; Satake et al., 2009; Simón-Sánchez et al., 2009). A large meta-analysis involving 13,708 cases and 95,282 controls identified 24 risk loci significantly associated with an increased risk of PD (Nalls et al., 2019b). In a recent GWAS, 17 novel PD risk loci were discovered among a cohort of 6,476 PD cases and 302,042 controls of European ancestry (Chang et al., 2017).

The overall impact of GWAS variants can be quantified using a PRS, which is specific to the ethnicity of the cohort and the characteristics of the studies. Using GWAS variants, it has been shown that individuals in the highest PRS quartile have a fourfold higher risk of PD compared to those in the lowest quartile (Nalls et al., 2019b). While most studies have focused on the European population, it is crucial to consider ethnic diversity in PD genomics research to gain a better understanding of the biological mechanisms underlying PD (Tam et al., 2019; Xue et al., 2018). Recently, several initiatives have been launched, such as the Michael J. Fox Foundation's (MJFF) program for ethnic diversity and the Global Parkinson's Genetics Program (GP2), which aim to expand research to understudied populations.

I.3.3. Environmental risk factors

Several environmental factors have been proposed to contribute to the development of sporadic PD. Positive correlations were found between factors such as depression, pesticide exposure, living in rural areas and head injuries, as well as factors such as diet (e.g., consumption of dairy products, soft drinks and red meat, for example). Negative correlations have been shown between the disease and some factors, such as smoking, drinking alcohol, coffee, and physical activity (Chairta et al., 2021; Hong et al., 2020; Wang et al., 2022).

Among the positive correlation factors, long-term exposure to pesticides, including insecticides, herbicides, and fungicides, has been identified as the most common risk factor for PD (Lai et al., 2002).

The neurotoxin 1-Methyl-4-Phenyl-1,2,3,6-TetrahydroPyridine (MPTP) was discovered in the 1980s to induce PD symptoms by damaging cells in the substantia nigra through inhibition of mitochondrial

complex I (Langston et al., 1983). Mitochondrial dysfunction has been identified as a key mechanism disrupted by pesticides in the pathogenesis of PD (Cochemé and Murphy, 2008). Furthermore, epidemiologic studies have revealed that prolonged exposure to certain metals and solvents can increase the risk of developing PD (Lock et al., 2013; Vellingiri et al., 2022).

I.3.4. Pathophysiology at the cellular and molecular level

Protein misfolding and aggregation are common mechanisms in most neurodegenerative diseases, including PD. In PD, there is a depletion of dopaminergic neurons in the *SNpc*, along with the aggregation of inclusion bodies. This accumulation occurs due to the decreased autophagy process (Sala et al., 2016). Moreover, various factors such as anomalies in the Endoplasmic Reticulum (ER) function, protein degradation pathways, intracellular trafficking, and calcium signaling accelerate the degeneration of dopamine neurons (Michel et al., 2016). The ER plays a crucial role in maintaining cellular homeostasis by ensuring proper protein folding and quality control. Synuclein accumulation disrupts vesicle trafficking between the ER and the Golgi apparatus, leading to ER stress (Gan et al., 2010). It is important to prevent ER mitochondrial dysfunction to protect against neurodegeneration and maintain normal physiological processes, thereby maintaining the normal transfer of calcium between these two compartments.

I.3.4.1. The role of α-synuclein

Lewy bodies and Lewy neurites, abnormal protein aggregates, are known to consist mainly of α-synuclein, which are characteristic pathological features of PD (Spillantini et al., 1997). Other neurodegenerative diseases also contain these α-synuclein aggregates, such as DLB and MSA (Spillantini and Goedert, 2018). Diseases resulting from α-synuclein aggregation are commonly referred to as α-synucleinopathies. α-Synuclein is a presynaptic protein encoded by the *SCNA* gene. It binds to synaptic vesicles and plays a role in neurotransmitter release by modulating synaptic vesicle storage (Gao et al., 2023).

Extensive research has revealed the impact of *SNCA* gene variants in PD. Several *SNCA* mutations, including A53T, A30G, A30P, E46K, H50Q, G51D, A53V, A53E, T72M, and E83Q, as well as gene duplications and triplications, have been implicated in familial PD (Gao et al., 2023). In addition, GWAS have shown an association between PD risk and common polymorphisms in the *SNCA* gene locus (Simón-Sánchez et al., 2009). The UPS and the lysosomal autophagy system, which are the main pathways responsible for protein degradation, are part of the α-synuclein clearance process (Cuervo et al., 2004; Tofaris et al., 2001; Webb et al., 2003). Dysfunction of these systems has been associated with PD, as evidenced by mutations in genes encoding proteins involved in the proteasome pathway or the lysosome-autophagy system (e.g., *ATP13A2*, *VPS35*), leading to familial forms of PD (Gan-Or et al., 2015; Zimprich et al., 2011). Among these pathways, the lysosomal autophagy system appears to play an important role in α-synuclein clearance, as inhibition of this system results

in α-synuclein accumulation (Vogiatzi et al., 2008). However, it should be noted that α-synuclein aggregation in Lewy bodies is not present in certain forms of familial PD, including cases associated with mutations in the *LRRK2* gene (Kalia and Lang, 2015). Therefore, the significance of these aggregations in patients is still under debate.

I.3.4.2. Mitochondrial dysfunction

The involvement of mitochondrial dysfunction in PD was initially recognized through exposure to the pesticide neurotoxin MPTP, which is converted into the neurotoxin 1-Methyl-4-PhenylPyridinium (MPP+), leading to parkinsonian syndrome in affected individuals (Langston et al., 1983). The oxidation of MPTP to MPP+ inhibits the activity of the mitochondrial electron transport chain and increases the production of Reactive oxygen species (ROS) (Langston et al., 1983). In patients with sporadic PD, a deficiency in mitochondrial complex I activity in the *SNpc* has been reported (Parker et al., 1989; Schapira et al., 1990). Moreover, various familial forms of PD-related syndromes result from mutations in genes that encode proteins involved in maintaining mitochondrial homeostasis. For instance, mutations in the *PINK1* and *PRKN* genes, crucial for initiating mitophagy (i.e., the degradation of damaged mitochondria), lead to early-onset autosomal recessive parkinsonian syndromes (Kitada et al., 1998; Valente et al., 2004). Similarly, mutations in the *PARK7* gene, which encodes *DJ-1*, result in an autosomal recessive familial parkinsonian syndrome (Healy et al., 2008). *DJ-1* in combination with *PINK1* preserves mitophagy and mitochondrial function during oxidative stress (Thomas et al., 2011). Thus, dysfunction in these pathways leads to the presence of abnormal mitochondria and an increase in oxidative stress responses (Joselin et al., 2012).

Although the link between α-synuclein aggregation and mitochondrial dysfunction is unclear, the *SNCA* p.A30P mutation has been shown to result in impaired neuronal activity, decreased mitochondrial respiration, energy deficit, increased sensitivity to rotenone and alterations in transcriptional patterns related to lipid metabolism (Barbuti et al., 2021).

Furthermore, overexpression of α-synuclein in neuronal cell models has been shown to alter mitochondrial shape, increase levels of free radicals, and reduce mitochondrial activity (Kamp et al., 2010; Parihar et al., 2009). These findings suggest that mitochondrial dysfunction occurs as a consequence of α-synuclein pathology. However, it has also been proposed that mitochondrial oxidative stress occurs early in a sequential pathogenic pathway, where high levels of oxidized dopamine lead to lysosomal dysfunction and α-synuclein accumulation (Burbulla et al., 2017). Consequently, while α-synuclein aggregation and mitochondrial dysfunction play significant roles in PD pathogenesis, the exact interplay between these cell-autonomous pathological processes remains uncertain.

I.3.4.3. Neuroinflammation

The pathogenic mechanism that has so far been described involves dysfunction protein homeostasis or intracellular pathways within the degradation of cells themselves. However, there is also evidence to suggest the involvement of non-cell autonomous mechanisms, specifically neuroinflammatory responses involving microglia and astrocytes (Hirsch and Hunot, 2009; Tansey and Goldberg, 2010). This is supported by the identification of a single nucleotide polymorphism in the human leukocyte antigen DQ1b (HLA-DQ1b) region associated with increased PD risk (Saiki et al., 2010). Furthermore, it has been suggested that *DJ-1*, *PRKN*, *PINK1*, and *LRRK2* play roles in the neuroinflammation associated with PD (Dias et al., 2013). *LRRK2* mutations, for instance, enhance the release of pro-inflammatory cytokines from activated microglia, leading to neurotoxicity (Badanjak et al., 2021; Gillardon et al., 2012). Loss-of-function mutations in *PRKN* contribute to inflammation-related degeneration of dopamine neurons (Frank-Cannon et al., 2008). *DJ-1*, by facilitating interaction between *STAT1* and *SHP1*, negatively modulates the inflammatory response of microglia and astrocytes (Kim et al., 2013). At autopsy examinations, active microglia have been observed in the substantia nigra of PD patients, indicating the presence of a neuroinflammatory response (Imamura et al., 2003; McGeer et al., 1988). Additionally, an increased density of astrocytes has been found in the substantia nigra of PD patients, although it remains unclear whether these findings represent a protective response or contribute to neurodegeneration (Damier et al., 1993). Targeting the neuroinflammatory response may hold promise as a potential therapeutic approach for PD treatment.

I.3.5. *GBA1* variants in Parkinson's Disease

Recent studies have revealed that variants in the *GBA1* gene play a significant role in the susceptibility to develop PD (Sidransky et al., 2009). Currently, *GBA1* is recognized as one of the genes most commonly linked to sporadic PD, and individuals carrying *GBA1* variants have an estimated risk of developing PD between 10% to 30% by the age of 80 (O'Regan et al., 2017).

I.3.5.1. Glucocerebrosidase and the *GBA1* gene

The *GBA1* gene is located on the long arm of chromosome 1 and consists of 11 exons and 10 introns (Horowitz et al., 1989). Its genomic coordinates are GRCh38 chr1:155234452-155244627 (https://www.ncbi.nlm.nih.gov/gene/2629). The *GBA1* gene covers a 7.6 kb sequence and encodes the lysosomal enzyme GCase (IUBMB enzyme nomenclature number EC 3.2.1.45) (Smith and Schapira, 2022). GCase cleaves the two GlycoSphingoLipids (GSLs), GlcCer and GlcSph, into glucose and ceramide, and glucose and sphingosine, respectively, within the lysosome (Smith and Schapira, 2022). The *GBA1* gene can be translated into at least two distinct mRNA sequences using different polyadenylation sites.

The mature GCase protein consists of 496 amino acids after the cleavage of the 39 amino acid signal peptides (Sorge et al., 1987). Structurally, the GCase protein has a three-domain structure. Its catalytic site is found in the third domain, characterized by a triose phosphate isomerase barrel structure (Dvir et al., 2003). Domain I comprise an antiparallel β-sheet structure with two disulfide bridges that facilitate correct protein folding. Domain II contains an immunoglobulin-like fold consisting of eight β-sheets. Domain III is a (β/α)8 barrel of Triosephosphate IsoMerase (TIM) (Dvir et al., 2003).

I.3.5.2. The pseudogene

The presence of a non-transcribed pseudogene called Glucosylceramidase beta 1 pseudogene (*GBAP1)* poses challenges in the genetic sequencing of the *GBA1* gene and sometimes hinders the identification of novel mutations (Horowitz et al., 1989). The *GBA1* pseudogene is located 16 kb downstream of the *GBA1* gene and shares 96% sequence homology with *GBA1*, constituting a 5.7 kb sequence (Figure **5**).

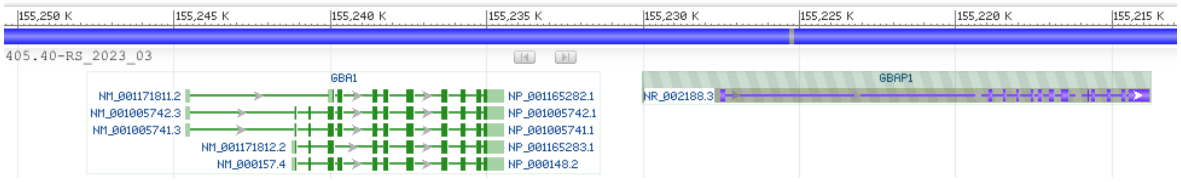


Figure 5 : Visualization of the *GBA1* gene and *GBAP1* pseudogene

The size difference between the two is caused by multiple Alu insertions in the intronic regions and 6.1-kb segment from the 5′ flanking sequence of the *GBAP1*. Two exonic deletions, a 4-bp deletion in exon 4 and a 55-bp deletion in exon 9, are believed to be the cause of the unfunctional pseudogene (Winfield et al., 1997).

The presence of *GBAP1* can lead to complex recombination events. The RecNciI allele, the first reported recombinant allele, carries three single nucleotide variants (p.L483P, p.A495P, and p.V499V) in exon 10 and corresponds to the *GBAP1* sequence inserted into *GBA1* (Latham et al., 1990). Subsequently, more recombinant alleles have been identified, typically involving RecNciI and a few additional variants derived from *GBAP1* (Tayebi et al., 2003).

The *MTX1* gene, which codes for the protein metaxin-1, is located right next to the 3' end of *GBAP1*. Metaxin-1 is a component of a preprotein import complex in the mitochondrial outer membrane (Armstrong et al., 1997).

I.3.5.3. Glucocerebrosidase pathway

The *GBA1* gene is transcribed into mRNA and transported from the nucleus to the ER, where GCase is synthesized (Do et al., 2019). It then binds to lysosomal membrane protein-2 (LIMP2), encoded by the *SCARB2* gene, facilitating the transfer of GCase through the Golgi to the endosome (Jović et al., 2012; Reczek et al., 2007). Subsequently, the endosome fuses with the lysosome, forming an autolysosome, and LIMP2 dissociates from GCase due to the low pH environment within the lysosome (Reczek et al., 2007). Finally, in the lysosome, GCase is activated by Saposin C (SAPC), a cleaved protein derived from prosaposin, and begins actively hydrolyzing its substrates (Do et al., 2019).

I.1.1.1. Association between *GBA1* variants and Parkinson's disease

GD is an autosomal recessive storage disorder caused by homozygous variants in the *GBA1* gene. It is characterized by the accumulation of glucosylceramide and reduced enzymatic activity of GCase in various cell types, particularly macrophages (Hruska et al., 2008). Based on the involvement of the Central Nervous System (CNS). GD is classified into three subtypes. Type 1 GD is the milder form of the disease, with the majority of patients living into adulthood. It is also known as "non-neuronopathic" because there is no obvious neurological involvement (Stirnemann et al., 2017). Type 2 and 3 GD are the "neuronopathic" forms of the disease. Most type 2 GD patients exhibit severe neurological symptoms and die earlier, within the first years of life. In type 3, the majority of patients live into adulthood despite severe neurological disorders (Grabowski, 2008).

The observation that first-degree relatives of GD patients frequently develop PD has led to the conclusion that having a single *GBA1* mutation is a risk factor for PD development (Goker-Alpan et al., 2004; Halperin et al., 2006). This runs against the conventional notion that type 1 GD does not affect the neurological system, and suggests that homozygous and compound heterozygous carriers of *GBA1* variants are more likely to develop PD.

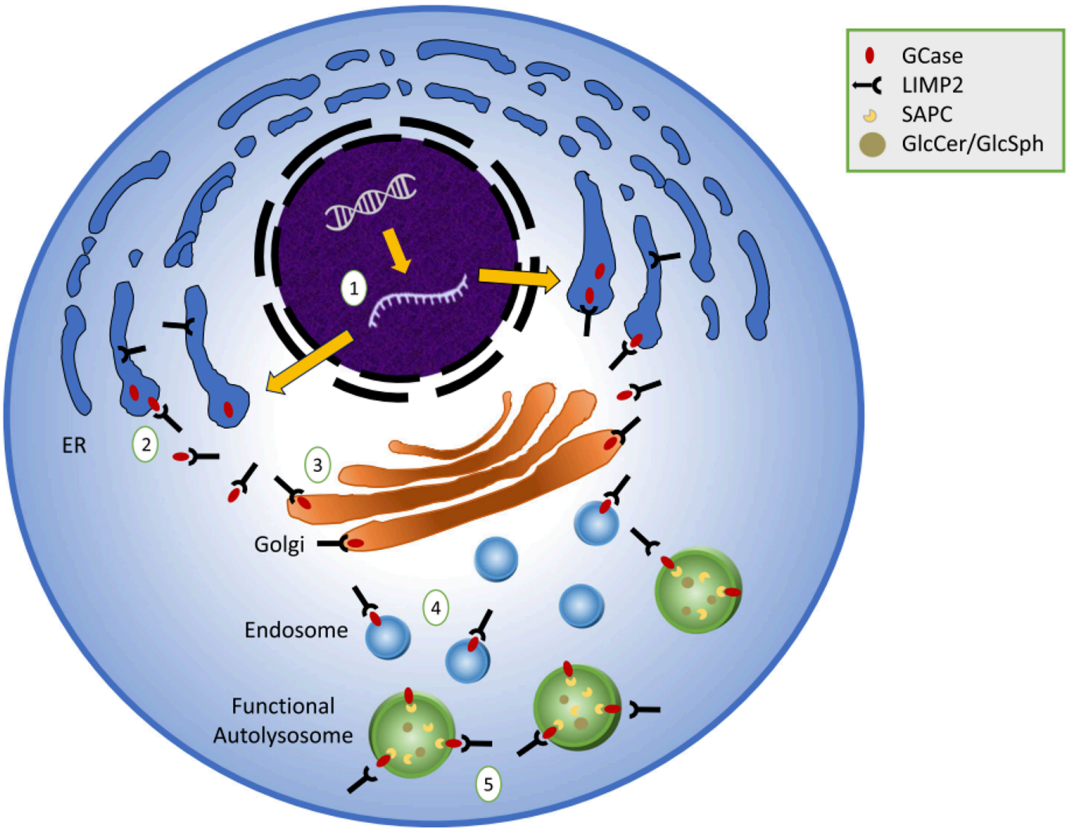


Figure 6 : GCase pathway in a functional cell

(Source: Copyright 2019 from Jenny Do et al, Molecular Neurodegeneration, Published by Springer Nature. Licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/))

I.1.1.2. Types of *GBA1* variants

Over 130 heterozygous *GBA1* variants have been identified as being associated with PD. Based on their historical links with neuropathic or non-neuropathic Gaucher disease, as well as their associations with PD risk and phenotypic severity, *GBA1* variants can be classified into three categories: severe variants, mild variants and at-risk variants.

Severe *GBA1* variants, such as p.L483P, p.L483R, p.A495P and p.R159W, are correlated with neuropathic GD. These variants have been shown to significantly increase the risk of PD and accelerate disease progression in PD. The mild *GBA1* variants that induce non-neuropathic GD, such as p.N409S, slightly increased the disease progression. The risk *GBA1* variants that are not pathogenic for GD but predispose to PD, notably p.E365K, p.T408M and p.E427K, predispose to more rapid motor and cognitive progression.

The effort to categorize *GBA1* variants faces complex challenges due to the existence of various classification methodologies, each developed by distinct groups of researchers. This complexity is increased by the range of criteria and factors used by the different variant classification systems, and for some variants, the type of variants proposed by one group differs from that of the other. Comparing the classification presented by Höglinger and colleagues in 2022 with the variants found in the *GBA1*-PD browser containing up-to-date information on *GBA1* variants, from Parlar SC and colleagues in 2023 (https://pdgenetics.shinyapps.io/GBA1Browser/), we noted that the p.R502H variant is classified as "Severe" by the Höglinger team and marked as "Unknown" by the *GBA1*-PD browser. Similarly, the p.Y244C variant is classified as "mild" by the *GBA1*-PD browser and as "severe" by the Höglinger team.

In response to these complexities, we proposed a classification approach that incorporates existing In silico bioinformatics databases for predicting the potential impact of variants on a protein or clinical level. For evaluating allele frequencies, we employed the Genome Aggregation Database (gnomAD r2.1), and variants with Minor Allele Frequencies (MAF) below 1% in the exomes and genomes database for the non-Finnish European (NFE) population were considered rare. To assess the relationships among human variants and phenotypes, we used two clinical databases, the Human Gene Mutation Database (HGMD) and ClinVar. To score the pathogenicity of the variants, we employed the Combined Annotation Dependent Depletion (CADD) and Rare Exome Variant Ensemble Learner (REVEL) scoring systems. CADD proposes ranking scores that predict the deleteriousness of variants, considering conservation and functional information; variants with scores equal to or greater than 20 are considered particularly deleterious. REVEL is an overall

method that predicts the pathogenicity of missense variants by integrating several scores. Scores greater than 0.7 are more likely to be pathogenic.

We subclassified *GBA1* variants into four categories based on their severity for PD:
(1) 'Severe' *GBA1* variants are Loss-Of-Function (LoF) variants defined as frameshift indels or "splice site (+/- 2 base pair)" variants OR missense variants annotated as "Pathogenic/Likely Pathogenic" in ClinVar and disease-causing mutation (DM) in HGMD with REVEL score > 0.7 AND CADD > 20 according to dbNFSP definition, (2) 'mild' *GBA1* variants are annotated "Pathogenic/Likely Pathogenic" OR "Conflicting interpretations of pathogenicity" in ClinVar with DM in HGMD with REVEL > 0.7 OR CADD > 20, (3) 'risk' *GBA1* variants are common variants annotated as "Pathogenic/Likely Pathogenic" OR "Conflicting interpretations of pathogenicity" in ClinVar, and (4) 'VUS' are Variant of Unknown Significant that did not fit into this classification scheme. We considered as pathogenic the severe, mild, and risk *GBA1* variants.

I.1.1.3. Frequency of *GBA1* variants

In contrast to the general population, where the prevalence of PD is around 3-4%, nearly 10% to 30% of *GBA1* carriers develop PD by the age of 80 (O'Regan et al., 2017; Rosenbloom et al., 2011). *GBA1* variants are more common in Ashkenazi Jews compared to other populations, and PD cases within this group often carry *GBA1* variants, with the most common p.N409S mutation (Gan-Or et al., 2008; Sidransky et al., 2009).
Early-onset PD, occurring before the age of 50 is associated with *GBA1* gene variants in approximately 25% of cases (Duran et al., 2013). The risk of developing PD varies among different *GBA1* variants, indicating a genotype-phenotype correlation. Homozygous L483P variants, for example, lead to type 3 GD, while the common p.N409S mutation is rarely associated with neuronopathic GD. Certain variants, such as L483P and D448H, have a significantly higher risk of developing PD compared to others, such as N409S (Alcalay et al., 2015; Gan-Or et al., 2008). *GBA1* gene variants represent the most significant genetic risk factor for PD, increasing the risk by 20 to 30 times (Lesage et al., 2011; Lesage and Brice, 2009; Sidransky et al., 2009; Stoker et al., 2018). Over 100 *GBA1* variations have been linked to PD, including recombination alleles, point mutations, splice-site variants, deletions, and insertions (Hruska et al., 2008; Huh et al., 2023; Lesage et al., 2011).

I.1.1.4. Comments on changing nomenclature

The *GBA1* gene variants are often denoted by the corresponding amino acid (aa) change, but there is a variation in how authors handle the signal peptide with 39 amino acids. Consequently, two parallel sets of coordinates for the same variants can confuse readers who are not familiar with these differences (Montfort et al., 2004). For instance, the variant p.N409S may historically still be qualified as p.N370S by some authors (Clark et al., 2007; Galvagnion et al., 2022; Wang et al., 2012).

Additionally, while the widely used transcript comprises 11 exons, some authors prefer a transcript with 12 exons.

Throughout this thesis, we will adopt a consistent nomenclature that includes the 39 aa signal peptides, and the exons will be numbered from 1 to 11, considering only the ones that encode the mature protein. This approach aims to eliminate confusion and maintain clarity in referring to *GBA1* gene variants.

I.1.1.5. Phenotype of *GBA1* variants carriers

GBA1 gene variants are typically associated with an earlier age of onset and a more severe clinical outcome in PD (Brockmann et al., 2015; Clark et al., 2007; Jesús et al., 2016; Nichols et al., 2009). Studies have demonstrated that patients with *GBA1* gene variants are more likely to experience cognitive decline, PD dementia, and a faster rate of motor progression (Cilia et al., 2016; Lunde et al., 2018; Oeda et al., 2015; Setó-Salvia et al., 2012; Winder-Rhodes et al., 2013). Various non-motor characteristics, including depression, urinary urgency, hyposmia, RBD, hallucinations, and constipation, are also more prevalent in *GBA1*-PD (Beavan et al., 2015; Goker-Alpan et al., 2010; Neumann et al., 2009). Some of these characteristics may indicate a prodromal stage of PD since they are more common in healthy individuals carrying *GBA1* variants compared to non-carriers (Beavan et al., 2015).

I.1.1.6. Biological mechanisms

The mechanisms by which *GBA1* variants contribute to PD are complex and not yet fully understood. A pathogenic link between the *GBA1* gene and α-synucleinopathies diseases has been demonstrated as *GBA1* variants are also present in DLB (Goker-Alpan et al., 2006; Mata et al., 2008; Nalls et al., 2013). Decreased GCase activity within the lysosome, leads to an accumulation of glucosylceramide and various lipid forms. This abnormal lipid accumulation can alter lipid membrane composition and potentially promote the aggregation of α-synuclein (Smith and Schapira, 2022) (Figure 7). This is supported by the increase of α-synuclein and glucosylceramide in dopaminergic induced Pluripotent Stem Cells (iPSCs) carrying *GBA1* variants (Schöndorf et al., 2014).

Alternative hypotheses propose that *GBA1* variants may exert a toxic gain-of-function effect. *GBA1* variants lead to the production of misfolded GCase protein, which remains trapped in the ER instead of being transported to the lysosome. This abnormal accumulation of GCase protein in the ER triggers the activation of ER stress pathways, such as the Unfolded Protein Response (UPR), which results in an inability to degrade α-synuclein (Bendikov-Bar and Horowitz, 2012; Fernandes et al., 2016; Schöndorf et al., 2014). Several experiments have shown that these variants disrupt the normal post-transcriptional modification of the GCase protein (Bendikov-Bar et al., 2011; Schmitz et al., 2005). As a result, misfolded GCase may cause a saturation of the ubiquitin-proteasome and ER-associated protein degradation systems, resulting in the inability to degrade other proteins, such as

α-synuclein (Bendikov-Bar and Horowitz, 2012; Fernandes et al., 2016; Schöndorf et al., 2014). Most pathogenic *GBA1* variants result in misfolding of GCase, leading to its retention in the ER and triggering ER stress and UPR (Fernandes et al., 2016).

A deficiency of GCase within the lysosome leads to impaired lysosomal function, disrupting the autophagic-lysosomal pathway. Consequently, lipid substrates such as GlcCer and GlcSph, as well as α-synuclein, accumulate within the cells. This accumulation can hinder the normal trafficking of newly synthesized GCase from the ER and Golgi to the lysosome, further aggravating lysosomal dysfunction. The compromised degradation of α-synuclein, due to faulty lysosomal and autophagic machinery, can result in an increased release of α-synuclein through exosomes. This mechanism allows the propagation of α-synuclein pathology throughout the brain, contributing to the progression of the disease (Smith and Schapira, 2022).

A lack of GCase and impaired Autophagy-Lysosome Pathway (ALP) can lead to impaired mitochondria clearance, resulting in the accumulation of defective mitochondria. Moreover, GCase deficiency has been associated with oxidative stress, reduced Adenosine triphosphate (ATP) production, and abnormal mitochondrial morphology. A limited number of studies have also reported mitochondrial dysfunction and ER stress in this context (Cleeter et al., 2013; Fernandes et al., 2016; Li et al., 2019; Osellame et al., 2013).

Neuroinflammation has been associated with GCase deficiency. The accumulation of lipids or α-synuclein can trigger the activation of microglia. In addition, α-synuclein released into the extracellular space can directly bind and activate microglia, further contributing to neuroinflammatory processes (Isik et al., 2023).

Furthermore, studies have demonstrated that increased GCase activity can reduce α-synuclein levels in iPSCs-derived neurons (Mazzulli et al., 2016; Sardi et al., 2013). This suggests the existence of a bidirectional loop, wherein decreased GCase activity leads to lysosomal dysfunction and the accumulation of glycosphingolipid substrates, ultimately resulting in the accumulation of pathological α-synuclein oligomers. This two-way communication between GCase and α-synuclein explains why there is decreased GCase activity in PD patients without *GBA1* and raises the possibility that inhibiting GCase could have an impact on α-synuclein pathology even in PD cases without *GBA1* (Gegg et al., 2012; Mazzulli et al., 2011).

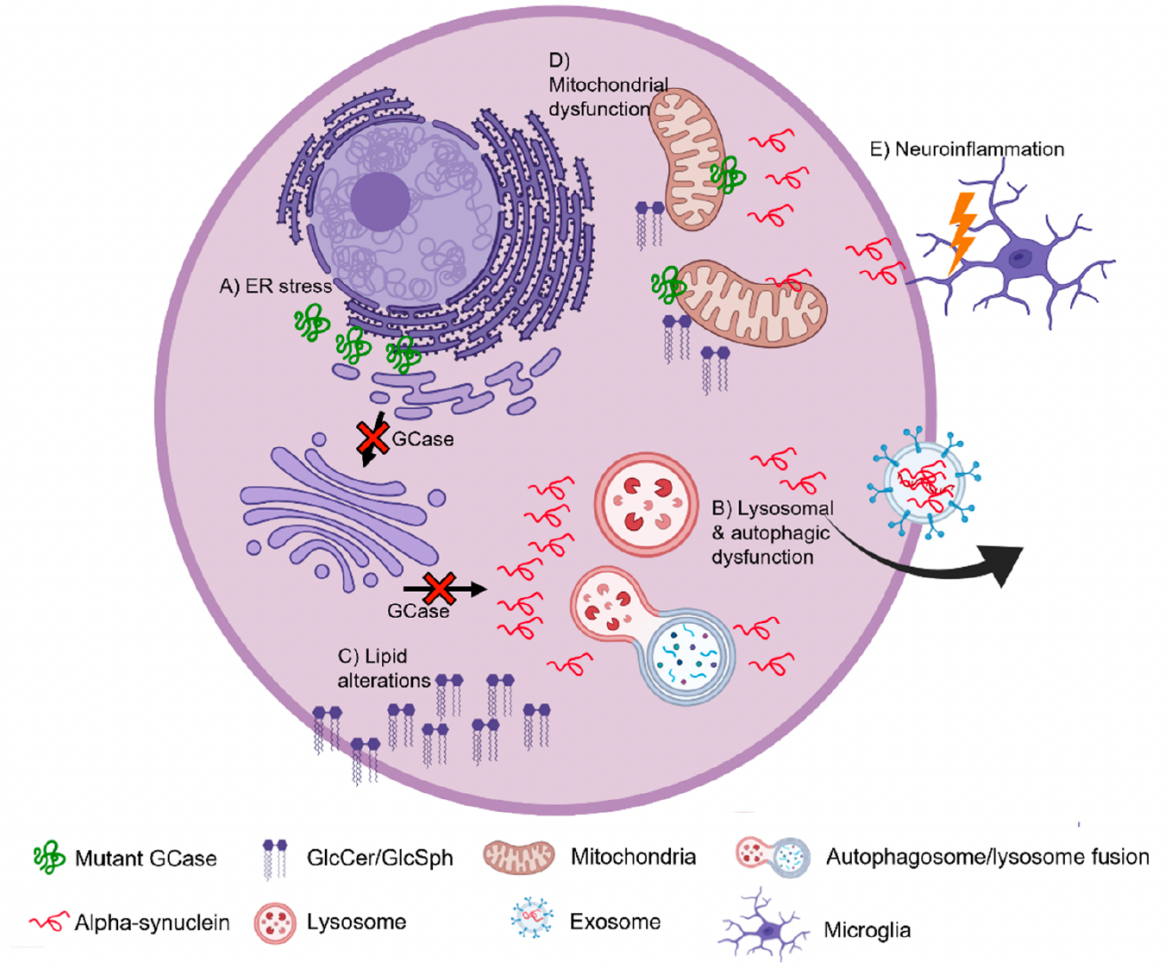


Figure 7 : Association between GCase, α-synuclein and Parkinson's disease
(A) Misfolded GCase protein triggers ER and activates ER stress pathways. (B). GCase deficiency alters the autophagic-lysosomal pathway, leading to the accumulation of GlcCer, GlcSph and α-synuclein. (C) Insufficient GCase activity leads to the accumulation of glycosphingolipids, altering lipid membrane composition and promoting α-synuclein aggregation. (D) Deficiency of GCase and the autophagy-lysosomal pathway leads to the accumulation of defective mitochondria. (E) The accumulation of lipids or α-synuclein may trigger microglial activation and contribute to neuroinflammatory processes. (Source: Copyright 2022 from Laura Smith, published by MDPI and available under an open access license. Licensed under the Creative Commons Attribution CC BY license)

I.1.2. Treatments for Parkinson's disease

Currently, there is no cure for PD, and available treatments focus on the reduction of motor symptoms. The main treatment involves dopaminergic drugs, such as L-DOPA, which helps restore dopamine levels in the striatum and remains the most effective option for reducing motor-related PD symptoms (Jankovic and Aguilar, 2008). However, long-term use of L-DOPA can lead to the development of L-DOPA-induced dyskinesia, which worsens over time. Other medications, including dopamine receptor agonists and inhibitors of enzymes involved in dopamine metabolism (such as MonoAmine Oxidase B (MAO-B) and Catechol-O-MethylTransferase (COMT)), may also be used to manage motor symptoms but can have side effects, such as hallucinations, orthostatic hypotension, nausea, vomiting, dyskinesia, and nausea (Kaakkola, 2000).

Deep Brain Stimulation (DBS) is a neurosurgical procedure in which electrodes are implanted in specific brain regions such as the subthalamic nucleus or globus pallidus (Volkmann, 2004). These electrodes generate electrical impulses to regulate abnormal brain activity and reduce tremors and dyskinesia. DBS may significantly improve quality of life and reduce motor impairments including dyskinesia (Schuepbach et al., 2013). DBS results can be influenced by a variety of factors, such as patient age, disease duration, presence of other pathologies, response to levodopa therapy, symptom severity, cognitive and psychiatric impairment, precise electrode placement and post-surgery program (Yoon et al., 2023). Recent studies showed that after the DBS surgery, *GBA1* mutation carriers generally have worse cognitive performance than non-carriers (Pal et al., 2022). In addition, compared to non-carriers, *PRKN* carriers have a greater decrease in the Levodopa-Equivalent Daily Dosage (LEDD) after the DBS (Lohmann et al., 2008). Patients with the *LRRK2* p.G2019S mutation also experience excellent motor improvements after DBS, however, those with the LRRK2 p.R1441G mutation had worse outcomes than non-carriers (Artusi et al., 2019).

I.1.2.1. *GBA1* pathway-specific treatments

As the *GBA1* mutation causes a reduction in GCase enzyme activity and contributes to PD, increasing GCase activity could be a potential solution, similar to the therapeutic approach for non-neuronopathic GD (Shemesh et al., 2015). However, Enzyme Replacement Therapy (ERT), which consists of administering functional GCase, is not effective for *GBA1*-associated PD as it cannot cross the blood-brain barrier and does not impact neurological symptoms (Revel-Vilk et al., 2018).

Substrate Reduction Therapy (SRT) targets the GlcCer and GlcSph, reducing their production and preventing accumulation resulting from GCase deficiency (Revel-Vilk et al., 2018). Unfortunately, a recent phase 2 clinical trial on the promising SRT drug venglustat failed to efficacy treat PD (Riboldi and Di Fonzo, 2019).

Emerging research has focused on Small Molecule Chaperones (SMCs) that have shown promise in increasing GCase levels and reducing ER stress by facilitating the proper folding of mutant GCase (Riboldi and Di Fonzo, 2019). Several SMCs, including isofagamine and ambroxol hydrochloride, have demonstrated the ability to restore GCase activity in preclinical and clinical studies, leading to improvements in motor performance, reduced microglial activation, and decreased synuclein aggregation (Ambrosi et al., 2015; Bendikov-Bar et al., 2013; Horowitz et al., 2016; McNeill et al., 2014; Richter et al., 2014).

Additionally, rapamycin, a macroautophagy inducer, has been shown to reduce α-synuclein accumulation and slow disease progression by targeting dysfunction in the lysosome-autophagy system (Cullen et al., 2011).

Gene therapy approaches, involving the introduction of the *GBA1* gene using viral vectors to increase GCase activity, have been conducted and are now closed, but the results have not yet been published (Sardi et al., 2013). These advancements hold potential for future treatments of *GBA1*-associated PD and the ClinicalTrials.gov platform provides more details on it.

Chapter 1 : Accurate long-read sequencing identified *GBA1* variants as a major genetic risk factor in the Luxembourg Parkinson's study

npj | parkinson's disease www.nature.com/npjparkd

ARTICLE OPEN

Check for updates

Accurate long-read sequencing identified *GBA1* as major risk factor in the Luxembourgish Parkinson's study

Sinthuja Pachchek[1✉], Zied Landoulsi[1], Lukas Pavelka[2,3], Claudia Schulte[4], Elena Buena-Atienza[5,6], Caspar Gross[5,6], Ann-Kathrin Hauser[4], Dheeraj Reddy Bobbili[1], Nicolas Casadei[5,6], Patrick May[1✉], Rejko Krüger[1,2,3✉] and on behalf of the NCER-PD Consortium*

Project description

The *GBA1* gene is an increasingly recognized risk factor for PD, and it continues to pose a significant challenge in understanding the genetic landscape of *GBA1*-associated PD. This challenge arises from the existence of the pseudogene *GBAP1* which shares 96% similarity with the *GBA1* gene. To address this, we employed a novel approach to comprehensively dissect the landscape of *GBA1*-related parkinsonism in Luxembourg.

Contributions

My contributions to this research project include the planning and execution of the entire study. The statistical analysis strategy was also my responsibility, with the help of senior researcher Dr. Zied Landoulsi, and I ensured that it corresponded to the objectives and research questions. I also led the effort to find and evaluate rare variants in the dataset by doing a thorough study of the genomic data. Moreover, I undertook the major tasks of exporting and curating clinical data, ensuring its quality and completeness for analysis. My involvement in all phases of the research, from design to statistical analysis, was essential to the validity and reliability of the results. I conducted the analysis and wrote the manuscript for publication, with substantial contributions from my colleague, Dr. Zied Landoulsi.

SP: Research project: Conception, Organization, Execution; Statistical Analysis: Design, Execution, Review and Critique; Manuscript Preparation: Writing of the first draft, Review and Critique; Genetic data: Analysis; Data collection: Exportation, Curation. ZL: Research project: Execution; Statistical Analysis: Design, Execution, Review and Critique; Manuscript Preparation: Writing of the first draft, Review and Critique; Data collection: Curation. LP: Research project: Execution; Statistical Analysis: Review and Critique; Manuscript Preparation: Review and Critique; Data collection: Participation, Exportation, Curation. CS, EBA, CG and AKH: Statistical Analysis: Review and Critique; Manuscript Preparation: Review and Critique; Genetic data: Sequencing Execution. DRB: Statistical Analysis: Review and Critique; Manuscript Preparation: Review and

Critique. NC: Manuscript Preparation: Review and Critique; Genetic data: Sequencing Execution. PM: Research project: Conception, Organization, Execution; Statistical Analysis: Design, Review and Critique; Manuscript Preparation: Review and Critique; Genetic data: Analysis; Data collection: Curation. RK: Research project: Conception, Organization, Execution; Statistical Analysis: Design, Review and Critique; Manuscript Preparation: Review and Critique; Data collection: Participation, Curation

The study was supervised by Prof. Dr. Rejko Küger and Dr. Patrick May.

1.1. Introduction

Identifying glucocerebrosidase *GBA1* variants as the main genetic risk factor for PD has led to a major switch in PD research. The association between *GBA1* variants and PD was initially observed over 20 years ago when GD patients carrying the *GBA1* mutation developed parkinsonian symptoms (Neudorfer et al., 1996). Subsequently, large multicenter studies confirmed the genetic link between *GBA1* and PD and showed that *GBA1* variants are the major genetic risk factor for PD (Nalls et al., 2014; Robak et al., 2017; Sidransky et al., 2009). The risk ratios for PD are similar for homozygous and heterozygous *GBA1* variants (Alcalay et al., 2014). Based on different studies and the depth of exome sequencing, the frequency of *GBA1* gene variants in PD patients ranges from 5% to 30% (Duran et al., 2013; Sidransky et al., 2009). According to clinical studies, *GBA1*-associated PD is characterized by an earlier age of disease onset, a higher frequency of non-motor symptoms, and cognitive impairment (Petrucci et al., 2020). Given their major impact on the clinical course and their very high frequency, a comprehensive knowledge of the pathogenic processes of *GBA1* variants could contribute to the development of targeted therapeutics treatments.

The *GBA1* gene is located on chromosome 1 and comprises 11 exons. Due to a pseudogene (*GBAP1*) located 6.9 kb downstream and showing an overall similarity of 96% with *GBA1*, it is challenging to identify *GBA1* variants. This increases to 98% between intron 8 and the 3′-Untranslated Region (UTR), where five identical regions are present (Zampieri et al., 2017). To date, over 100 distinct variants have been reported in GD patients (Grabowski and Horowitz, 1997), covering a range of genetic variants such as point mutations, deletions, insertions and complex alleles. These complex alleles result from genetic rearrangements between the functional gene and the pseudogene (Latham et al., 1990).

Knowing that none of the currently available short-read alignment sequencing methods has been able to fully resolve the identification of the complex alleles, this becomes inherently problematic when dealing with a highly similar pseudogene (Toffoli et al., 2022). Our study aims to accurately identify all *GBA1* coding variants in participants in the Luxembourg Parkinson's study. To this end, we used the *GBA1*-targeted sequencing method using the PacBio technology. Using Sanger sequencing as the gold standard for validation, we compared this approach with NeuroChip array genotyping data and short-read WGS data.

1.2. Copy of the preprint manuscript

Accurate long-read sequencing identified *GBA1* variants as a major genetic risk factor in the Luxembourg Parkinson's study

Sinthuja Pachchek[1,*]; Zied Landoulsi[1]; Lukas Pavelka[2,5] ; Claudia Schulte[3]; Elena Buena-Atienza[4]; Caspar Gross[4]; Ann-Kathrin Hauser[3]; Dheeraj Reddy Bobbili[1]; Nicolas Casadei[4]; Patrick May[1,*]; and Rejko Krüger[1,2,5,*] on behalf of the NCER-PD Consortium

[1]LCSB, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Sur-Alzette, Luxembourg.
[2] Parkinson Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg.
[3]Department of Neurodegeneration, Center of Neurology, Hertie Institute for Clinical Brain Research, German Center for Neurodegenerative Diseases, University of Tübingen, Tübingen, Germany.
[4]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany; NGS Competence Center Tübingen (NCCT), University of Tübingen, Tübingen, Germany.
[5]Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg.
*Corresponding authors

Corresponding authors:
Prof. Dr. Rejko Krüger, MD
rejko.krueger@lih.lu

Dr. Patrick May
patrick.may@uni.lu

Dr. Sinthuja Pachchek
sinthuja.pachchek@uni.lu

Running title: *GBA1* variants in the Luxembourg Parkinson's Study

Keywords: *GBA1*, glucocerebrosidase, Parkinson's disease, Genetics, long-read sequencing, Luxembourg

ABSTRACT

Heterozygous variants in the glucocerebrosidase *GBA1* gene are an increasingly recognized risk factor for Parkinson's disease (PD). Due to the *GBAP1* pseudogene, which shares 96% sequence homology with the *GBA1* coding region, accurate variant calling by array-based or short-read sequencing methods remains a major challenge in understanding the genetic landscape of *GBA1*-associated PD. We analysed 660 patients with PD, 100 patients with parkinsonism and 808 healthy controls from the Luxembourg Parkinson's study, sequenced using amplicon-based long-read DNA sequencing technology. We found that 12.1% (77/637) of PD patients carried *GBA1* variants, with 10.5% (67/637) of them carrying known pathogenic variants (including severe, mild, risk variants). In comparison, 5% (34/675) of the healthy controls carried *GBA1* variants, and among them, 4.3% (29/675) were identified as pathogenic variant carriers. We found four *GBA1* variants in patients with atypical parkinsonism. Pathogenic *GBA1* variants were 2.6-fold more frequently observed in PD patients compared to controls (OR=2.6; CI=[1.6,4.1]). Three novel variants of unknown significance (VUS) were identified. Using a structure-based approach, we defined a potential risk prediction method for VUS. This study describes the full landscape of *GBA1*-related parkinsonism in Luxembourg, showing a high prevalence of *GBA1* variants as the major genetic risk for PD. Although the long-read DNA sequencing technique used in our study may be limited in its effectiveness to detect potential structural variants, our approach provides an important advancement for highly accurate *GBA1* variant calling, which is essential for providing access to emerging causative therapies for *GBA1* carriers.

Keywords: *GBA1*, glucocerebrosidase, Parkinson's disease, Genetics, long-read sequencing, Luxembourg

INTRODUCTION

Heterozygous variants in the glucocerebrosidase (*GBA1*) gene, which encodes the enzyme β-glucocerebrosidase (GCase), are increasingly recognized as the most common genetic risk factor for the development of Parkinson's disease (PD). Homozygous variants in *GBA1* are causative for the most frequent autosomal-recessive lysosomal storage disorder, Gaucher disease (GD)[1]. GD is characterized by a deficiency of the enzyme GCase which is required to hydrolyse the β-glucosyl linkage of glucosylceramide lipide (GlcCer) in lysosomes to form glucose and ceramide[2].

Accurate variant calling in the *GBA1* gene is challenging due to the presence of the highly homogeneous untranslated pseudogene called *GBAP1*, which is located 16 kilobases downstream[3], and shares 96% sequence homology within the coding region[4]. In addition, recombination and structural chromosomal variation within and around the *GBA1* locus further complicate the analysis[5]. Complex alleles, which include several single nucleotide variants, are derived from recombination between the functional *GBA1* gene and the *GBAP1* pseudogene[6]. RecNciI is the most common recombinant allele, including the amino acid changes p.L483P and p.A495P, and the synonymous variant p.V499V[6].

Our study aimed to accurately assess all rare coding variants in the *GBA1* gene in all participants of the Luxembourg Parkinson's study[7], a case and control cohort including patients with PD and atypical parkinsonism. To assess the accuracy of the targeted *GBA1* DNA sequencing method using the Pacific Biosciences (PacBio)[8] technology, which targets only the *GBA1* gene without sequencing the *GBAP1* pseudogene, we compared this method with genotyping using the NeuroChip array[9] and short-read whole genome sequencing (WGS) data using Sanger sequencing as the gold standard for validation. We identified several types of pathogenic *GBA1* variants (severe, mild, and risk) and further characterized genotype-phenotype associations to better understand the influence of each variant type and their effect on disease severity.

RESULTS

Demographic and clinical characteristics

A total of 760 patients (660 PD patients (n_{PD}) and 100 patients with other forms of parkinsonism (n_{park})) and 808 healthy controls (n_{HC}) from the Luxembourg Parkinson's study (Figure 1) were genotyped using NeuroChip and screened for *GBA1* variants using targeted PacBio DNA sequencing method, while a subset of 72 patients was screened with WGS. Among the patients, 66.4% (n = 499) were male with a mean age at disease onset (AAO) of 63 ± 11.5 years (Supplementary Table 1). The control group consisted of 52.7% (n = 426) males with a mean age at assessment (AAA) of 59.3 ± 12.2 years. Due to their above 30-fold coverage provided by the long-read DNA sequencing, all samples were selected after successfully passing the MultiQC step (Supplementary Table 9). To ensure ethnic homogeneity and exclude other genetic factors that may bias the assessment of the

genetic contribution of *GBA1* to PD in the Luxembourgish population, we excluded carriers of mutations in other PD-causing genes (point mutations: n=10, n_{PD}=8,n_{HC}=2; CNV: n_{PD}=4) in PD-associated genes (no CNVs in *GBA1* were detected), first-degree family members (n=64, n_{PD}=8, n_{park}=2, n_{HC}=54), younger HC (< 60 AAA) with first-degree relatives having PD (n_{HC}=74), and individuals of non-European descent (n=6) from the cohort. The final cohort consisted of 735 patients (n_{PD}=637, n_{park}=98) and 675 HC with a mean AAO among the patients of 63.2 ± 11.3 years, whereas the mean AAA for HC was 61 ± 11.5 years. Based on Neurochip and WGS data, none of the *GBA1* carriers carried pathogenic variants in other PD-associated genes as defined by MDSGene[10].

Targeted PacBio DNA sequencing showed the highest specificity for detecting rare coding variants in *GBA1*

To measure the reliability of calling rare *GBA1* coding variants, we performed two types of comparison. Rare variants were here defined as variants with minor allele frequency (MAF) <1% in the European population. We compared the results from the PacBio, WGS and NeuroChip data for a subset of samples (n=72). We then compared the PacBio and NeuroChip data as they both covered the majority of samples (n=1568). We considered true positives to be the *GBA1* variants validated by Sanger sequencing. False-positive variants were those identified by the analysis method but not confirmed by Sanger sequencing. False-negative variants were not called by the analysis method but were later validated with Sanger sequencing (Supplementary Table 2). First, we evaluated 72 samples screened by all three methods (Figure 2). Using the *GBA1*-targeted PacBio DNA sequencing method and WGS in combination with the Gauchian[11] tool implemented in Dragen v4 (GBA caller option), we detected six individuals carrying *GBA1* variants (p.E365K (n=3), p.T408M (n=1), p.N409S (n=1), RecNciI (n=1)). The RecNil combines the three variants p.L483P, p.A495P, and p.V499V in one haplotype allele. All variants detected were confirmed by Sanger sequencing (true positive rate (TPR) of 100%). We did not identify any false positive variant calls. However, using the Dragen v.4 pipeline without the *GBA1* caller, relying only on the GATK best practices pipeline, the WGS method failed to detect the RecNciI recombinant allele in one individual (TPR of 83.3% (5/6)). Using Neurochip, we detected three potential *GBA1* variant carriers (p.T408M (n=1), p.N431S (n=1), p.A215D (n=1), but only one variant (p.T408M) was subsequently confirmed by Sanger sequencing (TPR of 16.6% (1/6), resulting in a false discovery rate (FDR) of 66.6% (2/3).

Next, we compared the results from 1568 samples screened with both, the *GBA1* targeted PacBio DNA sequencing method and the NeuroChip array (Figure 3). Using the GBA1-targeted PacBio DNA method, we detected 135 *GBA1* variants carriers, of which 100% were validated by Sanger sequencing. Using the NeuroChip array, we detected 47 potential *GBA1* variant carriers, among which only 36 were validated by Sanger sequencing (TPR of 26.7% (36/133), resulting in an FDR of 23.4% (11/47).

Classification of *GBA1* variants

Of the 1568 individuals sequenced using the *GBA1*-targeted PacBio DNA sequencing method, we identified 135 carriers of at least one *GBA1* variant (Supplementary Table 3-4). Based on the classification of Höglinger and colleagues in 2022[12], 25 were carriers of severe variants, 10 of mild variants, 72 of risk variants and 22 of VUS. The most common *GBA1* variants in PD patients were the risk variants p.E365K (n=23; 3.5%) and p.T408M (n=17; 2.6%).

GBA1 variants were mostly heterozygous missenses, one patient carried a heterozygous stop-gain variant p.R398*(rs121908309), two PD patients carried a homozygous missense variant p.E365K/p.E365K(rs2230288). We identified two HC carrying a pathogenic *LRRK2* variant and being positive for *GBA1* variant ($p.E365K_{GBA}$-$p.R1441C_{LRRK2}$; $p.K13R_{GBA}$-$p.G2019S_{LRRK2}$). We also detected nine different synonymous variants in exonic regions (Supplementary Table 4). The variant p.T408T(rs138498426) is a splice site variant (located within 2bp of the exon boundary) and is classified as VUS[12]. The remaining synonymous variants were not further analysed. Additionally, we identified 69 variants in intronic and UTRs regions of *GBA1* (Supplementary Table 5) with unclear pathogenic relevance, 35 of which were rare with MAF<1% in gnomAD for the Non-Finnish European population[10].

We classified the following combinations of multiple variants per individual as severe based on the classification of the respective associated pathogenic variants (Table 1): p.N409S-p.L483P, p.K13R-p.L483P, p.F252I-p.T408M, p.Y61H-p.T408M.

***GBA1* variant frequency**

To calculate the *GBA1* frequency in our study, we considered the individuals remaining after the exclusion step (735 patients and 675 HC). We detected 12.1% (n=77) *GBA1* variant carriers among 637 PD patients and 5% (34/675) in HC individuals. We found a frequency of 10.5% (67/637) of pathogenic variants in PD patients (severe, mild, risk) and 4.3% (29/675) in controls (Table 2). Four patients with parkinsonism had *GBA1* variants. Carriers of severe *GBA1* variants (n=21; 3.2%; OR=11.4; 95% CI=[2.6, 49]; p=0.0010) have a high risk of developing PD as defined by the indicated OR.

Genotype-phenotype associations in *GBA1*-PD patients

We characterized the clinical phenotype of severe (n=21), mild (n=7) and risk (n=39) *GBA1* carriers and non-carriers (n=554) only in unrelated PD patients excluding carriers with only one synonymous or VUS variant in individuals remaining after the filtering step. The AAO was similar between *GBA1* carriers (61.6±11.5) and non-carriers (62.6±11.6). Severe PD_{GBA1} variant carriers showed a trend towards younger AAO compared to mild and risk (severe: 58.6±13.1 vs mild: 65.4±17 vs risk: 62.5±9.3 years; p=0.29) (Table 3), with a significant risk to develop early onset PD (OR=4.02; p=0.0098). In contrast to non-carriers, we also observed that carriers of pathogenic variants have a

strong family history of PD (OR=0.74; *p*=0.0401). Compared to severe (6.4 ±4.7) and risk (4.4 ±4.9) variant carriers, the mild (1.7 ±1.4) variant carriers show a significantly shorter disease duration (Table 3).

We compared clinical features between PD patients carrying pathogenic *GBA1* variants and PD patients without *GBA1* variants (Supplementary Table 6). We found that in carriers the sense of smell was strongly impaired (uncorrected *p*=0.0210) and a higher rate of hallucinations (uncorrected *p*=0.0415). Next, we compared patients carrying variants from each category (severe, mild or risk) separately with PD patients without *GBA1* variants (Table 4). Carriers of severe *GBA1* variants showed more severe non-motor symptoms when compared to non-*GBA1* carriers, such as MDS-UPDRS Part I (uncorrected *p*=0.0074) and hallucinations (uncorrected *p*=0.0099), and also an impaired sense of smell as assessed by Sniffin' Stick test (uncorrected *p*=0.0405). To show the deleterious impact of the severe variants, we compared carriers of severe variants with patients carrying either mild or risk *GBA1* variants (Table 5). We observed that severe variants carriers have more severe gait disorder (uncorrected *p*=0.0188) and depression (uncorrected *p*=0.0074) and worse MDS-UPDRS Part I (uncorrected *p*=0.0019) and PDQ-39 (uncorrected *p*=0.0422). For all clinical features, there were no significant associations after the correction for multiple comparisons using FDR adjustment.

VUS and the Glucosylceramidase structure

We detected nine already reported VUS (p.K13R, p.Y61H, p.R78C, p.L213P, p.E427K, p.A495P, p.H529R, p.R534C, p.T408T) and three new VUS (p.A97G, p.A215 and p.R434C).

According to our strategy developed for the VUS classification of *GBA1* variants, where we assign the pathogenicity based on the REVEL, the CADD and the dbscSNV scores, as well as the presence or absence of the variants in the patients. We propose to subclassify the VUS p.Y61H, p.L213P, p.A215D, and p.R434C as probably pathogenic severe variants (Supplementary Table 7). The variant p.L213P changes the leucine into proline, which is known to be the 'helix breaker' amino acid that induces a bend into the protein structure[13] (Supplementary Figure 1). The p.L213P and p.A215D variants are in the catalytic site of the enzyme in the triose-phosphate isomerase (TIM) barrel structure. The p.Y61H variant (Figure 4.A) is located next to the residue position of the known severe variant p.C62W, and the patient carrying this variant had an AAO of 38 years, indicating an early-onset, probably severe form of PD. This patient has a family history of PD and reported that the paternal uncle and aunt were diagnosed with PD at the ages of 60 and 70, respectively. The p.R434C variant is close to a known severe (p.V433L) and mild (p.W432R, p.N435T) PD variants in the 3D structure. We compared the clinical scores obtained from carriers of known severe variants with the four carriers of probable severe VUS (p.Y61H, p.L213P, p.A215D, and p.R434C) (Supplementary Table 7). The z-score was used to determine the number of SD deviations from the mean for each clinical score. We observed that the PD patient carrying the p.L213P variant had a z-score that was

significantly different for MDS-UPDRS II (z-score=3.05) and MDS-UPDRS III (z-score=2.94) confirming its classification as a severe variant.

We propose to subclassify the variants p.H529R and p.R534C as probably mild variants, as they are both found only in PD patients. The variants p.K13R, p.R78C, p.E427K, and p.A495P are subclassified as probable risk variants. The variant p.K13R is located in the signal peptide region. The variant p.R78C was annotated as "PD susceptibility" in HGMD with deleterious impact in CADD. The variant p.E427K was annotated as associated to "parkinsonism" in ClinVar and "reduced activity" in HGMD. We propose to classify the variant p.A97G as probably benign because it is localized in a coil-bend structure and is not close to any known pathogenic variants.

The synonymous variant p.T408T was found in two cases and one healthy control individual. Two established splice-site prediction scores (dbscSNV: ada_score 0.9797 and rf_score 0.85) agreed in their prediction that the variant is likely to affect splicing. HGMD classified the variant as disease mutation (DM) (Supplementary Table 4). Therefore, we propose to classify the variant as a risk variant.

Overall, we propose to classify four VUS variants as probably severe pathogenic variants (p.Y61H, p.L213P, p.A215D, and p.R434C), two as probably mild pathogenic variants (p.H529R and p.R534C), five as probably pathogenic risk variants (p.K13R, p.R78C, p.E427K, p.A495P and p.T408T) and one as probably benign variants (p.A97G) (Figure 4.B).

DISCUSSION

Our study demonstrated in a large cohort the utility of targeted PacBio DNA sequencing for *GBA1* as a highly sensitive and specific method to identify known and novel *GBA1* variants and to overcome the problems posed by the presence of the *GBAP1* pseudogene by avoiding its amplification. The effectiveness of the targeted PacBio DNA sequencing method in investigating relevant genes with homologous pseudogenes has also been demonstrated in several other studies[13–16]. The PacBio method together with the WGS method combined with the new Gauchian tool showed a very high accuracy of 100% true positive validated variants. The comparative study that we performed with the different screening technologies to detect *GBA1* variants will help researchers to get a more accurate and comprehensive overview of *GBA1* variants. This implies a more critical evaluation of the results obtained by NeuroChip, which revealed a high proportion of false positive and negative results and those obtained by WGS, which will depend on the detection tool used for the complex *GBA1* region. Our study still has the limitation that we cannot fully exclude missing variants (false negatives) that could not be detected by all three methods used in our study. Long-read DNA sequencing excels in the detection of structural variants. However, the method employed in this study relies on a single amplicon, limiting its efficiency in detecting structural variants due to the generation and purification of amplicons of specific sizes only. However, we would like to

highlight the fact that the PacBio-based method can be a cost-effective (≃ 30€/sample for PacBio) alternative for the high-fidelity calling of *GBA1* variants. *GBA1* variants have been identified as the most common genetic risk factor for the development of PD. *GBA1* variants have typically been observed in 4%–12% of PD patients in different populations worldwide, with the highest prevalence of 20% described in Ashkenazi Jewish PD patients[17,18]. Large differences of prevalence were observed depending on the ethnicity of the cohort, the variants studied, and the sequencing method used. Previous studies looking only at coding regions reported frequencies of 14.3% in Italians[19] (n=874), 11.7% in southern Spanish[20] (n=532), 9.2% in New Zealanders of European descent[5] (n=229) and 8.3% in Irish[21] (n=314) (Supplementary Table 8). Our study describes the landscape of *GBA1* carriers in the Luxembourgish population showing a high prevalence (12.1%) of *GBA1* variants that could be the major genetic risk factor of PD in Luxembourg. Moreover, we observed a significantly higher proportion of pathogenic (severe, mild and risk) *GBA1* variants in PD patients compared to HC (10.4% *vs* 4.3%; OR=2.6; CI=[1.6,4.1], *p*=0.0001). Compared to previous studies, our study highlights that using the new PacBio sequencing method, the Luxembourg Parkinson's study showed a comparable frequency of PD_{GBA1} carriers reported so far in similarly sized Italian[19] and Spanish[20] cohorts (Supplementary Table 8). When comparing previous reports of *GBA1* variants in different populations, we want to highlight the fact that only cohorts that used full Sanger sequencing were able to detect the RecNciI recombinant allele so far. This once more emphasizes the accuracy of the PacBio sequencing methods for detecting rare and complex *GBA1* variants. Additionally, we confirmed that severe variants showed a higher OR than risk variants, which supports the concept of graded risk for different *GBA1* variants in PD_{GBA1} carriers[20].

The most prevalent *GBA1* variant in the Luxembourg Parkinson's study was p.E365K, and the frequency of this variant was similar to what was described in the Irish[21], Spanish[20] and New Zealand[5] populations. It is interesting to note that homozygous carriers of the p.E365K variant do not develop GD[22]. This variant is associated with PD, and multiple studies have found enrichments varying from 1.60 to 3.34[23–25]. Furthermore, carriers of the risk variants p.E365K and p.T408M could be associated with atypical parkinsonism, as these variants were the only ones also present in patients with DLB and PSP in our cohort. Whether this is simply related to the higher frequency of these risk variants in the general population or does have a specific impact on the phenotype needs to be determined in larger studies focusing on *GBA1* variants in atypical parkinsonism[26].

We present a concept for classifying VUS in the *GBA1* gene according to the localisation in relation to known variants in sequence and 3D structure, which may help to provide access to future targeted therapies for these patients. Here additional *in vitro* and *ex vivo* studies are needed to functionally validate the impact of these VUS on GCase function in neurons derived from stem cells or in enzyme-activity assays in cerebrospinal fluid of affected carriers of these VUS.

Additionally, we observed that the average AAO in PD was about four years younger in severe *GBA1* carriers compared to non-*GBA1* carriers. This was also observed in previous studies, which showed

that PD_{GBA1} patients generally have an earlier AAO compared to non-carriers with a median onset in the early fifties[27,28].

Recent studies have shown that PD_{GBA1} carriers have a higher prevalence of cognitive impairment[19,29,30] and non-motor symptoms including neuropsychiatric disturbances[19,20], autonomic dysfunction[29], and sleep disturbances such as RBD[31]. Although not significant after *p*-value adjustment, we found a similar trend and noticed that motor symptoms such as gait disorder, non-motor symptoms such as depression and hallucinations, were associated with a more aggressive clinical phenotype in severe *GBA1* carriers, supporting the effect of differential *GBA1* variant severity[20,32].

In conclusion, this study showed the utility of targeted PacBio DNA sequencing to identify known and novel *GBA1* variants with high accuracy. These findings offer important access to variant-specific counselling. Furthermore, our study describes the full landscape of *GBA1*-related PD in the current Luxembourgish population showing the high prevalence of *GBA1* variants as the major genetic risk in PD.

METHODS

Clinical Cohort

At the time of analysis, the Luxembourg Parkinson's study comprised 1568 participants (760 patients of parkinsonism and 808 healthy controls (HC) in the frame of the National Centre for Excellence in Research on Parkinson's disease program (NCER-PD).

All patients complied with the diagnostic criteria of typical PD or atypical parkinsonism as assessed by neurological examination following the United Kingdom Parkinson's Disease Society Brain Bank (UKPDSBB) diagnostic criteria[33]: 660 fulfilled the criteria for PD, 60 for progressive supranuclear palsy (PSP) including corticobasal syndrome as a subtype of PSP (PSP-CBS), 25 for Dementia with Lewy Body (DLB), 14 for Multiple System Atrophy (MSA), and one for Fronto-temporal dementia with parkinsonism (FTDP). All patients and HC underwent a comprehensive clinical assessment of motor and non-motor symptoms, neuropsychological profile and medical history along with comorbidities. The clinical symptoms assessed, and scales applied are defined in the Supplemental Information[34]. All individuals provided written informed consent. The patients were reassessed at regular follow-up visits every year and the HC every four years. We considered early-onset PD patients those with age at onset (AAO) equal to or younger than 45 years[35]. The genotype-phenotype analysis was based on the assessment of the first visit. The final diagnosis was taken according to the last visit. The study has been approved by the National Research Ethics Committee (CNER Ref: 201407/13 and 202304/03).

NeuroChip array

Genotyping was carried out on the InfiniumR NeuroChip Consortium Array[9] (v.1.0 and v1.1; Illumina, San Diego, CA USA). For rare variants analysis, standard quality control (QC) procedures were conducted, using PLINK v1.9[36], to remove variants if they had a low genotyping rate (<95%) and Hardy-Weinberg equilibrium *p*-value < $1x10^{-6}$. As an additional quality filter, we applied a study-wide allele frequency threshold of <1% in our cohort for rare variants. For further statistical analysis, we excluded individuals of non-European ancestry using PLINK 1.9 multidimensional scaling and merged our data with the 1000 genomes dataset[37]. We selected only samples of European ancestry excluding those with > +/- 3SD based on the first and the second principal components.

***GBA1*-targeted PacBio DNA long-read amplicon sequencing**

The targeted *GBA1* gene screening was performed by single-molecule real-time (SMRT) long read sequencing[8] using Sequel II instrument (PacBio). The targeted *GBA1* gene coordinates were chr1:155,232,501-155,241,415 (USCS GRCh38/hg38). Long-distance PCR was performed using *GBA1*-specific primer sequences (Forward: 5'-GCTCCTAAAGTTGTCACCCATACATG-3' and Reverse: 5'-CCAACCTTTCTTCCTTCTTCTCAA-3')[38] and the 2x KAPA HiFi Hot Start ReadyMix (Roche), which avoid *GBAP1* pseudogene amplification. For sample multiplexing, dual asymmetric barcoding was used based on a different 16-bp long index sequence upstream of each of the reverse and forward primers to allow the generation of uniquely barcoded amplicons in one-step PCR amplification. QC was performed prior to pooling. Pools of amplicons were purified with AMPure PacBio beads. A total of 1700 ng of purified amplicon pool was used as input for the SMRTbell library using the SMRTbell Express Template Prep Kit 2.0 (PacBio). Binding of the polymerase and diffusion loading on SMRTCell 8M was prepared according to SMRTLink instructions with CCS reads as sequencing mode (version SMRT Link: 9.0.0.92188). We generated high-quality consensus reads using the PacBio Sequel II sequencer on Circular Consensus Sequencing mode using the pbccs (v6.0.0) tool. The method replicates both strands of the target DNA[39]. We demultiplexed and mapped reads from each sample to the human reference genome GRCh38 using minimap2[40] from the pbmm2 package (v1.4.0) (https://github.com/PacificBiosciences/pbmm2). We used the MultiQC[41] tool and selected samples with more than 30-fold coverage. For variant calling, we used the DeepVariant[42] (1.0) with models optimized for CCS reads. Finally, we selected variants with quality above 30 (QUAL>30).

Whole genome sequencing

The TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA) and MGIEasy FS DNA Prep kit (BGI, China) were used according to the manufacturer's instructions to construct the WGS library. Paired-end sequencing was performed with the Illumina NovaSeq 6000[43] and on the MGI G400 sequencers. A QC of the raw data was performed using FastQC (version 0.11.9: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). To call the variants, we used the Bio-

IT Illumina Dynamic Read Analysis for GENomics (DRAGEN) DNA pipeline[44] v4[45] with standard parameters and with or without the '*GBA* caller' option, which uses the Gauchian tool. To select the high-quality variants, we annotated and selected variants using VariantAnnotator and SelectVariants modules of the Genome Analysis Toolkit (GATK 4)[46] pipeline and applied the following additional filtering steps: VariantFiltration module for SNVs (QD<2, FS>60, MQ<40, MQRankSum<−12, ReadPosRankSum<−8, DP<10.0, QUAL<30, VQSLOD<0, ABHet>0.75 or <0.25, SOR>3 and LOD<0), and insertions-deletions (QD<2, FS>200, QUAL<30, ReadPosRankSum<−20, DP<10 and GQ_MEAN<20).

Variant annotation and validation

Variant annotation was done with ANNOVAR[47], using the Genome Aggregation Database (gnomAD r2.1)[48], the Human Gene Mutation Database (HGMD)[49] and ClinVar[50], and the Combined Annotation Dependent Depletion (CADD)[51] and Rare Exome Variant Ensemble Learner (REVEL)[52] to score the pathogenicity of missense variants[53]. For variants in splice sites, we used the ada_score and rf_score from dbscSNV (version 1.1)[54]. Ada_score≥ 0.6 or rf_score≥ 0.6 indicate that the variant is likely to affect splicing.

Rare variants were selected according to MAF< 1% in gnomAD for the Non-Finnish European (NFE) population in the 'non-neuro' gnomAD subset. Then, exonic and splicing variants (+/- 2bp from the exon boundary) were selected for autosomal dominant (*LRRK2*, *SNCA*, *VPS35*, *GBA1*) and autosomal recessive (*PRKN*, *PINK1*, *PARK7*, *ATP13A2*) PD genes. Rare variants within these genes were then confirmed by Sanger sequencing[55].

CNVs in PD genes

To detect the presence of copy number variants (CNVs) in selected six PD genes (*PARK7*, *ATP13A2*, *PINK1*, *SNCA*, *GBA1*, and *PRKN*), we used the PennCNV tool (v1.0.5)[56] using the Neurochip array data applying the same filtering steps as previously described for CNV calls in PD[11]. The multiplex-ligation dependent probe amplification (MLPA) method, which exclusively targets the selected genes, was used to validate the CNVs. Six patients with each one CNVs in one of the six PD genes were found and no CNV in *GBA1* was found. To detect CNVs within the *GBA1* gene through the analysis of PacBio data, we employed the pbsv tool (version 2.9.0) (https://github.com/PacificBiosciences/pbbioconda), which is specifically designed for long-read data analysis from PacBio. This tool successfully identifies 59.46% of structural variants with precision[57,58].

***GBA1* variant nomenclature**

All variants in *GBA1* were annotated based on GRCh37 and were numbered according to the current variant nomenclature guidelines (http://varnomen.hgvs.org), based on the primary translation product (NM_001005742), which includes the 39-residue signal peptide.

GBA1 variant classification

GBA1 variants classification was done according to the PD literature based on the work of Höglinger and colleagues in 2022[12]. Exonic or splice-site variants that are not mentioned in the paper were subclassified as 'severe' *GBA1* variants if they were annotated as pathogenic in ClinVar, otherwise they were subclassified as variants with unknown significance (VUS)[51].

Statistical analysis

To assess the frequency of different *GBA1* variant types and to analyse the genotype-phenotype associations in the Luxembourg Parkinson's Study, we considered only unrelated individuals and kept only one proband per family. For cases, we kept the patient with the earliest AAO. To account for age-dependent penetrance, we excluded HC with first-degree relatives (parents, sibs, and offspring) with PD and an AAA of less than 60 years. This reduced the age difference between cases and HC. We also excluded carriers of rare variants or CNVs in PD-associated genes (except *GBA1*) and individuals of non-European ancestry. Thus, 1410 unrelated individuals (735 patients and 675 HC) were selected for the statistical analysis.

We used regression models to assess the effect of PD_{GBA1} carrier status on the clinical variables. In these models, the dependent variable was the clinical outcome, while the predictor was *GBA1* carrier status. We excluded individuals carrying only VUS or synonymous variants. To this aim, we performed three types of association tests: (1) all PD_{GBA1} pathogenic variant carriers (severe, mild and risk) *vs.* $PD_{GBA1\text{-non-carriers}}$, (2) for each sub-group of PD_{GBA1} pathogenic variant carriers *vs.* $PD_{GBA1\text{-non-carriers}}$, (3) severe PD_{GBA1} pathogenic variant carriers *vs* combined mild and risk PD_{GBA1} pathogenic variant carriers. The effect of each factor was expressed as the Beta (β) regression coefficient. The odds ratio (OR) along with a 95% confidence interval (CI) was used to assess whether a given exposure was a risk factor for a given outcome. Regression models were adjusted for AAA, sex, and disease duration. FDR-adjusted p-value < 0.05 was considered as statistically significant.

Structure-based evaluation of VUS

To evaluate VUS variants, we implemented a method to assign the pathogenicity based on the REVEL[53] and CADD[51] scores for missense variants and the dbscSNV scores (ada_score and rf_score) for splice variants according to the dbNFSP[54] definition, as well as whether the patients carried the variants. We reclassified a VUS (1) as 'severe' if the variant was present only in patients and with deleterious effect in all scores or present only in patients with early-onset PD, (2) as 'mild' if the variant was present only in patients and with tolerated effect in all scores, (3) as ' risk' if present in patients and HCs or with tolerated and deleterious effect in either score or annotated as 'PD susceptibility' in HGMD, and (4) as 'benign' if present only in HC.

We mapped the known pathogenic missense variants and newly identified VUS in our cohort together with all reported population variants from gnomAD onto the *GBA1* protein sequence and

the 3D structure. We used the X-ray structure of GCase at 2.0 Å resolution (PDB structure accession code 1ogs; https://www.rcsb.org/) (Supplementary Figure 2). Analysis of the 3D structure was carried out using PyMOL (http://www.pymol.org). VUS were evaluated as a risk variant if they were 2bp positions away in sequence or had a C-alpha distance of less than 5 Å in 3D from another known pathogenic variant similar to the approach used by Johannesen et al[59].

DATA AVAILABILITY

The dataset for this manuscript is not publicly available as it is linked to the Luxembourg Parkinson's Study and its internal regulations. Any requests for accessing the dataset can be directed to request.ncer-pd@uni.lu.

CODE AVAILABILITY

The code behind the analyses is available under the Apache-2.0 license. More information on how to request access to the data and the code is available at https://doi.org/10.17881/eqzd-a607.

ACKNOWLEDGEMENTS

Data used in the preparation of this manuscript were obtained from the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD).

The National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) is funded by the Luxembourg National Research Fund (FNR/NCER13/BM/11264123).

The work presented here was funded by the Luxembourg National Research (FNR/NCER13/BM/11264123), the PEARL program (FNR/P13/6682797 to RK), MotaSYN (12719684 to RK), MAMaSyn (INTER/LEIR/18/12719318 to RK), MiRisk-PD (C17/BM/11676395 to RK, PM), the FNR/DFG Core INTER (ProtectMove, FNR11250962 to PM), and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to RK, SP).

We would like to thank all participants of the Luxembourg Parkinson's Study for their important support of our research. Furthermore, we acknowledge the joint effort of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) Consortium members from the partner institutions Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health, Centre Hospitalier de Luxembourg, and Laboratoire National de Santé generally contributing to the Luxembourg Parkinson's Study as listed below: [list of authors].

ON BEHALF OF THE NCER-PD CONSORTIUM

Geeta ACHARYA[2], Gloria AGUAYO[2], Myriam ALEXANDRE[2], Muhammad ALI[1], Wim AMMERLANN[2], Giuseppe ARENA[1], Rudi BALLING[1], Michele BASSIS[1], Roxane BATUTU[3], Katy BEAUMONT[2], Regina BECKER[1], Camille BELLORA[2], Guy BERCHEM[3], Daniela BERG[11], Alexandre BISDORFF[5], Ibrahim BOUSSAAD[1], David BOUVIER[4], Kathrin BROCKMANN[11], Jessica CALMES[2], Lorieza CASTILLO[2], Gessica CONTESOTTO[2], Nancy DE BREMAEKER[3], Nico DIEDERICH[3], Rene DONDELINGER[5], Nancy E. RAMIA[1], Daniela ESTEVES[2], Guy FAGHERAZZI[2], Jean-Yves FERRAND[2], Katrin FRAUENKNECHT[4], Manon GANTENBEIN[2], Thomas GASSER[11], Piotr GAWRON[1], Soumyabrata GHOSH[1], Marijus GIRAITIS[2,3], Enrico GLAAB[1], Martine GOERGEN[3], Elisa GÓMEZ DE LOPE[1], Jérôme GRAAS[2], Mariella GRAZIANO[17], Valentin GROUES[1], Anne

GRÜNEWALD[1], Wei GU[1], Gaël HAMMOT[2], Anne-Marie HANFF[2, 20, 21], Linda HANSEN[1,3], Michael HENEKA[1], Estelle HENRY[2], Sylvia HERBRINK[6], Sascha HERZINGER[1], Michael HEYMANN[2], Michele HU[8], Alexander HUNDT[2], Nadine JACOBY[18], Jacek JAROSLAW LEBIODA[1], Yohan JAROSZ[1], Sonja JÓNSDÓTTIR[2], Quentin KLOPFENSTEIN[1], Jochen KLUCKEN[1,2,3], Rejko KRÜGER[1,2,3], Pauline LAMBERT[2], Zied LANDOULSI[1], Roseline LENTZ[7], Inga LIEPELT[11], Robert LISZKA[14], Laura LONGHINO[3], Victoria LORENTZ[2], Paula Cristina LUPU[2], Tainá M. MARQUES[1], Clare MACKAY[10], Walter MAETZLER[15], Katrin MARCUS[13], Guilherme MARQUES[2], Patricia MARTINS CONDE[1], Patrick MAY[1], Deborah MCINTYRE[2], Chouaib MEDIOUNI[2], Francoise MEISCH[1], Myriam MENSTER[2], Maura MINELLI[2], Michel MITTELBRONN[1,4], Brit MOLLENHAUER[12], Friedrich MÜHLSCHLEGEL[4], Romain NATI[3], Ulf NEHRBASS[2], Sarah NICKELS[1], Beatrice NICOLAI[3], Jean-Paul NICOLAY[19], Fozia NOOR[2], Marek OSTASZEWSKI[1], Clarissa P. C. GOMES[1], Sinthuja PACHCHEK[1], Claire PAULY[1,3], Laure PAULY[2, 20], Lukas PAVELKA[1,3], Magali PERQUIN[2], Rosalina RAMOS LIMA[2], Armin RAUSCHENBERGER[1], Rajesh RAWAL[1], Dheeraj REDDY BOBBILI[1], Kirsten ROOMP[1], Eduardo ROSALES[2], Isabel ROSETY[1], Estelle SANDT[2], Stefano SAPIENZA[1], Venkata SATAGOPAM[1], Margaux SCHMITT[2], Sabine SCHMITZ[1], Reinhard SCHNEIDER[1], Jens SCHWAMBORN[1], Raquel SEVERINO[2], Amir SHARIFY[2], Ekaterina SOBOLEVA[1], Kate SOKOLOWSKA[2], Hermann THIEN[2], Elodie THIRY[3], Rebecca TING JIIN LOO[1], Christophe TREFOIS[1], Johanna TROUET[2], Olena TSURKALENKO[2], Michel VAILLANT[2], Mesele VALENTI[2], Gilles VAN CUTSEM[1,3], Carlos VEGA[1], Liliana VILAS BOAS[3], Maharshi VYAS[1], Richard WADE-MARTINS[9], Paul WILMES[1], Evi WOLLSCHEID-LENGELING[1], Gelani ZELIMKHANOV[3]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [2]Luxembourg Institute of Health, Strassen, Luxembourg. [3]Centre Hospitalier de Luxembourg, Strassen, Luxembourg. [4]Laboratoire National de Santé, Dudelange, Luxembourg. [5]Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg. [6]Centre Hospitalier du Nord, Ettelbrück, Luxembourg. [7]Parkinson Luxembourg Association, Leudelange, Luxembourg. [8]Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [9]Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, UK. [10]Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK. 11Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Germany. [12]Paracelsus-Elena-Klinik, Kassel, Germany. [13]Ruhr-University of Bochum, Bochum, Germany. [14]Westpfalz-Klinikum GmbH, Kaiserslautern, Germany. [15]Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany. [16]Department of Neurology Philipps, University Marburg, Marburg, Germany. [17]Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg. [18]Private practice, Ettelbruck, Luxembourg. [19]Private practice, Luxembourg, Luxembourg. [20]Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [21]Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre+, Maastricht, the Netherlands

AUTHOR CONTRIBUTIONS

1. Research project: A. Conception, B. Organization, C. Execution;
2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique;
3. Manuscript Preparation: A. Writing of the first draft, B. Review and Critique;
4. Genetic data: A. Sequencing Execution, B. Analysis;
5. Data collection: A. Participation, B. Exportation, C. Curation

SP: 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B, 4B,5B, 5C
ZL: 1C, 2A , 2B, 2C,3A, 3B, 5C
LP: 1C, 2C, 3B, 5A, 5B, 5C
CS: 2C, 3B, 4A
EBA: 2C, 3B, 4A
CG: 2C, 3B, 4A
AKH: 2C, 3B, 4A
DRB: 2C, 3B
NC: 3B, 4A
PM: 1A, 1B, 1C, 2A, 2C, 3B, 4B, 5C
RK: 1A, 1B, 1C, 2A, 2C, 3B, 5A, 5C

CONFLICT OF INTEREST

The authors declare no competing interests.

REFERENCES

1. Hruska, K. S., LaMarca, M. E., Scott, C. R. & Sidransky, E. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum. Mutat.* **29**, 567–583 (2008).
2. Vieira, S. R. L. & Schapira, A. H. V. Glucocerebrosidase mutations and Parkinson disease. *J. Neural Transm.* **129**, 1105–1117 (2022).
3. Horowitz, M. *et al.* The human glucocerebrosidase gene and pseudogene: Structure and evolution. *Genomics* **4**, 87–96 (1989).
4. Do, J., McKinney, C., Sharma, P. & Sidransky, E. Glucocerebrosidase and its relevance to Parkinson disease. *Mol. Neurodegener.* **14**, 36 (2019).
5. Graham, O. E. E. *et al.* Nanopore sequencing of the glucocerebrosidase (GBA) gene in a New Zealand Parkinson's disease cohort. *Parkinsonism Relat. Disord.* **70**, 36–41 (2020).
6. Zimran, A. & Horowitz, M. RecTL: a complex allele of the glucocerebrosidase gene associated with a mild clinical course of Gaucher disease. *Am. J. Med. Genet.* **50**, 74–78 (1994).
7. Hipp, G. *et al.* The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. *Front. Aging Neurosci.* **10**, 326 (2018).
8. Korlach, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol.* **472**, 431–455 (2010).
9. Blauwendraat, C. *et al.* NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol. Aging* **57**, 247.e9-247.e13 (2017).
10. Lill, C. M. *et al.* Launching the movement disorders society genetic mutation database (MDSGene). *Mov. Disord. Off. J. Mov. Disord. Soc.* **31**, 607–609 (2016).
11. Toffoli, M. *et al.* Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene. *Commun. Biol.* **5**, 670 (2022).
12. Höglinger, G. *et al.* GBA-associated PD: chances and obstacles for targeted treatment strategies. *J. Neural Transm. Vienna Austria 1996* **129**, 1219–1233 (2022).
13. Qiao, W. *et al.* Long-Read Single Molecule Real-Time Full Gene Sequencing of Cytochrome P450-2D6. *Hum. Mutat.* **37**, 315–323 (2016).
14. Buermans, H. P. J. *et al.* Flexible and Scalable Full-Length CYP2D6 Long Amplicon PacBio Sequencing. *Hum. Mutat.* **38**, 310–316 (2017).
15. Borràs, D. M. *et al.* Detecting PKD1 variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Hum. Mutat.* **38**, 870–879 (2017).
16. Frans, G. *et al.* Conventional and Single-Molecule Targeted Sequencing Method for Specific Variant Detection in IKBKG while Bypassing the IKBKGP1 Pseudogene. *J. Mol. Diagn. JMD* **20**, 195–202 (2018).
17. Ruskey, J. A. *et al.* Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. *Eur. J. Med. Genet.* **62**, 65–69 (2019).
18. Gan-Or, Z. *et al.* Differential effects of severe vs mild GBA mutations on Parkinson disease. *Neurology* **84**, 880–887 (2015).
19. Petrucci, S. *et al.* GBA-Related Parkinson's Disease: Dissection of Genotype-Phenotype Correlates in a Large Italian Cohort. *Mov. Disord. Off. J. Mov. Disord. Soc.* **35**, 2106–2111 (2020).
20. Jesús, S. *et al.* GBA Variants Influence Motor and Non-Motor Features of Parkinson's Disease. *PloS One* **11**, e0167749 (2016).
21. Olszewska, D. A. *et al.* Association Between Glucocerebrosidase Mutations and Parkinson's Disease in Ireland. *Front. Neurol.* **11**, 527 (2020).

22. Duran, R. *et al.* The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* **28**, 232–236 (2013).
23. Ran, C. *et al.* Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. *Neurobiol. Aging* **45**, 212.e5-212.e11 (2016).
24. Davis, M. Y. *et al.* Association of GBA Mutations and the E326K Polymorphism With Motor and Cognitive Progression in Parkinson Disease. *JAMA Neurol.* **73**, 1217–1224 (2016).
25. Berge-Seidl, V. *et al.* The GBA variant E326K is associated with Parkinson's disease and explains a genome-wide association signal. *Neurosci. Lett.* **658**, 48–52 (2017).
26. Picillo, M. *et al.* Progressive Supranuclear Palsy-Like Phenotype in a GBA E326K Mutation Carrier. *Mov. Disord. Clin. Pract.* **4**, 444–446 (2017).
27. Blauwendraat, C. *et al.* Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. *Mov. Disord. Off. J. Mov. Disord. Soc.* **34**, 866–875 (2019).
28. Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* **361**, 1651–1661 (2009).
29. Brockmann, K. *et al.* GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study. *Mov. Disord. Off. J. Mov. Disord. Soc.* **30**, 407–411 (2015).
30. Setó-Salvia, N. *et al.* Glucocerebrosidase mutations confer a greater risk of dementia during Parkinson's disease course. *Mov. Disord. Off. J. Mov. Disord. Soc.* **27**, 393–399 (2012).
31. Krohn, L. *et al.* GBA variants in REM sleep behavior disorder: A multicenter study. *Neurology* **95**, e1008–e1016 (2020).
32. Brockmann, K. *et al.* Association between CSF alpha-synuclein seeding activity and genetic status in Parkinson's disease and dementia with Lewy bodies. *Acta Neuropathol. Commun.* **9**, 175 (2021).
33. Litvan, I. *et al.* SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders. *Mov. Disord.* **18**, 467–486 (2003).
34. Pavelka, L. *et al.* Age at onset as stratifier in idiopathic Parkinson's disease - effect of ageing and polygenic risk score on clinical phenotypes. *NPJ Park. Dis.* **8**, 102 (2022).
35. Gustavsson, E. K. *et al.* Genetic Identification in Early Onset Parkinsonism among Norwegian Patients. *Mov. Disord. Clin. Pract.* **4**, 499–508 (2017).
36. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
37. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
38. Leija-Salazar, M. *et al.* Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Mol. Genet. Genomic Med.* **7**, e564 (2019).
39. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
40. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma. Oxf. Engl.* **34**, 3094–3100 (2018).
41. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinforma. Oxf. Engl.* **32**, 3047–3048 (2016).
42. Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).
43. Modi, A., Vai, S., Caramelli, D. & Lari, M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. *Methods Mol. Biol. Clifton NJ* **2242**, 15–42 (2021).
44. Miller, N. A. *et al.* A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* **7**, 100 (2015).

45. Ji, J. *et al.* A semiautomated whole-exome sequencing workflow leads to increased diagnostic yield and identification of novel candidate variants. *Cold Spring Harb. Mol. Case Stud.* **5**, a003756 (2019).
46. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
47. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
48. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
49. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577–581 (2003).
50. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980-985 (2014).
51. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
52. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
53. Tian, Y. *et al.* REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Sci. Rep.* **9**, 12752 (2019).
54. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).
55. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).
56. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
57. Dierckxsens, N., Li, T., Vermeesch, J. R. & Xie, Z. A benchmark of structural variation detection by long reads through a realistic simulated model. 2020.12.25.424397 Preprint at https://doi.org/10.1101/2020.12.25.424397 (2020).
58. Sorrentino, E. *et al.* PacMAGI: A pipeline including accurate indel detection for the analysis of PacBio sequencing data applied to RPE65. *Gene* **832**, 146554 (2022).
59. Johannesen, K. M. *et al.* Genotype-phenotype correlations in SCN8A-related disorders reveal prognostic and therapeutic implications. *Brain J. Neurol.* **145**, 2991–3009 (2022).

TABLES

Table 1. Distribution of *GBA1* variants in the Luxembourg Parkinson's study.

Subclassification	nucleotide - protein changes	Subjects	PD n=660	Parkinsonism Patients n=100	Healthy controls n=808
Severe	c.115+1G>A	2	1		1
	p.P161S	2	2		
	p.G234W	1	1		
	p.G241R	2	2		
	p.H294Q	1			1
	p.R398*	1	1		
	p.G416S	1	1		
	p.L483P	6	5		1
	p.R502H	1	1		
	p.N409S; p.L483P	1	1		
	*RecNciI**	5	4		1
	p.K13R; p.L483P	1	1		
	p.F252I; p.T408M	1	1		
Mild	p.N409S	10	7		3
Risk	p.E365K	42	21+2[a]	1 DLB + 2 PSP	16
	p.E365K (*LRRK2*: p.R1441C)*	1			1
	p.T408M	28	15	1 DLB	12
	p.Y61H; p.T408M	1	1		
VUS	p.K13R	4	2		2
	p.K13R (*LRRK2*: p.G2019S)*	1			1
	p.Y61H	1	1		
	p.R78C	2			2
	p.A97G (new VUS)	1			1
	p.L213P	1	1		
	p.A215D (new VUS)	1	1		
	p.E427K	2	1		1
	p.R434C (new VUS)	1	1		
	p.H529R	1	1		
	p.R534C	1	1		
	p.A495P; p.V499V	3	1		2
	p.T408T	3	2		1

All variants were identified in the heterozygous state except in two individuals for p.E365K ([a] Homozygous state). Abbreviations: *GBA1*, glucocerebrosidase gene; PD, Parkinson's Disease and Parkinson's Disease with Dementia; PSP, Progressive Supranuclear Palsy; DLB, Dementia with Lewy Body; VUS, Variants of unknown significance; **RecNcil* (p.L483P; p.A495P; p.V499V); **LRRK2* mutation in brackets.

Table 2. Frequency of *GBA1* variants in the Luxembourg Parkinson's study.

Diagnosis	sub-classification of *GBA1* variants	Subjects	All GBA1-Carrier n (%)	Pathogenic *GBA1*-Carrier n (%)	OR (95%CI)	*p*-values
PD		637	**77 (12.1%)**	**67 (10.5%)**	**2.6 (1.6 to 4.1)**	**0.0001***
	severe			**21 (3.2%)**	**11.4 (2.6 to 49)**	**0.0010***
	mild			7 (1.1%)	3.7 (0.7 to 18)	0.1008
	risk			**39 (6%)**	**1.6 (1 to 2.8)**	**0.0537***
PSP	risk	59		2 (3.4%)	0.9 (0.2 to 3.9)	0.8941
DLB	risk	24		2 (8.3%)	2.2 (0.5 to 10)	0.2908
Healthy controls		675	34 (5%)	29 (4.3%)	-	-
	severe			2	-	-
	mild			2	-	-
	risk			25	-	-

Subject numbers result from excluding the first-degree family members interrelated in the cohort, the healthy controls of young age of assessment (< 60 AAA) with first-degree PD relatives, the CNV carriers, carrier of PD-causing variants (except *GBA1*) and the ethnic outliers. *GBA1*, glucocerebrosidase gene; PD, Parkinson's Disease and Parkinson's Disease with Dementia; PSP, Progressive Supranuclear Palsy; DLB, Dementia with Lewy Body. ORs are given with the 95% CI; Statistically significant results highlighted in bold and red with * (*p*-value < 0.05).

Table 3. Demographic data for the PD patients in the Luxembourg Parkinson's study separated by *GBA1* variant status.

Features	All pathogenic variants **(n = 67)**	Severe **(n = 21)**	Mild **(n = 7)**	Risk **(n = 39)**	Non carriers **(n = 554)**
AAA, mean (SD)	66.5 (±10.2) [OR=0.31; p=0.3977]	65.1 (±10.2) [OR=0.08; p=0.292]	67.1 (±15.6) [OR=0.59; p=0.8959]	67.1 (±9.2) [OR=0.57; p=0.7512]	67.6 (±10.7)
Sex, Male n (%)	40 (59.7%) [OR=0.71; p=0.1912]	13 (61.9%) [OR=0.78; p=0.5795]	5 (71.4%) [OR=1.19; p=0.8336]	22 (56.4%) [OR=0.62; p=0.151]	375 (67.7%)
AAO, mean (SD)	61.6 (±11.5) [OR=0.35; p=0.484]	58.6 (±13.1) [OR=0.02; p=0.1158]	65.4 (±17.0) [OR=16.16; p=0.5308]	62.5 (±9.3) [OR=0.9; p=0.9548]	62.6 (±11.6)
AAO < 45, N (%)	8 (11.9%) [OR=1.74; p=0.1767]	**5 (23.8%) [OR=4.02; p=0.0098*]**	2 (28.6%) [OR=5.14; p=0.0549]	1 (2.6%) [OR=0.34; p=0.2907]	40 (7.2%)
Disease Duration, mean (SD)	4.7 (±4.8) [OR=0.79; p=0.7303]	6.4 (±4.7) [OR=4.07; p=0.2238]	1.7 (±1.4) [OR=0.04; p=0.0981]	4.4 (±4.9) [OR=0.57; p=0.5103]	5.0 (±5.2)
Family History, N (%)	**25 (37.3%) [OR=1.74; p=0.0401*]**	8 (38.1%) [OR=1.8; p=0.2001]	2 (28.6% [OR=1.17; p=0.8508]	15 (38.5%) [OR=1.83; p=0.0782]	141 (25.5%)

Subject numbers result from excluding the first-degree family members interrelated in the cohort, the healthy controls with young age of assessment (< 60 AAA) or with first degree PD relatives, the CNV carriers, carrier of PD-causing variants (except GBA1), the ethnic outliers, synonymous and VUS variants carriers. Data are given as mean (SD) or N (%). Significance level for comparison is $p < 0.05$. AAA, age at assessment in years ; AAO, Age at onset in years.

Type of data	Clinical characteristics and scales	PD *GBA1* carrier SEVERE PD_{GBA1} (n=21)	missing values (%)	β (95%)	*p*-value	adj *p*-value	PD *GBA1* carrier MILD PD_{GBA1} (n=7)	missing values (%)	β (95%)	*p*-value	adj *p*-value	PD *GBA1* carrier RISK PD_{GBA1} (n=39)	missing Values (%)	β (95%)	*p*-value	adj *p*-value	PD GBA1 non-carrier N=554
Motor symptoms/scales	H&Y, mean (SD)	2.4 (±0.8)	3 (0.5%)	0.19 (-0.11 to 0.48)	0.2231	0.5354	2.0 (±0.6)	3 (0.5%)	0.0 (-0.50 to 0.51)	0.9917	0.9992	2.2 (±0.8)	3 (0.5%)	0.07 (-0.15 to 0.29)	0.5418	0.9471	2.2 (±0.8)
	MDS-UPDRS II, mean (SD)	12.6 (±4.4)	13 (2.3%)	0.49 (-2.75 to 3.72)	0.768	0.9678	10.1 (±6.4)	13 (2.3%)	0.92 (-4.67 to 6.51)	0.7476	0.9992	11.1 (±8.6)	13 (2.2%)	0.17 (-2.26 to 2.60)	0.8924	0.9748	11.4 (±8.3)
	MDS-UPDRS III, mean (SD)	34.8 (±15.7)	14 (2.4%)	-0.15 (-7.08 to 6.78)	0.966	0.9688	26.2 (±9.7)	13 (2.3%)	-4.62 (-16.79 to 7.54)	0.4564	0.9992	33.3 (±17.7)	12 (2.0%)	-0.37 (-5.30 to 4.57)	0.8846	0.9748	34.6 (±16.2)
	MDS-UPDRS IV, mean (SD)	3.0 (±4.5)	7 (1.2%)	0.87 (-0.46 to 2.20)	0.1991	0.512	0.1 (±0.4)	7 (1.2%)	-0.63 (-2.86 to 1.60)	0.5816	0.9992	1.2 (±2.3)	7 (1.2%)	-0.42 (-1.38 to 0.54)	0.395	0.8145	1.6 (±3.3)
	Dyskinesias, n (%)	5 (23.8%)	0	0.67 (-0.48 to 1.82)	0.2535	0.5368	0	-	-	-	-	4 (10.3%)	0	0.01 (-1.13 to 1.15)	0.9853	0.9928	64 (11.6%)
	Falls, n (%)	5 (23.8%)	0	0.46 (-0.62 to 1.53)	0.4054	0.6081	0	-	-	-	-	7 (17.9%)	0	0.17 (-0.74 to 1.08)	0.7103	0.9471	93 (16.8%)
	Gait Disorder, n (%)	16 (76.2%)	0	0.94 (-0.09 to 1.97)	0.075	0.512	3 (42.9%)	0	-0.24 (-1.76 to 1.29)	0.7627	0.9992	18 (46.2%)	0	-0.29 (-0.97 to 0.38)	0.3953	0.8145	307 (55.4%)
	FOG, n (%)	8 (38.1%)	0	0.66 (-0.33 to 1.65)	0.1932	0.512	0	-	-	-	-	7 (17.9%)	0	-0.19 (-1.12 to 0.74)	0.69	0.9471	123 (22.2%)
	Restless leg syndrome, n (%)	2 (9.5%)	0	0.03 (-1.47 to 1.53)	0.9688	0.9688	2 (28.6%)	0	1.61 (-0.09 to 3.3)	0.0632	0.8508	6 (15.4%)	0	0.7 (-0.23 to 1.63)	0.141	0.8145	46 (8.3%)
	Motor fluctuation, n (%)	5 (23.8%)	0	0.1 (-1.05 to 1.26)	0.8626	0.9678	0	-	-	-	-	5 (12.8%)	0	-0.22 (-1.25 to 0.82)	0.6834	0.9471	93 (16.8%)
Non-motor symptoms/scales	BDI, mean (SD)	12.4 (±5.7)	28 (4.9%)	2.09 (-0.91 to 5.1)	0.1721	0.512	8.0 (±3.5)	29 (5.2%)	-1.13 (-6.69 to 4.42)	0.6894	0.9992	8.0 (±5.7)	29 (4.9%)	-1.87 (-4.12 to 0.38)	0.1045	0.8145	9.9 (±7.1)
	MDS-UPDRS Part I, mean (SD)	**15.0 (±6.5)**	**15 (2.6%)**	**3.99 (1.06 to 6.92)**	**0.0074***	**0.1782**	8.6 (±2.4)	15 (2.7%)	-0.76 (-5.74 to 4.22)	0.7651	0.9992	9.5 (±6.8)	15 (2.5%)	-0.9 (-3.06 to 1.27)	0.417	0.8145	**10.5 (±7.0)**
	PDSS, mean (SD)	98.3 (±20.9)	44 (7.7%)	-4.62 (-15.31 to 6.08)	0.3978	0.6081	110.9 (±13.4)	43 (7.7%)	2.24 (-15.58 to 20.05)	0.8055	0.9992	104.0 (±24.9)	45 (7.6%)	-1.54 (-9.52 to 6.45)	0.7063	0.9471	105.0 (±24.6)
	SCOPA-AUT, mean (SD)	17.1 (±8.0)	32 (5.6%)	1.75 (-1.66 to 5.17)	0.3149	0.5967	13.1 (±7.1)	31 (5.5%)	-0.19 (-5.89 to 5.50)	0.9469	0.9992	14.1 (±7.8)	32 (5.4%)	-0.61 (-3.13 to 1.91)	0.6369	0.9471	14.9 (±8.1)
	Sniffin's stick test, mean (SD)	**6.4 (±3.6)**	**7 (1.2%)**	**-1.56 (-3.05 to -0.06)**	**0.0405***	**0.486**	6.6 (±2.1)	7 (1.2%)	-1.56 (-4.10 to 0.97)	0.2281	0.9992	7.4 (±4.0)	8 (1.3%)	-0.65 (-1.78 to 0.48)	0.2637	0.8145	**7.8 (±3.6)**
	SAS, mean (SD)	15.8 (±5.2)	33 (5.7%)	2.03 (-0.47 to 4.52)	0.1115	0.512	12.6 (±3.8)	32 (5.7%)	-1.48 (-5.66 to 2.69)	0.4866	0.9992	13.1 (±6.3)	34 (5.7%)	-0.76 (-2.64 to 1.12)	0.4299	0.8145	14.1 (±5.7)
	MoCA, mean (SD)	24.0 (±4.7)	12 (2.1%)	-0.78 (-2.58 to 1.02)	0.3947	0.6081	25.6 (±3.7)	12 (2.1%)	0.87 (-2.17 to 3.91)	0.5757	0.9992	24.9 (±4.0)	14 (2.4%)	0.14 (-1.22 to 1.49)	0.8443	0.9748	24.4 (±4.5)
	Constipation, n (%)	10 (47.6%)	0	0.06 (-0.83 to 0.94)	0.903	0.9678	5 (71.4%)	0	1.39 (-0.27 to 3.05)	0.1002	0.9018	14 (35.9%)	0	-0.32 (-1.01 to 0.37)	0.3666	0.8145	246 (44.4%)
	Dysphagia, n (%)	4 (19.0%)	0	-0.48 (-1.59 to 0.64)	0.4015	0.6081	1 (14.3%)	0	-0.52 (-2.66 to 1.61)	0.6295	0.9992	10 (25.6%)	0	-0.0 (-0.75 to 0.75)	0.9928	0.9928	145 (26.2%)
	Insomnia, n (%)	6 (28.6%)	0	-0.05 (-1.04 to 0.93)	0.914	0.9678	4 (57.1%)	0	1.44 (-0.09 to 2.97)	0.0658	0.8508	7 (17.9%)	0	-0.56 (-1.41 to 0.28)	0.1929	0.8145	151 (27.3%)
	Orthostatism, n (%)	5 (23.8%)	0	-0.36 (-1.39 to 0.67)	0.4922	0.7088	4 (57.1%)	0	1.4 (-0.12 to 2.91)	0.0709	0.8508	15 (38.5%)	0	0.44 (-0.24 to 1.12)	0.2021	0.8145	163 (29.4%)
	Urinary incontinence, n (%)	6 (28.6%)	0	-0.08 (-1.06 to 0.9)	0.8713	0.9678	2 (28.6%)	0	0.15 (-1.54 to 1.83)	0.865	0.9992	17 (43.6%)	0	0.65 (-0.03 to 1.32)	0.0615	0.8145	168 (30.3%)
	Hallucinations, n (%)	**8 (38.1%)**	**0**	**1.23 (0.29 to 2.16)**	**0.0099***	**0.1782**	2 (28.6%)	0	1.23 (-0.45 to 2.9)	0.1504	0.9024	6 (15.4%)	0	0.09 (-0.83 to 1.02)	0.8444	0.9748	**83 (15.0%)**
	Excessive daytime sleepiness, n (%)	9 (42.9%)	0	0.49 (-0.4 to 1.39)	0.2812	0.5624	0	-	-	-	-	14 (35.9%)	0	0.33 (-0.37 to 1.03)	0.3534	0.8145	170 (30.7%)
	ICD, n (%)	2 (9.5%)	0	-0.23 (-1.76 to 1.3)	0.7682	0.9678	0	-	-	-	-	4 (10.3%)	0	0.22 (-0.88 to 1.32)	0.6925	0.9471	53 (9.6%)
	Syncope, n (%)	2 (9.5%)	0	0.93 (-0.62 to 2.48)	0.2401	0.5368	1 (14.3%)	0	1.74 (-0.49 to 3.98)	0.1257	0.9024	3 (7.7%)	0	0.6 (-0.67 to 1.86)	0.3575	0.8145	26 (4.7%)
	RBDSQ, n (%)	10 (47.6%)	42 (7.3%)	0.65 (-0.28 to 1.59)	0.1727	0.512	1 (14.3%)	42 (7.5%)	-0.49 (-2.66 to 1.68)	0.6606	0.9992	14 (35.9%)	43 (7.3%)	0.42 (-0.31 to 1.15)	0.2631	0.8145	165 (29.8%)
Other clinical outcomes	LEDD (mg/day), mean (SD)	690.5 (±457.9)	19 (3.3%)	117.68 (-35.58 to 270.94)	0.1324	0.512	324.7 (±224.7)	20 (3.6%)	-68.22 (-349.43 to 212.99)	0.6345	0.9992	496.6 (±443.2)	20 (3.4%)	7.86 (-107.30 to 123.02)	0.8936	0.9748	513.5 (±404.9)
	PDQ-39, mean (SD)	52.0 (±26.3)	49 (8.5%)	9.46 (-1.34 to 20.27)	0.0863	0.512	26.1 (±17.7)	49 (8.7%)	-6.81 (-25.20 to 11.59)	0.4686	0.9992	34.9 (±27.2)	51 (8.6%)	-4.36 (-12.59 to 3.86)	0.2987	0.8145	39.3 (±26.7)
	DBS, n (%)	3 (14.3%)	0	1.38 (-0.21 to 2.98)	0.0882	0.512	0	-	-	-	-	1 (2.6%)	0	-0.07 (-2.25 to 2.11)	0.9477	0.9928	24 (4.3%)
Comorbidities	Diabetes, n (%)	2 (9.5%)	0	0.18 (-1.34 to 1.7)	0.8198	0.9678	1 (14.3%)	0	0.14 (-2.08 to 2.35)	0.9029	0.9992	5 (12.8%)	0	0.46 (-0.54 to 1.46)	0.3704	0.8145	55 (9.9%)
	Hypercholesterolemia, n (%)	6 (28.6%)	0	-0.44 (-1.42 to 0.54)	0.3784	0.6081	2 (28.6%)	0	-0.65 (-2.34 to 1.04)	0.4533	0.9992	17 (43.6%)	0	0.15 (-0.52 to 0.81)	0.6682	0.9471	226 (40.8%)
	Cardiovascular disease, n (%)	1 (4.8%)	0	-1.53 (-3.59 to 0.52)	0.1436	0.512	1 (14.3%)	0	-0.64 (-2.87 to 1.58)	0.5708	0.9992	8 (20.5%)	0	0.13 (-0.71 to 0.96)	0.7648	0.9748	116 (20.9%)
	Arterial hypertension, n (%)	9 (42.9%)	0	0.1 (-0.82 to 1.02)	0.8312	0.9678	2 (28.6%)	0	-0.88 (-2.6 to 0.85)	0.32	0.9992	12 (30.8%)	0	-0.57 (-1.28 to 0.15)	0.1197	0.8145	248 (44.8%)
	Traumatic Brain Injury, n (%)	5 (23.8%)	0	0.15 (-0.88 to 1.18)	0.7772	0.9678	0	-	-	-	-	6 (15.4%)	0	-0.43 (-1.33 to 0.46)	0.3421	0.8145	122 (22.0%)

Table 4. Clinical characteristics of PD classified by *GBA1* variant status.

Comparison of each type (severe, mild, risk) of *GBA1* variants and its association with clinical characterization. We used regression models (linear and logistic). Data are given as mean and standard deviation (SD) for continuous clinical outcomes and as percentage for binary clinical outcomes. Models adjusted for sex, age at assessment, and disease duration. Beta (β) regression coefficient is given with the 95% CI. Statistically significant results highlighted in bold with (*) sign and red (p-value < 0.05). Abbreviation : p-value, unadjusted p-value; adj p-value, corrected for multiple comparisons using FDR adjustment; AAO, age at onset; H&Y, Hoehn & Yahr; MDS-UPDRS, Movement Disorders Society - Unified Parkinson's Disease Rating Scale; FOG, freezing of gait; BDI, Beck Depression Inventory; PDSS, Parkinson's Disease Sleep Scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire; SAS, Starkstein apathy scale; MoCA, Montreal Cognitive Assessment; ICD, impulse control disorder; RBDSQ, REM Sleep Behavior Disorder Screening Questionnaire; LEDD, L-dopa equivalent daily dose (mg/day); PDQ-39, Parkinson's Disease quality of life Questionnaire; DBS, Presence of treatment by Deep Brain Stimulation

Table 5. The deleterious impact of severe *GBA1*-PD carriers in comparison with mild and risk and their clinical characteristics.

Type of data	Clinical characteristics and scales	PD_{GBA1} carrier		missing values (%)	β (95%)	*p*-value	adj *p*-value
		severe (n=21)	mild + risk (N=46)				
Motor symptoms/scales	H&Y, mean (SD)	2.4 (±0.8)	2.2 (±0.8)	0	0.14 (-0.24 to 0.52)	0.4719	0.7549
	MDS-UPDRS II, mean (SD)	12.6 (±4.4)	10.9 (±8.2)	0	0.24 (-3.18 to 3.65)	0.8922	0.9706
	MDS-UPDRS III, mean (SD)	34.8 (±15.7)	32.4 (±16.9)	3 (4.5%)	0.65 (-7.69 to 8.98)	0.8793	0.9706
	MDS-UPDRS IV, mean (SD)	3.0 (±4.5)	1.0 (±2.2)	0	1.33 (-0.17 to 2.82)	0.082	0.574
	Dyskinesias, n (%)	5 (23.8%)	4 (8.7%)	0	0.65 (-1.09 to 2.39)	0.4621	0.7549
	Falls, n (%)	5 (23.8%)	7 (15.2%)	0	0.24 (-1.29 to 1.77)	0.7594	0.9601
	Gait Disorder, n (%)	**16 (76.2%)**	**21 (45.7%)**	**0**	**1.49 (0.25 to 2.73)**	**0.0188***	**0.2193**
	FOG, n (%)	8 (38.1%)	7 (15.2%)	0	0.79 (-0.73 to 2.32)	0.3091	0.7549
	Restless leg syndrome, n (%)	2 (9.5%)	8 (17.4%)	0	-0.98 (-2.81 to 0.85)	0.2952	0.7549
	Motor fluctuation, n (%)	5 (23.8%)	5 (10.9%)	0	0.28 (-1.36 to 1.92)	0.7348	0.9601
Non-motor symptoms/scales	**BDI, mean (SD)**	**12.4 (±5.7)**	**8.0 (±5.4)**	**2 (3.0%)**	**4.03 (1.08 to 6.98)**	**0.0074***	**0.1295**
	MDS-UPDRS Part I, mean (SD)	**15.0 (±6.5)**	**9.3 (±6.3)**	**0**	**4.91 (1.8 to 8.02)**	**0.0019***	**0.0665**
	PDSS, mean (SD)	98.3 (±20.9)	105.1 (±23.5)	3 (4.5%)	-2.79 (-14.9 to 9.33)	0.6521	0.9129
	SCOPA-AUT, mean (SD)	17.1 (±8.0)	13.9 (±7.6)	2 (3.0%)	2.61 (-1.62 to 6.85)	0.2269	0.7542
	Sniffin's stick test, mean (SD)	6.4 (±3.6)	7.3 (±3.7)	1 (1.5%)	-0.86 (-2.75 to 1.02)	0.3695	0.7549
	SAS, mean (SD)	15.8 (±5.2)	13.0 (±6.0)	3 (4.5%)	2.02 (-0.94 to 4.97)	0.1817	0.7542
	MoCA, mean (SD)	24.0 (±4.7)	25.0 (±3.9)	2 (3.0%)	-0.24 (-2.4 to 1.92)	0.8292	0.9706
	Constipation, n (%)	10 (47.6%)	19 (41.3%)	0	0.12 (-1.02 to 1.26)	0.8394	0.9706
	Dysphagia, n (%)	4 (19.0%)	11 (23.9%)	0	-0.44 (-1.83 to 0.95)	0.5348	0.8138
	Insomnia, n (%)	6 (28.6%)	11 (23.9%)	0	-0.02 (-1.27 to 1.23)	0.9767	0.9836
	Orthostatism, n (%)	5 (23.8%)	19 (41.3%)	0	-0.61 (-1.83 to 0.61)	0.327	0.7549
	Urinary incontinence, n (%)	6 (28.6%)	19 (41.3%)	0	-0.76 (-1.97 to 0.44)	0.2156	0.7542
	Hallucinations, n (%)	8 (38.1%)	8 (17.4%)	0	1.03 (-0.24 to 2.3)	0.1127	0.6485
	Excessive daytime sleepiness, n (%)	9 (42.9%)	14 (30.4%)	0	0.41 (-0.71 to 1.52)	0.4745	0.7549
	ICD, n (%)	2 (9.5%)	4 (8.7%)	0	-0.02 (-1.89 to 1.85)	0.9836	0.9836
	Syncope, n (%)	2 (9.5%)	4 (8.7%)	0	0.1 (-1.72 to 1.92)	0.9151	0.9706
	RBDSQ, **n (%)**	10 (47.6%)	15 (32.6%)	4 (6.0%)	0.48 (-0.68 to 1.64)	0.4197	0.7549
Other clinical outcomes	LEDD (mg/day), mean (SD)	690.5 (±457.9)	473.1 (±422.4)	2 (3.0%)	66.4 (-107.09 to 239.89)	0.4531	0.7549
	PDQ-39, mean (SD)	**52.0 (±26.3)**	**33.5 (±25.9)**	**2 (3.0%)**	**12.77 (0.45 to 25.09)**	**0.0422***	**0.3692**
	DBS, n (%)	3 (14.3%)	1 (2.2%)	0	0.76 (-2.17 to 3.69)	0.6122	0.8928
Comorbidities	Diabetes, n (%)	2 (9.5%)	6 (13.0%)	0	-0.26 (-2.0 to 1.48)	0.7681	0.9601
	Hypercholesterolemia, n (%)	6 (28.6%)	19 (41.3%)	0	-0.47 (-1.63 to 0.69)	0.4269	0.7549
	Cardiovascular disease, n (%)	1 (4.8%)	9 (19.6%)	0	-1.71 (-3.92 to 0.5)	0.1297	0.6485
	Arterial hypertension, n (%)	9 (42.9%)	14 (30.4%)	0	0.66 (-0.49 to 1.81)	0.2586	0.7542
	Traumatic Brain Injury, n (%)	5 (23.8%)	6 (13.0%)	0	0.82 (-0.55 to 2.2)	0.2385	0.7542

We used regression models (linear and logistic). Data are given as mean and standard deviation (SD) for continuous clinical outcomes and as percentage for binary clinical outcomes. Models adjusted for sex, age at assessment, and disease duration. Beta (β) regression coefficient is given with the 95% CI. Statistically significant results highlighted in bold with (*) sign and red (p-value < 0.05). Abbreviation : p-value, unadjusted p-value; adj p-value, corrected for multiple comparisons using FDR adjustment; AAO, age at onset; H&Y, Hoehn & Yahr; MDS-UPDRS, Movement Disorders Society - Unified Parkinson's Disease Rating Scale; FOG, freezing of gait; BDI, Beck Depression Inventory; PDSS, Parkinson's Disease Sleep Scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire; SAS, Starkstein apathy scale; MoCA, Montreal Cognitive Assessment; ICD, impulse control disorder; RBDSQ, REM Sleep Behavior Disorder Screening Questionnaire; LEDD, L-dopa equivalent daily dose (mg/day); PDQ-39, Parkinson's Disease quality of life Questionnaire; DBS, Presence of treatment by Deep Brain Stimulation.

FIGURE

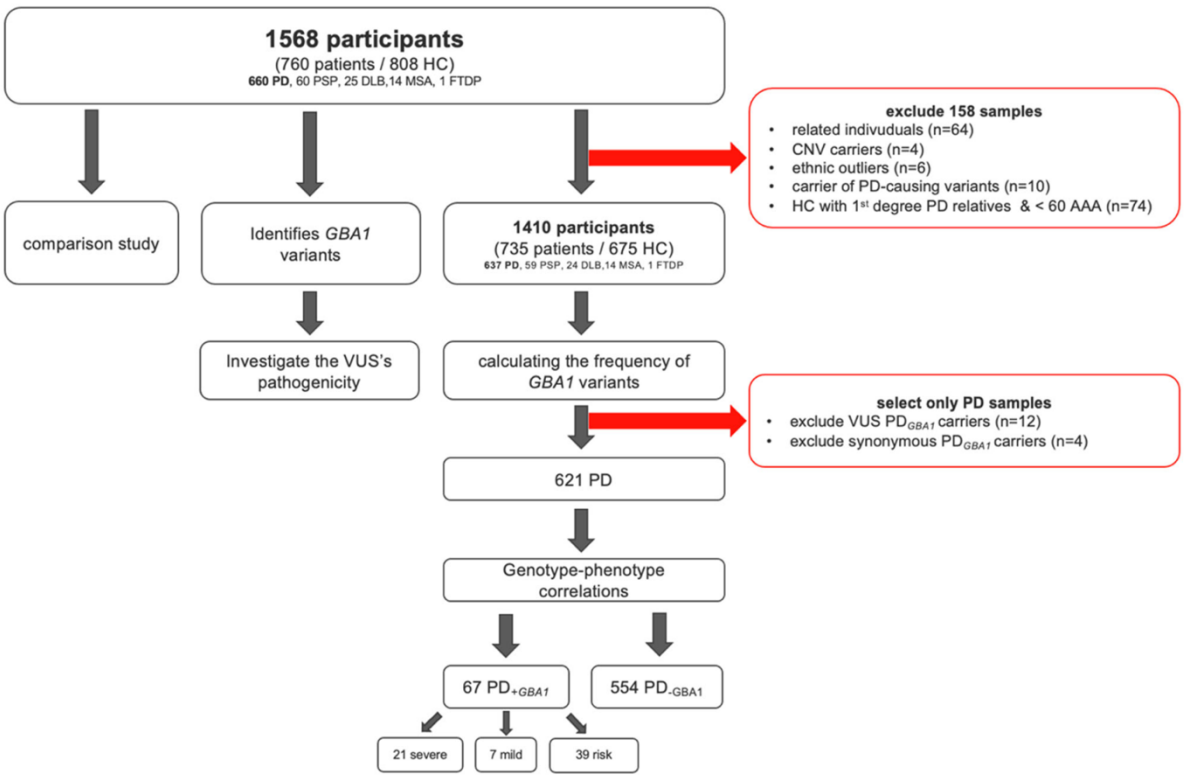


Figure 1: Description of the study dataset and methodology.

HC Healthy controls, PD Parkinson's Disease and Parkinson's Disease with Dementia, PSP Progressive Supranuclear Palsy, DLB Dementia with Lewy Body, MSA Multiple System Atrophy, FTDP Fronto-temporal dementia with parkinsonism, GBA1 glucocerebrosidase gene, VUS Variants of unknown significance, PD+GBA1 PD patients with GBA1 pathogenic variant, PD-GBA1 PD patients without GBA1 pathogenic variant, CNV copy number variants, AAA age at assessment.

a

GBA1 variants	Sample count	PacBio	WGS	WGS + Gauchian	NeuroChip
p.E365K	3	3 TP	3 TP	3 TP	-
p.T408M	1	TP	TP	TP	TP
p.N409S	1	TP	TP	TP	-
p.N431S	1	-	-	-	FP
p.A215D	1	-	-	-	FP
RecNcil*	1	TP	-	TP	-

b

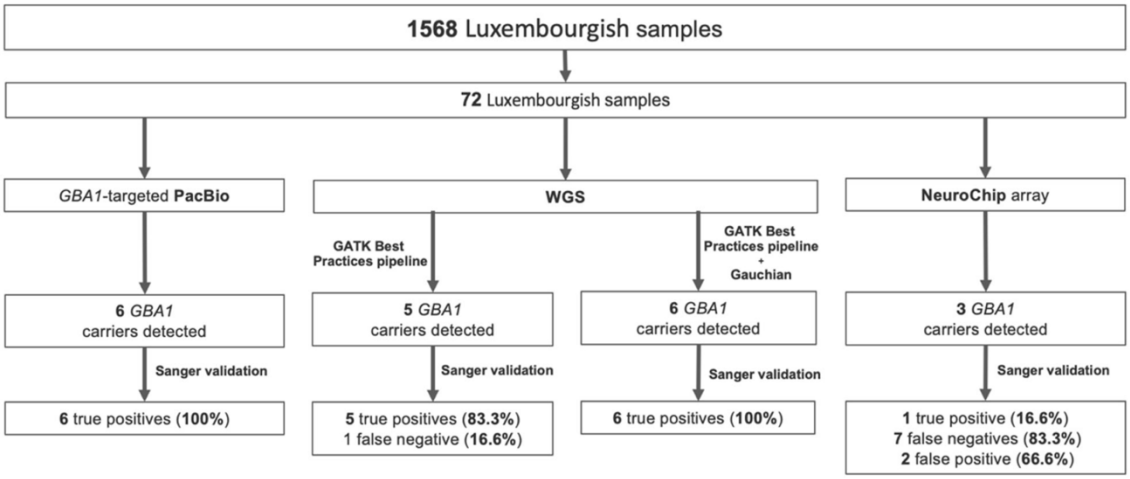


Figure 2. Comparison of variant calls from PacBio, WGS and NeuroChip genotyping data using 72 matched samples for the *GBA1* gene and validated by Sanger sequencing.

A)**Rec*Ncil (p.L483P; p.A495P; p.V499V)*; Sanger sequencing results : TP, true positive ; FP, false positive. Sample count gives total number of samples carrying the variant found by each method.

B) Comparative study of *GBA1* variants detection by the *GBA1*-targeted PacBio and NeuroChip array methods in the Luxembourg Parkinson's study. Due to overrepresented variants with the NeuroChip array, we applied for the detected variants a study-wide threshold of 1% in our cohort.

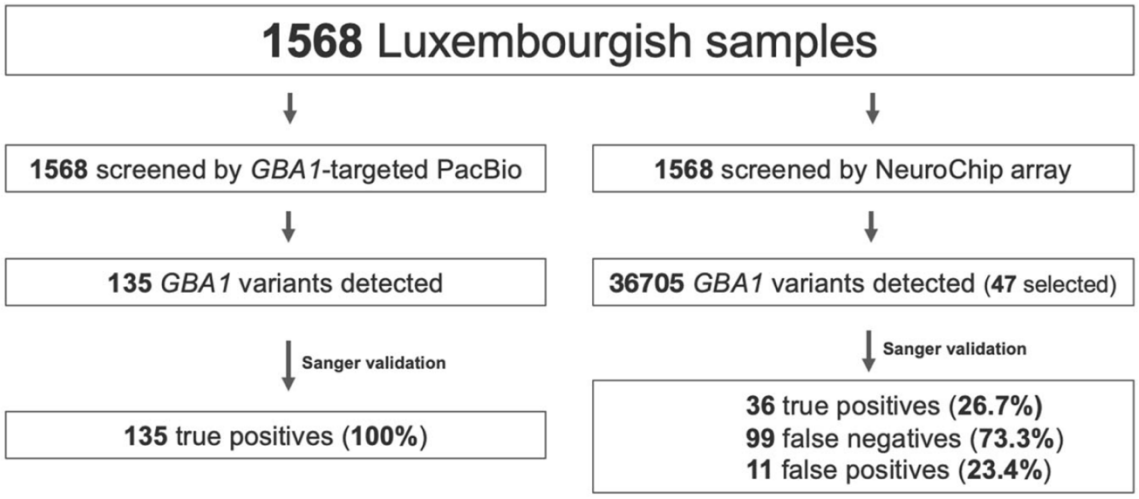
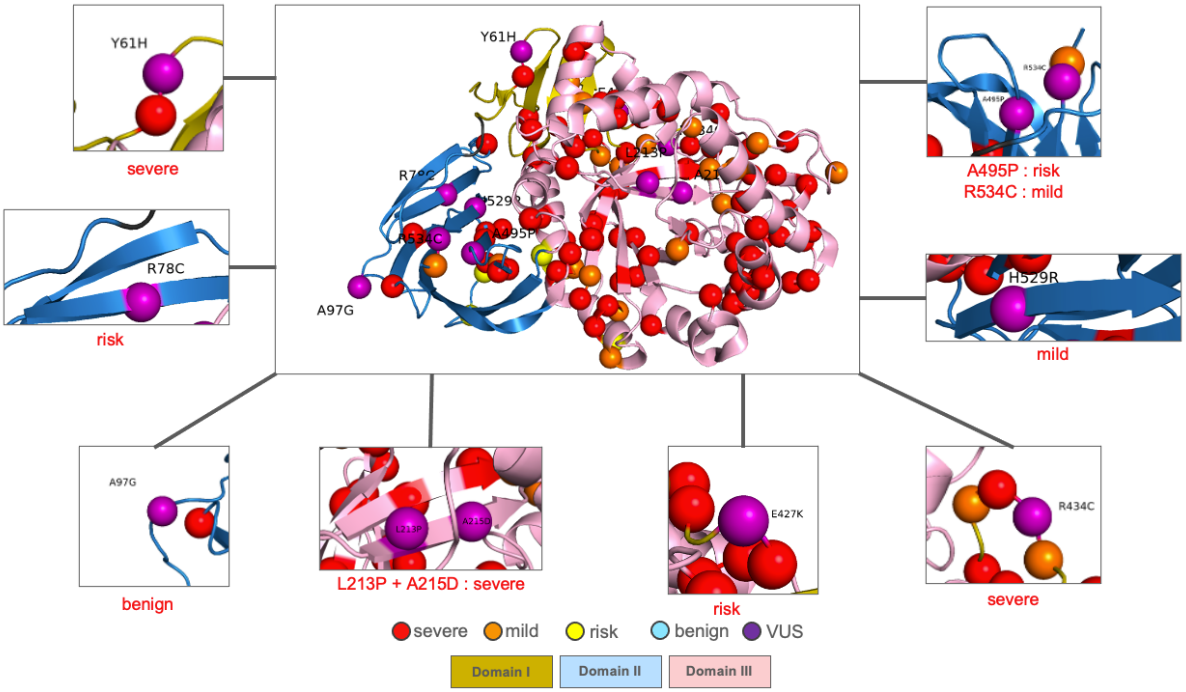


Figure 3: **Comparative study of *GBA1* variants detection by the GBA1-targeted PacBio and NeuroChip array methods in the Luxembourg Parkinson's study.**
Due to overrepresented variants with the NeuroChip array, we applied for the detected variants a study-wide threshold of 1% in our cohort.

A



B

Prediction	Subjects		protein change	nucleotide change	dbSNP	Höglinger et al.	ClinVar Significance	ClinVar	HGMD	HGMD interpretation	REVEL	CADD	GnomAD NFE	3D	domain
	PD (AA0)	HC (AAA)													
risk	2 (50/66)	3 (37/38/66)	K13R	c.A38G	rs150466109	VUS	Benign	GD	DM	GD	T	T	ultra-rare	/	/
severe	1 (38)		Y61H	c.T181C	rs1266341749	-	-	-	-	-	T	T	ultra-rare	coil-loop	I
risk		2 (56/58)	R78C	c.C232T	rs146774384	-	-	-	DM	PD susceptibility	T	D	ultra-rare	β-sheet	II
benign		1 (60)	A97G	c.C290G	-	-	-	-	-	-	T	T	-	coil-bend	II
severe	1 (56)		L213P	c.T638C	-	VUS	-	-	DM	PD	D	D	-	β-sheet	III
severe	1 (68)		A215D	c.C644A	-	-	-	-	DM	GD	D	D	-	β-sheet	III
risk	1 (65)	1 (70)	E427K	c.G1279A	rs149171124	VUS	Uncertain significance	Parkinsonism	DM	Reduced activity	T	T	ultra-rare	coil-turn	I
severe	1 (62)		R434C	c.C1300T	rs747284798	-	-	-	DM	GD 1	D	D	-	coil-loop	I
risk	1 (60)	2 (38/61)	A495P	c.G1483C	rs368060	-	Benign	GD	DM	-	T	T	ultra-rare	β-sheet	II
mild	1 (77)		H529R	c.A1586G	-	VUS	-	-	DM	PD	T	T	-	β-sheet	II
mild	1 (78)		R534C	c.C1600T	rs146519305	-	-	-	-	-	T	T	ultra-rare	coil-loop	II
risk	2 (61/64)	1 (61)	T408T	c.G1224A	rs138498426	VUS	Uncertain significance	GD	DM	PD	-	-	ultra-rare	buried residue	III

Figure 4. Sub-classification of VUS found in the Luxembourg Parkinson's study.

A) *GBA1* missense and stop gain variants mapped onto the three-dimensional structure of GCase. Domain 1 is shown in dark yellow, domain 2 in blue, and domain 3 in pink. Variants classified as severe are colored red, mild are colored orange, risk in yellow and VUS are colored purple. The 3D structure of GCase (PDB code 1ogs) was generated using PYMOL (http://www.pymol.org). B) Proposed sub-classification of identified VUSs with their score in a known database. *GBA1*, glucocerebrosidase gene; GD, Gaucher's disease; PD, Parkinson's disease. HGMD, The Human Gene Mutation Database; REVEL, Rare Exome Variant Ensemble Learner; CADD, Combined Annotation Dependent Depletion; gnomAD, The Genome Aggregation Database. DM, Disease causing mutation; D, Deleterious; T, Tolerate. Variants classified as severe are colored red, mild are colored orange, risk in yellow and VUS are colored purple.

1.3. Supplementary material

Supplementary. Clinical symptoms and scales

Clinical symptoms and scales Movement Disorder Society-Unified Parkinson's disease Rating Scale (MDS-UPDRS I-IV) and Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire (SCOPA-AUT) are used under the license number (14017_ND). Probable REM sleep behaviour disorder (pRBD) was based on the validated selfreporting questionnaire REM Sleep Behaviour Disorder Screening Questionnaire (RBDSQ) where possible pRBD was defined as RBDSQ ≥ 7[1]. Assessment of sleep quality was done via Parkinson's Disease Sleep Scale (PDSS)[2]. The calculation of levodopa equivalent daily dose (LEDD, reported in g/day) was based on established conversion factors[3]. Definition of constipation corresponds to the diagnostic criteria ROME III and information was acquired in a semi-structured interview[4]. The Hoehn and Yahr scale (H&Y) corresponds to the modified version of the scale[5]. Quality of life was assessed via Parkinson's disease Questionnaire 39 (PDQ-39)[6]. Depression symptoms were reflected by Beck Depression Inventory Version I (BDI)[7]. Olfactory function was examined with 16 items Sniffin' Stick test[8]. Cognitive performance was assessed via Montreal Cognitive Assessment (MoCA)[9]. Presence of recurrent orthostatic hypotension was assessed using a semi-structured interview inquiring for the symptoms of orthostatic hypotension, i.e. faintness, dizziness, light-headedness, vertigo, hearing disturbance, visual disturbance or syncope following the tilting, standing up or after a long standing relieved by sitting down or laying down. Symptoms included in the analysis (gait disorder, falls, freezing of gait (FOG), dyskinesia, motor fluctuations, excessive daily sleepiness, insomnia, dysphagia, urinary incontinence (corresponding to any type of urinary incontinence, i.e. stress, urge, overflow or mixed urinary incontinence), hallucinations and impulse control disorder (ICD)) were assessed during a semi-structured interview of the participant and/or the participant's proxy with a study physician and refer to the current motor and non-motor symptoms at the time of assessment. DBS, Presence of treatment by Deep Brain Stimulation. Restless leg syndrome corresponds to history or presence of rest less leg syndrome based on the accurate anamnestic description corresponding to the diagnostic features/criteria.

Supplementary Table 1. Comparison between AAO, AAA and Gender for the Luxembourgish cohort before and after the exclusion criteria.

	Before		After	
Features	**Cases (n = 760)**	**Controls (n = 808)**	**Cases (n = 735)**	**Controls (n = 675)**
AAA	67.6 (±10.5)	59.3 (±12.2)	67.9 (±10.4)	61 (±11.5)
AAO, mean (SD)	63 (±11.5)	-	63.2 (±11.3)	-
Sex, Male % (n)	505 (66.4%)	426 (52.7%)	489 (66.5%)	368 (54.5%)

We applied a t-test for AAA and AAO features and a Fisher test for the gender features. Exclusion criteria : the first-degree family members interrelated in the cohort, the healthy controls of young age of assessment (< 60 AAA) with first-degree PD relatives, the CNV carriers, carrier of PD-causing variants and the ethnic outliers.

Supplementary Table 2. PCR primers

PCR Primers targeted *GBA1* gene (Stone et al. 2000)

exon	Forward	reverse	length
1–5	CCTAAAGTTGTCACCCATAC	AGCAGACCTACCCTACAGTTT	2972
5–7	GACCTCAAATGATATACCTG	AGTTTGGGAGCCAGTCATTT	2049
8–11	TGTGTGCAAGGTCCAGGATCAG	ACCACCTAGAGGGGAAAGTG	1682

Nested Sequencing Primers targeted *GBA1* gene (partly Stone et al.2000)

exon	Forward	Reverse	Length
1	CCTAAAGTTGTCACCCATAC	aaattccagtgccaggattc	392
2	GAGAGTAGTTGAGGGGTGGA	CAAAGGACTATGAGGCAGAA	210
3	ATGTGTCCATTCTCCATGTC	GGTGATCACTGACACCATTT	323
4	GGTGTCAGTGATCACCATGG	ACGAAAAGTTTCAATGGCTCT	263
5	GCAAGTGATAAGCAGAGTCC	AGCAGACCTACCCTACAGTTT	280
6	CTCTGGGTGCTTCTCTCTTC	ACAGATCAGCATGGCTAAAT	271
7	TTGGCCGGATCATTCATGAC	AGTTTGGGAGCCAGTCATTT	342
8	TGTGTGCAAGGTCCAGGATCAG	TTTGCAGGAAGGGAGACTGG	294
9	CACAGGGCTGACCTACCCAC	GCTCCCTCGTGGTGTAGAGT	307
10	CAGGAGTTATGGGGTGGGTC	GAGGCACATCCTTAGAGGAG	329
11	GTGGGCTGAAGACAGCGTTGG	ACCACCTAGAGGGGAAAGTG	342

The first step is a long-range PCR which ensures that the pseudogene of *GBA1* is not amplified.

Supplementary Table 3. Exonic and splice-site *GBA1* variants found in the Luxembourg Parkinson's study.

Subclassification	protein change	Nucleotide change	dbSNP	Variant type	Höglinger et al.	ClinVar Significance	ClinVar	HGMD	REVEL	CADD	gnomAD NFE	3D	domain
severe	-	c.115+1G>A	rs104886460	splicing	severe	Pathogenic	PD/GD/DLB	DM	-	D	rare	/	/
	p.P161S	c.C481T	rs121908299	missense	-	Pathogenic	GD	DM	D	D	-	β-sheet	III
	p.G234W	c.G700T	-	missense	severe	-	-	DM	D	D	-	coil-loop	III
	p.G241R	c.G721A	rs409652	missense	severe	Pathogenic	GD	DM	D	T	rare	coil-turn	III
	p.F252I	c.T754A	rs381737	missense	severe	Pathogenic	GD	DM	D	T	rare	α-helix	III
	p.H294Q	c.T882G	rs367968666	missense	severe	Pathogenic	GD	DM	T	T		α-helix	III
	p.R398*	c.C1192T	rs121908309	nonsense	severe	Pathogenic	GD	DM	-	D	-	α-helix	III
	p.G416S	c.G1246A	rs121908311	missense	severe	Pathogenic	GD	DM	D	T	rare	β-sheet	III
	p.L483P*	c.T1448C	rs421016	missense	severe	Pathogenic Conflicting interpretations of pathogenicity	GD	DM	D	T	rare	β-sheet	II
	p.R502H	c.G1505A	rs80356772	missense	severe		GD	DM	D	D	rare	coil-loop	II
mild	p.N409S	c.A1226G	rs76763715	missense	mild	Pathogenic	PD/GD/DLB	DM	T	T	rare	α-helix	III
risk	p.E365K	c.G1093A	rs2230288	missense	risk	Benign	GD	DM	T	T	1.4%	α-helix	III
	p.T408M	c.C1223T	rs75548401	missense	risk	Conflicting interpretations of pathogenicity	PD/GD	DM	T	T	1.1%	buried residue	III
VUS	p.K13R	c.A38G	rs150466109	missense	VUS	Benign	GD	DM	T	T	rare	/	/
	p.Y61H	c.T181C	rs1266341749	missense	-	-	-	-	T	T	rare	coil-loop	I
	p.R78C	c.C232T	rs146774384	missense	-	-	-	DM	T	D	rare	β-sheet	II
	p.A97G	c.C290G	-	missense	-	-	-	-	T	T	-	coil-bend	II
	p.L213P	c.T638C	-	missense	VUS	-	-	DM	D	D	-	β-sheet	III
	p.A215D	c.C644A	-	missense	-	-	-	DM	D	D	-	β-sheet	III
	p.E427K	c.G1279A	rs149171124	missense	VUS	Uncertain significance	Parkinsonism	DM	T	T	rare	coil-turn	I
	p.R434C	c.C1300T	rs747284798	missense	-	-	-	DM	D	D	-	coil-loop	I
	p.A495P*	c.G1483C	rs368060	missense	-	Benign	GD	DM	T	T	rare	β-sheet	II
	p.H529R	c.A1586G	-	missense	VUS	-	-	DM	T	T	-	β-sheet	II
	p.R534C	c.C1600T	rs146519305	missense	-	-	-	-	T	T	rare	coil-loop	II
	p.T408T	c.G1224A	rs138498426	synonymous	VUS	Uncertain significance	GD	DM	-	-	rare	buried residue	III

Abbreviations: *GBA1*, glucocerebrosidase gene; GD, Gaucher's disease; PD, Parkinson's disease; DLB, Dementia with Lewy Bodies. HGMD, The Human Gene Mutation Database; REVEL, Rare Exome Variant Ensemble Learner; CADD, Combined Dependent Depletion; gnomAD, The Genome Aggregation Database. DM, Disease causing mutation; D, Deleterious; T, Tolerate; VUS, Variants of unknown significan; **RecNcil* (p.L483P; p.A495P; p.V499V).

Supplementary Table 4. Exonic synonymous *GBA1* variants in the Luxembourg Parkinson's study.

Subjects		nucleotide change	protein change	dbSNP	Exon	3D	gnomAD NFE	ClinVar	ClinVar Significance	HGMD
PD	HC									
5	3	c.G1497C	p.V499V*	rs1135675	11	β-sheet	rare	-	-	-
	1	c.C1473A	p.P491P	rs149257166	11	coil-turn	-	-	-	-
1		c.A1455G	p.A485A	rs199928507	11	β-sheet	rare	-	-	-
1		c.C228T	p.F76F	rs75954905	4	β-sheet	rare	-	-	-
2	1	c.G1224A	p.T408T	rs138498426	9	buried residue	rare	GD	Uncertain significance	PD
1		c.T1029C	p.Y343Y	-	9	coil-turn	-	-	-	-
2		c.C630T	p.P210P	rs201615998	7	coil-loop	rare	-	-	-
1		c.C585G	p.L195L	rs1157873928	6	π-helix	-	-	-	-
1		c.G105A	p.S35S	rs148001886	3	-	rare	-	-	-

All variants were identified in the heterozygous state.

Abbreviations: *GBA1*, glucocerebrosidase gene; GD, Gaucher's disease; PD, Parkinson's disease; HC, Healthy controls; HGMD, The Human Gene Mutation Database ; gnomAD, The Genome Aggregation Database; **RecNcil* (p.L483P; p.A495P; p.V499V).

Supplementary Table 5. Splicing, intronic and UTRs regions variants detected in the Luxembourg Parkinson's study by PacBio sequencing method.

N°	POS	REF	ALT	Region	Transcript	Nucleotide changes	GnomAD NFE	avsnp150	ClinVar	ClinVar Significance	PD	HC
1	155204345	C	T	UTR3	NM_001005742	c.*441G>A					1	0
2	155204541	G	A	UTR3	NM_001005743	c.*245C>T	0.0002				0	1
3	155204621	A	G	UTR3	NM_001005744	c.*165T>C	0.0393	rs375776699			81	79
4	155204684	A	G	UTR3	NM_001005745	c.*102T>C	0.0007	rs368275143			17	21
5	155204694	C	T	UTR3	NM_001005746	c.*92G>A	0.0025	rs708606	GD/DLB	Likely_benign	18	23
6	155204701	C	T	UTR3	NM_001005747	c.*85G>A					0	1
7	155205200	G	T	intronic			0.0018	rs183510604			0	2
8	155205203	G	A	intronic			0.0003	rs426516			4	1
9	155205300	G	A	intronic							3	1
10	155205359	G	A	intronic							0	1
11	155205378	C	T	intronic			0.0139	rs12752133			30	16
12	155205646	G	C	intronic			6.668e-05				1	0
13	155205669	G	T	intronic			0.9992	rs3115534	GD	Benign	752	806
14	155205709	A	G	intronic			0.0007	rs548435731			0	1
15	155205748	G	A	intronic			6.681e-05	rs1003268223			1	1
16	155205801	A	G	intronic							1	0
17	155205964	C	T	intronic							0	1
18	155206430	G	A	intronic			6.688e-05	rs767239225			0	1
19	155206542	TTGTGTGTGTA	T	intronic				rs998227221			0	1
20	155206578	G	GTA	intronic			0.0003	rs200655080			1	3
21	155206580	A	G	intronic			6.711e-05	rs146697312			1	0
22	155206863	C	G	intronic			6.704e-05	rs1026559493			1	0
23	155206981	C	T	intronic							1	0
24	155207030	G	A	intronic			0.0233	rs72704130			36	52
25	155207050	G	A	intronic				rs749925127			1	0
26	155207387	A	T	intronic			0.015	rs140335079	PD/GD/DLB	Benign	14	27
27	155207449	C	T	intronic				rs1006437355			2	0
28	155207550	G	T	intronic			0.0001				0	2
29	155207674	C	CAG	intronic							1	0
30	155207733	C	T	intronic			0.0396	rs28678003			82	79
31	155207846	T	C	intronic			0.0002	rs145066479			3	1
32	155207848	G	T	intronic			0.0043	rs183540501			4	2
33	155207866	A	T	intronic				rs529870563			1	2
34	155208167	G	A	intronic				rs566671462			0	2
35	155208495	G	A	intronic			0.0003	rs567935648			0	1
36	155208519	CT	C	intronic			6.673e-05				1	2
37	155208611	A	G	intronic			0.0003	rs569282073			2	1
38	155208624	A	G	intronic			0.0129	rs188328778			15	17
39	155208644	C	T	intronic			6.675e-05				0	1
40	155208647	T	C	intronic			0.7089	rs7416991			687	734
41	155208647	T	G	intronic			0.2911	rs7416991			316	349
42	155208647	T	G	intronic			0.2911	rs7416991			65	72
43	155208805	C	T	intronic			0.0				1	0
44	155208851	T	C	intronic			6.774e-05	rs149120852			0	1
45	155209078	C	T	intronic			0.0	rs1005434278			1	0
46	155209079	G	A	intronic			0.0001	rs114452199			3	2
47	155209082	A	G	intronic				rs899199374			1	2
48	155209251	G	A	intronic				rs991547343			1	0
49	155209297	C	T	intronic			0.0011	rs183903019			4	3
50	155209298	G	A	intronic			0.0002	rs776425625			0	2
51	155209594	G	A	intronic			0.0001	rs377315750			1	0
52	155209913	T	G	intronic			0.0005	rs199565854			0	2
53	155209938	G	A	intronic			0.0003	rs559516544			5	5
54	155209962	C	T	intronic			0.0115	rs114217696			14	12
55	155210030	G	A	intronic			0.0001	rs151028758			3	2
56	155210070	C	CA	intronic							0	1
57	155210146	C	T	intronic			0.0001				2	1
58	155210156	G	C	intronic							0	1
59	155210170	G	A	intronic			0.0002	rs962460364			0	1
60	155210570	T	C	intronic			0.001	rs2361534			0	2
61	155210613	C	T	intronic							1	1
62	155210641	A	C	intronic			0.0002	rs2070679			3	3
63	155210723	C	T	intronic							1	0
64	155210739	C	T	intronic							1	0
65	155210918	T	C	UTR5	NM_001005742	c.-15A>G	0.0013	rs41264927			3	2
66	155211027	C	T	UTR5	NM_000157	c.-124G>A					0	1
67	155211089	C	T	intronic							1	0
68	155211101	G	A	intronic			0.0	rs1007847984			0	1
69	155211106	T	C	intronic			0.0116	rs188978150		Uncertain_significance	30	27

Abbreviations: GD, Gaucher's disease; PD, Parkinson's disease; DLB, Dementia with Lewy Bodies.; gnomAD, The Genome Aggregation Database. DM, Disease causing mutation; FP, in vitro or in vivo functional polymorphism.

Supplementary Table 6.

Type of data	Clinical characteristics and scales	PD		missing values (%)	β (95%)	*p*-value	adj *p*-value
		GBA1 carrier					
		Yes (n=67)	No (N=554)				
Motor symptoms/scales	H&Y, mean (SD)	2.2 (±0.8)	2.2 (±0.8)	3 (0.5%)	0.1 (-0.07 to 0.27)	0.2700	0.9635
	MDS-UPDRS II, mean (SD)	11.4 (±7.3)	11.4 (±8.3)	13 (2.1%)	0.34 (-1.53 to 2.21)	0.7219	0.9701
	MDS-UPDRS III, mean (SD)	33.1 (±16.5)	34.6 (±16.2)	15 (2.4%)	-0.69 (-4.61 to 3.22)	0.7294	0.9701
	MDS-UPDRS IV, mean (SD)	1.6 (±3.2)	1.6 (±3.3)	7 (1.1%)	-0.03 (-0.79 to 0.72)	0.9296	0.9701
	Dyskinesias, n (%)	9 (13.4%)	64 (11.6%)	0	0.23 (-0.58 to 1.05)	0.5747	0.9701
	Falls, n (%)	12 (17.9%)	93 (16.8%)	0	0.19 (-0.51 to 0.9)	0.5950	0.9701
	Gait Disorder, n (%)	37 (55.2%)	307 (55.4%)	0	0.07 (-0.46 to 0.59)	0.8063	0.9701
	FOG, n (%)	15 (22.4%)	123 (22.2%)	0	0.07 (-0.6 to 0.74)	0.8402	0.9701
	Restless leg syndrome, n (%)	10 (14.9%)	46 (8.3%)	0	0.63 (-0.11 to 1.37)	0.0969	0.8721
	Motor fluctuation, n (%)	10 (14.9%)	93 (16.8%)	0	-0.17 (-0.95 to 0.6)	0.6628	0.9701
Non-motor symptoms/scales	BDI, mean (SD)	9.4 (±5.8)	9.9 (±7.1)	30 (4.8%)	-0.52 (-2.27 to 1.23)	0.5614	0.9701
	MDS-UPDRS Part I, mean (SD)	11.1 (±6.8)	10.5 (±7.0)	15 (2.4%)	0.65 (-1.04 to 2.33)	0.4550	0.9635
	PDSS, mean (SD)	103.0 (±22.8)	105.0 (±24.6)	46 (7.4%)	-2.06 (-8.23 to 4.11)	0.5126	0.9701
	SCOPA-AUT, mean (SD)	14.9 (±7.8)	14.9 (±8.1)	33 (5.3%)	0.15 (-1.82 to 2.12)	0.8789	0.9701
	Sniffin's stick test, mean (SD)	**7.0 (±3.7)**	**7.8 (±3.6)**	**8 (1.3%)**	**-1.03 (-1.91 to -0.15)**	**0.0210***	**0.7470**
	SAS, mean (SD)	13.9 (±5.8)	14.1 (±5.7)	35 (5.6%)	0.03 (-1.42 to 1.49)	0.9649	0.9701
	MoCA, mean (SD)	24.7 (±4.2)	24.4 (±4.5)	14 (2.3%)	-0.09 (-1.14 to 0.96)	0.8686	0.9701
	Constipation, n (%)	29 (43.3%)	246 (44.4%)	0	-0.01 (-0.54 to 0.51)	0.9577	0.9701
	Dysphagia, n (%)	15 (22.4%)	145 (26.2%)	0	-0.2 (-0.81 to 0.41)	0.5282	0.9701
	Insomnia, n (%)	17 (25.4%)	151 (27.3%)	0	-0.13 (-0.72 to 0.45)	0.6549	0.9701
	Orthostatism, n (%)	24 (35.8%)	163 (29.4%)	0	0.32 (-0.22 to 0.85)	0.249	0.9635
	Urinary incontinence, n (%)	25 (37.3%)	168 (30.3%)	0	0.38 (-0.16 to 0.92)	0.1718	0.8835
	Hallucinations, n (%)	**16 (23.9%)**	**83 (15.0%)**	**0**	**0.65 (0.03 to 1.28)**	**0.0415***	**0.7470**
	Excessive daytime sleepiness, n (%)	23 (34.3%)	170 (30.7%)	0	0.22 (-0.33 to 0.77)	0.4302	0.9635
	ICD, n (%)	6 (9.0%)	53 (9.6%)	0	-0.05 (-0.96 to 0.86)	0.9154	0.9701
	Syncope, n (%)	6 (9.0%)	26 (4.7%)	0	0.81 (-0.14 to 1.77)	0.0933	0.8721
	RBDSQ, n (%)	25 (37.3%)	165 (29.8%)	45 (7.2%)	0.43 (-0.14 to 1.0)	0.1401	0.8835
Other clinical outcomes	LEDD (mg/day), mean (SD)	543.4 (±442.6)	513.5 (±404.9)	21 (3.4%)	35.55 (-54.38 to 125.49)	0.4385	0.9635
	PDQ-39, mean (SD)	39.5 (±27.3)	39.3 (±26.7)	51 (8.2%)	-0.12 (-6.49 to 6.24)	0.9701	0.9701
	DBS, n (%)	4 (6.0%)	24 (4.3%)	0	0.63 (-0.64 to 1.9)	0.3285	0.9635
Comorbidities	Diabetes, n (%)	8 (11.9%)	55 (9.9%)	0	0.34 (-0.47 to 1.15)	0.4080	0.9635
	Hypercholesterolemia, n (%)	25 (37.3%)	226 (40.8%)	0	-0.11 (-0.64 to 0.43)	0.6959	0.9701
	Cardiovascular disease, n (%)	10 (14.9%)	116 (20.9%)	0	-0.29 (-1.02 to 0.44)	0.4303	0.9635
	Arterial hypertension, n (%)	23 (34.3%)	248 (44.8%)	0	-0.38 (-0.93 to 0.16)	0.1697	0.8835
	Traumatic Brain Injury, n (%)	11 (16.4%)	122 (22.0%)	0	-0.36 (-1.04 to 0.32)	0.2995	0.9635

We consider severe, mild, and risk *GBA1* variants as pathogenic mutations. We used regression models (linear and logistic). Data are given as mean and standard deviation (SD) for continuous clinical outcomes and as percentage for binary clinical outcomes. Models adjusted for sex, age at assessment, and disease duration. Beta (β) regression coefficient is given with the 95% CI. Statistically significant results highlighted in bold with (*) sign and red (p-value < 0.05). Abbreviation : p-value, unadjusted p-value; adj p-value, corrected for multiple comparisons using FDR adjustment; AAO, age at onset; H&Y, Hoehn & Yahr; MDS-UPDRS, Movement Disorders Society - Unified Parkinson's Disease Rating Scale; FOG, freezing of gait; BDI, Beck Depression Inventory; PDSS, Parkinson's Disease Sleep Scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire; SAS, Starkstein apathy scale; MoCA, Montreal Cognitive Assessment; ICD, impulse control disorder; RBDSQ, REM Sleep Behavior Disorder Screening Questionnaire; LEDD, L-dopa equivalent daily dose (mg/day); PDQ-39, Parkinson's Disease quality of life Questionnaire; DBS, Presence of treatment by Deep Brain Stimulation.

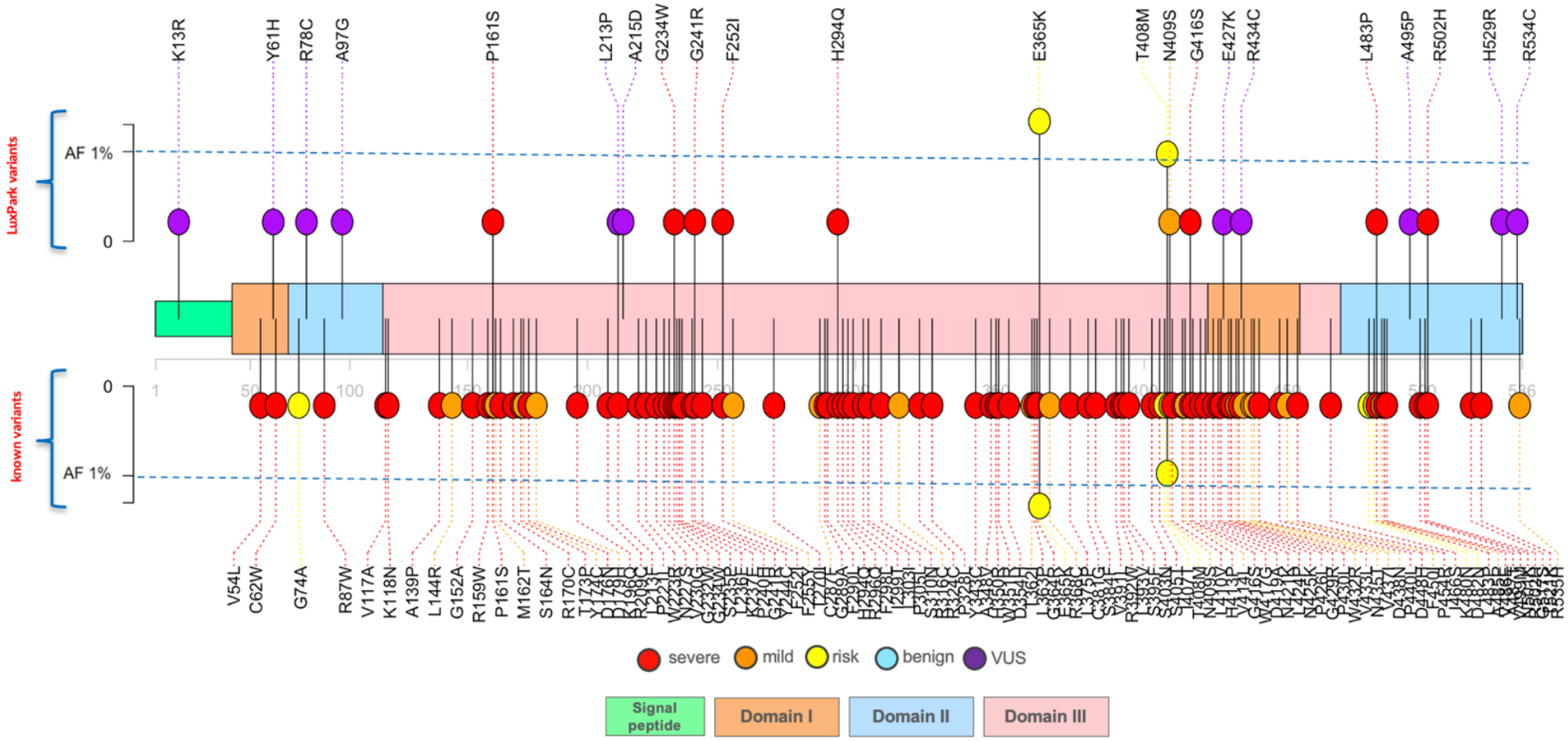
Supplementary Table 7. Comparison of clinical scores between known severe variant carriers and four probably severe VUS carriers

Type of data	Clinical characteristics and scales	Severe PD_{GBA1} carrier mean (SD) n=21	Probably severe VUS GBA carriers score (Z-score)			
			p.Y61H	p.L213P	p.A215D	p.R434C
	AAO	58.6 (±13.1)	**38 (-1.5)***	56 (-0.2)	68 (0.7)	62 (0.26)
	DD	6.4 (±4.7)	5 (-0.3)	**17 (2.2)***	**1 (-1.1)***	11 (0.9)
Motor symptoms/scales	H&Y	2.4 (±0.8)	2 (-0.5)	**4 (2)***	2 (-0.5)	3 (0.75)
	MDS-UPDRS II	12.6 (±4.4)	14 (0.32)	**26 (3.05)***	**6 (-1.5)***	**17 (1)***
	MDS-UPDRS III	34.8 (±15.7)	35 (0.01)	**81 (2.94)***	**14 (-1.32)***	44 (0.59)
	MDS-UPDRS IV	3.0 (±4.5)	**10 (1.56)***	**9 (1.33)***	1 (-0.44)	0 (-0.67)
Non-motor symptoms/scales	BDI	12.4 (±5.7)	9 (-0.6)	-	7 (-0.95)	16 (0.63)
	MDS-UPDRS Part I	15.0 (±6.5)	9 (-0.92)	**26 (1.69)***	17 (0.31)	**6 (-1.38)***
	PDSS	98.3 (±20.9)	95 (-0.16)	-	82 (-0.78)	**65 (-1.59)***
	SCOPA-AUT	17.1 (±8.0)	**8 (-1.14)***	-	12 (-0.64)	**27 (1.24)***
	Sniffin's stick test	6.4 (±3.6)	6 (-0.11)	7 (0.17)	8 (0.44)	3 (-0.94)
	SAS	15.8 (±5.2)	14 (-0.35)	-	17 (0.23)	19 (0.62)
	MoCA	24.0 (±4.7)	26 (0.43)	**18 (-1.28)***	20 (-0.85)	**18 (-1.28)***
Other clinical outcomes	LEDD (mg/day)	690.5 (±457.9)	**1378 (1.5)***	675 (-0.03)	675 (-0.03)	975 (0.62)
	PDQ-39	52.0 (±26.3)	42 (-0.38)	-	**14 (-1.44)***	48 (-0.15)

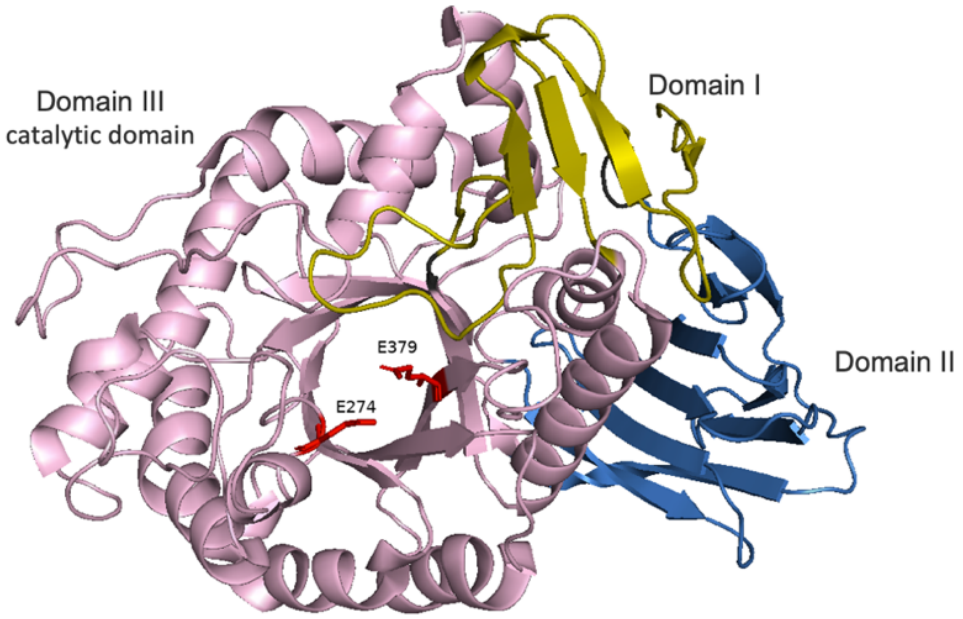
We compared the clinical scores obtained from known severe variants carriers with the four carriers of probably severe VUS (p.Y61H, p.L213P, p.A215D, and p.R434C). To calculate the z-score, we use the formula, $z=(X - \mu)/\sigma$; X is the clinical outcome value, μ is the mean and σ is the standard deviation. The z-score indicates the number of standard deviations from the mean. z-score deviated from the mean with more or less than 1 SD were highlighted in bold with (*) sign and red. Abbreviation: AAO, age at onset; H&Y, Hoehn & Yahr; MDS-UPDRS, Movement Disorders Society - Unified Parkinson's Disease Rating Scale; BDI, Beck Depression Inventory; PDSS, Parkinson's Disease Sleep Scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire; SAS, Starkstein apathy scale; MoCA, Montreal Cognitive Assessment; PDQ-39, Parkinson's Disease quality of life Questionnaire.

Supplementary Table 8. Frequency of *GBA1* variant in European Parkinson's disease population that performed full *GBA1* gene sequencing.

Population	PD (n)	Screening method	*GBA1* carrier frequency (%)	E365K	T408M	L483P	N409S	RecNcil	Other
Ashkenazi Jews[10]	735	Targeted NGS	18	1.6	0	0.3	11.8	0	4.2
Netherland[11]	3402	Long-range PCR	15	6.7	2.5	0.6	0.9	0	4.3
Italy[12]	874	Complete exon Sanger sequencing	14.3	1.7	0.6	2.3	3.3	0.8	5.3
Luxembourg (this cohort)	**644**	**PacBio**	**12**	**3.6**	**2.6**	**1.7**	**1.1**	**0.6**	**3**
Southern Spain[13]	532	High-resolution melting analysis (HRM)	11.7	3	0.9	2.4	0.9	0	4.3
New Zealand[14]	229	PCR amplicon + nanopore	9.2	4.8	3.1	0	0.4	0	1.3
Ireland[15]	314	Complete exon Sanger sequencing	8.3	4.1	1.9	0	0.9	0.9	0.5
Portugal[16]	230	X	8.3	0.9	0.9	1.3	2.2	0	0.9
Greece[17]	172	X	6.4	0.6	0	1.2	0	0	4.6



Supplementary Figure 1: Lolliplot graphics showing all *GBA1* variants found in the Luxembourg Parkinson's study through the whole protein sequences. Each lollipop represents a *GBA1* variant identified in this the Luxembourgish cohort (upper part) and known *GBA1* variants in ClinVar or literatures (bottom part). Variants classified as severe are colored red, mild are colored orange, risk in yellow and VUS are colored purple.



Supplementary Figure 2. The 3D structure of GCase (PDB code 1ogs) created using PYMOL (http://www.pymol.org). Domain I is shown in dark yellow with the antiparallel β sheet (residues 1–27 and 383–414), Domain II in blue with the immunoglobulin-like domain (residues 30–75 and 431–497), and Domain III in pink is the catalytic domain with the (b/a)8 (TIM) barrel structure (residues 76–381 and 416–430). The active site residues Glu274 and Glu379 are shown in red.

References to supplementary material

1. Nomura, T., Inoue, Y., Kagimura, T., Uemura, Y. & Nakashima, K. Utility of the REM sleep behavior disorder screening questionnaire (RBDSQ) in Parkinson's disease patients. *Sleep Med.* **12**, 711–713 (2011).
2. Chaudhuri, K. R. *et al.* The Parkinson's disease sleep scale: a new instrument for assessing sleep and nocturnal disability in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **73**, 629–635 (2002).
3. Tomlinson, C. L. *et al.* Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* **25**, 2649–2653 (2010).
4. Longstreth, G. F. *et al.* Functional bowel disorders. *Gastroenterology* **130**, 1480–1491 (2006).
5. Goetz, C. G. *et al.* Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov. Disord. Off. J. Mov. Disord. Soc.* **19**, 1020–1028 (2004).
6. Peto, V., Jenkinson, C., Fitzpatrick, R. & Greenhall, R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.* **4**, 241–248 (1995).
7. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch. Gen. Psychiatry* **4**, 561–571 (1961).
8. Hummel, T., Kobal, G., Gudziol, H. & Mackay-Sim, A. Normative data for the 'Sniffin' Sticks' including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. *Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil. Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg.* **264**, 237–243 (2007).
9. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
10. Ruskey, J. A. *et al.* Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. *Eur. J. Med. Genet.* **62**, 65–69 (2019).
11. den Heijer, J. M. *et al.* A Large-Scale Full GBA1 Gene Screening in Parkinson's Disease in the Netherlands. *Mov. Disord. Off. J. Mov. Disord. Soc.* **35**, 1667–1674 (2020).
12. Petrucci, S. *et al.* GBA-Related Parkinson's Disease: Dissection of Genotype-Phenotype Correlates in a Large Italian Cohort. *Mov. Disord. Off. J. Mov. Disord. Soc.* **35**, 2106–2111 (2020).
13. Jesús, S. *et al.* GBA Variants Influence Motor and Non-Motor Features of Parkinson's Disease. *PloS One* **11**, e0167749 (2016).
14. Graham, O. E. E. *et al.* Nanopore sequencing of the glucocerebrosidase (GBA) gene in a New Zealand Parkinson's disease cohort. *Parkinsonism Relat. Disord.* **70**, 36–41 (2020).
15. Olszewska, D. A. *et al.* Association Between Glucocerebrosidase Mutations and Parkinson's Disease in Ireland. *Front. Neurol.* **11**, 527 (2020).
16. Bras, J. *et al.* Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal. *Neurobiol. Aging* **30**, 1515–1517 (2009).
17. Kalinderi, K. *et al.* Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Greece. *Neurosci. Lett.* **452**, 87–89 (2009).

Chapter 2 : Genetic landscape of Parkinson's disease in Luxembourg

Genetic landscape of Parkinson's disease in Luxembourg

Zied Landoulsi [1†]; Sinthuja Pachchek [1†]; Dheeraj Reddy Bobbili [1], Lukas Pavelka[2,3], Patrick May[1*], and Rejko Krüger [1,2,3,*] on behalf of the NCER-PD Consortium

[1]LCSB, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Sur-Alzette, Luxembourg.
[2]Parkinson Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg.
[3]Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg.

*** Correspondence:**
Patrick May: patrick.may@uni.lu
Rejko Krüger: rejko.krueger@lih.lu

† These authors contributed equally to this work and shared the first authorship

Project description

For the first time, we investigated the genetic aspects of PD in a substantial case-control cohort of Parkinsonism patients from Luxembourg. Our primary objective was to create a comprehensive genetic profile of the cohort by examining rare SNVs and CNVs, and the impact of SNPs using Polygenic risk scores analysis (PRS).

Contributions

My contributions to this research were considerable and encompassed multiple critical aspects. Firstly, as an equal author of this manuscript, I actively contributed to the article's writing process, which clearly explained our results. I also led the efforts to find and evaluate rare variants in the dataset by conducting an in-depth study of the genomic data. In addition, I contributed to the analysis of the results, offering insightful comments that contributed to a better understanding of the significance of our findings.

ZL wrote the paper, conducted the analysis for PRS and CNVs, executed the statistical analysis and interpretation of results. SP contributed to the writing of the paper, performed rare variants screening analysis, and participated in the interpretation of results. D.R.B and L.P substantially participated in data collection, data exportation and data curation and critically revised the manuscript. R.K and P.M.: Conceived, organised, and co-executed the research project; participated in the interpretation of results; critically revised the manuscript. All authors contributed to the article and approved the submitted version.

The study was supervised by Prof. Dr. Rejko Küger and Dr. Patrick May.

2.1. Introduction

Over the last two decades, our understanding of PD genetics has seen remarkable advancements, representing a breakthrough in PD research. The identification of mutations in the genes such as *PARKN*, *DJ1/PARK7*, *PINK1*, *LRRK2*, *VPS35* and *SNCA* has been the major success in understanding and increasing the knowledge of the genetic monogenic causes of PD (Billingsley et al., 2018). Although PD has been associated with mutations in 15 genes responsible for monogenic forms of PD, these monogenic forms account for only approximately 30% of all PD cases (Billingsley et al., 2018).

The number of mutation carriers has increased considerably in recent years, mainly due to technological advancements, particularly next-generation sequencing methods, which make genetic testing more accessible.

PD patients are commonly stratified based on their age of onset (early or late onset, with a cut-off of 50 years of age), family history (familial or sporadic PD), and the presence of pathogenic variants (monogenic or idiopathic PD). These stratification factors, combined with identified genetic causes, facilitate a more personalized and targeted approach to understanding and managing the disease.

A combined effect of more than 90 loci with low-risk factors may affect a disease's susceptibility. These loci, which include genes like *SNCA* and *LRRK2*, have been discovered by GWAs (Blauwendraat et al., 2020a; Nalls et al., 2019b). While CNVs have received comparatively less focus than the more prevalent SNVs, their presence has also been detected. In genes associated with PD, deletions and duplications were identified using a targeting method for candidate genes (La Cognata et al., 2017; Pankratz et al., 2011; Toft and Ross, 2010), such as *PRKN*, *SNCA*, *PINK1*, *PARK7* and *ATP13A2*, or genome-wide burden analysis (Liu et al., 2013; Sarihan et al., 2021).

Even with ongoing worldwide efforts in genetic analysis, further development is required in the areas of early diagnosis and prognosis, causative treatments and innovative therapeutic approaches. Therefore, it is important to produce reliable evidence on the genetic etiology and epidemiology of PD.

In this study, we carried out a comprehensive and detailed genetic analysis of participants in the Luxembourg Parkinson's study, who underwent in-depth phenotyping. Our investigation focused on the identification of rare SNVs and CNVs and the effects of common SNPs using PRS.

2.2. Copy of the preprint manuscript

Genetic landscape of Parkinson's disease in Luxembourg

Zied Landoulsi [1†]; Sinthuja Pachchek [1†]; Dheeraj Reddy Bobbili [1], Lukas Pavelka[2,3], Patrick May[1*], and Rejko Krüger [1,2,3,*] on behalf of the NCER-PD Consortium

[1]LCSB, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Sur-Alzette, Luxembourg.
[2]Parkinson Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg.
[3]Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg.

*** Correspondence:**
Patrick May: patrick.may@uni.lu

Rejko Krüger: rejko.krueger@lih.lu

†These authors contributed equally to this work and share first authorship

Keywords: Parkinson's disease, genetics, Luxembourg, polygenic risk score, copy number variants.

ABSTRACT

Objectives: To explore the genetic architecture of PD in the Luxembourg Parkinson's Study including cohorts of healthy people and patients with Parkinson's disease (PD) and atypical parkinsonism (AP).

Methods: 809 healthy controls, 680 PD and 103 AP were genotyped using the Neurochip array. We screened and validated rare single nucleotide variants (SNVs) and copy number variants (CNVs) within seven PD-causing genes (*LRRK2*, *SNCA*, *VPS35*, *PRKN*, *PARK7*, *PINK1* and *ATP13A2*)). Polygenic risk scores (PRSs) were generated using the latest genome-wide association study for PD. We then estimated the role of common variants in PD risk by applying gene-set-specific PRSs.

Results: We identified 60 rare SNVs in seven PD-causing genes, nine of which were pathogenic in *LRRK2*, *PINK1* and *PRKN*. Eleven rare CNVs were detected in *PRKN* including seven duplications and four deletions. The majority of *PRKN* SNVs and CNVs carriers were heterozygous and not differentially distributed between cases and controls. We identified new digenetic variants in genes associated with PD. The PRSs were significantly associated with PD and identified specific molecular pathways related to protein metabolism and signal transduction as drivers of PD risk.

Conclusion: We performed a comprehensive genetic characterization of the deep-phenotyped individuals of the Luxembourgish Parkinson's Study. Heterozygous SNVs and CNVs in *PRKN* were not associated with higher PD risk. In particular, we reported novel digenic variants in PD related genes and rare *LRRK2* SNVs in AP patients. Our findings will help future studies to unravel the genetic complexity of PD.

INTRODUCTION

Parkinson's disease (PD) is the fastest growing neurodegenerative disorder, affecting more than 8.5 million people (Dorsey and Bloem, 2018). The main pathological hallmarks of PD include loss of dopaminergic neurons in the substantia nigra and the presence of intraneural Lewy bodies, with motor and non-motor symptoms (Bloem et al., 2021). The etiology of sporadic PD is complex and influenced by both environmental and genetic factors. Familial monogenic forms defined by rare and pathogenic variants in autosomal dominant (e.g., *SNCA*, *LRRK2*, *VPS35*) or recessive (*PRKN*, *PINK1*, *PARK7*) PD-related genes, account for less than 10% of Mendelian cases (Lesage and Brice, 2009). The contribution of genetics in the remaining patients with sporadic forms of PD is not yet well defined. Moreover, the presence of heterozygous variants in the *GBA1* gene has emerged as a common risk factor for PD, estimated to occur in about 4-12% of PD patients (Peiris et al., 2023). Disease susceptibility may be influenced by a combined effect of more than 90 common low-risk genetic loci defined by large genome-wide association studies (Blauwendraat et al., 2020; Nalls et al., 2019), including those in the *SNCA* and *LRRK2* genes. Although less explored than common single-nucleotide variants (SNVs), copy number variants (CNVs) have been reported, especially in PD-associated genes where pathogenic deletions and duplications have been identified using either a gene candidate approach (*PRKN*, *SNCA*, *PINK1*, *PARK7* and *ATP13A2)* (La Cognata et al., 2017; Pankratz et al., 2011; Toft and Ross, 2010) or genome-wide burden analysis (Liu et al., 2013; Sarihan et al., 2021).

Despite ongoing global scientific efforts in genetic analysis, improvements are still needed in terms of early diagnosis and prognosis, causative treatments and new therapeutic approaches. As the population ages, the number of PD patients will increase dramatically. It is therefore important to generate reliable evidence on the epidemiology and genetic etiology of PD to enable precision medicine and prevention for neurodegeneration in PD. In particular, three genetic discoveries that have led to new therapeutic approaches (targeting α-synuclein, glucocerebrosidase and *LRRK2* pathway) are now in clinical development (Sardi et al., 2018).

We had previously performed a comprehensive screening of *GBA1* gene variants (Peiris et al., 2023).

Here, we sought to genetically characterize patients with PD or atypical parkinsonism (AP) in the Luxembourg Parkinson's Study screening for rare SNVs, CNVs, and estimated the effect of common SNVs using polygenic risk scores (PRS).

MATERIALS AND METHODS

Cohort characteristics

A total of 1592 individuals (783 cases and 809 neurologically healthy controls) were recruited from March 2015 to December 2022 as part of the Luxembourg Parkinson's Study, a large longitudinal monocentric study within the framework of the NCER-PD (National Centre for Excellence in Research in PD (Hipp et al., 2018; Pavelka et al., 2022)). The most up-to-date diagnostic status of the participants was used at the time of export (July 2023). Assignment of diagnosis was based on the following diagnostic criteria: for PD UKPDSBB (Litvan et al., 2003); for progressive supranuclear palsy (PSP) Institute of Neurological Disorders and Stroke/Society criteria (Litvan et al., 1996); for frontotemporal dementia with parkinsonism (FTD-P) (Neary et al., 1998); for multiple system atrophy (MSA) (Gilman et al., 2008); for dementia with Lewy bodies (DLB) (McKeith et al., 2005). Of these individuals, 680 fulfilled the criteria for PD and 103 for AP (52 for PSP, 26 for DLB, 14 for MSA, 10 for corticobasal syndrome (CBS), and one for FTD-P). All subjects gave written informed consent. The study was approved by the National Research Ethics Committee (CNER Ref: 201407/13).

Genotyping and quality controls

Samples were genotyped with the customized NeuroChip array, which was designed to contain tagging rare and common variants associated with neurodegenerative diseases (v.1.0 and v1.1; Illumina, San Diego, CA, (Blauwendraat et al., 2017)). These disease-targeted variants include loci from the largest completed meta-analysis of PD cases and controls, which identified many of the known PD mutations and additional rare high-risk variants. Using PLINK v1.9 (Chang et al., 2015), we performed two rounds of quality control (QC). The first round included the following steps: samples with call rates < 95%, missingness rates > 5%, Hardy-Weinberg equilibrium P-value < 1e-6 and whose genetically determined sex deviated from the sex reported in clinical data were excluded from the analysis. We also removed samples exhibiting excess heterozygosity (F statistic > 0.2). After these steps, the remaining samples were used for rare variant screening and validation process. Next, we performed a second round of QC steps where the filtered variants were checked for relatedness (using KING (Manichaikul et al., 2010)) and samples with first-degree relatedness were excluded. To determine the genetic ancestry, we calculated the first ten principal components (PCs) using PLINK and merged our data with the 1000 genomes dataset. We selected only samples of European ancestrys, excluding those with a value > 3SD based on the first and the second PCs. Samples passing QC were then imputed using the Haplotype Reference Consortium r1.1 2016 on a local instance of the Michigan Imputation Server (Das et al., 2016) and filtered for imputation quality (R2 > 0.3).

Variant annotation and rare variant screening

We annotated the variants with ANNOVAR (v 2020-06-08, (Wang et al., 2010)). We searched for variants within a list of "PD genes" including the major causal genes according to MDS gene classification (https://www.mdsgene.org) (1. Dominant forms of classical parkinsonism: *LRRK2*, *SNCA*, *VPS35*; 2. Recessive forms of early-onset parkinsonism: *PRKN*, *PARK7*, and *PINK1*; 3. Atypical parkinsonism: *ATP13A2*). We then selected rare nonsynonymous and splicing (+/-2bp) rare variants based on the minor alleles frequency (MAF) <1% in the Genome Aggregation Database (gnomAD r2.1 (Karczewski et al., 2020)) exomes and genomes for the non-Finnish European (NFE) population. We performed Sanger sequencing to confirm all rare variants within these PD genes. The pathogenicity of the validated rare variants was assigned based on ClinVar (Landrum et al., 2014), the MDSgenes pathogenicity score (https://www.mdsgene.org), the Combined Annotation Dependent Depletion (CADD (Rentzsch et al., 2019)) and the Rare Exome Variant Ensemble Learner REVEL scores (Ioannidis et al., 2016). CADD provides ranking scores that predict the deleteriousness of variants, considering conservation and functional information, and variants with scores equal to or greater than a CADD score of 20 are in the 1% most deleterious. REVEL is an ensemble method for predicting the pathogenicity of missense variants by integrating multiple scores. Scores range from 0 to 1 and variants with higher scores are more likely to be pathogenic. Scores greater than 0.5 are predicted to be 'likely disease causing', as 75.4% of known disease mutations but only 10.9% of neutral variants have a score greater than 0.5 (Ioannidis et al., 2016).

Copy number variant calling

We generated a custom population B-allele frequency (BAF) and GC wave-adjusted log R ratio (LRR) intensity file using GenomeStudio (v2.0.5 Illumina) for all the samples that passed genotyping QC steps and used PennCNV (v1.0.5, (Wang et al., 2007)) to detect CNVs. Only autosomal CNV were targeted for CNV calling, as calls from sex chromosomes are often of poor quality. Adjacent CNV calls were merged into one single call if the number of overlapping markers between them was less than 20% of the total number when the two segments were combined. We conducted an intensity-based QC to exclude samples with low-quality data. After intensity-based QC, all samples had an LRR standard deviation < 0.25, an absolute value of the waviness factor < 0.05 and a BAF drift < 0.01. Spurious CNV calls in known problematic genomic regions (provided by PennCNV) were also removed prior the analysis. We excluded additional samples with a total number of CNVs calls greater than 80 (this threshold corresponds to the median + 3 SDs of the total number of CNVs per sample). Called CNVs were removed from the dataset if they spanned < 20 SNPs, were < 20 kilobases (kb) in length and had a SNP density < 0.0001 (number of markers/length of CNVs). Additionally, SNV density was not considered for CNVs spanning ≥ 20 SNPs and ≥1 Mb in length. CNVs were then annotated for gene content using refGene including gene name and the corresponding exonic coordinates in the hg19 assembly using ANNOVAR (v 2020-06-08). We then

searched for CNVs in the same list of "PD genes" used to screen for rare SNVs. We assessed the frequency of CNVs based on complete overlap with CNVs of the same copy number reported in gnomAD-SV (Collins et al., 2020) and in the Database of Genomic Variants (DGV (MacDonald et al., 2014)). We evaluated the clinical impact of the detected CNVs using the CNV-ClinViewer (Macnee et al., 2023), which integrates clinical interpretation of CNVs according to the ACMG guideline and the ClassifyCNV scores. Selected CNVs were validated using the multiplex ligation-dependent probe amplification (MLPA) assay.

Polygenic risk score calculation

PRSs were calculated for healthy controls and PD cases using the R package PRSice2 (Choi and O'Reilly, 2019) with default parameters. PRSs for each sample were generated using the imputed genotyping data from the Luxembourg Parkinson's Study and the summary statistics of 90 genomewide significant SNVs that were previously reported to be associated with PD risk in the largest PD genome-wide association study (GWAS) statistics to date (Nalls et al., 2019). PRSice2 implements the clumping and thresholding method. The criteria for linkage disequilibrium (LD) clumping of SNPs were pairwise LD r2 < 0.1 within a 250 kb window. PRSs were computed at different GWAS P-value thresholds (from 5e-08 to 5e-01). PRSice2 identified the best P-value threshold for selecting variants that explained the maximum variance in the target sample. The predictive accuracy of the PRS model was determined by two methods: by the observed phenotypic variance (PRS model fit, R2) calculated by PRSice2 and by the area under the receiver operating curve (AUC, pROC R package). The phenotypic variance R2 was adjusted for a PD prevalence of 0.005 (Bandres-Ciga et al., 2020). The PRS distributions between healthy controls and PD cases were compared using the Wilcoxon ranksum test. PRSs for curated gene-sets were generated using the *msigdb* function implemented in PRSice2, based on a collection of 3090 canonical pathways from the molecular signature database (MSigDB, https://www.gsea-msigdb.org/, "c2.cp.v2023.1.Hs.symbols.gmt") with an MAF threshold of 0.01. The summary statistics of PD GWAS from Nalls et al.2019 (excluding 23&me data) were used as the base dataset. The mapping file "Homo_sapiens.GRCh37.87.gtf" was used as the universal background for gene-set analysis. Resulting gene-sets with a p-value less than 0.05, corrected for Bonferroni multiple testing, were considered significant. In order to understand which biological processes were associated with PD after excluding known risk factors, we performed the same analysis after removing the 90 PD GWAS hits (Nalls et al., 2019) and additional SNVs that were located 1 Mb upstream and downstream. We used a logistic regression model to calculate the odds ratio (OR) to assess whether PRS could predict PD risk. Age, sex and the first five PCs were included as covariates.

RESULTS

Cohort description

After the quality control procedure, the final dataset for the Luxembourg Parkinson's study comprised 1490 individuals (667 PD cases, 99 atypical PD cases and 724 healthy controls). Detailed demographic data are summarized in Table 1. The control group had a mean age at assessment of 65.8 ± 11.6 years. The PD patients had a mean age of onset (AAO) of 62.3 ± 11.8 years. To illustrate the ethnic composition of our cohort, we performed PCA using 1000 Genome populations as a reference (The 1000 Genomes Project Consortium, 2012) and showed that all our samples clustered strongly with the European ancestry (supplementary figure 1).

Rare variants in PD-related genes

We screened for rare (gnomAD NFE MAF < 1%) exonic and splice region variants in seven PD causal genes and validated these findings by Sanger sequencing. We identified 60 rare variants (59 missense and one frameshift) in all PD-related genes except for *SNCA* (Table 2), in 119 individuals including 52 controls, 57 PD and 10 AP patients, representing 7.9% of the total cohort (Table 3). All carriers were heterozygous, except two PD patients that were homozygous for *LRRK2* p.I723V and *PINK1* p.L369P, respectively.

Among the 29 rare variants identified in *LRRK2,* five variants have a CADD score > 20 and REVEL score > 0.5 (p.R1325Q, p.R1441S, p.R1441C, p.M1869T, and p.G2019S) showing high evidence for pathogenicity. Three of these variants were classified as pathogenic for PD in ClinVar (Table 2) and were present in nine individuals representing 0.60% of the total cohort (Table 3). Among these variants, five PD patients carried the extensively studied pathogenic variant p.G2019S while two PD patients carried the pathogenic p.R1441C and p.R1441S variant.

Two control individuals with family history of PD (Table 3) had rare *LRRK2* variants. One control individual carried the variant p.G2019S (38 years old) and has a high probability of developing PD (Healy et al., 2008). Another control individual (77 years old) carried the p.R1441C, although this variant is described as highly penetrant (more than 90% of carriers had PD by the age of 75 (Haugarvoll and Wszolek, 2009)).

In the autosomal recessive PD-causing genes (*PRKN*, *PARK7*, *PINK1*, and *ATP13A2*), we identified 28 heterozygous rare variant carriers and only one homozygous rare variant carrier (*PINK1* p.L369P, Table 2). The distribution of these variants was similar between cases and controls (27 PD, six AP and 29 controls, p-value = 0.39, Table 3). Four controls and ten patients had a first-degree family-history of PD. The age of the control individuals carrying these heterozygous variants ranged from 52 to 85 years (mean = 67.8 years). The AAO of the PD patients carrying these heterozygous variants ranged from 39 to 87 (mean = 65.5). One PD patients was younger than 40 years (carrying *PINK1* p.A383T), all the others were older than 50 years.

According to ClinVar, two pathogenic *PRKN* variants (p.R275W and p.Q34fs) were found in three PD patients, one PSP patient and three controls (all heterozygous, representing 0.46% of the total cohort, Table 3). Three *PRKN* variants (p.M192L, p.R256C and p.R275W) were predicted to be likely pathogenic with CADD and REVEL scores above the selected threshold. However, we noted the occurence of heterozygous p.R256C in three controls (age 80, 81 and 85years) and one PD patient (AAO = 52 years), which is classified as 'probably pathogenic' according to the MDSgenes pathogenicity score. For *PINK1* we found no pathogenic variant classified in ClinVar. However, p.R279H, p.A339T and p.L369P are 'probably pathogenic' according to the MDSgene pathogenicity scores, but only when homozygous. Two of these variants (p.R279H, p.L369P) and p.M318L were classified as 'likely pathogenic' based on CADD and REVEL scores. In addition, the *PARK7* p.A104S and *ATP13A2* p.R172H, p.S277C, p.P358L, p.R924H and p.R980H heterozygous variants had higher CADD and REVEL scores but were not reported to be pathogenic for PD in ClinVar or MDSgene. Overall, we described nine pathogenic variants from databases of clinical interest in *LRRK2, PRKN* and *PINK1* in a total of 26 samples (13 PD, 1 PSP and 12 controls, all heterozygous) representing 1.7% of the total cohort (Table 3). Given the zygosity of the variants, only variants in *LRRK2* can be responsible for the disease. AP patients were heterozygous carriers of probably benign variants in *LRRK2* and *PINK1* and a pathogenic variant in *PRKN*.

An extensive screening of *GBA1* variants was previously performed by our team (Peiris et al., 2023) using *GBA1*-targeted PacBio sequencing in individuals from the Luxembourg Parkinson's study (660 PD patients, 100 patients with other forms of parkinsonism and 808 controls). We identified 21 rare *GBA1* variants (20 missense and one splice site) in 37 PD patients and 16 controls (representing 5.6% of PD patients and 1.9% controls), which were validated by Sanger sequencing. Eleven rare variants were classified as pathogenic while the others were classified as variants of unknown significance (VUS). For the samples that were both genotyped and screened by targeted *GBA1*-sequencing, we found that none of carriers of rare variants in the studied PD-causing genes, identified within the NeuroChip, harboured an additional pathogenic *GBA1* variant.

Rare copy number variants in PD-related genes

We initially detected 25299 CNVs, including 13862 duplications and 11437 deletions in 728 controls and 757 PD cases. After all QC and filtering steps, the final number of CNVs was 1079 CNVs, including 737 duplications and 342 deletions in 373 controls and 366 cases. CNV analysis showed that almost half of the samples (49.7%) carried at least one QC-passed CNV. The length of the CNVs in the entire cohort ranged from 20 kb to 3.0 megabases (Mb) with a median size of 160 kb. The characteristics of our CNV analysis are shown in Table 4.

We then explored CNVs overlapping known PD genes and identified 15 CNVs in 18 samples (six controls and 12 PD cases) that were exclusively in the *PRKN* gene (Table 5). None of the *PRKN* CNV carriers had a rare variant in the same gene. We tested the presence of five CNVs by MLPA.

As MLPA only covers exonic regions of *PRKN*, three MLPA results were consistent with PennCNV results (Table 5). One duplication was located in exon 2 rather than in a nearby intronic region and one duplication was found to be homozygous covering exon1 rather than heterozygous covering exon 2 (Table 5). After MLPA validation, of the 15 *PRKN* CNVs, eight were single copy deletions, six were single copy duplications and only one was a probably pathogenic homozygous duplication in a late-onset PD patient (AAO = 69 years). Among the PD cases, three *PRKN* heterozygous CNV carriers had an AAO ≤ 50 years (including one patient diagnosed with a juvenile form of PD at the age of 18). One CNV was detected in four samples, while the others were detected in only a single sample (Table 5). Eleven CNVs were considered as rare, since they were not reported in DGV and were spanning structural variants reported in European descent gnomAD_SV dataset with a frequency of less than 1% (Table 5). No clear clinical impact was observed for all the *PRKN* CNVs (uncertain significance in CNVClinViewer).

Rare digenic variants

Eight individuals (five PD cases, one with PSP and two controls) carried two variants in two different PD-related genes (Table 3). The AAO of the patients ranged from 52 to 71 (mean = 64.3). In particular, in autosomal recessive PD genes, pathogenic *PRKN* p.R256C and *PINK1* p.A339T (in heterozygous state) were detected in the same individual with another probably benign variant. One PD patient (AAO = 62 years) carried the heterozygous *PRKN* deletion (chr6:162279763-162406957) and also the benign *LRRK2* variant p.R1514Q. Moreover, two controls were carriers of two variants in *PRKN-ATP13A2* (81 years old) and in *PRKN-PINK1* (70 years old) respectively (Table 3).

Combining rare single nucleotide and copy number variants in *PRKN*

The number of heterozygous rare pathogenic *PRKN* SNVs (p.Q34fs, p.R256C and p.R275W) was not significantly different between controls (n=6, 0.82%) and PD cases (n=4, 0.5%, p-value = 0.6). If we consider all the rare heterozygous CNV deletions as pathogenic loss-of-function variations together with the homozygous duplication, we counted seven PD cases each carrying one rare pathogenic deletion. Overall, the number of heterozygous pathogenic SNVs and CNVs was slightly higher in PD (n=11, 1.64%) than in controls (n=6, 0.82%), but the difference is still not significant (p-value = 0.16).

The sample size is too small to examine a significant burden of these rare variants on PD risk.

Polygenic risk scores

Using significant common SNVs from the largest PD GWAS summary statistics (Nalls et al., 2019), we calculated the PRS in the Luxembourg Parkinson's study for 724 controls and 667 PD patients. The PRS model was calculated based on 75 clumped SNPs that showed the best prediction at the GWAS p-value threshold of 5e-08 and an observed phenotypic variance R2 of 5.3% (1.9% after adjustment for PD prevalence of 5e-03, empirical p-value=9.9e-05) with an AUC of 62.8%. We

found a significant association of PRS with higher PD risk (OR= 1.70[1.50-1.93], p=5.9e-17). The distribution of PRS scores in PD cases and healthy controls was significantly different (Wilcoxon test p-value < 2.2e-16, Figure 1.A). Individuals with the 5% and 10% of highest PRS values had a 9.5-fold [3.9-26.3] (p=1.4e08) and 5.6-fold [3.3-9.7] (p=1.8e-12) increased risk, respectively, compared to individuals with the lowest 5% and 10% PRS values (Figure 1.A). Out of the 3090 canonical pathways gene-sets representing the most important biological processes and diseases, 17 gene-sets were significantly associated with PD risk (Bonferroni adjusted p-value < 0.05, Figure 1.B, Supplementary Table 1). Among the enriched pathways, the majority were associated with PD (showing the highest R2 values, Figure 1.B) and PD pathogenesis (α-synuclein, Parkin and ubiquitination related pathways), Alzheimer disease (AD), signal transduction and metabolism of proteins. No gene-set remained significant after excluding the 90 PD GWAS hits region (1Mb upstream and downstream each locus), indicating the absence of other risk loci acting independently of the known ones.

DISCUSSION

The current report is a comprehensive genetic description of participants recruited within the monocentric case-control Luxembourg Parkinson's study, including patients with PD and atypical parkinsonism, with previously described recruitment design and clinical characteristics (Hipp et al., 2018; Pavelka et al., 2022). Previous long-read sequencing of *GBA1* gene in our cohort revealed that 12.1% of PD patients carried *GBA1* variants (Peiris et al., 2023). Analysing now the complete Neurochip genotyping data, we investigated the potential effect of rare variants, common low-risk variants and CNVs on the PD pathogenesis. Our findings are consistent with those previously reported in European ancestry datasets.

In the *LRRK2* gene, two well-established pathogenic SNVs were found in five PD patients and two controls with a frequency similar to previous European ancestry datasets (Correia Guedes et al., 2010; Shu et al., 2019). Pathogenic and probably pathogenic variants in the *ATP13A2*, *PARK7*, *PRKN* and *PINK1* genes associated with autosomal recessive PD were found in the heterozygous state, except in one PD patient. The latter carried a homozygous pathogenic *PINK1* variant (p.L369P) and had an AAO of 32 years (Arena and Valente, 2017) reported a similar finding, where homozygous variants in *PINK1* associated with early-onset PD (EOPD) were present in the patient before the age of 45 years. In our study, the number of heterozygous SNVs in the recessive *PARK7*, *ATP13A2, PRKN* and *PINK1* genes was not significantly different between PD cases and controls. Controls carrying these variants were over 50 years of age. The majority had no family history of PD and most of PD patients were not of young onset (AAO > 50 years). Patients with AP carried probably benign heterozygous

variants, mainly in *LRRK2* and *PINK1*. Pathogenic *LRRK2* variants have been described in patients with primary tauopathies, although at a low frequency (Sanchez-Contreras et al., 2017; Wen et al., 2021). In particular, *LRRK2* has recently emerged as a genetic risk factor associated with PSP progression (Jabbari et al., 2021).

We called CNVs from the genotyping data of individuals in the Luxembourg Parkinson's study and after a stringent quality control and filtering steps, we screened for CNVs overlapping PD causal genes. We identified 12 PD patients who carried CNVs exclusively in the *PRKN* gene, of which three CNVs were validated by MLPA and were reported in patients having a disease AAO of 50 years. Especially, we described a heterozygous exon1-4 duplication in a patient with EOPD (AAO of 18 years) who did not present any rare variant in the PD-related genes studied here. Moreover, we validated by MLPA a homozygous duplication of *PRKN* exon1 in another PD patient with a late disease-onset (69 years). Duplications of *PRKN* exons were previously reported as 'likely pathogenic' (Schüle et al., 2015). Indeed, both homozygous and compound heterozygous *PRKN* deletions and duplications have previously been associated with early-onset and familial forms of PD (Ahmad et al., 2023; Elfferich et al., 2011; Huttenlocher et al., 2015; Kim et al., 2012). This was recently reproduced in a large CNV study of 4800 clinical exome sequencing reports (Pennings et al., 2023). In a Latin American PD cohort, CNVs in *PRKN* were significantly associated with disease progression, with a prevalence of 5.6% in EOPD cases (Sarihan et al., 2021).

We found that six PD cases and one PSP case carried digenic variants in two different PD-related genes (*LRRK2-PRKN*, *LRRK2-PINK1, PINK1-ATP13A2*) with AAO greater than 50 years. Hitherto, only a few studies have identified digenic variants of PD-related genes (*LRRK2-PRKN* (Dächsel et al., 2006), *PINK1-PARK7* (Tang et al., 2006) or *PRKN-PINK1* (Hayashida et al., 2021)). A detailed familial and clinical study could be carried for every individual, to confirm that the combination of these heterozygous variants, in the context of a digenic inheritance, may point out the phenotype observed in PD and PSP cases.

In our study, we did not find a significant overrepresentation of rare heterozygous SNVs and CNVs in *PRKN.* In particular, heterozygous pathogenic *PRKN* variants were not significantly more frequent in controls than in PD cases. Homozygous or compound heterozygous variants in this gene are the most common cause of EOPD, (Kilarski et al., 2012)), while heterozygous loss of *PRKN* function may be a potential risk factor for developing PD (Castelo Rueda et al., 2021; Huttenlocher et al., 2015; Klein et al., 2007; Lubbe et al., 2021) and therefore identifying individuals at increased risk might be useful in the prodromal phase. However, this role of heterozygous *PRKN* is still under debate, as previous reports suggested a lack of association with PD (Kay et al., 2010). Recently, in a larger association study, Yu and colleagues fully sequenced *PRKN* in a PD cases/controls cohort from European ancestry, including 1965 late-onset and 553 early-onset, and concluded that heterozygous SNVs or CNVs in *PRKN* are not associated with EOPD (Yu et al., 2021). They reported that 1.52% of PD and 1.8% of controls were carriers. Here, using a SNP array based on CNVs and SNPs

screening, we showed similar percentages (1.64% of PD and 0.82% of controls) with non-significant differences between controls and mainly late-onset PD cases excluding a role of heterozygous Parkin mutations in late-onset PD.

Potential neuroprotective PD therapies and clinical trials are now targeting specific PD subtypes based on genetic markers causing or increasing the disease risk, such as therapies targeting *LRRK2*, *GBA1* and alpha-synuclein (Sardi et al., 2018). Parkin-proved disease is characterized by a slow motor progression, preserved cognition and a limited increase in dopaminergic medication over time (Menon et al., 2023). Moreover, severe loss of dopaminergic neurons was observed in homozygous *PRKN* carriers without Lewy bodies formation, which is one of the major markers of idiopathic PD (Mata, 2004). Confirming the potential role of heterozygous *PRKN* variants in the pathogenesis of PD will be crucial, despite the lack of data describing PD conversion of individuals carrying these genetic risk factors.

Beyond the effects of rare variants, we estimated the total cumulative contribution of common low risk SNVs by calculating the PRS. Our PRS model of disease risk showed an expected trend similar to previous reports showing that PRS discriminates PD cases from unaffected individuals (Dehestani et al., 2021). Several polygenic analyses have become standard tools for assessing the risk for complex disorders and an accurate method for predicting disease status and identifying high-risk individuals (Lewis and Vassos, 2020). Next, we looked up at how thousands of biological pathways might contribute to the risk of developing PD. In addition to pathways already associated with PD and AD, molecular processes underlying proteins metabolism, signal transduction and post-translational protein modification were among the most important contributors to PD risk. The metabolic dysfunction, energy failure and redox imbalance observed in PD were considered obvious features to qualify PD as a complex metabolic disorder (Anandhan et al., 2017). In addition, disruption of any stage in the protein life cycle could engender PD pathology (Langston and Cookson, 2020). Comparing our results with a previous large-scale gene set-specific PRS studies that reported the involvement of multiple processes in the aetiology of PD (Bandres-Ciga et al., 2020), similar molecular processes were found here. However, other processes such as immune response, synaptic transmission and endosomal-lysosomal dysfunction were not highlighted which may be due to the smaller sample size in our dataset. Pathway PRSs are expected to provide important insights into the complex heterogeneity of PD and how patients respond to treatment, by generating biologically traceable therapeutic targets from polygenic signals (Choi et al., 2023).We are aware that our study has several limitations: (1) the sample size was not large enough to have sufficient statistical power to perform further analysis, such as GWAS for PD risk or AAO, genome-wide CNV burden or human leukocyte antigen (HLA) association; (2) not all known variants associated with PD can be accurately assessed by the NeuroChip and we might have missed some mutated alleles, even though we confirmed all the identified variants by Sanger sequencing; and (3) we used best practices to call

CNVs from genotyping data (Sarihan et al., 2021), and thus we will always miss small CNVs that are systematically filtered out. Moreover, we could validate only few of the called CNVs with MLPA.

CONCLUSION

In conclusion, our study has successfully performed a comprehensive genetic baseline characterization of the Luxembourgish PD case-control cohort, investigating rare variants, CNVs and PRSs. Our findings do not support an association between PD risk and rare heterozygous *PRKN* variants. We also described a possible role of *LRRK2* in AP and new possible digenic inheritance patterns in PD. Together with other studies in different European populations, our findings will advance the understanding of PD pathogenesis and genetics and could redefine the development of future therapeutic targets and therapies.

DATA AVAILABILITY STATEMENT

Genetic and clinical data for this manuscript are not publicly available as they are linked to the internal regulations of the Luxembourg Parkinson's Study. Requests for accessing the datasets can be directed to request.ncer-pd@uni.lu.

AUTHOR CONTRIBUTION

ZL wrote the paper, conducted the analysis for PRS and CNVs, executed the statistical analysis and interpretation of results. SP contributed to the writing of the paper, performed rare variants screening analysis, and participated in interpretation of results. D.R.B and L.P substantially participated in data collection, data exportation and data curation and critically revised the manuscript. R.K and P.M.: Conceived, organised, and co-executed the research project; participated in interpretation of results; critically revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) is funded by the Luxembourg National Research Fund (FNR/NCER13/BM/11264123), the PEARL program (FNR/P13/6682797 to R.K), MotaSYN (12719684 to R.K), MAMaSyn (to R.K), MiRisk-PD (C17/BM/11676395 to R.K, P.M), the FNR/DFG Core INTER (ProtectMove, FNR11250962 to P.M, Z.L), and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to R.K, S.P).

ACKNOWLEDGMENTS

We would like to thank all participants of the Luxembourg Parkinson's Study for their important support of our research. Furthermore, we acknowledge the joint effort of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) Consortium members from the partner institutions Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health, Centre Hospitalier de Luxembourg, and Laboratoire National de Santé generally contributing to the Luxembourg Parkinson's Study as listed below:

Geeta ACHARYA 2, Gloria AGUAYO 2, Myriam ALEXANDRE 2, Muhammad ALI 1, Wim AMMERLANN 2, Giuseppe ARENA 1, Rudi BALLING 1, Michele BASSIS 1, Katy BEAUMONT 2, Regina BECKER 1, Camille BELLORA 2, Guy BERCHEM 3, Daniela BERG 11, Alexandre BISDORFF 5, Ibrahim BOUSSAAD 1, Kathrin BROCKMANN 11, Jessica CALMES 2, Lorieza CASTILLO 2, Gessica CONTESOTTO 2, Nico DIEDERICH 3, Rene DONDELINGER 5, Daniela ESTEVES 2, Guy FAGHERAZZI 2, Jean-Yves FERRAND 2, Manon GANTENBEIN 2, Thomas GASSER 11, Piotr GAWRON 1, Soumyabrata GHOSH 1, Marijus GIRAITIS 2,3, Enrico GLAAB 1, Elisa GÓMEZ DE LOPE 1, Jérôme GRAAS 2, Mariella GRAZIANO 17, Valentin GROUES 1, Anne GRÜNEWALD 1, Wei GU 1, Gaël HAMMOT 2, Anne-Marie HANFF 2, Linda HANSEN 1,3, Michael HENEKA 1, Estelle HENRY 2, Sylvia HERBRINK 6, Sascha HERZINGER 1, Michael HEYMANN 2, Michele HU 8, Alexander HUNDT 2, Nadine JACOBY 18, Jacek JAROSLAW LEBIODA 1, Yohan JAROZ 1, Sonja JÓNSDÓTTIR 2, Quentin KLOPFENSTEIN 1, Jochen KLUCKEN 1,2,3, Rejko KRÜGER 1,2,3, Pauline LAMBERT 2, Zied LANDOULSI 1, Roseline LENTZ 7, Inga LIEPELT 11, Robert LISZKA 14, Laura LONGHINO 3, Victoria LORENTZ 2, Paula Cristina LUPU 2, Clare MACKAY 10, Walter MAETZLER 15, Katrin MARCUS 13, Guilherme MARQUES 2, Tainá M. MARQUES 1, Patricia MARTINS CONDE 1, Patrick MAY 1, Deborah MCINTYRE 2, Chouaib MEDIOUNI 2, Francoise MEISCH 1, Myriam MENSTER 2, Maura MINELLI 2, Michel MITTELBRONN 1,4, Brit MOLLENHAUER 12, Friedrich MÜHLSCHLEGEL 4, Romain NATI 3, Ulf NEHRBASS 2, Sarah NICKELS 1, Beatrice NICOLAI 3, Jean-Paul NICOLAY 19, Fozia NOOR 2, Marek OSTASZEWSKI 1, Clarissa P. C. GOMES 1, Sinthuja PACHCHEK 1, Claire PAULY 1,3, Laure PAULY 1, Lukas PAVELKA 1,3, Magali PERQUIN 2, Rosalina RAMOS LIMA 2, Armin RAUSCHENBERGER 1, Rajesh RAWAL 1, Dheeraj REDDY BOBBILI 1, Kirsten ROOMP 1, Eduardo ROSALES 2, Isabel ROSETY 1, Estelle SANDT 2, Stefano SAPIENZA 1, Venkata SATAGOPAM 1, Margaux SCHMITT 2, Sabine SCHMITZ 1, Reinhard SCHNEIDER 1, Jens SCHWAMBORN 1, Amir SHARIFY 2, Ekaterina SOBOLEVA 1, Kate SOKOLOWSKA 2, Hermann THIEN 2, Elodie THIRY 3, Rebecca TING JIIN LOO 1, Christophe TREFOIS 1, Johanna TROUET 2, Olena TSURKALENKO 2, Michel VAILLANT 2, Mesele VALENTI 2, Carlos VEGA 1, Liliana VILAS BOAS 3, Maharshi VYAS 1, Richard WADEMARTINS 9, Paul WILMES 1, Evi WOLLSCHEID-LENGELING 1, Gelani ZELIMKHANOV 3

1 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

2 Luxembourg Institute of Health, Strassen, Luxembourg

3 Centre Hospitalier de Luxembourg, Strassen, Luxembourg

4 Laboratoire National de Santé, Dudelange, Luxembourg

5 Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg

6 Centre Hospitalier du Nord, Ettelbrück, Luxembourg

7 Parkinson Luxembourg Association, Leudelange, Luxembourg

8 Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK 9 Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK

10 Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK

11 Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Tübingen, Germany

12 Paracelsus-Elena-Klinik, Kassel, Germany

13 Ruhr-University of Bochum, Bochum, Germany

14 Westpfalz-Klinikum GmbH, Kaiserslautern, Germany

15 Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany

16 Department of Neurology Philipps, University Marburg, Marburg, Germany

17 Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg

18 Private practice, Ettelbruck, Luxembourg

19 Private practice, Luxembourg, Luxembourg

SUPPLEMENTARY MATERIAL

Supplementary figure 1
Supplementary table 1

REFERENCES

Ahmad, A., Nkosi, D., Iqbal, M.A., 2023. PARK2 Microdeletion or Duplications Have Been Implicated in Different Neurological Disorders Including Early Onset Parkinson Disease. Genes 14, 600. https://doi.org/10.3390/genes14030600

Anandhan, A., Jacome, M.S., Lei, S., Hernandez-Franco, P., Pappa, A., Panayiotidis, M.I., Powers, R., Franco, R., 2017. Metabolic Dysfunction in Parkinson's Disease: Bioenergetics, Redox Homeostasis and Central Carbon Metabolism. Brain Research Bulletin 133, 12–30. https://doi.org/10.1016/j.brainresbull.2017.03.009

Arena, G., Valente, E.M., 2017. PINK1 in the limelight: multiple functions of an eclectic protein in human health and disease: Functions of PINK1 in human pathology. J. Pathol. 241, 251–263. https://doi.org/10.1002/path.4815

Bandres-Ciga, S., Saez-Atienzar, S., Kim, J.J., Makarious, M.B., Faghri, F., Diez-Fairen, M., Iwaki, H., Leonard, H., Botia, J., Ryten, M., Hernandez, D., Gibbs, J.R., Ding, J., Gan-Or, Z., Noyce, A., Pihlstrom, L., Torkamani, A., Soltis, A.R., Dalgard, C.L., The American Genome Center, Scholz, S.W., Traynor, B.J., Ehrlich, D., Scherzer, C.R., Bookman, M., Cookson, M., Blauwendraat, C., Nalls, M.A., Singleton, A.B., the International Parkinson Disease Genomics Consortium, 2020. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. Acta Neuropathol 140, 341–358. https://doi.org/10.1007/s00401-020-02181-3

Blauwendraat, C., Faghri, F., Pihlstrom, L., Geiger, J.T., Elbaz, A., Lesage, S., Corvol, J.-C., May, P., Nicolas, A., Abramzon, Y., Murphy, N.A., Gibbs, J.R., Ryten, M., Ferrari, R., Bras, J., Guerreiro, R., Williams, J., Sims, R., Lubbe, S., Hernandez, D.G., Mok, K.Y., Robak, L., Campbell, R.H., Rogaeva, E., Traynor, B.J., Chia, R., Chung, S.J., International Parkinson's Disease Genomics Consortium (IPDGC), COURAGE-PD Consortium, Hardy, J.A., Brice, A., Wood, N.W., Houlden, H., Shulman, J.M., Morris, H.R., Gasser, T., Krüger, R., Heutink, P., Sharma, M., Simón-Sánchez, J., Nalls, M.A., Singleton, A.B., Scholz, S.W., 2017. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. Neurobiol Aging 57, 247.e9-247.e13. https://doi.org/10.1016/j.neurobiolaging.2017.05.009

Blauwendraat, C., Nalls, M.A., Singleton, A.B., 2020. The genetic architecture of Parkinson's disease. The Lancet Neurology 19, 170–178. https://doi.org/10.1016/S1474-4422(19)30287-X

Bloem, B.R., Okun, M.S., Klein, C., 2021. Parkinson's disease. The Lancet 397, 2284–2303. https://doi.org/10.1016/S0140-6736(21)00218-X

Castelo Rueda, M.P., Raftopoulou, A., Gögele, M., Borsche, M., Emmert, D., Fuchsberger, C., Hantikainen, E.M., Vukovic, V., Klein, C., Pramstaller, P.P., Pichler, I., Hicks, A.A., 2021. Frequency of Heterozygous Parkin (PRKN) Variants and Penetrance of Parkinson's Disease Risk Markers in the Population-Based CHRIS Cohort. Front. Neurol. 12, 706145. https://doi.org/10.3389/fneur.2021.706145

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci 4, 7. https://doi.org/10.1186/s13742-015-0047-8

Choi, S.W., García-González, J., Ruan, Y., Wu, H.M., Porras, C., Johnson, J., Bipolar Disorder Working group of the Psychiatric Genomics Consortium, Hoggart, C.J., O'Reilly, P.F., 2023. PRSet: Pathway-based polygenic risk score analyses and software. PLoS Genet 19, e1010624. https://doi.org/10.1371/journal.pgen.1010624

Choi, S.W., O'Reilly, P.F., 2019. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience 8, giz082. https://doi.org/10.1093/gigascience/giz082

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., Watts, N.A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C.W., Huang, Y., Brookings, T., Sharpe, T., Stone, M.R., Valkanas, E., Fu, J., Tiao, G., Laricchia, K.M., Ruano-Rubio, V., Stevens, C., Gupta, N., Cusick, C., Margolin, L., Genome Aggregation Database Production Team, Alföldi, J., Armean, I.M., Banks, E., Bergelson, L., Cibulskis, K., Collins, R.L., Connolly, K.M., Covarrubias, M., Cummings, B., Daly, M.J., Donnelly, S., Farjoun, Y., Ferriera, S., Francioli, L., Gabriel, S., Gauthier, L.D., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Karczewski, K.J., Laricchia, K.M., Llanwarne, C., Minikel, E.V., Munshi, R., Neale, B.M., Novod, S., O'Donnell-Luria, A.H., Petrillo, N., Poterba, T., Roazen, D., Ruano-Rubio, V., Saltzman, A., Samocha, K.E., Schleicher, M., Seed, C., Solomonson, M., Soto, J., Tiao, G., Tibbetts, K., Tolonen, C., Vittal, C., Wade, G., Wang, A., Wang, Q., Ware, J.S., Watts, N.A., Weisburd, B., Whiffin, N., Genome Aggregation Database Consortium, Salinas, C.A.A., Ahmad, T., Albert, C.M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E.J., Boehnke, M., Bonnycastle,

L.L., Bottinger, E.P., Bowden, D.W., Bown, M.J., Chambers, J.C., Chan, J.C., Chasman, D., Cho, J., Chung, M.K., Cohen, B., Correa, A., Dabelea, D., Daly, M.J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P.T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S.J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M.M., Hultman, C.M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.-J., Kim, Y.J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H.M., Kugathasan, S., Kwak, S.H., Laakso, M., Lehtimäki, T., Loos, R.J.F., Lubitz, S.A., Ma, R.C.W., MacArthur, D.G., Marrugat, J., Mattila, K.M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Meigs, J.B., Melander, O., Metspalu, A., Neale, B.M., Nilsson, P.M., O'Donovan, M.C., Ongur, D., Orozco, L., Owen, M.J., Palmer, C.N.A., Palotie, A., Park, K.S., Pato, C., Pulver, A.E., Rahman, N., Remes, A.M., Rioux, J.D., Ripatti, S., Roden, D.M., Saleheen, D., Salomaa, V., Samani, N.J., Scharf, J., Schunkert, H., Shoemaker, M.B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P.F., Suvisaari, J., Tai, E.S., Teo, Y.Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Vawter, M.P., Ware, J.S., Watkins, H., Weersma, R.K., Wessman, M., Wilson, J.G., Xavier, R.J., Taylor, K.D., Lin, H.J., Rich, S.S., Post, W.S., Chen, Y.-D.I., Rotter, J.I., Nusbaum, C., Philippakis, A., Lander, E., Gabriel, S., Neale, B.M., Kathiresan, S., Daly, M.J., Banks, E., MacArthur, D.G., Talkowski, M.E., 2020. A structural variation reference for medical and population genetics. Nature 581, 444–451. https://doi.org/10.1038/s41586-020-2287-8

Correia Guedes, L., Ferreira, J.J., Rosa, M.M., Coelho, M., Bonifati, V., Sampaio, C., 2010. Worldwide frequency of G2019S LRRK2 mutation in Parkinson's disease: A systematic review. Parkinsonism & Related Disorders 16, 237–242. https://doi.org/10.1016/j.parkreldis.2009.11.004

Dächsel, J.C., Mata, I.F., Ross, O.A., Taylor, J.P., Lincoln, S.J., Hinkle, K.M., Huerta, C., Ribacoba, R., Blazquez, M., Alvarez, V., Farrer, M.J., 2006. Digenic parkinsonism: Investigation of the synergistic effects of PRKN and LRRK2. Neuroscience Letters 410, 80–84. https://doi.org/10.1016/j.neulet.2006.06.068

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. Nat Genet 48, 1284–1287. https://doi.org/10.1038/ng.3656

Dehestani, M., Liu, H., Gasser, T., 2021. Polygenic Risk Scores Contribute to Personalized Medicine of Parkinson's Disease. JPM 11, 1030. https://doi.org/10.3390/jpm11101030

Dorsey, E.R., Bloem, B.R., 2018. The Parkinson Pandemic—A Call to Action. JAMA Neurol 75, 9. https://doi.org/10.1001/jamaneurol.2017.3299

Elfferich, P., Verleun-Mooijman, M.C., Maat-Kievit, J.A., van de Warrenburg, B.P.C., Abdo, W.F., Eshuis, S.A., Leenders, K.L., Hovestadt, A., Zijlmans, J.C.M., Stroy, J.-P.M., van Swieten, J.C., Boon, A.J.W., van Engelen, K., Verschuuren-Bemelmans, C.C., Lesnik-Oberstein, S.A.J., Tassorelli, C., Lopiano, L., Bonifati, V., Dooijes, D., van Minkelen, R., 2011. Breakpoint mapping of 13 large parkin deletions/duplications reveals an exon 4 deletion and an exon 7 duplication as founder mutations. Neurogenetics 12, 263–271. https://doi.org/10.1007/s10048-011-0302-9

Gilman, S., Wenning, G.K., Low, P.A., Brooks, D.J., Mathias, C.J., Trojanowski, J.Q., Wood, N.W., Colosimo, C., Durr, A., Fowler, C.J., Kaufmann, H., Klockgether, T., Lees, A., Poewe, W., Quinn, N., Revesz, T., Robertson, D., Sandroni, P., Seppi, K., Vidailhet, M., 2008. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 71, 670–676. https://doi.org/10.1212/01.wnl.0000324625.00404.15

Haugarvoll, K., Wszolek, Z.K., 2009. Clinical features of LRRK2 parkinsonism. Parkinsonism & Related Disorders 15, S205–S208. https://doi.org/10.1016/S1353-8020(09)70815-6

Hayashida, A., Li, Y., Yoshino, H., Daida, K., Ikeda, A., Ogaki, K., Fuse, A., Mori, A., Takanashi, M., Nakahara, T., Yoritaka, A., Tomizawa, Y., Furukawa, Y., Kanai, K., Nakayama, Y., Ito, H., Ogino, M., Hattori, Y., Hattori, T., Ichinose, Y., Takiyama, Y., Saito, T., Kimura, T., Aizawa, H., Shoji, H., Mizuno, Y., Matsushita, T., Sato, M., Sekijima, Y., Morita, M., Iwasaki, A., Kusaka, H., Tada, M., Tanaka, F., Sakiyama, Y., Fujimoto, T., Nagara, Y., Kashihara, K., Todo, H., Nakao, K., Tsuruta, K., Yoshikawa, M., Hara, H., Yokote, H., Murase, N., Nakamagoe, K., Tamaoka, A., Takamiya, M., Morimoto, N., Nokura, K., Kako, T., Funayama, M., Nishioka, K., Hattori, N., 2021. The identified clinical features of Parkinson's disease in homo-, heterozygous and digenic variants of PINK1. Neurobiology of Aging 97, 146.e1-146.e13. https://doi.org/10.1016/j.neurobiolaging.2020.06.017

Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., Ferreira, J.J., Tolosa, E., Kay, D.M., Klein, C., Williams, D.R., Marras, C., Lang, A.E., Wszolek, Z.K., Berciano, J., Schapira, A.H., Lynch, T., Bhatia, K.P., Gasser, T., Lees, A.J., Wood, N.W., 2008. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated

Parkinson's disease: a case-control study. The Lancet Neurology 7, 583–590. https://doi.org/10.1016/S1474-4422(08)70117-0

Hipp, G., Vaillant, M., Diederich, N.J., Roomp, K., Satagopam, V.P., Banda, P., Sandt, E., Mommaerts, K., Schmitz, S.K., Longhino, L., Schweicher, A., Hanff, A.-M., Nicolai, B., Kolber, P., Reiter, D., Pavelka, L., Binck, S., Pauly, C., Geffers, L., Betsou, F., Gantenbein, M., Klucken, J., Gasser, T., Hu, M.T., Balling, R., Krüger, R., 2018. The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. Front. Aging Neurosci. 10, 326. https://doi.org/10.3389/fnagi.2018.00326

Huttenlocher, J., Stefansson, H., Steinberg, S., Helgadottir, H.T., Sveinbjörnsdóttir, S., Riess, O., Bauer, P., Stefansson, K., 2015. Heterozygote carriers for CNVs in *PARK2* are at increased risk of Parkinson's disease. Hum. Mol. Genet. 24, 5637–5643. https://doi.org/10.1093/hmg/ddv277

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J., Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C., Hsieh, C.-L., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittemore, A.S., Sieh, W., 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics 99, 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Jabbari, E., Koga, S., Valentino, R.R., Reynolds, R.H., Ferrari, R., Tan, M.M.X., Rowe, J.B., Dalgard, C.L., Scholz, S.W., Dickson, D.W., Warner, T.T., Revesz, T., Höglinger, G.U., Ross, O.A., Ryten, M., Hardy, J., Shoai, M., Morris, H.R., Mok, K.Y., Murphy, D.P., Al-Sarraj, S., Troakes, C., Gentleman, S.M., Allinson, K.S.J., Jaunmuktane, Z., Holton, J.L., Lees, A.J., Morris, C.M., Compta, Y., Gelpi, E., Van Swieten, J.C., Rajput, A., Ferguson, L., Cookson, M.R., Gibbs, J.R., Blauwendraat, C., Ding, J., Chia, R., Traynor, B.J., Pantelyat, A., Viollet, C., Traynor, B.J., Pletnikova, O., Troncoso, J.C., Rosenthal, L.S., Boxer, A.L., Respondek, G., Arzberger, T., Roeber, S., Giese, A., Burn, D.J., Pavese, N., Gerhard, A., Kobylecki, C., Leigh, P.N., Church, A., T.M. Hu, M., 2021. Genetic determinants of survival in progressive supranuclear palsy: a genome-wide association study. The Lancet Neurology 20, 107–116. https://doi.org/10.1016/S1474-4422(20)30394-X

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Genome Aggregation Database Consortium, Aguilar Salinas, C.A., Ahmad, T., Albert, C.M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E.J., Boehnke, M., Bonnycastle, L.L., Bottinger, E.P., Bowden, D.W., Bown, M.J., Chambers, J.C., Chan, J.C., Chasman, D., Cho, J., Chung, M.K., Cohen, B., Correa, A., Dabelea, D., Daly, M.J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P.T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S.J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M.M., Hultman, C.M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.-J., Kim, Y.J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H.M., Kugathasan, S., Kwak, S.H., Laakso, M., Lehtimäki, T., Loos, R.J.F., Lubitz, S.A., Ma, R.C.W., MacArthur, D.G., Marrugat, J., Mattila, K.M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Meigs, J.B., Melander, O., Metspalu, A., Neale, B.M., Nilsson, P.M., O'Donovan, M.C., Ongur, D., Orozco, L., Owen, M.J., Palmer, C.N.A., Palotie, A., Park, K.S., Pato, C., Pulver, A.E., Rahman, N., Remes, A.M., Rioux, J.D., Ripatti, S., Roden, D.M., Saleheen, D., Salomaa, V., Samani, N.J., Scharf, J., Schunkert, H., Shoemaker, M.B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P.F., Suvisaari, J., Tai, E.S., Teo, Y.Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Vawter, M.P., Ware, J.S., Watkins, H., Weersma, R.K., Wessman, M., Wilson, J.G., Xavier, R.J., Neale, B.M., Daly, M.J., MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kay, D.M., Stevens, C.F., Hamza, T.H., Montimurro, J.S., Zabetian, C.P., Factor, S.A., Samii, A., Griffith, A., Roberts, J.W., Molho, E.S., Higgins, D.S., Gancher, S., Moses, L., Zareparsi, S., Poorkaj, P., Bird, T., Nutt, J., Schellenberg, G.D., Payami, H., 2010. A comprehensive analysis of deletions,

multiplications, and copy number variations in PARK2. Neurology 75, 1189–1194. https://doi.org/10.1212/WNL.0b013e3181f4d832

Kilarski, L.L., Pearson, J.P., Newsway, V., Majounie, E., Knipe, M.D.W., Misbahuddin, A., Chinnery, P.F., Burn, D.J., Clarke, C.E., Marion, M.-H., Lewthwaite, A.J., Nicholl, D.J., Wood, N.W., Morrison, K.E., Williams-Gray, C.H., Evans, J.R., Sawcer, S.J., Barker, R.A., Wickremaratchi, M.M., Ben-Shlomo, Y., Williams, N.M., Morris, H.R., 2012. Systematic Review and UK-Based Study of *PARK2 (parkin), PINK1, PARK7 (DJ-1)* and *LRRK2* in early-onset Parkinson's disease: *PARK2, PINK1, PARK7, LRRK2* in EOPD. Mov. Disord. 27, 1522–1529. https://doi.org/10.1002/mds.25132

Kim, S.Y., Seong, M.W., Jeon, B.S., Kim, S.Y., Ko, H.S., Kim, J.Y., Park, S.S., 2012. Phase analysis identifies compound heterozygous deletions of the PARK2 gene in patients with early-onset Parkinson disease. Clinical Genetics 82, 77–82. https://doi.org/10.1111/j.1399-0004.2011.01693.x

Klein, C., Lohmann-Hedrich, K., Rogaeva, E., Schlossmacher, M.G., Lang, A.E., 2007. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. The Lancet Neurology 6, 652–662. https://doi.org/10.1016/S1474-4422(07)70174-6

La Cognata, V., Morello, G., D'Agata, V., Cavallaro, S., 2017. Copy number variability in Parkinson's disease: assembling the puzzle through a systems biology approach. Hum Genet 136, 13–37. https://doi.org/10.1007/s00439-016-1749-4

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., Maglott, D.R., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42, D980-985. https://doi.org/10.1093/nar/gkt1113

Langston, R.G., Cookson, M.R., 2020. Pathways of protein synthesis and degradation in PD pathogenesis, in: Progress in Brain Research. Elsevier, pp. 217–270. https://doi.org/10.1016/bs.pbr.2020.01.002

Lesage, S., Brice, A., 2009. Parkinson's disease: from monogenic forms to genetic susceptibility factors. Hum Mol Genet 18, R48-59. https://doi.org/10.1093/hmg/ddp012

Lewis, C.M., Vassos, E., 2020. Polygenic risk scores: from research tools to clinical instruments. Genome Med 12, 44. https://doi.org/10.1186/s13073-020-00742-5

Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R.C., Goetz, C.G., Golbe, L.I., Grafman, J., Growdon, J.H., Hallett, M., Jankovic, J., Quinn, N.P., Tolosa, E., Zee, D.S., 1996. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP International Workshop*. Neurology 47, 1–9. https://doi.org/10.1212/WNL.47.1.1

Litvan, I., Bhatia, K.P., Burn, D.J., Goetz, C.G., Lang, A.E., McKeith, I., Quinn, N., Sethi, K.D., Shults, C., Wenning, G.K., 2003. SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders. Mov Disord. 18, 467–486. https://doi.org/10.1002/mds.10459

Liu, X., Cheng, R., Ye, X., Verbitsky, M., Kisselev, S., Mejia-Santana, H., Louis, E.D., Cote, L.J., Andrews, H.F., Waters, C.H., Ford, B., Fahn, S., Marder, K., Lee, J.H., Clark, L.N., 2013. Increased rate of sporadic and recurrent rare genic copy number variants in P arkinson's disease among A shkenazi J ews. Mol Genet Genomic Med 1, 142–154. https://doi.org/10.1002/mgg3.18

Lubbe, S.J., Bustos, B.I., Hu, J., Krainc, D., Joseph, T., Hehir, J., Tan, M., Zhang, W., Escott-Price, V., Williams, N.M., Blauwendraat, C., Singleton, A.B., Morris, H.R., for International Parkinson's Disease Genomics Consortium (IPDGC), 2021. Assessing the relationship between monoallelic *PRKN* mutations and Parkinson's risk. Human Molecular Genetics 30, 78–86. https://doi.org/10.1093/hmg/ddaa273

MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucl. Acids Res. 42, D986–D992. https://doi.org/10.1093/nar/gkt958

Macnee, M., Pérez-Palma, E., Brünger, T., Klöckner, C., Platzer, K., Stefanski, A., Montanucci, L., Bayat, A., Radtke, M., Collins, R.L., Talkowski, M., Blankenberg, D., Møller, R.S., Lemke, J.R., Nothnagel, M., May, P., Lal, D., 2023. CNV-ClinViewer: enhancing the clinical interpretation of large copy-number variants online. Bioinformatics 39, btad290. https://doi.org/10.1093/bioinformatics/btad290

Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.-M., 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Mata, I.F., 2004. Parkin genetics: one model for Parkinson's disease. Human Molecular Genetics 13, 127R – 133. https://doi.org/10.1093/hmg/ddh089

McKeith, I.G., Dickson, D.W., Lowe, J., Emre, M., O'Brien, J.T., Feldman, H., Cummings, J., Duda, J.E., Lippa, C., Perry, E.K., Aarsland, D., Arai, H., Ballard, C.G., Boeve, B., Burn, D.J., Costa, D., Del Ser, T., Dubois, B., Galasko, D., Gauthier, S., Goetz, C.G., Gomez-Tortosa, E., Halliday, G., Hansen, L.A., Hardy, J., Iwatsubo, T., Kalaria, R.N., Kaufer, D., Kenny, R.A., Korczyn, A., Kosaka, K., Lee, V.M.Y., Lees, A., Litvan, I., Londos, E., Lopez, O.L., Minoshima, S., Mizuno, Y., Molina, J.A.,

Mukaetova-Ladinska, E.B., Pasquier, F., Perry, R.H., Schulz, J.B., Trojanowski, J.Q., Yamada, M., for the Consortium on DLB, 2005. Diagnosis and management of dementia with Lewy bodies: Third report of the DLB consortium. Neurology 65, 1863–1872. https://doi.org/10.1212/01.wnl.0000187889.17253.b1

Menon, P.J., Sambin, S., Criniere-Boizet, B., Courtin, T., Casse, F., Tesson, C., Ferrien, M., Flamand-Roze, E., Lejeune, F.-X., Lanore, A., Mangone, G., Mariani, L.-L., Aasly, J., Or, Z.G., Yu, E., Dauvilliers, Y., Zimprich, A., Fernandez, I.A., Pastor, P., Di Fonzo, A., Bhatia, K., Magrinelli, F., Houlden, H., Real, R., Narendra, D., Lin, H.-P., Jovanovic, C., Koks, S., Lynch, T., Gallagher, A., Vandenberghe, W., Gasser, T., Morris, H., Klein, C., Brockmann, K., Corti, O., Brice, A., Lesage, S., Corvol, J.-C., 2023. Genotype-Phenotype correlation in Parkin-Parkinson's disease (P3-11.012), in: Sunday, April 23. Presented at the 2023 Annual Meeting Abstracts, Lippincott Williams & Wilkins, p. 2970. https://doi.org/10.1212/WNL.0000000000202927

Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., Bras, J., Young, E., von Coelln, R., Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrøm, L., Siitonen, A., Iwaki, H., Leonard, H., Faghri, F., Gibbs, J.R., Hernandez, D.G., Scholz, S.W., Botia, J.A., Martinez, M., Corvol, J.-C., Lesage, S., Jankovic, J., Shulman, L.M., Sutherland, M., Tienari, P., Majamaa, K., Toft, M., Andreassen, O.A., Bangale, T., Brice, A., Yang, J., Gan-Or, Z., Gasser, T., Heutink, P., Shulman, J.M., Wood, N.W., Hinds, D.A., Hardy, J.A., Morris, H.R., Gratten, J., Visscher, P.M., Graham, R.R., Singleton, A.B., Adarmes-Gómez, A.D., Aguilar, M., Aitkulova, A., Akhmetzhanov, V., Alcalay, R.N., Alvarez, I., Alvarez, V., Bandres-Ciga, S., Barrero, F.J., Bergareche Yarza, J.A., Bernal-Bernal, I., Billingsley, K., Blauwendraat, C., Blazquez, M., Bonilla-Toribio, M., Botía, J.A., Boungiorno, M.T., Bras, J., Brice, A., Brockmann, K., Bubb, V., Buiza-Rueda, D., Cámara, A., Carrillo, F., Carrión-Claro, M., Cerdan, D., Chelban, V., Clarimón, J., Clarke, C., Compta, Y., Cookson, M.R., Corvol, J.-C., Craig, D.W., Danjou, F., Diez-Fairen, M., Dols-Icardo, O., Duarte, J., Duran, R., Escamilla-Sevilla, F., Escott-Price, V., Ezquerra, M., Faghri, F., Feliz, C., Fernández, M., Fernández-Santiago, R., Finkbeiner, S., Foltynie, T., Gan-Or, Z., Garcia, C., García-Ruiz, P., Gasser, T., Gibbs, J.R., Gomez Heredia, M.J., Gómez-Garre, P., González, M.M., Gonzalez-Aramburu, I., Guelfi, S., Guerreiro, R., Hardy, J., Hassin-Baer, S., Hernandez, D.G., Heutink, P., Hoenicka, J., Holmans, P., Houlden, H., Infante, J., Iwaki, H., Jesús, S., Jimenez-Escrig, A., Kaishybayeva, G., Kaiyrzhanov, R., Karimova, A., Kia, D.A., Kinghorn, K.J., Koks, S., Krohn, L., Kulisevsky, J., Labrador-Espinosa, M.A., Leonard, H.L., Lesage, S., Lewis, P., Lopez-Sendon, J.L., Lovering, R., Lubbe, S., Lungu, C., Macias, D., Majamaa, K., Manzoni, C., Marín, J., Marinus, J., Marti, M.J., Martinez, M., Martínez Torres, I., Martínez-Castrillo, J.C., Mata, M., Mencacci, N.E., Méndez-del-Barrio, C., Middlehurst, B., Mínguez, A., Mir, P., Mok, K.Y., Morris, H.R., Muñoz, E., Nalls, M.A., Narendra, D., Noyce, A.J., Ojo, O.O., Okubadejo, N.U., Pagola, A.G., Pastor, P., Perez Errazquin, F., Periñán-Tocino, T., Pihlstrom, L., Plun-Favreau, H., Quinn, J., R'Bibo, L., Reed, X., Rezola, E.M., Rizig, M., Rizzu, P., Robak, L., Rodriguez, A.S., Rouleau, G.A., Ruiz-Martínez, J., Ruz, C., Ryten, M., Sadykova, D., Scholz, S.W., Schreglmann, S., Schulte, C., Sharma, M., Shashkin, C., Shulman, J.M., Sierra, M., Siitonen, A., Simón-Sánchez, J., Singleton, A.B., Suarez-Sanmartin, E., Taba, P., Tabernero, C., Tan, M.X., Tartari, J.P., Tejera-Parrado, C., Toft, M., Tolosa, E., Trabzuni, D., Valldeoriola, F., van Hilten, J.J., Van Keuren-Jensen, K., Vargas-González, L., Vela, L., Vives, F., Williams, N., Wood, N.W., Zharkinbekova, N., Zharmukhanov, Z., Zholdybayeva, E., Zimprich, A., Ylikotila, P., Shulman, L.M., von Coelln, R., Reich, S., Savitt, J., Agee, M., Alipanahi, B., Auton, A., Bell, R.K., Bryc, K., Elson, S.L., Fontanillas, P., Furlotte, N.A., Huber, K.E., Hicks, B., Jewett, E.M., Jiang, Y., Kleinman, A., Lin, K.-H., Litterman, N.K., McCreight, J.C., McIntyre, M.H., McManus, K.F., Mountain, J.L., Noblin, E.S., Northover, C.A.M., Pitts, S.J., Poznik, G.D., Sathirapongsasuti, J.F., Shelton, J.F., Shringarpure, S., Tian, C., Tung, J., Vacic, V., Wang, X., Wilson, C.H., Anderson, T., Bentley, S., Dalrymple-Alford, J., Fowdar, J., Gratten, J., Halliday, G., Henders, A.K., Hickie, I., Kassam, I., Kennedy, M., Kwok, J., Lewis, S., Mellick, G., Montgomery, G., Pearson, J., Pitcher, T., Sidorenko, J., Silburn, P.A., Vallerga, C.L., Visscher, P.M., Wallace, L., Wray, N.R., Xue, A., Yang, J., Zhang, F., 2019. 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 18, 1091–1102. https://doi.org/10.1016/S1474-4422(19)30320-5

Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M., Boone, K., Miller, B.L., Cummings, J., Benson, D.F., 1998. Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. Neurology 51, 1546–1554. https://doi.org/10.1212/WNL.51.6.1546

Pankratz, N., Dumitriu, A., Hetrick, K.N., Sun, M., Latourelle, J.C., Wilk, J.B., Halter, C., Doheny, K.F., Gusella, J.F., Nichols, W.C., Myers, R.H., Foroud, T., DeStefano, A.L., the PSG–PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories, 2011. Copy Number

Variation in Familial Parkinson Disease. PLoS ONE 6, e20988. https://doi.org/10.1371/journal.pone.0020988

Pavelka, L., Rauschenberger, A., Landoulsi, Z., Pachchek, S., May, P., Glaab, E., Krüger, R., on behalf of the NCER-PD Consortium, Acharya, G., Aguayo, G., Alexandre, M., Ali, M., Allen, D., Ammerlann, W., Balling, R., Bassis, M., Beaumont, K., Becker, R., Bellora, C., Berchem, G., Berg, D., Bisdorff, A., Brockmann, K., Calmes, J., Castillo, L., Contesotto, G., Diederich, N., Dondelinger, R., Esteves, D., Fagherazzi, G., Ferrand, J.-Y., Gantenbein, M., Gasser, T., Gawron, P., Ghosh, S., Glaab, Enrico, Gomes, C., De Lope, E.G., Goncharenko, N., Graas, J., Graziano, M., Groues, V., Grünewald, A., Gu, W., Hammot, G., Hanff, A.-M., Hansen, L., Hansen, M., Heneka, M., Henry, E., Herbrink, S., Herenne, E., Herzinger, S., Heymann, M., Hu, M., Hundt, A., Jacoby, N., Lebioda, J.J., Jaroz, Y., Klopfenstein, Q., Krüger, Rejko, Lambert, P., Landoulsi, Zied, Lentz, R., Liepelt, I., Liszka, R., Longhino, L., Lorentz, V., Lupu, P.C., Mackay, C., Maetzler, W., Marcus, K., Marques, G., Marques, T., May, Patrick, Mcintyre, D., Mediouni, C., Meisch, F., Menster, M., Minelli, M., Mittelbronn, M., Mollenhauer, B., Mommaerts, K., Moreno, C., Moudio, S., Mühlschlegel, F., Nati, R., Nehrbass, U., Nickels, S., Nicolai, B., Nicolay, J.-P., Oertel, W., Ostaszewski, M., Pachchek, Sinthuja, Pauly, C., Pauly, L., Pavelka, Lukas, Perquin, M., Lima, R.R., Rauschenberger, Armin, Rawal, R., Bobbili, D.R., Rosales, E., Rosety, I., Rump, K., Sandt, E., Satagopam, V., Schlesser, M., Schmitt, M., Schmitz, S., Schneider, R., Schwamborn, J., Sharify, A., Soboleva, E., Sokolowska, K., Terwindt, O., Thien, H., Thiry, E., Loo, R.T.J., Trefois, C., Trouet, J., Tsurkalenko, O., Vaillant, M., Valenti, M., Boas, L.V., Vyas, M., Wade-Martins, R., Wilmes, P., 2022. Publisher Correction: Age at onset as stratifier in idiopathic Parkinson's disease – effect of ageing and polygenic risk score on clinical phenotypes. npj Parkinsons Dis. 8, 112. https://doi.org/10.1038/s41531-022-00378-9

Peiris, S.P., Landoulsi, Z., Pavelka, L., Schulte, C., Buena-Atienza, E., Gross, C., Hauser, A.-K., Bobbili, D.R., Casadei, N., May, P., Krüger, R., 2023. Accurate long-read sequencing identified GBA variants as a major genetic risk factor in the Luxembourg Parkinson's study (preprint). Genetic and Genomic Medicine. https://doi.org/10.1101/2023.03.29.23287880

Pennings, M., Meijer, R.P.P., Gerrits, M., Janssen, J., Pfundt, R., De Leeuw, N., Gilissen, C., Gardeitchik, T., Schouten, M., Voermans, N., Van De Warrenburg, B., Kamsteeg, E.-J., 2023. Copy number variants from 4800 exomes contribute to ~7% of genetic diagnoses in movement disorders, muscle disorders and neuropathies. Eur J Hum Genet 31, 654–662. https://doi.org/10.1038/s41431-023-01312-0

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., Kircher, M., 2019. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research 47, D886–D894. https://doi.org/10.1093/nar/gky1016

Sanchez-Contreras, M., Heckman, M.G., Tacik, P., Diehl, N., Brown, P.H., Soto-Ortolaza, A.I., Christopher, E.A., Walton, R.L., Ross, O.A., Golbe, L.I., Graff-Radford, N., Wszolek, Z.K., Dickson, D.W., Rademakers, R., 2017. Study of *LRRK2* variation in tauopathy: Progressive supranuclear palsy and corticobasal degeneration: *LRRK2* in Autopsy-Confirmed Tauopathies. Mov Disord. 32, 115–123. https://doi.org/10.1002/mds.26815

Sardi, S.P., Cedarbaum, J.M., Brundin, P., 2018. Targeted Therapies for Parkinson's Disease: From Genetics to the Clinic: Targeted Therapies for PD. Mov Disord. 33, 684–696. https://doi.org/10.1002/mds.27414

Sarihan, E.I., Pérez-Palma, E., Niestroj, L., Loesch, D., Inca-Martinez, M., Horimoto, A.R.V.R., Cornejo-Olivas, M., Torres, L., Mazzetti, P., Cosentino, C., Sarapura-Castro, E., Rivera-Valdivia, A., Dieguez, E., Raggio, V., Lescano, A., Tumas, V., Borges, V., Ferraz, H.B., Rieder, C.R., Schumacher-Schuh, A.F., Santos-Lobato, B.L., Velez-Pardo, C., Jimenez-Del-Rio, M., Lopera, F., Moreno, S., Chana-Cuevas, P., Fernandez, W., Arboleda, G., Arboleda, H., Arboleda-Bustos, C.E., Yearout, D., Zabetian, C.P., Thornton, T.A., O'Connor, T.D., Lal, D., Mata, I.F., Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD), Micheli, F., Gatto, E., Tumas, V., Borges, V., Ferraz, H.B., Rieder, C.R.M., Shumacher-Schuh, A., Santos-Lobato, B.L., Chaná, P., Velez-Pardo, C., Jimenez-Del-Rio, M., Lopera, F., Arboleda, G., Arboleda, H., Orozco, J.L., Moreno, S., Fernandez, W., Arboleda-Bustos, C.E., Fornaguera, J., Guillén, A.H., Acosta, G.T., Chang-Castello, J., Muñoz, B.A., Medina, Alex, Ferrera, A., Martinez-Ramirez, D., Rodriguez, M., Cornejo-Olivas, M., Mazzetti, P., Sarapura, H., Rivera, A., Torres, L., Cosentino, C., Medina, Angel, Viñuela, A., Dieguez, E., Raggio, V., Lescano, A., Amorin, I., 2021. Genome-Wide Analysis of Copy Number Variation in Latin American Parkinson's Disease Patients. Mov Disord 36, 434–441. https://doi.org/10.1002/mds.28353

Schüle, B., Hatchwell, E., Eis, P.S., Langston, J.W., 2015. Comparative genomic hybridization solves a 14-year-old *PARKIN* mystery. Ann Neurol. 78, 663–664. https://doi.org/10.1002/ana.24478

Shu, L., Zhang, Y., Sun, Q., Pan, H., Tang, B., 2019. A Comprehensive Analysis of Population Differences in LRRK2 Variant Distribution in Parkinson's Disease. Front. Aging Neurosci. 11, 13. https://doi.org/10.3389/fnagi.2019.00013

Tang, B., Xiong, H., Sun, P., Zhang, Y., Wang, D., Hu, Z., Zhu, Z., Ma, H., Pan, Q., Xia, J., Xia, K., Zhang, Z., 2006. Association of PINK1 and DJ-1 confers digenic inheritance of early-onset Parkinson's disease. Human Molecular Genetics 15, 1816–1825. https://doi.org/10.1093/hmg/ddl104

The 1000 Genomes Project Consortium, 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65. https://doi.org/10.1038/nature11632

Toft, M., Ross, O.A., 2010. Copy number variation in Parkinson's disease. Genome Med 2, 62. https://doi.org/10.1186/gm183

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674. https://doi.org/10.1101/gr.6861907

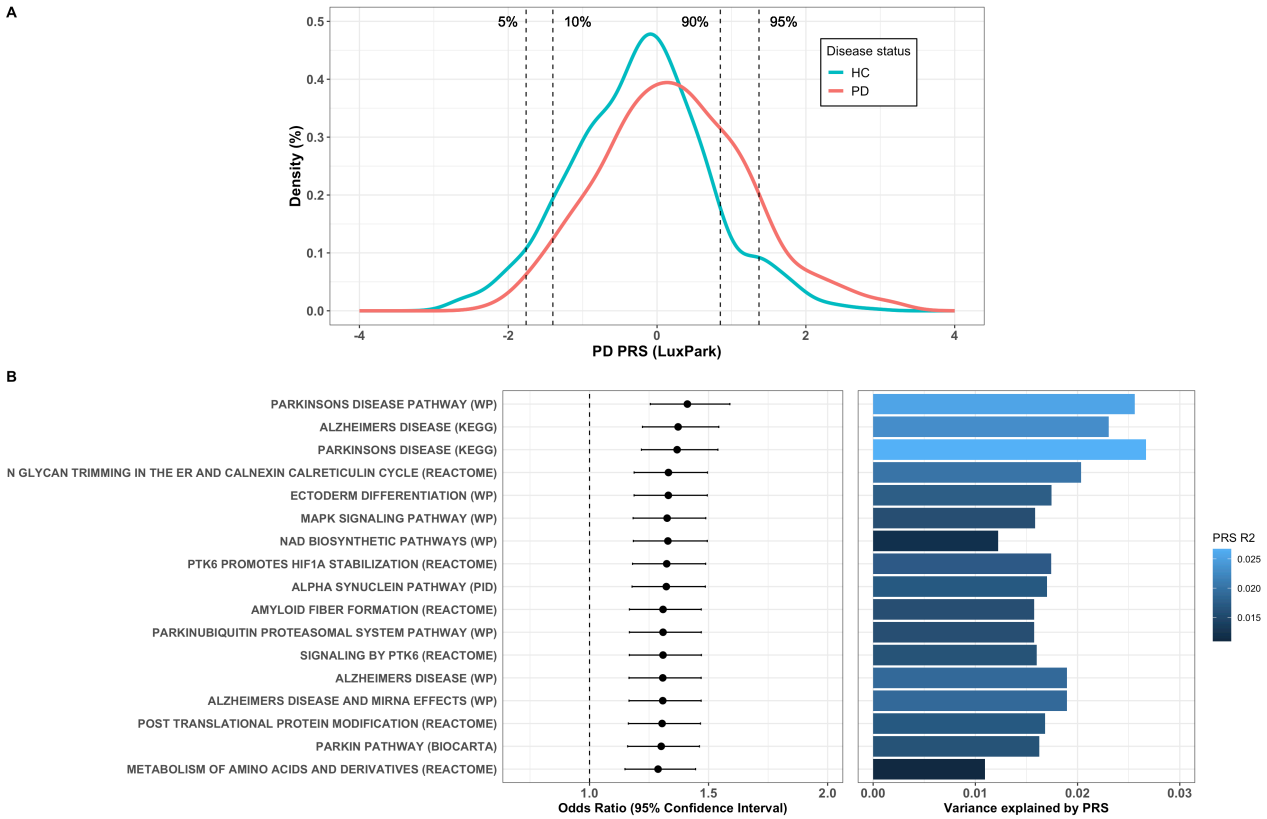
Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164. https://doi.org/10.1093/nar/gkq603

Wen, Y., Zhou, Y., Jiao, B., Shen, L., 2021. Genetics of Progressive Supranuclear Palsy: A Review. JPD 11, 93–105. https://doi.org/10.3233/JPD-202302

Yu, E., Rudakou, U., Krohn, L., Mufti, K., Ruskey, J.A., Asayesh, F., Estiar, M.A., Spiegelman, D., Surface, M., Fahn, S., Waters, C.H., Greenbaum, L., Espay, A.J., Dauvilliers, Y., Dupré, N., Rouleau, G.A., Hassin-Baer, S., Fon, E.A., Alcalay, R.N., Gan-Or, Z., 2021. Analysis of Heterozygous *PRKN* Variants and Copy-Number Variations in Parkinson's Disease. Mov Disord 36, 178–187. https://doi.org/10.1002/mds.28299

Figures

Figure 1: A. Distribution of the polygenic risk score (PRS) between Parkinson's disease (PD) patients and controls. **B.** Forest plots showing PRS odds ratio (OR) and 95% confidence interval for the significant canonical pathways associated with PD risk (left panel) and the estimation of variance explained by PRS (right panel).



Tables

Table 1: Demographic data of cases and healthy controls from the Luxembourg Parkinson's study. Counts, means and percentage were calculated after genotyping data quality controls. PD (Parkinson's disease),PSP (progressive supranuclear palsy), LBD (Lewy body dementia), MSA (Multiple System Atrophy), CBS (corticobasal syndrome) and FTD-P (Fronto-temporal dementia with parkinsonism).

	N	Sex (% male)	Age at assessment(mean±SD)	Age at onset (mean±SD)	Family history of PD (n 1st degree relative, %)
Controls	724	54.2	65.8±11.6		132 (18.2%)
PD	667	68.0	73.0±11.0	62.3±11.8	97 (14.5%)
PSP	50	62.0	76.4±6.7	67.6±7.4	5 (10%)
LBD	25	68.0	77.8±9.8	70.5±9.4	3 (12%)
MSA	13	69.2	75.2±7.7	65.8±8.4	1 (7.7%)
CBS	10	30.0	77.1±7.8	69.2±8.6	1 (10%)
FTD-P	1	0	69	58	0 (0%)

Table2: Rare single nucleotide variants in PD related genes in the Luxembourg Parkinson's study. Rare single nucleotide variants (SNVs) were screened from Neurochip array and validated by Sanger sequencing. We included CliVvar prediction only for Parkinson's disease (benign includes both 'benign' and 'likely benign' prediction, Pathogenic includes both 'pathogenic' and 'likely pathogenic' prediction). CADD (Combined Annotation Dependent Depletion). REVEL (Rare Exome Variant Ensemble Learner).

Gene	c.DNA	Protein	CADD	REVEL	ClinVar prediction for PD	MDSgene pathogenicity	rsid
ATP13A2	c.C35T	p.T12M	21.9	0.497	*NA*	*NA*	rs151117874
	c.C3170T	p.A1057V	25.8	0.07	*NA*	*NA*	rs201610681
	c.G515A	p.R172H	31	0.781	*NA*	*NA*	rs776601823
	c.C746T	p.A249V	10.24	0.473	*NA*	*NA*	rs145515028
	c.A829T	p.S277C	24.2	0.729	*NA*	*NA*	rs538497077
	c.C1073T	p.P358L	23	0.555	*NA*	*NA*	rs757503427
	c.G1229A	p.R410Q	23	0.312	*NA*	*NA*	rs190746040
	c.C1294G	p.L432V	14.16	0.24	*NA*	*NA*	rs149372969
	c.G2771A	p.R924H	32	0.967	*NA*	*NA*	rs564643512
	c.A2836T	p.I946F	22.5	0.292	*NA*	*NA*	rs55708915
	c.G2939A	p.R980H	21.2	0.63	*NA*	*NA*	rs150748722
	c.A3361T	p.T1121S	10.04	0.149	*NA*	*NA*	rs41273151
LRRK2	c.A382G	p.S128G	20.4	0.044	Uncertain significance	*NA*	rs187299177
	c.A784G	p.M262V	0.001	0.013	Uncertain significance	*NA*	rs182233369
	c.C856G	p.L286V	23	0.173	Uncertain significance	*NA*	rs200437744
	c.G1000A	p.E334K	22.4	0.194	Conflicting interpretations of pathogenicity	*NA*	rs78501232
	c.A1004G	p.N335S	10.47	0.053	*NA*	*NA*	rs989570613
	c.G1108A	p.A370T	25.7	0.186	*NA*	*NA*	rs200189771
	c.A2167G	p.I723V	12.73	0.045	Benign	*NA*	rs10878307
	c.G2291A	p.S764N	8.77	0.01	*NA*	*NA*	rs774818561
	c.G2378T	p.R793M	23.5	0.305	Conflicting interpretations of pathogenicity	*NA*	rs35173587
	c.C2594T	p.S865F	23.1	0.149	Benign	*NA*	rs142700458
	c.G2769C	p.Q923H	14.21	0.262	Conflicting interpretations of pathogenicity	*NA*	rs58559150
	c.G2918A	p.S973N	23.4	0.13	*NA*	*NA*	rs75148313
	c.G3451A	p.A1151T	20.4	0.029	Uncertain significance	*NA*	rs74985840
	c.T3477G	p.S1159R	20.7	0.165	*NA*	*NA*	rs200965490
	c.G3683C	p.S1228T	16.12	0.307	Uncertain significance	*NA*	rs60185966
	c.G3974A	p.R1325Q	32	0.553	Conflicting interpretations of pathogenicity	*NA*	rs72546338
	c.C4321T	p.R1441C	23.3	0.727	Pathogenic	pathogenic	rs33939927
	c.C4321A	p.R1441S	22.6	0.66	Pathogenic	*NA*	rs33939927
	c.G4541A	p.R1514Q	22.3	0.1	Benign	*NA*	rs35507033
	c.T4939A	p.S1647T	13.97	0.086	Benign	*NA*	rs11564148
	c.T5606C	p.M1869T	22.8	0.514	Conflicting interpretations of pathogenicity	*NA*	rs35602796
	c.G5822A	p.R1941H	23	0.24	Uncertain significance	*NA*	rs77428810

	c.G6055A	p.G2019S	31	0.97	Pathogenic	pathogenic	rs34637584
	c.G6688A	p.E2230K	20.7	0.049	Uncertain significance	*NA*	rs201317931
	c.C7067T	p.T2356I	19.59	0.154	Conflicting interpretations of pathogenicity	*NA*	rs113511708
	c.T7169C	p.V2390A	14.66	0.165	Uncertain significance	*NA*	rs376230626
	c.G7183A	p.E2395K	23.1	0.168	*NA*	*NA*	rs78964014
	c.G7224A	p.M2408I	14.27	0.055	*NA*	*NA*	rs60545352
	c.G7483A	p.V2495I	16.34	0.022	Benign	*NA*	rs150062967
PARK7	c.G310T	p.A104S	25.7	0.72	*NA*	*NA*	rs774005786
	c.G535A	p.A179T	16.87	0.127	Uncertain significance	*NA*	rs71653622
PINK1	c.A377G	p.Q126R	14.44	0.353	*NA*	*NA*	rs775809722
	c.G836A	p.R279H	26	0.522	Uncertain significance	Possibly pathogenic	rs74315358
	c.A952T	p.M318L	23.9	0.596	Uncertain significance	*NA*	rs139226733
	c.G1015A	p.A339T	23.5	0.386	Conflicting interpretations of pathogenicity	Probably pathogenic	rs55831733
	c.T1106C	p.L369P	26.5	0.83	*NA*	Probably pathogenic	rs1195888869
	c.G1147A	p.A383T	13.23	0.405	Conflicting interpretations of pathogenicity	*NA*	rs45515602
	c.G1231A	p.G411S	20.7	0.429	Conflicting interpretations of pathogenicity	*NA*	rs45478900
	c.G1426A	p.E476K	14.31	0.127	Benign	Benign	rs115477764
	c.G1609A	p.A537T	24.1	0.297	Uncertain significance	*NA*	rs771032673
PRKN	c.101_102del	p.Q34fs	*NA*	*NA*	Pathogenic	*NA*	*NA*
	c.C245A	p.A82E	0.765	0.559	Conflicting interpretations of pathogenicity	*NA*	rs55774500
	c.G500A	p.S167N	15.67	0.164	Benign	*NA*	rs1801474
	c.A574C	p.M192L	22.5	0.519	Benign	*NA*	rs9456735
	c.C766T	p.R256C	32	0.811	Uncertain significance	Probably pathogenic	rs150562946
	c.C823T	p.R275W	29.5	0.747	Pathogenic	Definitely pathogenic	rs34424986
VPS35	c.A110G	p.N37S	21.1	0.054	Uncertain significance	*NA*	rs777006799
	c.G151A	p.G51S	22.6	0.236	Benign	*NA*	rs193077277

Table 3: Number and phenotypes of rare variant carriers in PD related genes. (* homozygous carriers, One PD patient was homozygous for *LRRK2* p.I723V and one for *PINK1* p.L369P)

n PD	n atypical parkinsonism	n Controls	*LRRK2*	*VPS35*	*PRKN*	*PINK1*	*PARK7*	*ATP13A2*
2	0	0	p.A1151T	*NA*	*NA*	*NA*	*NA*	*NA*
0	1 PSP	0	p.A370T	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	1	p.E2230K	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	2	p.E334K	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.G2019S	*NA*	p.A82E	*NA*	*NA*	*NA*
3	0	1	p.G2019S	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.G2019S, p.Q923H	*NA*	*NA*	*NA*	*NA*	*NA*
4	1 CBS	1	p.I723V*	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.L286V	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	2	p.M1869T	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	2	p.E2395K	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.M2408I	*NA*	p.R256C	*NA*	*NA*	*NA*
0	1 PSP	0	p.M262V	*NA*	*NA*	p.A383T	*NA*	*NA*
1	0	0	p.N335S	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	2	p.R1325Q		*NA*	*NA*	*NA*	*NA*
1	0	1	p.R1441C		*NA*	*NA*	*NA*	*NA*
1	0	0	p.R1441S		p.S167N	*NA*	*NA*	*NA*
1	0	0	p.R1514Q	p.N37S	*NA*	*NA*	*NA*	*NA*
5	1 CBS	2	p.R1514Q	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	1	p.R1941H	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	1	p.R793M	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.S1159R	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	1	p.S1228T	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	1	p.S128G	*NA*	*NA*	*NA*	*NA*	*NA*
3	0	0	p.S1647T	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.S764N	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.S865F	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.S973N	*NA*	*NA*	*NA*	*NA*	*NA*
0	1 DLB	3	p.T2356I	*NA*	*NA*	*NA*	*NA*	*NA*
1	0	0	p.V2390A	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	1	p.V2495I	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	1	*NA*	*NA*	p.R256C	*NA*	*NA*	p.R172H
1	0	0	*NA*	*NA*	*NA*	*NA*	*NA*	p.A1057V
1	0	2	*NA*	*NA*	*NA*	*NA*	*NA*	p.A249V
2	0	2	*NA*	*NA*	*NA*	*NA*	*NA*	p.I946F
1	0	2	*NA*	*NA*	*NA*	*NA*	*NA*	p.L432V
0	0	1	*NA*	*NA*	*NA*	*NA*	*NA*	p.P358L
1	0	0	*NA*	*NA*	*NA*	*NA*	*NA*	p.R410Q
1	0	0	*NA*	*NA*	*NA*	*NA*	*NA*	p.R924H
0	0	1	*NA*	*NA*	*NA*	*NA*	*NA*	p.R980H
0	0	1	*NA*	*NA*	*NA*	*NA*	*NA*	p.S277C
1	0	0	*NA*	*NA*	*NA*	p.A339T	*NA*	p.T1121S
0	0	2	*NA*	*NA*	*NA*	*NA*	*NA*	p.T12M
1	0	1	*NA*	*NA*	*NA*	*NA*	p.A104S	*NA*
1	0	1	*NA*	*NA*	*NA*	*NA*	p.A179T	*NA*
1	0	2	*NA*	*NA*	*NA*	p.A339T	*NA*	*NA*
2	0	1	*NA*	*NA*	*NA*	p.A383T	*NA*	*NA*
1	1 DLB	0	*NA*	*NA*	*NA*	p.A537T	*NA*	*NA*
1	0	0	*NA*	*NA*	*NA*	p.E476K	*NA*	*NA*
0	1 PSP / 1 DLB	0	*NA*	*NA*	*NA*	p.G411S	*NA*	*NA*
1	0	0	*NA*	*NA*	*NA*	p.L369P*	*NA*	*NA*
0	0	1	*NA*	*NA*	p.A82E	p.M318L	*NA*	*NA*
1	1 DLB	2	*NA*	*NA*	*NA*	p.M318L	*NA*	*NA*
1	0	0	*NA*	*NA*	*NA*	p.Q126R	*NA*	*NA*
0	0	1	*NA*	*NA*	*NA*	p.R279H	*NA*	*NA*
0	0	2	*NA*	*NA*	p.R256C	*NA*	*NA*	*NA*
3	0	1	*NA*	*NA*	p.A82E	*NA*	*NA*	*NA*
0	0	1	*NA*	*NA*	p.M192L	*NA*	*NA*	*NA*
1	0	1	*NA*	*NA*	p.Q34fs	*NA*	*NA*	*NA*
2	1 PSP	2	*NA*	*NA*	p.R275W	*NA*	*NA*	*NA*
0	0	1	*NA*	*NA*	p.S167N	*NA*	*NA*	*NA*
1	0	1	*NA*	p.G51S	*NA*	*NA*	*NA*	*NA*

Table 4: Summary of CNV calls from the Luxembourg Parkinson's study. Details of CNVs called following QC steps in cases and controls. CNV (Copy number variants), PD (Parkinson's disease), PSP (progressive supranuclear palsy), LBD (Lewy body dementia), MSA (Multiple System Atrophy), CBS (corticobasal syndrome) and FTDP (Fronto-temporal dementia with parkinsonism), Kb (Kilobases).

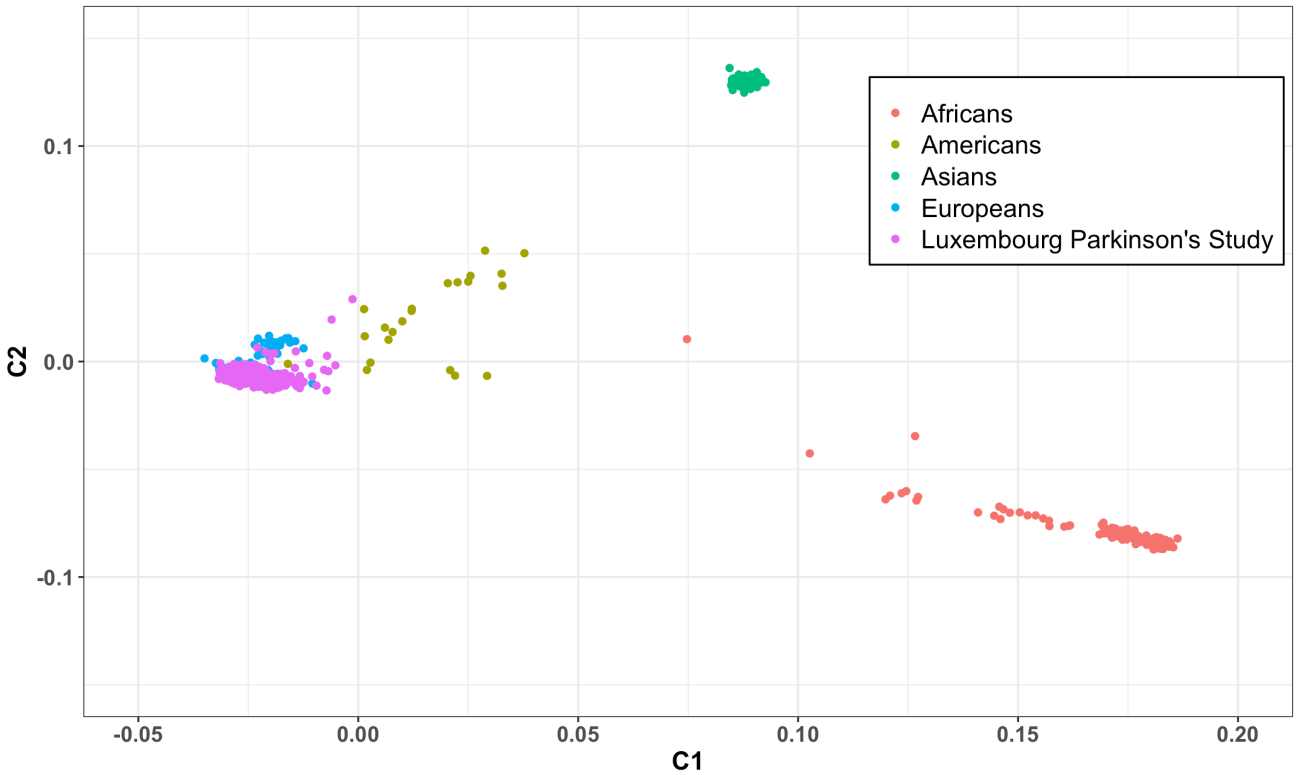
	Controls	PD	PSP	LBD	MSAP	CBS	FTDP
Number of samples	728	656	51	27	13	9	1
CNV carriers (n, %)	373 (51.2 %)	322 (49.0 %)	21 (41.1 %)	15 (55.5 %)	4 (30.8 %)	4 (44.4 %)	0 (0 %)
Number of CNVs	544	480	25	18	7	5	0
Duplication	377	323	17	12	6	2	0
Deletions	167	157	8	6	1	3	0
CNVs per sample (Mean, SD)	1.8 (0.9)	1.8 (0.9)	1.2 (0.4)	1.3 (0.7)	2.1 (0.9)	1.4 (0.5)	0
Mean size of CNVs (Kb, Mean, SD)	284 (343)	283 (335)	338 (435)	307 (326)	225 (143)	260 (348)	0
Number of SNPs per CNV (Mean, SD)	51.8 (61.6)	55.2 (118)	51.8 (61.6)	55.6 (57.6)	32.3 (12.8)	36.8 (15.6)	0
filtered out CNVs	11875	10598	1069	350	230	95	3
filtered out duplication	6472	5896	462	95	136	62	2
filtered out deletions	5403	4702	607	255	94	33	1

Table 5: CNVs detected in any of the PD genes in the Luxembourg Parkinson's study. CN: Copy Number, CN=1 (one copy deletion) and CN=3 (one copy duplication). DGV frequency: the number of individuals carrying overlapping CNVs in DGV (Database of Genomic Variants). N spanning gnomAD_SV: number of samples carrying complete overlapping CNVs with the same copy number in gnomAD-SV (the genome aggregation database-structural variants). gnomAD_SV Freq: the highest frequency of the complete overlapping CNVs. AAA (Age at assessment for healthy controls) AAO (Age at Onset for PD. cases). MLPA Multiplex ligation-dependent probe amplification.

CNV	Length (base)	n SNPs	CN	Gene	CNV region in *PRKN*	DGV Freq	n spanning gnomAD_ SV	gnomAD_ SV Freq	n Sample	Diagnostic	AAA /AAO	MLPA	MLPA results
chr6:162724247-162855426	131179	30	3	*PRKN*	intronic (exon 2 – exon3)	0	2	9.20E-05	4	Control	60	no	*NA*
										Control	50	no	*NA*
										Control	53	no	*NA*
										PD	39	yes	exon2 one copy duplication
chr6:161692286-161772949	80663	30	3	*PRKN*	exon 12	0	0	0	1	Control	69	no	*NA*
chr6:162199157-162875706	676549	153	1	*PRKN*	exon 2-7	0	73	0.14	1	PD	61	no	*NA*
chr6:162279763-162406957	127195	37	1	*PRKN*	exon 6	0	4	0.01	1	PD	62	no	*NA*
chr6:162305402-162674093	368691	90	1	*PRKN*	exon 4-6	0	22	0.14	1	Control	35	no	*NA*
chr6:162445941-162513967	68026	21	1	*PRKN*	exon 5	0	2	0.006	1	PD	63	no	*NA*
chr6:162541706-162750426	208720	55	1	*PRKN*	exon 3-4	0	23	0.14	1	PD	65	no	*NA*
chr6:162646892-163007394	360502	91	3	*PRKN*	exon 2-3	0	10	0.001	1	PD	66	no	*NA*
chr6:162653609-163029097	375488	60	3	*PRKN*	exon 2-3	0	11	0.001	1	PD	18	yes	exon1-4 one copy duplication
chr6:162664364-162750426	86062	29	1	*PRKN*	exon 3	0	9	0.008	1	PD	56	no	*NA*
chr6:162724247-162889975	165728	57	1	*PRKN*	exon 2	0	27	0.008	1	PD	75	no	*NA*
chr6:162736336-163054293	317957	58	3	*PRKN*	exon 2	0	10	0.001	1	Control	64	no	*NA*
chr6:162744935-162914986	170051	52	3	*PRKN*	exon 2	0	4	0.001	1	PD	69	yes	exon1 two copies duplication
chr6:162747573-162855426	107853	22	1	*PRKN*	exon 2	0	15	0.001	1	PD	42	yes	exon2 one copy deletion
chr6:162945539-163176151	230612	29	3	*PRKN, PACRG*	exon 1	0	2	4.00E-04	1	PD	52	yes	exon1 one copy duplication

2.3. Supplementary material

Supplementary Figures:



Supplementary Figure 1. PCAs of the Luxembourg Parkinson's Study individuals (in purple) and the 1000 Genomes population for comparison (color-coded by different populations).

Supplementary Table 1: Canonical pathways significantly associated with PD risk in the Luxembourg Parkinson's study. PRS R2: variance explained by polygenic risk score, SE: standard error, P-value (after Bonferroni correction), N SNP: Number of SNPs included in each gene-set analysis

Gene set	N genes	Beta	SE	P-value	PRS R2	N SNPs
PARKINSONS DISEASE PATHWAY (WP)	71	0.344278	0.060147	3.10E-05	0.025608	16
ALZHEIMERS DISEASE (KEGG)	166	0.316265	0.059383	3.10E-04	0.023056	11
PARKINSONS DISEASE (KEGG)	129	0.313221	0.059654	4.60E-04	0.026692	42
N GLYCAN TRIMMING IN THE ER AND CALNEXIN CALRETICULIN CYCLE (REACTOME)	34	0.286283	0.058815	3.40E-03	0.020331	356
ECTODERM DIFFERENTIATION (WP)	139	0.285898	0.05872	3.40E-03	0.017453	5
MAPK SIGNALING PATHWAY (WP)	244	0.282071	0.058492	4.30E-03	0.015836	33
NAD BIOSYNTHETIC PATHWAYS (WP)	22	0.284523	0.059372	4.90E-03	0.012237	9
PTK6 PROMOTES HIF1A STABILIZATION (REACTOME)	6	0.280816	0.058907	5.90E-03	0.017434	7
ALPHA SYNUCLEIN PATHWAY (PID)	31	0.279482	0.05915	7.10E-03	0.017014	17
AMYLOID FIBER FORMATION (REACTOME)	105	0.269046	0.058621	1.40E-02	0.015743	4
PARKIN-UBIQUITIN PROTEASOMAL SYSTEM PATHWAY (WP)	71	0.269046	0.058621	1.40E-02	0.015743	4
SIGNALING BY PTK6 (REACTOME)	53	0.268987	0.058748	1.50E-02	0.016014	5
ALZHEIMERS DISEASE (WP)	261	0.268217	0.058881	1.60E-02	0.018947	7
ALZHEIMERS DISEASE AND MIRNA EFFECTS (WP)	320	0.268217	0.058881	1.60E-02	0.018947	7
POST TRANSLATIONAL PROTEIN MODIFICATION (REACTOME)	1419	0.26604	0.059004	2.00E-02	0.016832	66
PARKIN PATHWAY (BIOCARTA)	8	0.263226	0.058883	2.40E-02	0.016246	10
METABOLISM OF AMINO ACIDS AND DERIVATIVES (REACTOME)	366	0.252922	0.058534	4.90E-02	0.010947	6

Chapter 3 : Association of polygenic risk score and pathogenic *GBA1* variants in the Luxembourg Parkinson's study

Association of polygenic risk and severity of *GBA1* pathogenic variants in Parkinson's Disease

Emadeldin Hassanin[1,2*], Zied Landoulsi[1*], Sinthuja Pachchek[1*], Peter Krawitz[2], Rejko Krüger[1,3,4], Carlo Maj[5], Patrick May[1#], and Dheeraj Reddy Bobbili[1#], on behalf of the NCER-PD Consortium

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Sur-Alzette, Luxembourg
[2]Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany
[3]Transversal Translational Medicine, Luxembourg Institute of Health, Strassen, Luxembourg
[4]Department of Neurology, Centre Hospitalier de Luxembourg, Strassen, Luxembourg
[5]Centre for Human Genetics, University of Marburg, Marburg, Germany
[*] These authors contributed equally to this work and shared the first authorship
[#] These authors contributed equally to this work and shared the last authorship

Corresponding author:
Dheeraj R. Bobbili, Ph.D.: dheeraj.bobbili@uni.lu

Project description

Our research primarily centered on investigating the correlation between the PRS and the severity of pathogenic *GBA1* variants concerning PD patients.

Contributions

My contributions to this research were considerable and encompassed multiple critical aspects. Firstly, as an equal author of this manuscript, I actively contributed to the article's writing process, which clearly explained our results. I also led the efforts to find and evaluate rare variants in the dataset by conducting an in-depth study of the genomic data. In addition, I actively participated in the analysis of the results, offering insightful comments that contributed to a better understanding of the significance of our findings.

E.H.: Research Project: Conception, Organization, Execution; Statistical Analysis: Design, Execution; Manuscript Preparation: Writing of the First Draft, Review and Critique. S.P.: Research Project: Organization, Execution; Data: acquisition Curation; Manuscript Preparation: Writing of the First Draft, Review and Critique. Z.L.: Research Project: Organization; Data: acquisition Curation; Manuscript Preparation: Review and Critique. P.K.: Manuscript Preparation: Review and Critique. R.K.: Data: acquisition; Manuscript Preparation: Review and Critique. C.M.: Research Project: Conception, Organization, Execution; Manuscript Preparation: Review and Critique. P.M.: Research Project: Conception, Organization, Execution; Statistical Analysis: Design; Data: acquisition Curation; Manuscript Preparation: Writing of the First Draft, Review and Critique. D.R.B.: Research Project: Conception, Organization, Execution; Statistical Analysis: Design; Manuscript Preparation: Writing of the First Draft, Review and Critique.

The study was supervised by Prof. Dr. Rejko Küger, Dr. Patrick May and Dr. Dheeraj Reddy Bobbili.

3.1. Introduction

In recent years, GWAS have led to the discovery of an increasing number of risk loci associated with PD. Although each of these loci individually exerts only a small effect on disease risk, the PRS combines the cumulative impact of multiple genetic variations to assess an individual's susceptibility to a specific trait or disease. To determine a score that reflects an individual's overall genetic risk for a given disease, thousands of genome-wide SNPs need to be analyzed (Nalls et al., 2019a). Recent research has shown that PRS can provide important information about the genetic architecture of PD. Higher PRS suggests a higher genetic susceptibility to the disease, while a lower score indicates a lower genetic risk.

Recent studies have established a strong association between PD-PRS and PD risk. PD-PRS correlated with Age at onset (AAO) (Ibanez et al., 2017), motor progression (Pihlstrøm et al., 2016) and cognitive symptoms (Iwaki et al., 2019).

GBA1 variants are the most frequent genetic risk factor for PD, which ranges from 10% to 30%, and age is the penetrance factor (Anheim et al., 2012). Having previously established the classification of *GBA1* variants identified in the Luxembourg Parkinson's study, the present study aims to explore the combined effect of PD specific PRS and the severity of pathogenic *GBA1* variants in predicting the risk of PD. In this study, we generated the PRS using 42 SNPs identified in a meta-analysis of PD (Chang et al., 2017). A study was carried out with the *GBA1* carriers and the PRS, and it was shown that *GBA1* penetrance and PRS are linked to common polymorphisms in the PD risk loci *SNCA* and *CTSB* (Blauwendraat et al., 2020b). The identification of factors influencing the penetrance of *GBA1* carriers could be used to stratify carriers and for personalized preventative treatment.

PRSs are still an emerging field of research and their application in clinical settings is not yet widespread. PRSs were exclusively developed and evaluated on populations with European ancestry and they only consider the effects of common genetic variants, leaving out rare, CNVs or indels variants, which remain unaccounted for.

Similar to many other complex diseases, PD results from a mix of genetic, environmental, and lifestyle factors, and the predictive power of PRS is not yet complete. To employ PRSs to predict PD risk and potentially guide personalized prevention and treatment strategies, further validation and improvement through additional research are needed.

3.2. Copy of the preprint manuscript

Association of polygenic risk and severity of GBA1 pathogenic variants in Parkinson's Disease

Emadeldin Hassanin[1,2*], Zied Landoulsi[1*], Sinthuja Pachchek[1*], Rejko Krüger[1,3,4], Peter Krawitz[2], Carlo Maj[5], Patrick May[1#], and Dheeraj Reddy Bobbili[1#], on behalf of the NCER-PD Consortium

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Sur-Alzette, Luxembourg
[2]Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany
[3]Transversal Translational Medicine, Luxembourg Institute of Health, Strassen, Luxembourg
[4]Department of Neurology, Centre Hospitalier de Luxembourg, Strassen, Luxembourg
[5]Centre for Human Genetics, University of Marburg, Marburg, Germany

[*] These authors contributed equally to this work and shared the first authorship
[#] These authors contributed equally to this work and shared the last authorship

Corresponding author:

Dheeraj R. Bobbili, Ph.D.
Luxembourg Centre for Systems Biomedicine, University Luxembourg
6 avenue du Swing, L-4367 Belvaux, Luxembourg
Email address: dheeraj.bobbili@uni.lu

Keywords

Parkinson's disease; polygenic risk score; GBA1; genetics

Abstract

Background: It is well established that pathogenic variants in the *GBA1*, when present in a heterozygous state, are a prevalent risk factor for Parkinson's Disease (PD). Despite this knowledge, the variables that impact the manifestation of PD in these carriers remain unclear. Our study intends to assess the role of polygenic risk scores (PRSs) in affecting the likelihood of PD onset among carriers of *GBA1* pathogenic variants (*GBA1*$_{PVs}$), spanning different degrees of pathogenicity.

Methods: PRSs for all participants in our study were determined using known genome-wide significant loci and then categorized into three classes: "High", "Intermediate" and "Low" based on their tertiles. The *GBA1*$_{PVs}$ were further classified into "severe", "mild" and "risk" categories according to their reported and predicted severity. We then stratified each participant according to their PRS and *GBA1*$_{PVs}$ carrier status to investigate their influence on PD risk. We used logistic regression models to calculate odds ratios (OR). Our analyses were conducted on data from the Luxembourgish Parkinson Study, which served as a discovery cohort, and we subsequently confirmed our findings in a separate validation cohort.

Results: In the discovery cohort, we noted that carriers of severe, mild or risk *GBA1*$_{PVs}$ who fell into the highest PRS tertile showed a significantly higher OR (OR: 1.53; 95% CI, 1.43-1.63) for PD, when compared to carriers within the lowest PRS tertile (OR: 1.14; 95% CI, 1.04-1.24). Intriguingly, non-carriers within the discovery cohort possessing a high PRS demonstrated an OR of 1.22 (95% CI, 1.20-1.23), a risk level equivalent to that of risk *GBA1*$_{PVs}$ carriers having an intermediate PRS (OR: 1.17; 95% CI, 0.97-1.34). When we examined the replication cohort, we identified comparable trends.

Conclusions: Our results underscore the significant role that PRSs play in modulating PD risk among carriers of risk *GBA1*$_{PVs}$. It appears that carriers with severe *GBA1*$_{PVs}$ are generally at a higher risk of developing PD, irrespective of their PRS status. However, it is crucial to mention that these findings necessitate confirmation through further large-scale investigations.

KEY MESSAGES

WHAT IS ALREADY KNOWN ON THIS TOPIC ?

- There's significant variability in the incidence of Parkinson's disease (PD) among individuals carrying *GBA1* pathogenic variants (*GBA1*$_{PVs}$).
- Polygenic risk scores (PRSs) may be useful in stratifying PD risk, especially in individuals with common *GBA1* variants (p.E365K, p.T408M, or p.N409S).

WHAT THIS STUDY ADDS

- Our research extends the current understanding by considering a wider range of *GBA1*$_{PVs}$ by providing a focused analysis of the severity of *GBA1*$_{PVs}$ (risk, mild, severe) and including both common and rare variants.
- Our findings indicate that PRS may influence PD risk among *GBA1*$_{PVs}$ carriers, particularly for risk and mild variants, whereas, severe variants consistently exhibit a higher PD risk (twofold or greater), regardless of PRS status.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

The combined understanding of PRS and *GBA1*$_{PVs}$ severity may improve the accuracy of predicting PD risk in individuals carrying *GBA1*$_{PVs}$, thereby informing future research, clinical practice, and policy decisions.

Introduction

Parkinson's Disease (PD) is a complex disorder in which genetics plays a crucial role in the disease etiology[1]. The presence of heterozygous variants in the *GBA1* gene, which encodes for a lysosomal enzyme called glucocerebrosidase, has emerged as a prevalent risk factor for PD, with an estimated risk of developing PD between 10% to 30% by the age of 80[2]. *GBA1 variant* carriers are linked to an earlier onset of symptoms and a more severe clinical outcome in PD patients[3]. Apart from the known pathogenic mutations linked in the homozygous state to Gaucher's disease (GD), there are several "non-pathogenic" variants of the *GBA1* gene that are not associated with GD but are observed at higher frequencies in individuals with PD[4]. These include the commonly observed p.E365K and p.T408M missense variants, classified as risk variants for PD [5].

The effort to classify *GBA1* variants is beset with complexity, given the numerous existing classification systems, each developed by separate research teams. This complexity is amplified by the classification frameworks suggested by Höglinger et al. in 2022a and the *GBA1*-PD disease browser[6]. The task is rendered even more daunting due to the multitude of criteria and factors that these diverse systems utilize in their categorization of *GBA1* variations.

Furthermore, the contribution of common genetic variants to the development of PD has been previously highlighted [7]. Genome-wide association studies (GWAS) have identified an increasing number of common PD risk loci [8,9]. While each of these common variants alone has a relatively modest effect on overall disease risk, they confer a significantly high disease risk when combined in polygenic risk scores (PRSs) [10]. PRSs are generated on an individual basis and, in particular, when individuals in the highest PRS percentile are compared with the rest of the population, the effect can be similar to that of monogenic variants [11]. Recently, the PRS has gained importance as a risk stratification tool in many complex diseases [12,13]. In PD, studies have been conducted on PRS and its effect on the disease status [14], age of onset [15], and carrier status [16].

The polygenic background has been found to modulate the risk of PD in carriers of *GBA1* variants (*i.e.,* p.E365K, p.T408M or p.N409S) [17] and in carriers of the *LRRK2* p.G2019S variant [16]. PRS was found to be significantly associated with increased penetrance of the *LRRK2* p.G2019 variant, especially in young carriers. It has been reported that common polymorphisms in the PD risk loci *SNCA* and *CTSB* are associated with the penetrance of *GBA1* and PRS and may even modify the penetrance[18].

Building on previous research, our study extends the analysis of the effect of PRS on carriers of *GBA1* pathogenic variants ($GBA1_{PVs}$) within the Luxembourg Parkinson's Study. We subsequently validated our findings using data from the UK Biobank cohort. Our investigation focused on understanding how the synergistic effect of PRS and the severity of $GBA1_{PVs}$ (risk, mild or severe) influences the risk of developing PD. Crucially, this study also proposes a comprehensive framework for classifying *GBA1* variants to standardize interpretations and facilitate future research.

Methods

The Luxembourg Parkinson's study

As a discovery cohort, we used the Luxembourg Parkinson's Study[5,19], a longitudinal population case-control PD study from Luxembourg. Genotyping was performed using the NeuroChip platform[16], and *GBA1* variants were identified using the *GBA1*-targeted PacBio method[21]. To establish the diagnosis of PD, the neurological examination was conducted based on the diagnostic guidelines of the United Kingdom Parkinson's Disease Society Brain Bank[22]. Ethical approval for the Luxembourg Parkinson's study was obtained from the National Ethics Board (CNER Ref: 201407/13 and 202304/03) and the Data Protection Committee (CNPD Ref: 446/2017). Follow-up visits were performed regularly, with PD patients being reassessed annually and healthy controls being reassessed every four years. The final diagnosis was determined based on the last visit. Written informed consent was obtained from all participants.

To ensure the integrity of our study cohort, we implemented several exclusion criteria. First, individuals with non-European ancestry, carriers of pathogenic variants or copy number variants (CNVs) in genes associated with PD (based on the MDSGene[23] related gene list, excluding *GBA1*). We selected only unrelated individuals, and within each family, we selected only one proband. Priority was given to the patient with the earliest age of onset (AAO). In addition, HC with first-degree relatives with PD were excluded. Then, we also excluded individuals who carried variants of uncertain significance (VUS) as described below or synonymous variants in the *GBA1* gene.

The UK biobank cohort

As an independent replication cohort, we used the data from the UK Biobank (UKB). UKB is a long-term prospective study recruiting volunteers mainly from England, Scotland, and Wales, with more than 500,000 participants aged between 40-69 years. The dataset is available for research purposes, and all participants provided documented consent [24]. All participants were genotyped by using the UKB Axiom Array. Nearly 850,000 variants were genotyped and >90 million variants were imputed using the Haplotype Reference Consortium and UK10K + 1000 Genomes reference panels. Participants underwent whole exome sequencing (WES) using the IDT xGen Exome Research Panel v1.0 [25]. PD diagnosis was based on participant self-reported or the International Classification of Diseases (ICD-10) diagnosis codes. This included individuals with self-reported code 1262 or ICD-10 code of G20 in hospital records. We only included individuals with both WES and genotype data. We included only individuals who were identified

as belonging to European ancestry by the UK Biobank. We excluded outliers with putative sex chromosome aneuploidy (field 22019), high heterozygosity or missing genotype rates (field 22027), and discordant reported versus genotypic sex (field 22001). The analysis was restricted to unrelated individuals to the second degree. We excluded one from each pair of related individuals if the genetic relationship was closer than the second degree, defined as kinship coefficient > 0.0884 as calculated by the UKB. Characteristics of both discovery and replication cohorts are shown in Table 1.

Annotation

All the *GBA1* variants were annotated by ANNOVAR [26] using the RefSeq gene annotation, the dbNSFP[27] prediction and conservation scores, the genome-wide CADD scores[28], the REVEL score[29], HGMD[30] and ClinVar[31].

All the *GBA1* variants were numbered according to the current variant nomenclature guidelines (http://varnomen.hgvs.org). The numbering system followed the guidelines based on the primary translation product (NM_001005742), which includes the 39-residue signal peptide. This standardized approach to variant annotation ensured consistency and facilitated accurate interpretation of the genetic variations within the *GBA1* gene.

A framework to classify *GBA1* variants

We restricted our analyses to exonic and "splice-site" variants excluding synonymous variants. We further subclassified the variants by their severity of increased risk into 'risk', *'mild'*, and *'severe'* impact on PD:

1. 'severe' $GBA1_{PVs}$ are defined as variants belonging to any of the following categories:
    a. Loss-of-function (LoF) variants are defined as frameshift indels or "splice site (+/- 2 base pair)" variants.
    b. Missense variants annotated as "Pathogenic/Likely Pathogenic" in ClinVar and disease-causing mutation (DM) in HGMD with REVEL > 0.7 and CADD > 20.
2. 'mild' $GBA1_{PVs}$ are defined as pathogenic or conflicting interpretations of pathogenicity in ClinVar with DM in HGMD with REVEL > 0.7 or CADD > 20, and
3. 'risk' $GBA1_{PVs}$ are common variants annotated as pathogenic or conflicting interpretations of pathogenicity in ClinVar.

We considered severe, mild, and risk variants as $GBA1_{PVs}$.

Polygenic risk score

To generate the PRS, we used the summary statistics from a list of 42 SNPs discovered from a previous meta-analysis of PD [32]. Common variants within the *GBA1* locus, along with the *LRRK2* p.G2019, were excluded for the PRS calculation.

There were no overlapping SNPs between the *GBA1*$_{PVs}$ and the 42 GWAS SNPs. Hence, we proceeded with the 42 SNPs for PRS calculation using PRsice2 [33], taking into account allele-flipping and using the '*–no-regress'* and *'–no-clumping'* options along with the default parameters.

Statistical analysis

To investigate the potential association between PRS, *GBA1*$_{PVs}$ carrier status and disease risk, we conducted logistic regression using disease occurrence as the outcome variable. Individuals were stratified based on their PRS and *GBA1*$_{PVs}$ carrier status. First, individuals were divided into three groups according to the distribution of their PRSs. Those with a PRS in the lower tertile were assigned to the 'low' PRS group, while those with a PRS in the upper tertile were assigned to the 'high' PRS group. Those in the middle tertile were assigned to the 'intermediate' PRS group.

For the two cohorts, the predicted odds ratios (ORs) were derived from a logistic regression model using the formula below, conditioning on covariates such as sex, age, and the first four principal components, similar to previous studies[34–36]. To assess the combined effect of PRS and *GBA1*$_{PVs}$ carriers, we reclassified the three PRS groups into six groups based on carrier status. Carriers were defined as positive (and encoded as *GBA1*$_{PVs(+)}$ $Car = pos$) or negative (encoded as *GBA1*$_{PVs(-)}$) whether the individual carried a *GBA1*$_{PVs}$ or not. For example, individuals with high PRS and carrying a *GBA1*$_{PVs}$ are encoded as PRS$_{high}$*GBA1*$_{PVs(+)}$. We then fitted the following logistic regression model:

$$logit\big(P(Y=1)\big) = \beta_0 + \beta_{PRS_{low}GBA1_{\mathrm{PVs(+)}}}\, PRS_{low}GBA1_{\mathrm{PVs(+)}} +$$
$$\beta_{PRS_{low}GBA1_{PVs(-)}}\, PRS_{low}GBA1_{PVs(-)} + \beta_{PRS_{int}GBA1_{PVs(-)}}\, PRS_{int}GBA1_{\mathrm{PVs(+)}} +$$
$$\beta_{PRS_{high}GBA1_{PVs(-)}}\, PRS_{high}GBA1_{PVs(-)} + \beta_{PRS_{high}GBA1_{\mathrm{PVs(+)}}}\, PRS_{high}GBA1_{\mathrm{PVs(+)}} + \beta_{sex}\, sex +$$
$$\beta_{age}\, age + \sum_{k=1}^{4} \beta_{PC_k}\, PC_k$$

Then, we further categorized the *GBA1*$_{PVs}$ carriers' participants into three groups: (1) severe *GBA1*$_{PVs}$ carriers (2) mild *GBA1*$_{PVs}$ carriers and 3) risk *GBA1*$_{PVs}$ carriers. This classification allowed to evaluate the association between different levels of *GBA1*$_{PVs}$ severity and the risk of developing the disease of interest. The reference category is then given by non-carriers with intermediate PRS. We used R v4.2.2 for all statistical analyses.

Results

For the discovery cohort, a total of 1142 individuals were selected, including 631 PD with a mean of AAO of 62.6 years, and 511 HC with a mean of AAA of 71.5 years. For the validation cohort, we restricted our analysis to the 179,095 individuals from the UKB for whom WES and genotyping data were available. Overall, we identified 92 individuals carrying *GBA1*$_{PVs}$ (nPD=68 and nHC=24) in the discovery cohort (Table 1 and Supplementary Table S.1).

Descriptive statistics of the study population after the filtering steps are shown in Table 1. In the replication cohort, we identified 770 PD patients with a mean age at onset (AAO) of 64.6 years. We considered 178,325 participants as controls with a mean age at assessment (AAA) of 64.1 years. We identified 8,544 carriers of 33 *GBA1*$_{PVs}$ (nPD=63 and nHC=8481) in the replication cohort (Table 1 and Supplementary Table S.1).

A complete overview of the study framework is shown in Figure 1. A detailed classification of the variants and their associated classification according to our study and previous studies have been provided in Supplementary 2.

Stratification of participants according to the PD prevalence and *GBA1* carrier status

We calculated the PRS using a panel of 42 SNPs to investigate the influence of the genetic background on PD risk in PD *GBA1*$_{PVs}$ carriers. Our analysis revealed that the PRS was significantly higher in PD patients compared to healthy controls in both cohorts ($P < 0.01$).

We further investigated the association between PRS and rare *GBA1* on the risk of developing PD. ORs were calculated across PRS tertiles for both non-carriers and carriers in both cohorts.

In the Luxembourg Parkinson's Study which served as the discovery cohort, *GBA1*$_{PVs}$ carriers in the highest PRS tertile had a higher OR for getting PD (OR :1.54; 95% CI, 1.44-1.64), compared to carriers in the lowest PRS tertile (OR :1.14; 95% CI, 1.04-1.24).

In the UKB cohort, we observed the same trend where *GBA1*$_{PVs}$ carriers in the highest PRS tertile had a higher OR for getting PD (OR: 2.19; 95% CI, 1.91-2.51), compared to carriers in the lowest PRS tertile (OR: 1.01; 95% CI, 0.85-1.20) (Figure 2).

To gain a deeper understanding of how disease severity is influenced, we further examined the combined effect of PRS and the severity types of *GBA1*$_{PVs}$, which were categorized into three groups: severe, mild, and risk variants.

Among individuals carrying risk *GBA1*$_{PVs}$ variants, the risk of developing PD increased according to their PRS tertile. In the Luxembourg Parkinson's Study, risk *GBA1*$_{PVs}$ carriers had ORs ranging from 0.97 (95%

CI, 0.77-1.17) to 1.45 (95% CI, 1.25 -1.65) depending on their PRS tertile. While mild *GBA1*$_{PVs}$ carriers had ORs ranging from 1.46 (95% CI, 0.68-2.25) to 1.76 (95% CI, 1.02 -2.55). Similar trends were observed in the UKB cohort, with ORs ranging from 0.92 (95% CI, 0.82-1.17) to 2.05 (95% CI, 1.75-2.39) for risk *GBA1*$_{PVs}$ carriers, and ranging from 1.49 (95% CI, 0.75-2.94) to 2.90 (95% CI, 1.48-5.67) for mild *GBA1*$_{PVs}$ carriers.

PRS and *GBA1*$_{PVs}$ are independent risk factors for PD risk.

In both cohorts, we found an association between the severity of *GBA1*$_{PVs}$ and a higher OR for PD (Figure 3). Carriers with severe *GBA1*$_{PVs}$ tended to have higher OR of PD (twofold or higher), regardless of their PRS status. In the Luxembourg Parkinson's Study ORs ranged from 1.96 (95% CI, 1.51-2.06) to 2.15 (95% CI, 1.96-2.12), while in the UKB cohort, ORs ranged from 2.28 (95% CI, 0.962-5.39) to 4.57 (95% CI, 1.94-10.70) for severe *GBA1*$_{PVs}$ carriers (Figure 3).

Notably, in the Luxembourg Parkinson's Study, non-carriers with a high PRS had an OR of 1.22 (95% CI, 1.20-1.23), which was comparable to the risk observed in risk *GBA1*$_{PVs}$ carriers with an intermediate PRS (OR: 1.17; 95% CI, 0.97-1.34). Similarly, non-carriers with a high PRS in the UKB cohort had an OR of 1.40 (95% CI, 1.39-1.42), which was comparable to the risk observed in risk *GBA1*$_{PVs}$ carriers with an intermediate PRS (OR: 1.30; 95% CI, 1.10-1.50).

Discussion

We examined the effect of the PRS for the risk of developing PD in carriers of *GBA1*$_{PVs}$ within two different cohorts. Our results suggest that the risk of developing PD conferred by *GBA1*$_{PVs}$ may be influenced by the PRS, which is thought to act by affecting multiple pathways associated with PD [35]. Our research confirms that a combination of high PRS and *GBA1*$_{PVs}$ may increase the risk of developing PD. Consistent with previous studies, we found that carriers of common 'risk' *GBA1* variants (specifically, p.E365K, p.T408M, or p.N409S) with a high PRS have a higher PD risk compared to those with a low PRS. Importantly, our analysis extends to different types of rare *GBA1*$_{PVs}$, further classified into severe, mild and risk. To strengthen our conclusions, we confirmed our findings in an independent replication cohort.

The divergent criteria and factors that different systems use to categorize *GBA1* variants contribute to the complexity and difficulty of the classification process. This represents a significant obstacle for researchers in this field and requires further attention to standardize the classification of *GBA1* variants. In response to

these challenges, we have followed a systematic classification of *GBA1* variants based on their pathogenicity and predicted severity, taking into account their frequency in PD patients and the general population, their biochemical effect, and their association with Gaucher disease.

Our results highlight a noteworthy observation: the distinct effect of PRS on individuals carrying "risk" versus "severe" and "mild" *GBA1*$_{PVs}$. In individuals carrying risk *GBA1*$_{PVs}$, a high PRS shifts their OR to a higher value, indicating an increased risk of PD (Figure 3). This emphasizes the role of polygenic risk in regulating the likelihood of developing the disease in those carrying less severe pathogenic variants. While carriers of severe *GBA1*$_{PVs}$ have an increased risk of PD (twofold or more), even when they have a low PRS, indicating the dominant effect of severe *GBA1*$_{PVs}$ despite their high-risk polygenic background. These findings emphasize the complexity of the combined effect of polygenic risk and *GBA1*$_{PVs}$, and the need to consider both aspects in the genetic assessment of PD risk.

In the light of our findings, we echo previous recommendations to perform genetic analysis prior to clinical trials to stratify patients based on PD-related genetic variants [17,38,39]. In addition, it may be useful to systematically generate genotyping data together with whole exome sequencing or disease-specific targeted gene sequencing data to be able to calculate the PRS as part of the initial diagnosis. This could contribute to ongoing efforts to reduce healthcare costs and would increase the size of available cohorts. Furthermore, it would also improve the feasibility of incorporating PRS information into standard clinical practice and encourage the development of multifactorial models that integrate genetic and non-genetic risk factors, similar to existing models such as the BOADICEA model in breast cancer [20].

It is important to consider that the results of PRS studies can be influenced by several factors, including the study design and recruitment strategies employed [35]. In the context of this study, the observed results were somewhat more pronounced in the replication cohort compared to the discovery cohort. This difference is likely due to differences in study design and recruitment approaches between the two cohorts. The replication cohort, the UKB cohort, is population-based and has yielded more clear-cut results, whereas the discovery cohort, the Luxembourgish Parkinson's Study, is a PD-specific case-control cohort and has shown similar trends but to a lesser extent [35,36]. Additionally, it is worth noting that there was a significant discrepancy in family history between PD cases and controls within UKB, whereas no significant association in family history was observed in the Luxembourgish Parkinson's Study. This discrepancy further emphasizes the influence of different study designs and recruitment strategies on the observed results and highlights the importance of carefully considering these factors when interpreting the study findings.

This study has several limitations. It was conducted only among participants of European ancestry. The current findings need to be validated in cohorts of different ethnicities to ensure generalizability. Population bias is a major limitation of most genetic studies to date [37]. In addition, UKB is not a PD-specific disease cohort and does not control for age and sex between cases and controls, we have attempted to address this by using age and sex as covariates during regression analysis. Furthermore, the sample size is limited in certain PRS tertiles or *GBA1*$_{PVs}$ severity categories, which may lead to an overestimation of effect sizes and should be considered carefully when interpreting the results.

Despite these limitations, this study provides a valuable contribution to our understanding of the genetic factors influencing of *GBA1*$_{PVs}$ and the risk of developing PD. Overall, the study of two independent cohorts highlights the complexity of the combined effect between polygenic background and *GBA1* pathogenic variants and moreover the severity of these variants in the development of PD. We find that the polygenic risk can influence the risk of developing PD both in the general population and among carriers of mild *GBA1* pathogenic variants. While carriers of severe *GBA1* pathogenic variants tend to have a higher risk of developing PD, regardless of their PRS. Additionally, the severity of *GBA1* pathogenic variants further modulates this risk, highlighting the complex genetic architecture of PD. These findings support the potential utility of incorporating polygenic risk scores into clinical risk stratification for PD.

Further, the use of the classification framework we developed as part of this study may streamline future research by providing a more consistent description of the role of *GBA1* variants in PD. This, in turn, will facilitate easier comparison and integration of research findings across the international scientific community.

Deepening our understanding of this complex genetic landscape, including the classification of variant severity, can set the stage for more personalized therapeutic strategies and targeted preventive measures in the future. As we continue to unravel the genetic intricacies of PD, we move closer to a future where treatment is increasingly tailored to individual genetic profiles, optimizing outcomes and improving patients' lives.

Ethics approval and consent to participate

UK Biobank analyses were conducted via application 52446 using a protocol approved by the Partners HealthCare Institutional Review Board. All study participants provided written informed consent.

Ethical approval for the Luxembourg Parkinson's study was obtained from the National Ethics Board (CNER Ref: 201407/13 and 202304/03) and the Data Protection Committee (CNPD Ref: 446/2017).

Consent for publication

Not applicable.

Competing interests

No potential conflicts (financial, professional, or personal) relevant to the manuscript.

Availability of data and materials

Data used to prepare this article were obtained from the UK Biobank, and the National Centre of Excellence in Research: Early diagnosis and stratification of Parkinson's Disease (NCER-PD or the Luxembourg Parkinson's study) (https://www.parkinson.lu/). Restrictions apply to the availability of these data for the UK biobank, which were used under license for the current study (Project ID: 52446).
The dataset for the Luxembourg Parkinson's Study is not publicly available. Any requests for accessing the dataset can be directed to request.ncer-pd@uni.lu.

Funding

The Fonds National de Recherche (FNR) supported P.M., R.K and S.P as part of the National Centre of Excellence in Research on Parkinson's disease (FNR/NCER13/BM/11264123), the DFG Research Units FOR2715 (INTER/DFG/17/11583046) to P.M, FOR2488 (INTER/DFG/19/14429377) to P.M and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to R.K, S.P). The FNR supported D.R.B through the Industrial fellowship program of Luxembourg (FNR14323864). C.M. and E.H. are supported by the BONFOR-program of the Medical Faculty, University of Bonn (O-147.0002).

Authors' contributions

(1) Research Project: A. Conception, B. Organization, C. Execution; (2) Statistical Analysis: A. Design, B. Execution; (3) Data: A. acquisition B. Curation (4) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique.

E.H.: 1A, 1B, 1C, 2A, 2B, 4A, 4B
S.P.: 1B, 1C, 3A, 3B, 4A, 4B
Z.L.: 1B, 3A, 3B, 4B
P.K.: 4B
R.K.: 3A, 4B
C.M.: 1A, 1B,1C, 4B
P.M.: 1A, 1B, 1C, 2A, 3A, 3B, 4A, 4B
D.R.B.: 1A, 1B, 1C, 2A, 4A, 4B

Acknowledgements

Data used to prepare this article were obtained from the UK biobank, and the National Centre of Excellence in Research: Early diagnosis and stratification of Parkinson's Disease (NCER-PD) (https://www.parkinson.lu/). Parts of the computational analysis were done on the High-Performance Computing cluster of the University of Luxembourg (https://hpc.uni.lu/).

References

1. Blauwendraat, C., Nalls, M. A. & Singleton, A. B. The genetic architecture of Parkinson's disease. *Lancet Neurol* **19**, 170–178 (2020).
2. O'Regan, G., deSouza, R.-M., Balestrino, R. & Schapira, A. H. Glucocerebrosidase Mutations in Parkinson Disease. *Journal of Parkinson's Disease* **7**, 411–422 (2017).
3. Brockmann, K. *et al.* GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study. *Mov Disord* **30**, 407–411 (2015).
4. Stoker, T. B. *et al.* Impact of GBA1 variants on long-term clinical progression and mortality in incident Parkinson's disease. *J Neurol Neurosurg Psychiatry* **91**, 695–702 (2020).
5. Pachchek, S. Accurate long-read sequencing identified GBA variants as a major risk factor in the Luxembourgish cohort. (2023).
6. Parlar, S. C., Grenn, F. P., Kim, J. J., Baluwendraat, C. & Gan-Or, Z. Classification of GBA1 Variants in Parkinson's Disease: The GBA1-PD Browser. *Mov Disord* **38**, 489–495 (2023).
7. Bandres-Ciga, S., Diez-Fairen, M., Kim, J. J. & Singleton, A. B. Genetics of Parkinson disease: an introspection of its journey towards precision medicine. *Neurobiol Dis* **137**, 104782 (2020).
8. Liu, X. *et al.* Genome-Wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC Medical Genetics* **12**, 104 (2011).
9. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-genome wide association study. *Lancet Neurol* **18**, 1091–1102 (2019).
10. Bobbili, D. R., Banda, P., Krüger, R. & May, P. Excess of singleton loss-of-function variants in Parkinson's disease contributes to genetic risk. *J Med Genet* **57**, 617–623 (2020).
11. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219–1224 (2018).
12. Weale, M. E. *et al.* Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. *Am J Cardiol* **148**, 157–164 (2021).
13. Lee, A. *et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine* **21**, 1708–1718 (2019).
14. Ibanez, L. *et al.* Parkinson disease polygenic risk score is associated with Parkinson disease status and age at onset but not with alpha-synuclein cerebrospinal fluid levels. *BMC Neurol* **17**, 198 (2017).
15. Blauwendraat, C. *et al.* Parkinson disease age at onset GWAS: defining heritability, genetic loci and α-synuclein mechanisms. *Mov Disord* **34**, 866–875 (2019).
16. Iwaki, H. *et al.* Penetrance of Parkinson's Disease in LRRK2 p.G2019S Carriers Is Modified by a Polygenic Risk Score. *Mov Disord* **35**, 774–780 (2020).
17. Blauwendraat, C. *et al.* Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain* **143**, 234–248 (2020).
18. Dehestani, M., Liu, H. & Gasser, T. Polygenic Risk Scores Contribute to Personalized Medicine of Parkinson's Disease. *J Pers Med* **11**, 1030 (2021).
19. Hipp, G. *et al.* The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. *Front Aging Neurosci* **10**, 326 (2018).
20. Blauwendraat, C. *et al.* NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging* **57**, 247.e9-247.e13 (2017).
21. Accurate long-read sequencing identified GBA variants as a major genetic risk factor in the Luxembourg Parkinson's study. https://www.researchsquare.com (2023) doi:10.21203/rs.3.rs-2751593/v1.
22. Litvan, I. *et al.* SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders. *Mov Disord.* **18**, 467–486 (2003).
23. Lill, C. M. *et al.* Launching the movement disorders society genetic mutation database (MDSGene). *Mov Disord* **31**, 607–609 (2016).

24. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
25. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).
26. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
27. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. *Hum Mutat* **37**, 235–241 (2016).
28. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2019).
29. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877–885 (2016).
30. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD®): 2003 update. *Human Mutation* **21**, 577–581 (2003).
31. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985 (2014).
32. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511–1516 (2017).
33. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, giz082 (2019).
34. Hassanin, E. *et al.* Clinically relevant combined effect of polygenic background, rare pathogenic germline variants, and family history on colorectal cancer incidence. *BMC Medical Genomics* **16**, 42 (2023).
35. Fahed, A. C. *et al.* Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* **11**, 3635 (2020).
36. Hassanin, E. *et al.* Breast and prostate cancer risk: The interplay of polygenic risk, rare pathogenic germline variants, and family history. *Genetics in Medicine* **24**, 576–585 (2022).
37. Bandres-Ciga, S. *et al.* Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. *Acta Neuropathol* **140**, 341–358 (2020).
38. Leonard, H. *et al.* Genetic variability and potential effects on clinical trial outcomes: perspectives in Parkinson's disease. *J Med Genet* **57**, 331–338 (2020).
39. Arena, G. *et al.* Polygenic risk scores validated in patient-derived cells stratify for mitochondrial subtypes of Parkinson's disease. 2023.05.12.23289877 Preprint at https://doi.org/10.1101/2023.05.12.23289877 (2023).
40. Dueñas, N. *et al.* Ability of a polygenic risk score to refine colorectal cancer risk in Lynch syndrome. 2023.04.20.23288850 Preprint at https://doi.org/10.1101/2023.04.20.23288850 (2023).
41. Jenkins, M. A. *et al.* Assessment of a Polygenic Risk Score for Colorectal Cancer to Predict Risk of Lynch Syndrome Colorectal Cancer. *JNCI Cancer Spectr* **5**, pkab022 (2021).
42. Privé, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *The American Journal of Human Genetics* **109**, 12–23 (2022).

Tables

Cohort	Luxembourg Parkinson's study		UK Biobank	
Diagnostic	Cases	Controls	Cases	Controls
Participants, n	631	511	770	178,325
Male, n (%)	426 (67.5)	272 (53.1)	463 (60.1)	80,259 (45.0)
Female, n (%)	205 (32.5)	239 (46.9)	307 (39.9)	89,066 (55.0)
Age at Onset, mean (SD)	62.6 (11.5)	-	64.6 (8.6)	-
Age at assessment, mean (SD)	73.2 (10.7)	71.5 (7.12)	62.8 (5.3)	64.1 (8.0)
Carriers, n (%)	68 (10.8)	24 (4.7)	63 (8.2)	8,481 (4.8)
Family history of PD, n (%)	161 (25.5)	117 (22.8)	90 (11.7)	7,596 (4.3)

Table 1: Characteristics of the Luxembourg Parkinson's study and the UK Biobank cohort

Figures

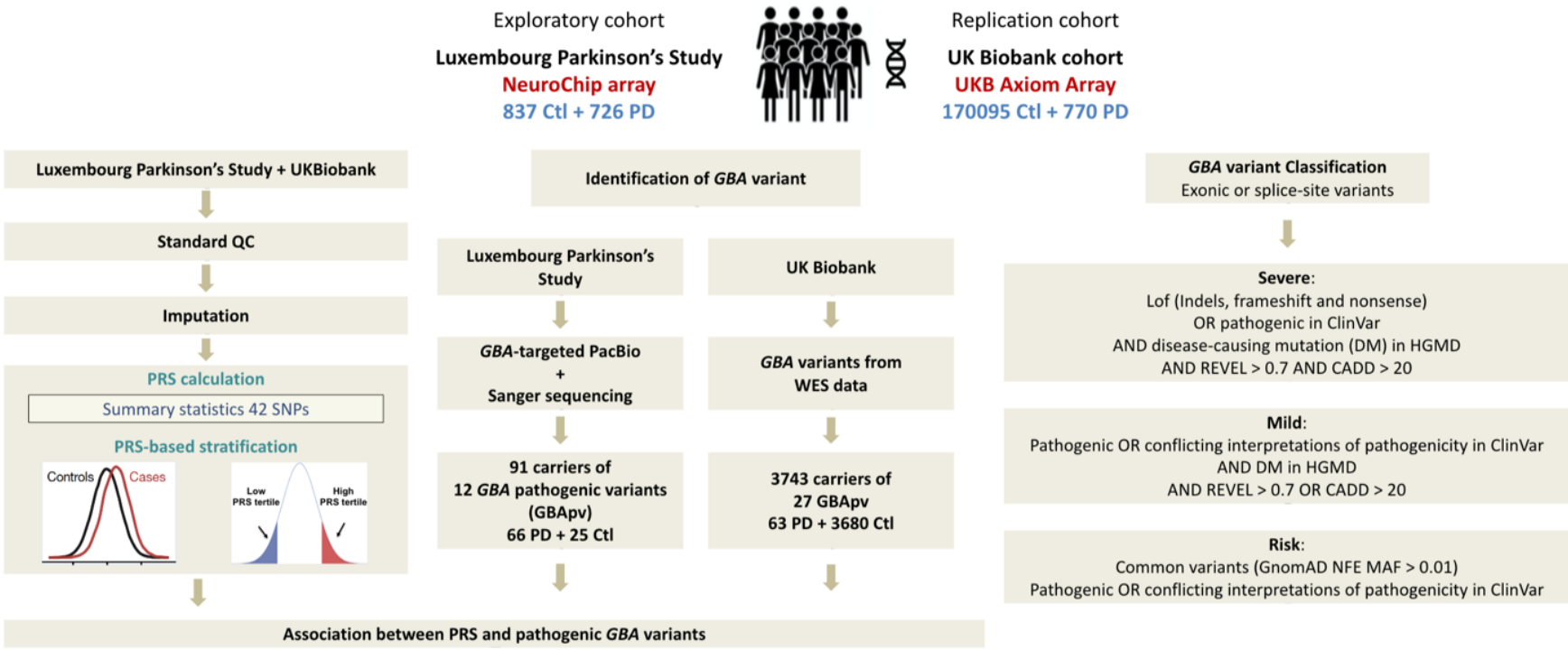


Figure 1: Overview of the study describing the framework for GBA1 variant classification, PRS stratification and statistical analysis.

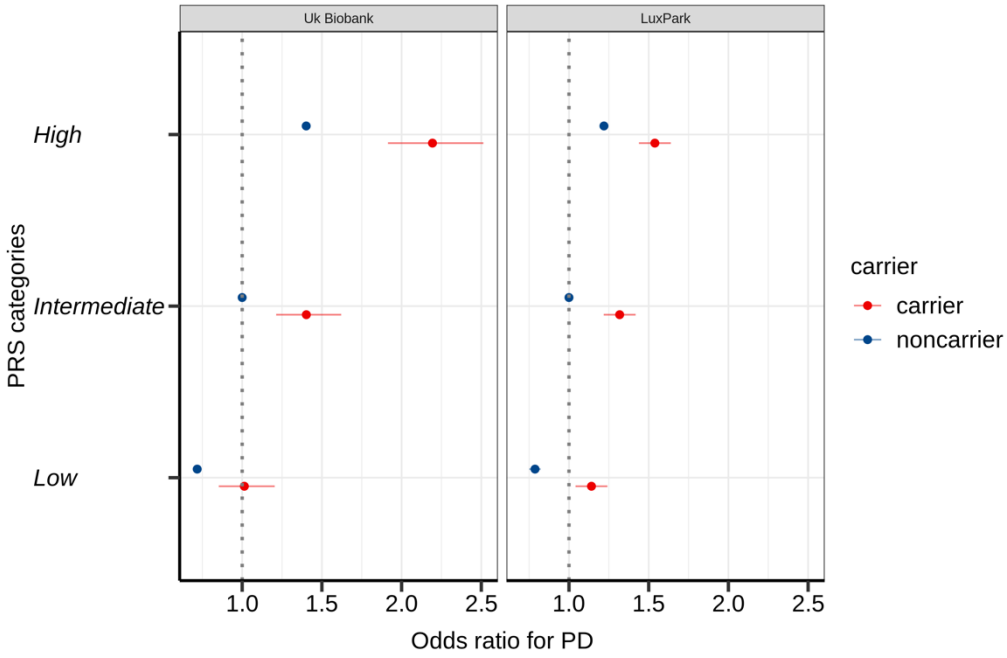


Figure 2: Parkinson's disease (PD) risk stratified by polygenic risks scores (PRS) tertiles and *GBA1* carrier status.
Predicted odds ratio for PD were estimated from logistic models, while conditioning on the sex, age and the first four ancestry principal components. Noncarriers with median PRS served as the reference group. Carriers and noncarriers were categorized into tertiles based on their PRS.

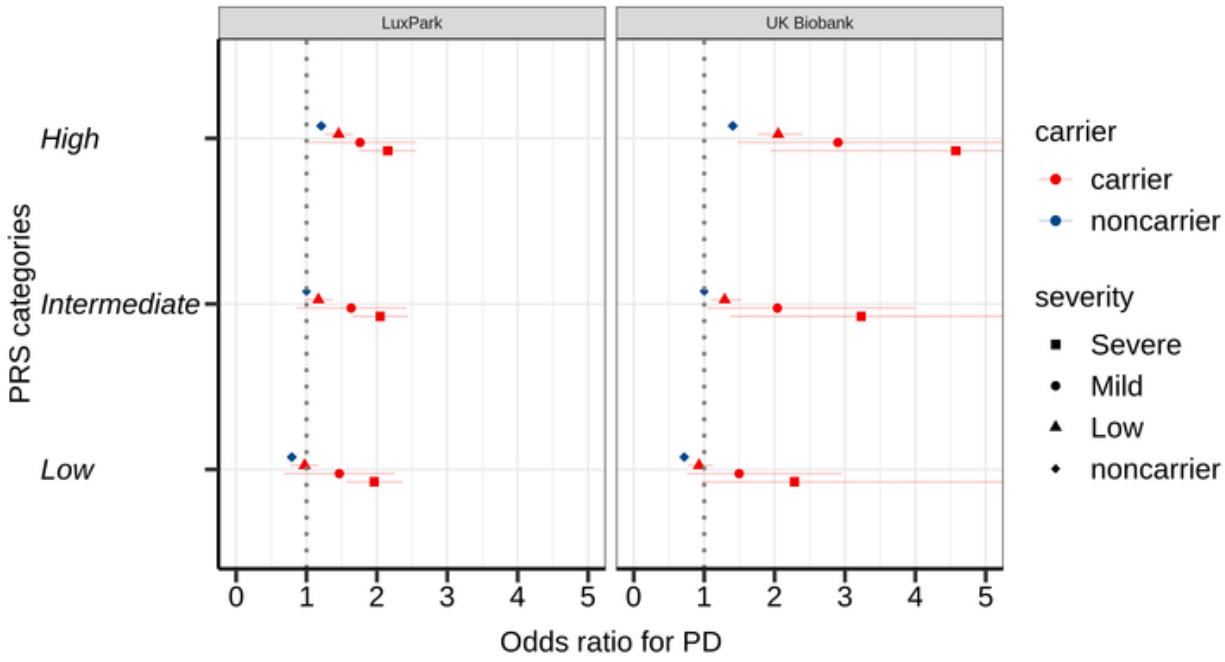


Figure 3: Parkinson's disease risk stratified by polygenic risk scores tertiles and the status and severity of rare pathogenic GBA1 variants. X-axis limits are truncated between (0,5).

3.3. Supplementary material

NCER-PD Consortium Members

Geeta ACHARYA [2], Gloria AGUAYO [2], Myriam ALEXANDRE [2], Muhammad ALI [1], Wim AMMERLANN [2], Giuseppe ARENA [1], Rudi BALLING [1], Michele BASSIS [1], Katy BEAUMONT [2], Regina BECKER [1], Camille BELLORA [2], Guy BERCHEM [3], Daniela BERG [11], Alexandre BISDORFF 5, Ibrahim BOUSSAAD [1], Kathrin BROCKMANN [11], Jessica CALMES [2], Lorieza CASTILLO [2], Gessica CONTESOTTO [2], Nico DIEDERICH [3], Rene DONDELINGER [5], Daniela ESTEVES [2], Guy FAGHERAZZI [2], Jean-Yves FERRAND [2], Manon GANTENBEIN [2], Thomas GASSER [11], Piotr GAWRON [1], Soumyabrata GHOSH [1], Marijus GIRAITIS [2,3], Enrico GLAAB [1], Elisa GÓMEZ DE LOPE [1], Jérôme GRAAS [2], Mariella GRAZIANO [17], Valentin GROUES [1], Anne GRÜNEWALD [1], Wei GU [1], Gaël HAMMOT [2], Anne-Marie HANFF [2], Linda HANSEN [1],3, Maxime HANSEN [1],[3], Michael HENEKA [1], Estelle HENRY [2], Sylvia HERBRINK [6], Sascha HERZINGER [1], Michael HEYMANN [2], Michele HU [8], Alexander HUNDT [2], Nadine JACOBY [18], Jacek JAROSLAW LEBIODA [1], Yohan JAROZ [1], Quentin KLOPFENSTEIN [1], Jochen KLUCKEN [1], [2],3, Rejko KRÜGER [1, 2,3], Pauline LAMBERT [2], Zied LANDOULSI [1], Roseline LENTZ [7], Inga LIEPELT [11], Robert LISZKA [14], Laura LONGHINO [3], Victoria LORENTZ [2], Paula Cristina LUPU [2], Clare MACKAY [10], Walter MAETZLER [15], Katrin MARCUS [13], Guilherme MARQUES [2], Tainá M MARQUES [1], Patricia MARTINS CONDE [1], Patrick MAY [1], Deborah MCINTYRE [2], Chouaib MEDIOUNI [2], Francoise MEISCH [1], Myriam MENSTER [2], Maura MINELLI [2], Michel MITTELBRONN [1,4], Brit MOLLENHAUER [12], Friedrich MÜHLSCHLEGEL [4], Romain NATI [3], Ulf NEHRBASS [2], Sarah NICKELS [1], Beatrice NICOLAI [3], Jean-Paul NICOLAY [19], Marek OSTASZEWSKI [1], Clarissa P. da C. GOMES [1], Sinthuja PACHCHEK [1], Claire PAULY [1,3], Laure PAULY [1], Lukas PAVELKA [1,3], Magali PERQUIN [2], Rosalina RAMOS LIMA [2], Armin RAUSCHENBERGER [1], Rajesh RAWAL [1], Dheeraj REDDY BOBBILI [1], Kirsten ROOMP [1], Eduardo ROSALES [2], Isabel ROSETY [1], Estelle SANDT [2], Stefano SAPIENZA [1], Venkata SATAGOPAM [1], Margaux SCHMITT [2], Sabine SCHMITZ [1], Reinhard SCHNEIDER [1], Jens SCHWAMBORN [1], Jean-Edouard SCHWEITZER [1], Amir SHARIFY [2], Ekaterina SOBOLEVA [1], Kate SOKOLOWSKA [2], Olivier TERWINDT [1],3, Hermann THIEN [2], Elodie THIRY [3], Rebecca TING JIIN LOO [1], Joana TORRE [2], Christophe TREFOIS [1], Johanna TROUET [2], Olena TSURKALENKO [2], Michel VAILLANT [2], Mesele VALENTI [2], Carlos VEGA [1], Liliana VILAS BOAS [3], Maharshi VYAS [1], Richard WADE-MARTINS [1], Paul WILMES [1], Evi WOLLSCHEID-LENGELING [1], Gelani ZELIMKHANOV [3]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[2]Luxembourg Institute of Health, Strassen, Luxembourg
3 Centre Hospitalier de Luxembourg, Strassen, Luxembourg
4 Laboratoire National de Santé, Dudelange, Luxembourg
5 Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg
6 Centre Hospitalier du Nord, Ettelbrück, Luxembourg
7 Parkinson Luxembourg Association, Leudelange, Luxembourg
8 Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
9 Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
10 Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK
11 Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Tübingen, Germany
12 Paracelsus-Elena-Klinik, Kassel, Germany
13 Ruhr-University of Bochum, Bochum, Germany
14 Westpfalz-Klinikum GmbH, Kaiserslautern, Germany
15 Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany
16 Department of Neurology Philipps, University Marburg, Marburg, Germany
17 Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg
18 Private practice, Ettelbruck, Luxembourg
19 Private practice, Luxembourg, Luxembourg

Subclassification	protein change	Nucleotide change	Number of carriers				dbSNP
			Luxembourg Parkinson's study		UK Biobank		
			PD	HC	PD	HC	
	-	c.115+1G>A	1				rs104886460
	p.R159Q	c.G476A				3	rs79653797
	p.P161S	c.C481T	2				rs121908299
	p.R170C	c.C508T				3	rs398123530
	p.R202X	c.C604T				1	rs1009850780
	p.S212X	c.C635G			4		rs1671872221
	p.N227S	c.A680G				6	rs364897
	p.G234E	c.G701A				1	rs74462743
	p.G234W	c.G700T	1				-
	p.G241R	c.G721A	2			3	rs409652
	p.F252I	c.T754A	1				rs381737
severe	p.H294Q	c.T882G		1	1	17	rs367968666
	p.R296Q	c.G887A				20	rs140955685
	p.P305L	c.C914T				6	-
	p.R398X	c.C1192T	1			4	rs121908309
	p.G416S	c.G1246A	1			7	rs121908311
	p.D419N	c.G1255A				1	-
	p.D448H	c.G1342C				7	rs1064651
	p.L483P	c.T1448C	10*				rs421016
	p.N501K	c.C1503G				2	-
	p.R502C	c.C1504T				65	rs80356771
	p.R502H	c.G1505A	1			1	rs80356772
	p.L29Afs*18	c.84dupG				1	-
	p.F255Y	c.T764A				5	rs74500255
	p.S310G	c.A928G				1	rs1057942
	p.R316C	c.C946T				1	rs1264734195
mild	p.T362I	c.C1085T				3	rs76539814
	p.R368C	c.C1102T				2	rs374306700
	p.N409S	c.A1226G	6	2	6	291	rs76763715
	p.R535H	c.G1604A				1	rs75822236
risk	p.E365K	c.G1093A	24	13	33	2114	rs2230288
	p.T408M	c.C1223T	16	9	19	1114	rs75548401

Supplementary Table S.1: Pathogenic *GBA1* variants used in this study.
*include RecNcil (p.L483P; p.A495P; p.V499V); Abbreviations: PD, Parkinson's disease; HC, healthy controls.

Chapter 4 : Whole Genome Sequencing reveals new monogenic forms of Parkinson's disease in idiopathic familial cases within the Luxembourg Parkinson's Study

Project description

To unravel the complex genetic landscape underlying idiopathic PD. In this study, we used high-resolution WGS to explore potential candidate genes associated with PD in a cohort of 67 patients with familial PD. The results of this study are therefore primarily exploratory.

Contributions

As part of this preliminary investigation, I performed a comprehensive genetic analysis and subsequent interpretation in collaboration with Dr. Zied Landoulsi and Dr. Dheeraj Reddy Bobbili.

Abstract

Introduction: Several loci associated with increased risk of Parkinson's disease (PD) have been identified by GWAS. However, it is essential to conduct further investigations to explore the potential involvement of rare variants within these loci, which could describe the full picture of the interaction between common risk and deleterious rare variants.

Methods: We used the Whole Genomic Sequencing method to study the rare variants in 607 PD-related genes identified by NCBI and ParkinsonsUK-UCL. We analyzed the genome of 67 patients with familial PD in the Luxembourg Parkinson's Study. We have filtered out rare variants where the frequency of the minor allele is less than 1%. Rare variants were classified into four groups: loss-of-function, non-synonymous, splice-site, and non-coding variants. We retained only variants with REVEL and mvPPT scores above 0.7 or MISTIC annotated as damaging. We used the SpliceAI tool to analyze splice site variants and retained only variants with a score higher than 0.5. The intronic variants were checked for enrichment in Deoxyribonuclease (DNase) I hypersensitive regions.

Results: We identified one loss-of-function, 141 nonsynonymous and eight rare copy number variants.

Conclusion: Our research suggests that some of the effects of PD risk loci may be due to rare variants that could potentially contribute to PD pathogenicity. Further analysis and confirmation of the identified variants in a larger cohort are needed, as this is an exploratory study.

4.1. Introduction

PD is a multifaceted polygenic disorder influenced by a combination of genetic and environmental factors. Familial cases account for 30% of PD cases, while the vast majority, 90%, are sporadic (Papapetropoulos et al., 2007). Recent research has identified numerous genetic loci associated with both the onset and progression of PD (Blauwendraat et al., 2020a). Notably, more than 90 risk loci, including those in the *SNCA* and *LRRK2* genes, have been identified as prevalent genetic contributors to PD (Nalls et al., 2014). Further investigations are required to determine the existence of rare variants that contribute to PD risk.

Recognizing the importance of genetics in unraveling the complexity of PD, in the previous chapters we mainly analyzed genotyping and *GBA1* target sequencing data. We were able to identify and confirm with high resolution the *GBA1* variants and some rare variants already tagged in the NeuroChip. Although these techniques offer advantages in terms of time and cost efficiency, they could not explain the genetic risk in certain idiopathic PD patients attributed to other genetic factors. Only WGS could provide a comprehensive overview to unravel the genetic complexity of Parkinson's disease.

To this aim, we present a comprehensive analysis of 67 patients with familial PD, screened by the high-resolution WGS, recruited as part of the Luxembourg Parkinson's study (Hipp et al., 2018; Pachchek, 2023).

4.2. Materials and Methods

4.2.1. Samples selection

As part of the National Center for Excellence in Research on Parkinson's disease program (NCER-PD), 67 patients with familial PD from the Luxembourg Parkinson's Disease Study underwent a short-read WGS data. Patients got thorough clinical and neuropsychological profiling (Pavelka et al., 2022). Following the requirements of the institutional review board, all patients gave their written consent. The study has been approved by the National Research Ethics Committee (CNER Ref: 201407/13).

4.2.2. Whole genome sequencing and quality control

The library preparations and sequencing method are detailed in the article from Pachchek et al., 2023. SNVs and CNVs detection were carried out using Illumina's Dragen DNA pipeline from the Bio-IT Illumina Dynamic Read Analysis for GENomics platform (DRAGEN; v3.8; Illumina) with the

standard parameters. The reads were aligned to the human reference genome (GRCh37, hg19 build). We selected only variants tagged as "PASS" by the Dragen DNA pipeline. To produce the final multisample Variant Call Format (VCF), all gVCFs were pooled using the parameter "--enable-joint-genotyping true" in Dragen. To select high-quality variants, the resulting VCF file was then filtered using the American College of Medical Genetics and Genomics (ACMG) guidelines and GATK best practices QC (Richards et al., 2015). Selected variants were annotated using Annotate Variation tool (Annovar) (Wang et al., 2010) by integrating several clinical and functional genome annotations, including those from the Single Nucleotide Polymorphism database (dbSNP) (Sherry et al., 2001), ClinVar (Landrum et al., 2014), gnomAD (Gudmundsson et al., 2021), the MISsense deleTeriousness predICtor (MISTIC) (Chennen et al., 2020), REVEL (Ioannidis et al., 2016) and the Pathogenicity Prediction Tool for missense variants (mvPPT) (Tong et al., 2022). We removed common variant blocks and defective genes from our analysis. Additionally, we excluded all SNVs and CNVs that overlapped with telomeres, centromeres, segmental duplications, immunoglobulin regions and T-cell receptor loci.

4.2.3. Classification of rare SNVs based on the variant type

We excluded common variants with Minor Allele Frequencies (MAF) above 1% in gnomAD in the Non-Finnish European (NFE) population. Rare variants in exonic and splicing regions (+/- 2 bp) of PD causal genes (*LRRK2*, *SNCA*, *VPS35*, *GBA1, PRKN*, *PINK1*, *PARK7*, *ATP13A2)* were selected and confirmed by Sanger sequencing. Based on the rare variant's positions within a gene region and their predicted effect on transcript and protein, the rare variants were classified into five groups: 1) Loss-of-Function (LoF) (stop gain/loss of codons, splicing and frameshift variants), 2) non-synonymous, 3) splice-site analysis variants (synonymous and intronic variants), and 4) non-coding variants (intergenic, down/upstream, UTRs). Only variants with a REVEL or mvPPT score above 0.7 or MISTIC equal to "D" for Deleterious were retained.

We investigated 607 susceptibility PD-related genes identified by the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/gene) and ParkinsonsUK-UCL (https://www.ebi.ac.uk/QuickGO/targetset/ParkinsonsUK-UCL).

4.2.4. Splicing site alteration analysis

SpliceAI (Jaganathan et al., 2019) was used to process rare intronic and synonymous variants to predict the splicing consequences based on the genomic sequence and to identify an alternative donor and acceptor site within 50 bp of the variant. SpliceAI was based on the transcript sequences of the GENCODE reference (Cotto et al., 2023). It provides four delta scores for each type of splicing abnormality (donor loss/gain, acceptor loss/gain). A variant was considered to have a significant impact on splicing if it had at least one score higher than 0.5 (Jaganathan et al., 2019).

4.2.5. Non-coding variants

To assess whether intronic variants are enriched in DNase I hypersensitive sites, which represent open chromatin regions accessible to transcription factors, we downloaded the wgEncodeRegDnaseClustered table from the DNAse Clusters (V3), containing DNaseI Hypersensitive Sites in ENCODE (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeRegDnaseClustered) and the wgEncodeRegTfbsClustered table from TFBS clusters (V3) from ENCODE data (https://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeRegTfbsClustered/).

4.2.6. Calling CNVs

CNVs were called using Dragen (see section 4.1.2). CNVs in autosomes were further filtered based on 50% or more overlap with CNVs reported in the Genome Aggregation Database Structural Variants (gnomAD-SV) (Collins et al., 2020) and the Database of Genomic Variants (DGV) (MacDonald et al., 2014). We applied the following criteria to characterize common structural variants in gnomAD-SV and DGV: CNV length greater than 50 bp (Conrad et al., 2010; MacDonald et al., 2014), variants occurring in at least 100 individuals, allele frequency ≥ 0.01. Common CNVs were eliminated using BEDTools (Quinlan and Hall, 2010). Suspected CNVs were evaluated visually using the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013).

4.3. Results

We carried out WGS in 67 individuals to find potential genes in familial PD patients (Figure 4.1). We used rare variant categorization to find and rank the variant in PD-related genes.

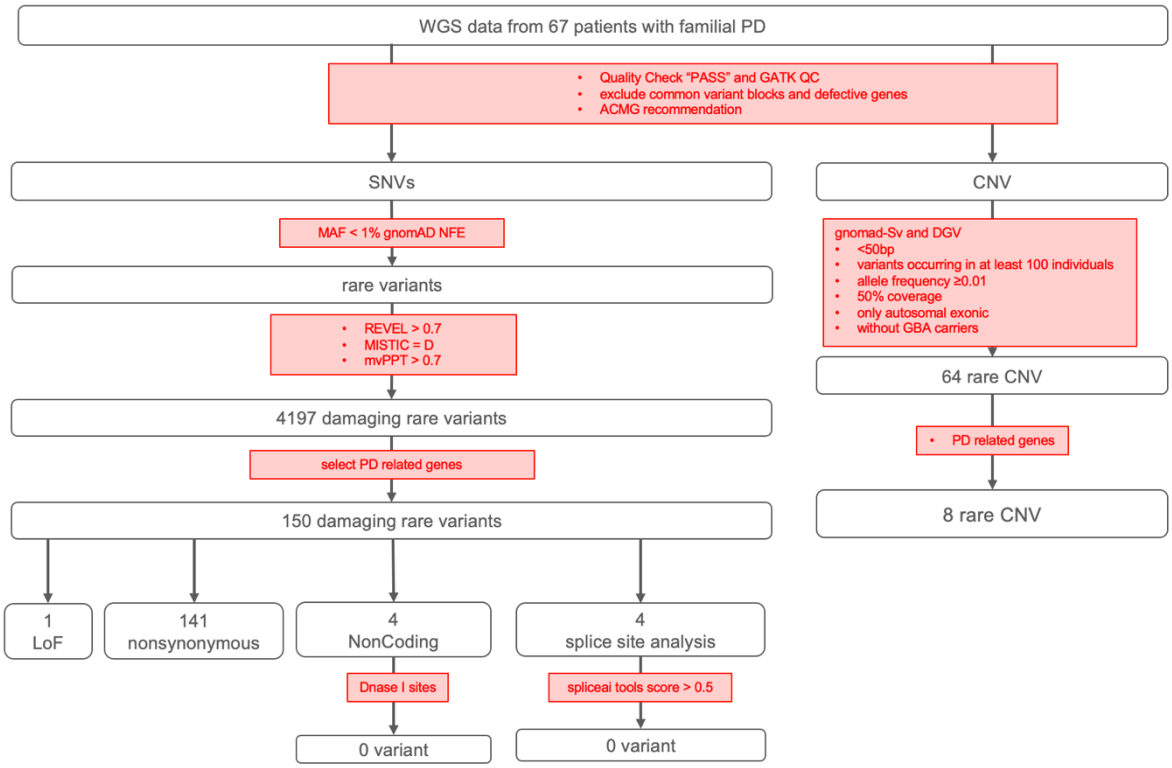


Figure 4. 1 : Analysis workflow for SNVs and CNVs analysis.

4.3.1. Demographic data

Among the familial PD patients, 65% (n = 44) were male with a mean AAO of 60 ± 12.8 years (range, 45–98) and a mean DD at the first visit of 5 ± 5.6 years. The mean of AAA is 72 ± 11.8 years (Table 4. 1). The average AAO is 59.6 years (range, 33–88).

Of the participants, 34 had a first-degree relative with PD, while 33 had a second-degree relative affected. Two siblings were present in this cohort (subjects 8 and 9), with AAO of 55 and 46 years respectively (Table 4. 1).

4.1.1. Identification of rare variants in candidate PD genes

We performed WGS on familial PD patients to identify potential candidate genes and variants. A total of 4197 rare variants, predicted to be deleterious or damaging, by *in silico* prediction tools, were selected after QC filtered with REVEL, MISTIC and mvPPT. After selecting the variants present in the 607 PD-related genes, identified by NCBI and ParkinsonsUK-UCL, a total of 150 variants remained. The following groups of variants were assessed: 1) loss-of-function (n=1), 2) nonsynonymous (n=141), 3) non-coding variants (n=4) and 4) splice site analysis variants (n=4).

We observed that the p.I11T variant of the *DERL1* gene, which is annotated as damaging by MISTIC and has a high score for REVEL and mvpptv1, is carried by eight PD patients. One of them also carries the pathogenic *GBA1* variant RecNcil.

With only six patients carrying *GBA1* variants together with another rare variant in a PD-related gene, this study is underpowered to evaluate digenic pathogenicity, but we believed that a descriptive and exploratory analysis was necessary for these carriers. We identified six pathogenic *GBA1* variants carriers, including four risk variants carriers (n_{T408M} = 3; n_{E365K} = 3), one severe RecNcil complex recombination variant carrier and one mild variants carrier (N409S) (Table 4. 1). Carriers of severe and mild pathogenic variants (p.N409S and RecNcil) of *GBA1* also carry a rare deleterious variant in PD-related genes in *DERL1*, *PANK2*, *MC1R*, *LINGO1* and *PDE8B*. Carriers of risk *GBA1* variants also carried rare deleterious variants in *OCA2*, *ESR2*, *LRRK1*, *CYP1A2*, *TPCN2*, *SCN8A*, *DRD3*, *VEGFA*, *NCL*, *SREBF1* (Table 4.2).

None of the individuals were related to each other, except for the one set of siblings reported in this study. Both siblings carry a synonymous variant in the *GBA1* gene (p.P201P (het)). The siblings do not carry similar rare deleterious variants in PD-related genes.

For the splice site analysis, we identified four variants to be evaluated with the SpliceAI tool, but none of them have a score above 0.5 nor are classified as having a significant impact on the splicing.

For the non-coding intronic variants, we checked whether the four variants found by our analyses were enriched in DNase I hypersensitive sites (DHS) and TFBS. None of them are enriched in DHS or TFBS. One patient (subject 1) had a *PRKN* gene CNV duplication, which was described in Chapter 2.

Table 4. 1 : Demographic and genetic data of the 67 familial Parkinson's disease cases

Subject	Gender	AAO	AAA visit1	DD visit1	Pathogenic variants carriers	Family History of PD	PD relatives information
1	M	39	61	22	pathogenic CNV PRKN exon 2 duplication het	2nd	Grandfather
2	M	77	80	2	severe pathogenic *GBA1* RecNciI	2nd	Cousin
3	M	45	48	3	mild pathogenic *GBA1* p.N409S (het)	2nd	Cousin
4	F	58	60	2	risk pathogenic *GBA1* p.E365K (het)	2nd	Aunt
5	F	49	54	5	risk pathogenic *GBA1* p.E365K (het)	2nd	Uncle
6	M	63	64	1	risk pathogenic *GBA1* p.E365K (het)	2nd	Aunt & Grandfather
7	M	65	76	11	risk pathogenic *GBA1* p.T408M (het)	2nd	Grandfather
8	F	55	57	2	synonymous *GBA1* p.P210P (het)	1st (sibling)	Brother
9	M	46	53	8	synonymous *GBA1* p.P210P (het)	1st (sibling)	Sister
10	F	66	69	3		1st	Brother
11	F	64	65	0		1st	Father
12	F	52	58	7		1st	Father
13	F	64	68	4		1st	Father
14	F	51	53	2		1st	Father
15	F	57	63	6		1st	Great-Grandfather & Sister
16	F	69	69	0		1st	Mother
17	F	78	80	2		1st	Mother
18	F	45	53	8		1st	Mother
19	F	47	65	18		1st	Mother
20	F	40	43	3		1st	Mother & Cousin
21	F	67	73	6		1st	Sister & Brother
22	F	60	68	8		2nd	Aunt
23	F	75	87	12		2nd	Aunt & Aunt's daughter
24	F	57	72	15		2nd	Cousin
25	F	74	78	4		2nd	Cousin
26	F	48	70	22		2nd	Cousin
27	F	51	70	19		2nd	Nephew & Grandmother & aunt
28	F	52	55	3		2nd	Uncle
29	F	80	81	1		2nd	Uncle & Aunt & Cousin
30	M	53	58	5		1st	Brother
31	M	82	83	1		1st	Brother
32	M	68	70	2		1st	Brother & cousin
33	M	88	91	4		1st	Brother & Cousin
34	M	56	58	2		1st	Father
35	M	57	63	6		1st	Father
36	M	64	77	13		1st	Father
37	M	67	67	0		1st	Father & Aunt
38	M	79	82	3		1st	Father & twin Brother
39	M	59	70	11		1st	Mother
40	M	33	58	25		1st	Mother
41	M	58	69	11		1st	Mother
42	M	66	78	12		1st	Mother
43	M	56	60	4		1st	Mother
44	M	76	76	1		1st	Mother
45	M	64	68	4		1st	Mother
46	M	84	86	2		1st	Mother & Sister
47	M	69	70	0		1st	Sister
48	M	68	68	0		1st	Sister
49	M	57	59	2		1st	Sister
50	M	69	72	3		2nd	2 grandfathers
51	M	48	50	2		2nd	Aunt
52	M	37	53	16		2nd	Aunt
53	M	56	62	6		2nd	Aunt & Uncle
54	M	66	68	2		2nd	Cousin
55	M	60	64	3		2nd	Grandfather
56	M	51	56	5		2nd	Grandfather
57	M	49	53	3		2nd	Grandmother
58	M	52	54	2		2nd	Grandmother
59	M	51	52	1		2nd	Grandmother
60	M	57	61	4		2nd	Grandmother
61	M	62	62	0		2nd	Great-Grandfather
62	M	64	67	3		2nd	Half sister
63	M	59	60	1		2nd	maternal Grandmother & Uncle
64	M	37	38	1		2nd	Uncle
65	M	57	63	6		2nd	Uncle
66	M	46	48	1		2nd	Uncle
67	M	78	88	9		2nd	Uncle

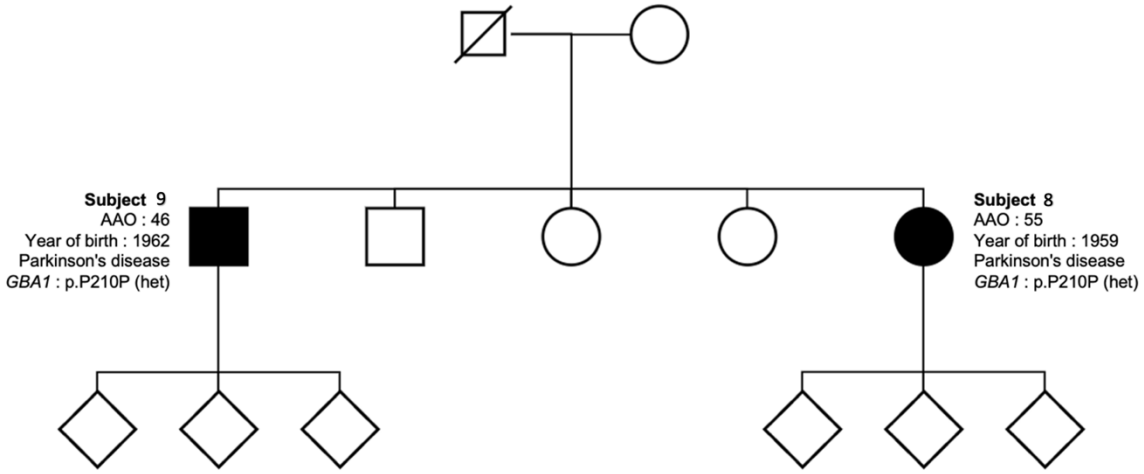


Figure 4. 2 : Sibling pedigree

Table 4. 2 : Identified rare variants in candidate Parkinson's disease genes

Gene	Protein changes	rsID	REVEL	mvpptv 1	MISTIC	Groups	Samples counts	Pathogenic carriers	Subject n°
DERL1	p.I11T		0,60	0,66	D	nonsynonymous	8	Subject n°2 severe pathogenic *GBA1*	2 / 9 / 12 / 24 / 25 / 31 / 40 / 42
ATF6B	p.H598L	rs147955878	0,69	0,00	D	nonsynonymous	3		19 / 45 / 59
PANK2	p.D46V	rs148036492	0,16	0,00	D	nonsynonymous	2	Subject n°3 mild pathogenic *GBA1*	3 / 10
OCA2	p.V419I	rs121918166	0,75	0,02	D	nonsynonymous	2	Subject n°4 risk pathogenic *GBA1*	4 / 49
MC1R	p.R142H	rs11547464	0,76		D	nonsynonymous	2	Subject n°2 severe pathogenic *GBA1*	2 / 26
FZD6	p.A182T	rs147788385	0,50	0,01	D	nonsynonymous	2		10 / 45
NOTCH2	p.F1209V	rs147223770	0,85	0,01	D	nonsynonymous	2		36 / 62
WNT1	p.G252R	rs200151492	0,58	0,02	D	nonsynonymous	2		32 / 42
WFS1	p.E776V	rs56002719	0,98	0,02	D	nonsynonymous	2		67 / 34
PRDX3	p.G173C	rs11554923	0,64	0,01	D	nonsynonymous	2		20 / 23
COL13A1	p.A130T	rs144774788	0,31	0,00	D	nonsynonymous	2		39 / 49
RNF19A	p.K61Q	rs149891168	0,45	0,01	D	nonsynonymous	2		50 / 65
WFS1	p.R818C	rs35932623	0,76	0,01	D	nonsynonymous	2		28 / 60
WNT7A	p.D168G		0,95	0,61	D	nonsynonymous	1	Subject n°1 pathogenic CNV PRKN	1
EDEM1	p.H582R	rs939309875	0,93	0,90	D	nonsynonymous	1	Subject n°1 pathogenic CNV PRKN	1
HMOX1	p.R183L		0,77	0,91	B	nonsynonymous	1	Subject n°1 pathogenic CNV PRKN	1
SNCB	p.P109H	rs104893937	0,59	0,01	D	nonsynonymous	1	Subject n°1 pathogenic CNV PRKN	1
SETD1A	p.R1288C	rs147948503	0,36	0,01	D	nonsynonymous	1	Subject n°1 pathogenic CNV PRKN	1
LINGO1	p.T197M	rs1014789586	0,53	0,09	D	nonsynonymous	1	Subject n°3 mild pathogenic *GBA1*	3
PDE8B	p.M1?		0,53			LoF	1	Subject n°3 mild pathogenic *GBA1*	3
ESR2	p.W348C	rs113652563	0,77	0,05	D	nonsynonymous	1	Subject n°4 risk pathogenic *GBA1*	4
SCN8A	p.T668I	rs758253791	0,72	0,11	D	nonsynonymous	1	Subject n°4 risk pathogenic *GBA1*	4
TPCN2	p.V179M	rs371461781	0,23	0,01	D	nonsynonymous	1	Subject n°4 risk pathogenic *GBA1*	4
CYP1A2	p.R431W	rs28399424	0,93	0,02	D	nonsynonymous	1	Subject n°5 risk pathogenic *GBA1*	5
LRRK1	p.V697I	rs141102052	0,33	0,00	D	nonsynonymous	1	Subject n°5 risk pathogenic *GBA1*	5
VEGFA	p.E236K		0,83		D	nonsynonymous	1	Subject n°6 risk pathogenic *GBA1*	6
DRD3	p.C181Y		0,66	0,94	D	nonsynonymous	1	Subject n°6 risk pathogenic *GBA1*	6
SREBF1	p.R1019W	rs755965677	0,29	0,05	D	nonsynonymous	1	Subject n°7 risk pathogenic *GBA1*	7
NCL		rs148967659	0,03		D	noncoding	1	Subject n°7 risk pathogenic *GBA1*	7
GAK	p.H428R		0,88	0,36	D	nonsynonymous	1		10
RET	p.R590L	rs55947360	0,84	0,49	D	nonsynonymous	1		10
POLG	p.R562W	rs756952607	0,79	0,08	D	nonsynonymous	1		36
STXBP1	p.A177T		0,73	0,76	D	nonsynonymous	1		36
PRKAG3	p.D485N	rs149508864	0,32		D	nonsynonymous	1		36
USP24	p.L2267F		0,22	0,76	B	nonsynonymous	1		36
TH	p.E153K	rs754247629	0,46	0,12	D	nonsynonymous	1		32
POLG	p.G268A	rs61752784	0,97	0,02	D	nonsynonymous	1		41
MFN2	p.D499H		0,61	0,52	B	nonsynonymous	1		41
CELSR3	p.S2832G		0,42	0,21	D	nonsynonymous	1		46
NEDD9	p.P136Q	rs34265420	0,33	0,01	D	nonsynonymous	1		46
UNC13C	p.R171H	rs137929834	0,60	0,00	D	nonsynonymous	1		50
TLR2	p.F217S	rs139227237	0,53	0,00	B	nonsynonymous	1		50
UNC13B		rs546244395			D	splice site	1		50
PLCG2	p.S110C	rs1018511099	0,59	0,76	D	nonsynonymous	1		53
SNX2	p.D24V	rs143461530	0,51	0,02	B	nonsynonymous	1		53
SCN8A	p.V1148M	rs536452913	0,57	0,03	D	nonsynonymous	1		31
AMBRA1	p.R505H	rs145466300	0,56	0,01	B	nonsynonymous	1		31
ESR2	p.R221G	rs78851986	0,79	0,04	D	nonsynonymous	1		67
LINGO1	p.P135L		0,43	0,59	D	nonsynonymous	1		67
CSMD1	p.L2851V		0,27	0,07	D	nonsynonymous	1		67
LINC01195		rs146444255			D	noncoding	1		67
NCL		rs757546451	0,01		D	noncoding	1		23
BMP2	p.P76T		0,92	0,95	D	nonsynonymous	1		56
SIPA1L2	p.R1690Q	rs34461614	0,32	0,00	D	nonsynonymous	1		56
OPA1	p.I346M	rs143319805	0,94	0,03	D	nonsynonymous	1		49
TBC1D24	p.R242C	rs398122965	0,86		B	nonsynonymous	1		49
POLG	p.D1219G	rs776506626	0,85	0,22	D	nonsynonymous	1		49
ABCA1	p.P85L	rs145183203	0,84	0,01	D	nonsynonymous	1		24
EP300	p.P899T	rs148884710	0,29	0,00	D	nonsynonymous	1		24
MFN1	p.R460Q	rs145789238	0,79	0,22	D	nonsynonymous	1		55
UNC13C	p.T1095N		0,62	0,19	D	nonsynonymous	1		55
CP	p.T841R	rs56033670	0,60	0,00	D	nonsynonymous	1		55
SMPD1	p.E473V	rs142787001	0,82	0,03	D	nonsynonymous	1		63
AAK1	p.N279K		0,32	0,73	D	nonsynonymous	1		63
MTA1	p.E115G		0,83	0,87	D	nonsynonymous	1		27
ARHGEF7	p.L9R	rs144138131	0,38	0,01	D	nonsynonymous	1		27
ITIH4	p.G418S	rs141154056	0,94	0,02	D	nonsynonymous	1		45
FLOT1	p.M94T		0,86	0,80	D	nonsynonymous	1		45
PRODH	p.L333P	rs2904551	0,73	0,01	D	nonsynonymous	1		45
HP	p.Q15R		0,68	0,09	D	nonsynonymous	1		45
TRIM32	p.T526S		0,29	0,48	D	nonsynonymous	1		45
ABCA5	p.V241I	rs544106230	0,45	0,01	D	nonsynonymous	1		65
COL13A1	p.P165L		0,33	0,07	D	nonsynonymous	1		65
PIAS4		rs551647074	0,11		D	noncoding	1		65
GRIN2A	p.T663A		0,10	0,71	D	nonsynonymous	1		65
PRSS23	p.M261I	rs139206479	0,68	0,01	D	nonsynonymous	1		47
NFATC1	p.A9V	rs146472413	0,50	0,02	D	nonsynonymous	1		47
LCN2	p.L6P	rs139418967	0,48	0,01	D	nonsynonymous	1		47
PCLO	p.N608I	rs754954951	0,64	0,44	D	nonsynonymous	1		22
ACE	p.R109Q	rs149412997	0,66	0,06	D	nonsynonymous	1		66
MFN2	p.V705I	rs142271930	0,17	0,00	D	nonsynonymous	1		66
UNC13C	p.R303T		0,55	0,12	D	nonsynonymous	1		9
GBA2	p.S740R	rs200894732	0,64	0,01	B	nonsynonymous	1		44
TRIB3	p.R181C	rs149447454	0,47	0,04	D	nonsynonymous	1		44
PVALB	p.G35S	rs571435563	0,79	0,55	D	nonsynonymous	1		19
TF	p.G544E	rs121918677	0,39	0,06	D	nonsynonymous	1		19
WFS1	p.R685C	rs112967046	0,80	0,06	D	nonsynonymous	1		21
USP25	p.D276V		0,68	0,68	D	nonsynonymous	1		21

CASP9	p.H237P	rs146054764	0,60	0,12	D	nonsynonymous	1		21
ABCA1	p.R1839S		0,49	0,17	D	nonsynonymous	1		21
LRRK1	p.V854M	rs368404584	0,46	0,01	D	nonsynonymous	1		21
HP	p.D338H	rs189115161	0,42	0,01	D	nonsynonymous	1		21
COL13A1	p.A30E		0,24	0,05	D	nonsynonymous	1		21
NEDD9	p.G50S	rs373441149	0,64	0,01	D	nonsynonymous	1		12
UNC13B	p.M1184T	rs769161004	0,61	0,09	D	nonsynonymous	1		12
GDNF	p.R93W	rs36119840	0,79	0,02	D	nonsynonymous	1		59
WFS1	p.C360Y	rs147157374	0,16	0,21	D	nonsynonymous	1		42
LRRK1	p.R1943Q	rs376808685	0,08	0,01	D	nonsynonymous	1		28
CREB3L3	p.V179M	rs548714946	0,49	0,01	D	nonsynonymous	1		62
PSEN1	p.S43N		0,04	0,18	D	nonsynonymous	1		62
RHOT1		rs754371798			D	splice site	1		62
MCCC1	p.S80N	rs774565207	0,85	0,36	D	nonsynonymous	1		51
TPCN2	p.S218T		0,81		D	nonsynonymous	1		51
CASP9	p.N168N		0,55			splice site	1		51
ACE	p.T162M	rs3730043	0,22		D	nonsynonymous	1		51
ATP5F1A	p.D89G	rs148515768	0,79	0,01	D	nonsynonymous	1		17
APP	p.L8P		0,62	0,67	D	nonsynonymous	1		8
DNAJC13	p.T95K		0,54	0,64	D	nonsynonymous	1		8
ACE	p.G259S	rs571848794	0,73		B	nonsynonymous	1		26
CASP4	p.F374S	rs140485344	0,51	0,01	B	nonsynonymous	1		34
A2M	p.L18R	rs201671036	0,26	0,01	D	nonsynonymous	1		34
SMPD1	p.R291H	rs1803161	0,75	0,09	D	nonsynonymous	1		60
CYP2D6	p.Y304C	rs202102799	0,51	0,00	D	nonsynonymous	1		14
TPCN2	p.V163M	rs79490424	0,32	0,00	D	nonsynonymous	1		14
MTHFR	p.R429C	rs200138092	0,84	0,11	D	nonsynonymous	1		64
WFS1	p.D339H		0,83	0,95	D	nonsynonymous	1		64
ANG	p.P136L	rs121909543	0,73	0,15	D	nonsynonymous	1		64
SERPINA1	p.A84T	rs111850950	0,66	0,01	D	nonsynonymous	1		64
MAP3K10	p.P168Q	rs36102209	0,59	0,00	D	nonsynonymous	1		64
ABCB1	p.Q1107P	rs55852620	0,44	0,00	D	nonsynonymous	1		64
SPR	p.S103C	rs748740519	0,54	0,02	B	nonsynonymous	1		57
CSMD1	p.G2183R	rs371824064	0,52	0,02	B	nonsynonymous	1		18
DNAJC6	p.R330Q	rs375848137	0,64	0,03	D	nonsynonymous	1		52
RET	p.V450A		0,57	0,58	B	nonsynonymous	1		52
WNT7A	p.R157H	rs199592697	0,31	0,01	D	nonsynonymous	1		52
MAP3K10	p.R304C	rs574172025	0,56	0,01	D	nonsynonymous	1		15
RET	p.L56M	rs145633958	0,51	0,01	B	nonsynonymous	1		15
RET	p.E369K	rs377767402	0,45	0,04	D	nonsynonymous	1		33
NOD2	p.R284W	rs104895427	0,63	0,02	B	nonsynonymous	1		29
CELSR3	p.D2049N	rs186494511	0,33	0,01	D	nonsynonymous	1		29
TRIM32	p.R408C	rs3747835	0,52	0,01	D	nonsynonymous	1		48
CP	p.R793H	rs115552500	0,34	0,00	D	nonsynonymous	1		48
EDEM1	p.R89R				D	splice site	1		48
CSMD1	p.G1226S	rs745828621	0,61	0,02	D	nonsynonymous	1		13
PANK2	p.E241D		0,50	0,41	D	nonsynonymous	1		13
SEPTIN14	p.S200F	rs190406908	0,31	0,01	D	nonsynonymous	1		13
GDNF	p.D150N	rs76466003	0,25	0,01	D	nonsynonymous	1		13
RIMS2	p.D608A	rs61753731	0,77	0,00	D	nonsynonymous	1		37
CREB3L3	p.E379K	rs150431015	0,30	0,00	D	nonsynonymous	1		37
HTRA2	p.G399S	rs72470545	0,79	0,01	D	nonsynonymous	1		61
BRINP1	p.R607Q	rs150796528	0,18	0,01	D	nonsynonymous	1		11
TYR	p.P406L	rs104894313	0,92	0,01	D	nonsynonymous	1		61
TRAP1	p.I200V	rs113476582	0,77	0,01	B	nonsynonymous	1		32
CSMD1	p.P2261A	rs190894161	0,63	0,01	D	nonsynonymous	1		67
APOA1	p.F95Y	rs138407155	0,58	0,01	D	nonsynonymous	2		9 / 41
VTI1B	p.R176H	rs45548534	0,56	0,01	D	nonsynonymous	2		23 / 16
SORL1	p.D2065V	rs140327834	0,56	0,01	B	nonsynonymous	1		34
CUL9	p.R2316W	rs143984539	0,33	0,01	D	nonsynonymous	1		10
PLCG2	p.N571S	rs75472618	0,29	0,01	D	nonsynonymous	2		30/ 33

The RecNcil combines three *GBA1* variants p.L483P, p.A495P, and p.V499V. D, Deleterious; B, Benign.

4.1.2. CNVs analysis

We identified 64 rare CNVs in 105 genes, including eight CNVs in PD-related genes, two of which carry the pathogenic risk *GBA1* variants. Three CNVs carriers (*PRKN*, *CHRNA7* and *HSPA6*) have earlier AAO. The duplication in the *PRKN* gene observed in one patient was previously validated by the MLPA technique and presented in the previous Chapter 2.

chr:start-end	TYPE CNV	Gene	Pathogenic carriers	AAO (mean)	Actual age	Subject n°
6:162717494-162916407	**DUP**	**PRKN**		**39**	**68**	**1**
15:32016461-32515085	DUP	CHRNA7		47	71	19
1:161492664-161509871	DEL	HSPA6		49	61	5
10:135240915-135381133	DEL	CYP2E1;SYCE1		58	76	41
7:151500618-151697683	DUP	GALNTL5;LOC644090;PRKAG2		66	84	42
15:32069982-32514085	DUP	CHRNA7		78	94	67
9:28644821-28728423	DEL	LINGO2	risk pathogenic *GBA1*	58	67	4
12:112180652-112314835	DUP	ACAD10;ALDH2;MAPKAPK5	risk pathogenic *GBA1*	65	81	7

Table 4. 3 : Identified CNVs in Parkinson's disease related genes

4.1. Discussion

Genome-wide association studies have revealed numerous loci associated with the risk of PD. However, to fully understand the increased susceptibility, it is essential to further investigate the rare variants present at these loci. Some of the recently identified GWAS loci and PD-related genes have not yet been characterized at the molecular level to identify specific variants or causative genes (Grenn et al., 2020b). We explored whether rare variants with possible functional significance are also present in PD-related genes using WGS in the Luxembourg Parkinson's study cohort.

We identified 141 nonsynonymous rare variants in PD-related genes carried by 62 patients, including six individuals carrying pathogenic *GBA1* variants. If we assume that the pathogenic *GBA1* variants and the one *PRKN* CNVs are the cause of PD and explore only idiopathic PDs, then only 51 idiopathic PD patients carried 144 rare deleterious variants in PD-related genes. In the idiopathic patients, we identified five probable pathogenic rare deleterious variants (*STXBP1*: p.A177T; *MTA1*: p.E115G; *FLOT1*: p.M94T), that are annotated as deleterious in MISTIC and have high scores in REVEL and mvpptv (> 0.7). These carriers have an AAO of 64, 51 and 64 respectively. There is no additional information provided regarding the pathogenicity of these variants and these variants need to be further investigated.

The patient carrying the LoF variant (*PDE8B*: p.M1?) also carried the mild pathogenic *GBA1* variant (p.N409S). Because this patient had an early AAO of 45, it is difficult to assess whether the penetrance is due to *GBA1* or *PDE8B*. Since the first methionine aa of the gene is mutated in the

PDE8B gene, this mutation likely leads to protein chain termination, which will have a more deleterious effect compared to the mild *GBA1* mutation.

In our study, we present the identification of eight significant CNVs found in genes associated with PD (*PRKN*, *CHRNA7*, *HSPA6*). The CNV in the *PRKN* gene has been successfully validated using the MLPA technique, and existing research has highlighted the potential pathogenic influence of these indels in the PD background (Lubbe et al., 2021).
Using AAO as a marker to assess the pathogenicity of indels, our results highlight a duplication of the *CHRNA7* gene in a carrier with an AAO of 47 years and a deletion of the *HSPA6* gene identified in a carrier with an AAO of 49 years. These observations suggest an increased predisposition to PD associated with these specific genetic irregularities. To fully understand the mechanisms behind these associations and their implications for disease progression, further in-depth investigations are essential. However, before any definitive conclusions can be drawn, it is imperative to determine the presence or absence of these CNVs in control samples. Further studies to correlate these candidate CNVs with underlying phenotypes remain essential. Unfortunately, there were no additional relatives of PD patients who could be tested for co-segregation with the disease.

To clarify their potential impact on disease pathogenesis, larger follow-up studies of these variations, evaluation of family segregation with PD, and functional studies are needed. The main limitation of this study is its lack of power, mainly due to the inclusion of only 67 familial PD cases. Despite the small sample size, previous similar studies have reported interesting results. For example, Trinh et al. reported interesting variants in *SPG7* based on a smaller exome sequencing study of 50 early-onset PD patients (Trinh et al., 2019). Despite their limitations, smaller discovery cohorts may still be helpful, although further replication in larger cohorts is needed. This study is particularly interesting in terms of how low-frequency variants, which are more likely to have a negative functional impact, may affect PD.

Conclusions and Perspectives

At approximately 60 years of age, PD affects approximately 2 in 10,000 individuals, and the risk increases as people get older. Aging is the most important risk factor for Parkinson's disease because it correlates with the incidence of the disease (Dauer and Przedborski, 2003). There are treatments available for PD symptoms. However, none of them can stop or slow down the progression of the disease. The discovery of new therapeutic approaches is challenging due to the lack of a recognized cause for the disease. Although some of the pathological mechanisms contributing to cellular dysfunction in PD have been identified, such as the accumulation of toxic α-synuclein oligomers, degeneration of dopaminergic cells in the *SNpc*, oxidative stress and inflammation, the exact cause of the disease remains elusive. Most cases of PD are idiopathic, influenced by both environmental and genetic factors that contribute to an individual's susceptibility to the disease, making understanding the etiology of idiopathic PD a major challenge (Horowitz and Greenamyre, 2010).

The objective of this doctoral thesis was to investigate the genetic architecture of PD in the Luxembourg population to identify known and novel genetic loci associated with this progressive neurodegenerative disorder. The results describe the outcomes of the first genetic analysis of PD-related genes performed in the Luxembourgish population.
Various methods have been used to detect rare variants and study their impact on PD patients, with a particular focus on the *GBA1* gene, which has been identified as a significant genetic risk factor for developing PD.

In **Chapter 1**, based on long-read sequencing technology, we demonstrated the high prevalence of *GBA1* variants in PD patients in Luxembourg (12.1%), reinforcing the idea that *GBA1* is a major genetic risk factor for PD. We showed the effectiveness of targeted PacBio sequencing in detecting known and novel *GBA1* variants with high sensitivity and specificity. We demonstrated that severe variants are more likely to be associated with a higher risk of PD than risk variants, thereby justifying the different levels of pathogenicity proposed in the *GBA1* variant classification. With our study, we have also proposed a concept for categorizing variants of VUS in the *GBA1* gene, based on their localization relative to known variants within the gene sequence and three-dimensional structure. This classification approach has the potential to facilitate the identification of future targeted therapies for patients carrying such VUS.

We then examined the impact of genotype-phenotype correlations on clinical outcomes to find an association between variant carriers and their symptoms and to predict a prodromal phase that might characterize *GBA1* carriers. We found that individuals with severe *GBA1* variants in PD had an average AAO that was approximately four years younger than those without *GBA1* variants. We observed that in severe *GBA1* carriers, non-motor symptoms such as depression and hallucinations,

as well as motor symptoms such as gait disturbance, were associated with a more severe clinical spectrum, highlighting the impact of different degrees of *GBA1* variant severity.

To unravel the precise mechanisms by which *GBA1* variants contribute to the pathogenesis of PD, it is essential to investigate their functional consequences. Understanding the impact on how these variants impact glucocerebrosidase activity, lysosomal function, and α-synuclein aggregation will reveal the fundamental biological processes involved in the development of PD. To this end, we can use patient-derived iPSCs, differentiate them into neurons or macrophages and assess glucocerebrosidase activity in these cells, providing a more relevant pathological context for understanding the impact of the variant on GCase activity. Another approach is to model the GCase protein. Using structural data derived from X-ray crystallography, we can build a model of the GCase protein. This modeling allows us to make predictions regarding how specific variants might impact critical aspects of the enzyme, such as the active site, substrate binding or protein stability. This will help to assess the potential functional consequences resulting from the presence of *GBA1* variants.

Furthermore, exploring intronic variants within the *GBA1* gene is crucial to gain a comprehensive understanding of the full spectrum of genetic variants in this gene. These intronic variants, located in non-coding regions of the gene, are traditionally considered non-functional. Recent research suggested that the non-coding regions may play significant roles (Beutler et al., 1992; Malekkou et al., 2020). Intronic variants may alter numerous regulatory elements, such as enhancers or splicing motifs, and affect the GCase synthesis and processing (Kuo et al., 2021; Ohnmacht et al., 2020). Therefore, investigating intronic variants may reveal novel regulatory processes that affect *GBA1* expression and eventually lead to PD susceptibility. Understanding the molecular mechanisms underlying these intronic variants will be crucial to guide future research and to develop therapeutic strategies that specifically target the *GBA1*-related molecular pathways.

The identification of *GBA1* variants as an important genetic risk factor for PD in the Luxembourg Parkinson's disease study provides new possibilities for therapeutic development. We are contributing to a clinical trial focusing on preventing dementia in individuals with *GBA1*-associated PD, led by our collaborator, Dr Kathrin Brockman from the Hertie Institute for Clinical Brain Research in Tübingen, Germany. Notably, PD patients carrying a pathogenic variant in the *GBA1* gene experience a significantly faster development of cognitive impairment. This cognitive decline is attributed to the abnormal aggregation of several proteins within neurons, in particular α-synuclein proteins. Interestingly, individuals with a mutation in the *GBA1* gene show a pronounced aggregation of α-synuclein instead of the more typical β-amyloid and tau aggregates seen in other forms of the disease with dementia. The clinical trial attempts to reduce cognitive decline in PD patients with a *GBA1* mutation by using Prasinezumab, an antibody designed to block the propagation of α-synuclein aggregates between neurons. This treatment approach holds promise for positively

impacting the outlook for PD patients and offering hope for effective therapies targeting the cognitive aspects of PD in carriers of *GBA1* variants.

An imperative next step for future research is to investigate how risk variants impact biology. In this pursuit, the use of longitudinal cohorts in which biomaterials are repeatedly collected is of great value. Ideally, such cohorts would be initiated at the prodromal stage of the disease and continue until the patient's passing, with brain donation for further study. This approach has the potential to provide valuable insights into biomarkers that accurately reflect the underlying pathological mechanisms and serve as measurable indicators for assessing the effectiveness of targeted interventions. In line with this, we also conducted a longitudinal study focusing on *GBA1* carrier and non-carrier PD and attempted to categorize VUSs based on their clinical outcomes.

To obtain meaningful results in the future and to increase statistical power, the Luxembourgish Parkinson's study should enroll more patients and perform more extensive burden analyses at the level of individual variants, genes and gene sets.

Moreover, collaborative efforts and data-sharing initiatives will be the key to accelerating progress in genetic studies. By establishing large-scale international collaborations and harmonizing data sets, we will ultimately achieve more robust and reproducible results, leading to a better understanding of the complex genetics involved in this disease.

In **Chapter 2**, after identifying *GBA1* variants that might not be detected by the WGS and NeuroChip array screening method, our focus shifted to identifying other genes responsible for PD in our cohort. We carefully screened for the presence of pathogenic variants in genes associated with PD. Specifically, we identified nine carriers with pathogenic variants in the *LRRK2* gene and one carrier with a pathogenic variant in the *PINK1* gene. We also identified one carrier of CNVs in the *PRKN* gene.

Exploring the monogenic history of PD patients is crucial as it can reveal important details about the underlying genetic causes of the condition. While PD is often considered to be a complex multifactorial disorder influenced by a combination of genetic and environmental factors, an important proportion of cases have a monogenic inheritance. By identifying monogenic cases, clinicians can make more accurate diagnoses, enabling early, targeted interventions that can potentially slow down disease progression or improve symptoms. To gain a comprehensive understanding, it is essential to expand our investigation beyond the *GBA1* gene and consider not only the seven PD-causal genes but also explore additional PD-related genes and those involved in lysosomal pathways, such as *MAPT, SAPC* and *SCARB2.*

By expanding our investigation to include additional PD-related genes and exploring rare pathogenic variants, we can improve our comprehension of the genetic complexity of the disease. By integrating a broader genetic perspective, we can identify new genetic contributors and better understand the molecular pathways and mechanisms underlying PD. Currently, the genotyping data from Neurochip is insufficient to explore rare variants. Therefore, incorporating data from WES or WGS, which provides a more comprehensive view of the genetic landscape, becomes imperative in identifying and characterizing rare pathogenic variants and their potential relevance to PD. To achieve this, we collaborated with the Global Parkinson's Genetics Program (GP2) as part of the Complex-Hub initiative and used the new Neuro Booster array. We aimed to identify rare and common variants associated with PD.

The search for rare risk variants faces limitations due to the feasibility of the employed techniques and the sample size. In addition, rare risk variants may have consequences through altered expression or alternative splicing rather than through changes in protein sequence, thus necessitating investigation of the non-coding regions of these genes. Furthermore, although many common variants have been identified in recent years through GWAs analysis, only variants in the *SNCA* and *MAPT* genes have been consistently replicated (Grenn et al., 2020a; Pan et al., 2023). We are curious to explore whether any loci emerge as more significant contributors in our cohort, once we get the necessary samples to conduct a GWAS study.

In **Chapter 3**, we focused on investigating the impact of *GBA1* variants and the PRS of each individual on the susceptibility to develop PD. The PRS can be used to classify individuals into high- and low-risk categories, which can then be used to perform stratified studies that use medications that are successful for specific types of PD in proportion to genetic risk. Our analysis revealed that the risk of developing PD in association with *GBA1* pathogenic variants might be influenced by the PRS, which is thought to exert its effects by influencing multiple pathways associated with PD. We found supportive evidence that the combined presence of a high PRS and *GBA1* pathogenic variants may significantly increase the risk of developing PD. Notably, among individuals carrying both severe and mild *GBA1* pathogenic variants, the presence of a high PRS significantly increased their risk, suggesting an increased susceptibility to PD. These findings underscore the role of polygenic risk in modulating disease susceptibility, particularly in individuals carrying less severe pathogenic variants. A similar study also showed the same trend: The higher PRS was also significantly associated with *GBA1* variant carriers (Blauwendraat et al., 2020b).

We wanted to further analyze how PRS relates to clinical features, such as cognitive decline. A similar study had been conducted on the effect of PRS on clinical outcomes and showed an

association between PD motor outcomes and PRS (Pihlstrøm et al., 2016). The PRS was also found to be significantly associated with earlier AAO. Significant correlations were found between Levodopa-Induced Dyskinesia (LID) and higher PRS (Aradi and Hauser, 2020).

Another study evaluated the combination of PRS, environmental and lifestyle factors associated with PD and showed an association between five factors (age, head injury, family history, depression, and Body Mass Index (BMI)) (Chairta et al., 2021). Although an adequate number of participants is needed for this study, it will probably help us to differentiate between high and low-risk PRS in terms of clinical features.

In **Chapter 4**, we investigated potential PD candidate genes that may play a role in increased susceptibility to PD. We focused on 67 familial PD cases screened by WGS. Through this analysis, we identified a total of 142 rare variants in genes associated with PD, as well as eight CNVs that could potentially contribute to the development of PD.

We are actively involved in the GP2 Monogenic-Hub program with over 200 PD samples. This program recruits and studies PD patients with a potential monogenic factor, such as those with an early onset (< 50 years), a positive family history or an atypical clinical presentation. Our collaboration in this program aims to identify new genes or variants that may be linked to the risk of PD. Through this collaboration effort, we have access to genetic data from samples belonging to the same families, enabling us to conduct a *de novo* analysis. This approach allows us to investigate novel genetic variants and their potential impact on PD, contributing to a deeper understanding of the genetic factors that influence the risk and development of PD.

In conclusion, our comprehensive genetic analyses, including the exploration of rare variants, PRS, and the interplay between clinical outcomes, have yielded significant progress in unraveling the genetic landscape of PD in the Luxembourgish Parkinson's disease study. The identification of important genetic risk factors, such as the *GBA1* gene, has underscored the complex genetic nature of PD and highlighted the need for further research. Studying the genetics of common neurodegenerative diseases such as PD is of paramount importance to fully understand the underlying causes of the disease and to develop innovative therapies that may slow disease progression or even provide a cure. While familial cases of PD caused by inherited gene variants may offer relatively clear avenues for investigation, the challenge becomes greater when dealing with idiopathic cases, where multiple factors are at play, making it difficult to pinpoint a single direct cause.

To date, numerous genes and loci identified by GWAS have been implicated in the development of PD. Investigation of genetic risk factors must continue, along with a concerted and focused effort to

understand the consequences of these findings at the molecular and biological levels. By combining genetic insights with functional knowledge, we can lay the foundation for therapeutic strategies that directly target the underlying causes of the disease, offering hope for more effective and personalized treatments for individuals affected by PD.

Patient-derived cell models, particularly iPSCs with isogenic controls, provide a powerful framework for analyzing and understanding the effects of specific genetic variables in a controlled experimental setting. These models have important implications for drug discovery, understanding of disease mechanisms, and personalized treatment.

Bibliography

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A., 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073. https://doi.org/10.1038/nature09534

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R., 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393

Abbas, N., Lücking, C.B., Ricard, S., Dürr, A., Bonifati, V., De Michele, G., Bouley, S., Vaughan, J.R., Gasser, T., Marconi, R., Broussolle, E., Brefel-Courbon, C., Harhangi, B.S., Oostra, B.A., Fabrizio, E., Böhme, G.A., Pradier, L., Wood, N.W., Filla, A., Meco, G., Denefle, P., Agid, Y., Brice, A., 1999. A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. French Parkinson's Disease Genetics Study Group and the European Consortium on Genetic Susceptibility in Parkinson's Disease. Hum. Mol. Genet. 8, 567–574. https://doi.org/10.1093/hmg/8.4.567

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. Molecular Biology of the Cell, 4th ed. Garland Science.

Alcalay, R.N., Dinur, T., Quinn, T., Sakanaka, K., Levy, O., Waters, C., Fahn, S., Dorovski, T., Chung, W.K., Pauciulo, M., Nichols, W., Rana, H.Q., Balwani, M., Bier, L., Elstein, D., Zimran, A., 2014. Comparison of Parkinson risk in Ashkenazi Jewish patients with Gaucher disease and GBA heterozygotes. JAMA Neurol. 71, 752–757. https://doi.org/10.1001/jamaneurol.2014.313

Alcalay, R.N., Levy, O.A., Waters, C.C., Fahn, S., Ford, B., Kuo, S.-H., Mazzoni, P., Pauciulo, M.W., Nichols, W.C., Gan-Or, Z., Rouleau, G.A., Chung, W.K., Wolf, P., Oliva, P., Keutzer, J., Marder, K., Zhang, X., 2015. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. Brain J. Neurol. 138, 2648–2658. https://doi.org/10.1093/brain/awv179

Alkan, C., Coe, B.P., Eichler, E.E., 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12, 363–376. https://doi.org/10.1038/nrg2958

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21, 30. https://doi.org/10.1186/s13059-020-1935-5

Ambrosi, G., Ghezzi, C., Zangaglia, R., Levandis, G., Pacchetti, C., Blandini, F., 2015. Ambroxol-induced rescue of defective glucocerebrosidase is associated with increased LIMP-2 and saposin C levels in GBA1 mutant Parkinson's disease cells. Neurobiol. Dis. 82, 235–242. https://doi.org/10.1016/j.nbd.2015.06.008

Anheim, M., Elbaz, A., Lesage, S., Durr, A., Condroyer, C., Viallet, F., Pollak, P., Bonaïti, B., Bonaïti-Pellié, C., Brice, A., French Parkinson Disease Genetic Group, 2012. Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. Neurology 78, 417–420. https://doi.org/10.1212/WNL.0b013e318245f476

Aradi, S.D., Hauser, R.A., 2020. Medical Management and Prevention of Motor Complications in Parkinson's Disease. Neurother. J. Am. Soc. Exp. Neurother. 17, 1339–1365. https://doi.org/10.1007/s13311-020-00889-4

Armstrong, L.C., Komiya, T., Bergman, B.E., Mihara, K., Bornstein, P., 1997. Metaxin is a component of a preprotein import complex in the outer membrane of the mammalian mitochondrion. J. Biol. Chem. 272, 6510–6518. https://doi.org/10.1074/jbc.272.10.6510

Artusi, C.A., Dwivedi, A.K., Romagnolo, A., Pal, G., Kauffman, M., Mata, I., Patel, D., Vizcarra, J.A., Duker, A., Marsili, L., Cheeran, B., Woo, D., Contarino, M.F., Verhagen, L., Lopiano, L., Espay, A.J., Fasano, A., Merola, A., 2019. Association of Subthalamic Deep Brain Stimulation With Motor, Functional, and Pharmacologic Outcomes in Patients With Monogenic Parkinson Disease. JAMA Netw. Open 2, e187800. https://doi.org/10.1001/jamanetworkopen.2018.7800

Asi, Y.T., Simpson, J.E., Heath, P.R., Wharton, S.B., Lees, A.J., Revesz, T., Houlden, H., Holton, J.L., 2014. Alpha-synuclein mRNA expression in oligodendrocytes in MSA. Glia 62, 964–970. https://doi.org/10.1002/glia.22653

Badanjak, K., Fixemer, S., Smajić, S., Skupin, A., Grünewald, A., 2021. The Contribution of Microglia to Neuroinflammation in Parkinson's Disease. Int. J. Mol. Sci. 22, 4676. https://doi.org/10.3390/ijms22094676

Balding, D.J., 2006. A tutorial on statistical methods for population association studies. Nat. Rev. Genet. 7, 781–791. https://doi.org/10.1038/nrg1916

Bandres-Ciga, S., Diez-Fairen, M., Kim, J.J., Singleton, A.B., 2020. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. Neurobiol. Dis. 137, 104782. https://doi.org/10.1016/j.nbd.2020.104782

Barbuti, P.A., Ohnmacht, J., Santos, B.F.R., Antony, P.M., Massart, F., Cruciani, G., Dording, C.M., Pavelka, L., Casadei, N., Kwon, Y.-J., Krüger, R., 2021. Gene-corrected p.A30P SNCA patient-derived isogenic neurons rescue neuronal branching and function. Sci. Rep. 11, 21946. https://doi.org/10.1038/s41598-021-01505-x

Beavan, M., McNeill, A., Proukakis, C., Hughes, D.A., Mehta, A., Schapira, A.H.V., 2015. Evolution of prodromal clinical markers of Parkinson disease in a GBA mutation-positive cohort. JAMA Neurol. 72, 201–208. https://doi.org/10.1001/jamaneurol.2014.2950

Beilina, A., Rudenko, I.N., Kaganovich, A., Civiero, L., Chau, H., Kalia, S.K., Kalia, L.V., Lobbestael, E., Chia, R., Ndukwe, K., Ding, J., Nalls, M.A., International Parkinson's Disease Genomics Consortium, North American Brain Expression Consortium, Olszewski, M., Hauser, D.N., Kumaran, R., Lozano, A.M., Baekelandt, V., Greene, L.E., Taymans, J.-M., Greggio, E., Cookson, M.R., 2014. Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. Proc. Natl. Acad. Sci. U. S. A. 111, 2626–2631. https://doi.org/10.1073/pnas.1318306111

Bendikov-Bar, I., Horowitz, M., 2012. Gaucher disease paradigm: from ERAD to comorbidity. Hum. Mutat. 33, 1398–1407. https://doi.org/10.1002/humu.22124

Bendikov-Bar, I., Maor, G., Filocamo, M., Horowitz, M., 2013. Ambroxol as a pharmacological chaperone for mutant glucocerebrosidase. Blood Cells. Mol. Dis. 50, 141–145. https://doi.org/10.1016/j.bcmd.2012.10.007

Bendikov-Bar, I., Ron, I., Filocamo, M., Horowitz, M., 2011. Characterization of the ERAD process of the L444P mutant glucocerebrosidase variant. Blood Cells. Mol. Dis. 46, 4–10. https://doi.org/10.1016/j.bcmd.2010.10.012

Benskey, M.J., Perez, R.G., Manfredsson, F.P., 2016. The contribution of alpha synuclein to neuronal survival and function - Implications for Parkinson's disease. J. Neurochem. 137, 331–359. https://doi.org/10.1111/jnc.13570

Beutler, E., West, C., Gelbart, T., 1992. Polymorphisms in the human glucocerebrosidase gene. Genomics 12, 795–800. https://doi.org/10.1016/0888-7543(92)90311-f

Bhidayasiri, R., Rattanachaisit, W., Phokaewvarangkul, O., Lim, T.T., Fernandez, H.H., 2019. Exploring bedside clinical features of parkinsonism: A focus on differential diagnosis. Parkinsonism Relat. Disord. 59, 74–81. https://doi.org/10.1016/j.parkreldis.2018.11.005

Billingsley, K.J., Bandres-Ciga, S., Saez-Atienzar, S., Singleton, A.B., 2018. Genetic risk factors in Parkinson's disease. Cell Tissue Res. 373, 9–20. https://doi.org/10.1007/s00441-018-2817-y

Biskup, S., Moore, D.J., Celsi, F., Higashi, S., West, A.B., Andrabi, S.A., Kurkinen, K., Yu, S.-W., Savitt, J.M., Waldvogel, H.J., Faull, R.L.M., Emson, P.C., Torp, R., Ottersen, O.P., Dawson, T.M., Dawson, V.L., 2006. Localization of LRRK2 to membranous and vesicular structures in mammalian brain. Ann. Neurol. 60, 557–569. https://doi.org/10.1002/ana.21019

Blauwendraat, C., Faghri, F., Pihlstrom, L., Geiger, J.T., Elbaz, A., Lesage, S., Corvol, J.-C., May, P., Nicolas, A., Abramzon, Y., Murphy, N.A., Gibbs, J.R., Ryten, M., Ferrari, R., Bras, J., Guerreiro, R., Williams, J., Sims, R., Lubbe, S., Hernandez, D.G., Mok, K.Y., Robak, L., Campbell, R.H., Rogaeva, E., Traynor, B.J., Chia, R., Chung, S.J., International Parkinson's Disease Genomics Consortium (IPDGC), COURAGE-PD Consortium, Hardy, J.A., Brice, A., Wood, N.W., Houlden, H., Shulman, J.M., Morris, H.R., Gasser, T., Krüger, R., Heutink, P., Sharma, M., Simón-Sánchez, J., Nalls, M.A., Singleton, A.B., Scholz, S.W., 2017.

NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. Neurobiol. Aging 57, 247.e9-247.e13. https://doi.org/10.1016/j.neurobiolaging.2017.05.009

Blauwendraat, C., Nalls, M.A., Singleton, A.B., 2020a. The genetic architecture of Parkinson's disease. Lancet Neurol. 19, 170–178. https://doi.org/10.1016/S1474-4422(19)30287-X

Blauwendraat, C., Reed, X., Krohn, L., Heilbron, K., Bandres-Ciga, S., Tan, M., Gibbs, J.R., Hernandez, D.G., Kumaran, R., Langston, R., Bonet-Ponce, L., Alcalay, R.N., Hassin-Baer, S., Greenbaum, L., Iwaki, H., Leonard, H.L., Grenn, F.P., Ruskey, J.A., Sabir, M., Ahmed, S., Makarious, M.B., Pihlstrøm, L., Toft, M., van Hilten, J.J., Marinus, J., Schulte, C., Brockmann, K., Sharma, M., Siitonen, A., Majamaa, K., Eerola-Rautio, J., Tienari, P.J., 23andMe Research Team, Pantelyat, A., Hillis, A.E., Dawson, T.M., Rosenthal, L.S., Albert, M.S., Resnick, S.M., Ferrucci, L., Morris, C.M., Pletnikova, O., Troncoso, J., Grosset, D., Lesage, S., Corvol, J.-C., Brice, A., Noyce, A.J., Masliah, E., Wood, N., Hardy, J., Shulman, L.M., Jankovic, J., Shulman, J.M., Heutink, P., Gasser, T., Cannon, P., Scholz, S.W., Morris, H., Cookson, M.R., Nalls, M.A., Gan-Or, Z., Singleton, A.B., 2020b. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. Brain J. Neurol. 143, 234–248. https://doi.org/10.1093/brain/awz350

Bloem, B.R., Okun, M.S., Klein, C., 2021. Parkinson's disease. Lancet Lond. Engl. 397, 2284–2303. https://doi.org/10.1016/S0140-6736(21)00218-X

Braak, H., Del Tredici, K., Rüb, U., de Vos, R.A.I., Jansen Steur, E.N.H., Braak, E., 2003. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol. Aging 24, 197–211. https://doi.org/10.1016/s0197-4580(02)00065-9

Bragg, D.C., Mangkalaphiban, K., Vaine, C.A., Kulkarni, N.J., Shin, D., Yadav, R., Dhakal, J., Ton, M.-L., Cheng, A., Russo, C.T., Ang, M., Acuña, P., Go, C., Franceour, T.N., Multhaupt-Buell, T., Ito, N., Müller, U., Hendriks, W.T., Breakefield, X.O., Sharma, N., Ozelius, L.J., 2017. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. Proc. Natl. Acad. Sci. U. S. A. 114, E11020–E11028. https://doi.org/10.1073/pnas.1712526114

Brockmann, K., Srulijes, K., Pflederer, S., Hauser, A.-K., Schulte, C., Maetzler, W., Gasser, T., Berg, D., 2015. GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study. Mov. Disord. Off. J. Mov. Disord. Soc. 30, 407–411. https://doi.org/10.1002/mds.26071

Brooks, J., Ding, J., Simon-Sanchez, J., Paisan-Ruiz, C., Singleton, A.B., Scholz, S.W., 2009. Parkin and PINK1 mutations in early-onset Parkinson's disease: comprehensive screening in publicly available cases and control. J. Med. Genet. 46, 375–381. https://doi.org/10.1136/jmg.2008.063917

Burbulla, L.F., Song, P., Mazzulli, J.R., Zampese, E., Wong, Y.C., Jeon, S., Santos, D.P., Blanz, J., Obermaier, C.D., Strojny, C., Savas, J.N., Kiskinis, E., Zhuang, X., Krüger, R., Surmeier, D.J., Krainc, D., 2017. Dopamine oxidation mediates mitochondrial and lysosomal dysfunction in Parkinson's disease. Science 357, 1255–1261. https://doi.org/10.1126/science.aam9080

Carvalho, C.M.B., Lupski, J.R., 2016. Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. 17, 224–238. https://doi.org/10.1038/nrg.2015.25

Chairta, P.P., Hadjisavvas, A., Georgiou, A.N., Loizidou, M.A., Yiangou, K., Demetriou, C.A., Christou, Y.P., Pantziaris, M., Michailidou, K., Zamba-Papanicolaou, E., 2021. Prediction of Parkinson's Disease Risk Based on Genetic Profile and Established Risk Factors. Genes 12, 1278. https://doi.org/10.3390/genes12081278

Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., Fan, X., Wen, J., Handsaker, R.E., Fairley, S., Kronenberg, Z.N., Kong, X., Hormozdiari, F., Lee, D., Wenger, A.M., Hastie, A.R., Antaki, D., Anantharaman, T., Audano, P.A., Brand, H., Cantsilieris, S., Cao, H., Cerveira, E., Chen, C., Chen, X., Chin, C.-S., Chong, Z., Chuang, N.T., Lambert, C.C., Church, D.M., Clarke, L., Farrell, A., Flores, J., Galeev, T., Gorkin, D.U., Gujral, M., Guryev, V., Heaton, W.H., Korlach, J., Kumar, S., Kwon, J.Y., Lam, E.T., Lee, J.E., Lee, J., Lee, W.-P., Lee, S.P., Li, S., Marks, P., Viaud-Martinez, K., Meiers, S., Munson, K.M., Navarro, F.C.P., Nelson, B.J.,

Nodzak, C., Noor, A., Kyriazopoulou-Panagiotopoulou, S., Pang, A.W.C., Qiu, Y., Rosanio, G., Ryan, M., Stütz, A., Spierings, D.C.J., Ward, A., Welch, A.E., Xiao, M., Xu, W., Zhang, C., Zhu, Q., Zheng-Bradley, X., Lowy, E., Yakneen, S., McCarroll, S., Jun, G., Ding, L., Koh, C.L., Ren, B., Flicek, P., Chen, K., Gerstein, M.B., Kwok, P.-Y., Lansdorp, P.M., Marth, G.T., Sebat, J., Shi, X., Bashir, A., Ye, K., Devine, S.E., Talkowski, M.E., Mills, R.E., Marschall, T., Korbel, J.O., Eichler, E.E., Lee, C., 2019. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. 10, 1784. https://doi.org/10.1038/s41467-018-08148-z

Chan, N.C., Chan, D.C., 2011. Parkin uses the UPS to ship off dysfunctional mitochondria. Autophagy 7, 771–772. https://doi.org/10.4161/auto.7.7.15453

Chang, D., Nalls, M.A., Hallgrímsdóttir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., International Parkinson's Disease Genomics Consortium, 23andMe Research Team, Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., Hinds, D., Behrens, T.W., Singleton, A.B., Bhangale, T.R., Graham, R.R., 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat. Genet. 49, 1511–1516. https://doi.org/10.1038/ng.3955

Chennen, K., Weber, T., Lornage, X., Kress, A., Böhm, J., Thompson, J., Laporte, J., Poch, O., 2020. MISTIC: A prediction tool to reveal disease-relevant deleterious missense variants. PloS One 15, e0236962. https://doi.org/10.1371/journal.pone.0236962

Cilia, R., Tunesi, S., Marotta, G., Cereda, E., Siri, C., Tesei, S., Zecchinelli, A.L., Canesi, M., Mariani, C.B., Meucci, N., Sacilotto, G., Zini, M., Barichella, M., Magnani, C., Duga, S., Asselta, R., Soldà, G., Seresini, A., Seia, M., Pezzoli, G., Goldwurm, S., 2016. Survival and dementia in GBA-associated Parkinson's disease: The mutation matters. Ann. Neurol. 80, 662–673. https://doi.org/10.1002/ana.24777

Clark, L.N., Ross, B.M., Wang, Y., Mejia-Santana, H., Harris, J., Louis, E.D., Cote, L.J., Andrews, H., Fahn, S., Waters, C., Ford, B., Frucht, S., Ottman, R., Marder, K., 2007. Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. Neurology 69, 1270–1277. https://doi.org/10.1212/01.wnl.0000276989.17578.02

Clayton-Smith, J., O'Sullivan, J., Daly, S., Bhaskar, S., Day, R., Anderson, B., Voss, A.K., Thomas, T., Biesecker, L.G., Smith, P., Fryer, A., Chandler, K.E., Kerr, B., Tassabehji, M., Lynch, S.-A., Krajewska-Walasek, M., McKee, S., Smith, J., Sweeney, E., Mansour, S., Mohammed, S., Donnai, D., Black, G., 2011. Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. Am. J. Hum. Genet. 89, 675–681. https://doi.org/10.1016/j.ajhg.2011.10.008

Cleeter, M.W.J., Chau, K.-Y., Gluck, C., Mehta, A., Hughes, D.A., Duchen, M., Wood, N.W., Hardy, J., Mark Cooper, J., Schapira, A.H., 2013. Glucocerebrosidase inhibition causes mitochondrial dysfunction and free radical damage. Neurochem. Int. 62, 1–7. https://doi.org/10.1016/j.neuint.2012.10.010

Cochemé, H.M., Murphy, M.P., 2008. Complex I is the major site of mitochondrial superoxide production by paraquat. J. Biol. Chem. 283, 1786–1798. https://doi.org/10.1074/jbc.M708597200

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., Watts, N.A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C.W., Huang, Y., Brookings, T., Sharpe, T., Stone, M.R., Valkanas, E., Fu, J., Tiao, G., Laricchia, K.M., Ruano-Rubio, V., Stevens, C., Gupta, N., Cusick, C., Margolin, L., Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Taylor, K.D., Lin, H.J., Rich, S.S., Post, W.S., Chen, Y.-D.I., Rotter, J.I., Nusbaum, C., Philippakis, A., Lander, E., Gabriel, S., Neale, B.M., Kathiresan, S., Daly, M.J., Banks, E., MacArthur, D.G., Talkowski, M.E., 2020. A structural variation reference for medical and population genetics. Nature 581, 444–451. https://doi.org/10.1038/s41586-020-2287-8

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W.C., Robson, S., Stirrups, K., Valsesia, A.,

Walter, K., Wei, J., Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., Hurles, M.E., 2010. Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712. https://doi.org/10.1038/nature08516

Coppedè, F., 2012. Genetics and epigenetics of Parkinson's disease. ScientificWorldJournal 2012, 489830. https://doi.org/10.1100/2012/489830

Cotto, K.C., Feng, Y.-Y., Ramu, A., Richters, M., Freshour, S.L., Skidmore, Z.L., Xia, H., McMichael, J.F., Kunisaki, J., Campbell, K.M., Chen, T.H.-P., Rozycki, E.B., Adkins, D., Devarakonda, S., Sankararaman, S., Lin, Y., Chapman, W.C., Maher, C.A., Arora, V., Dunn, G.P., Uppaluri, R., Govindan, R., Griffith, O.L., Griffith, M., 2023. Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. Nat. Commun. 14, 1589. https://doi.org/10.1038/s41467-023-37266-6

Cuervo, A.M., Stefanis, L., Fredenburg, R., Lansbury, P.T., Sulzer, D., 2004. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. Science 305, 1292–1295. https://doi.org/10.1126/science.1101738

Cullen, V., Sardi, S.P., Ng, J., Xu, Y.-H., Sun, Y., Tomlinson, J.J., Kolodziej, P., Kahn, I., Saftig, P., Woulfe, J., Rochet, J.-C., Glicksman, M.A., Cheng, S.H., Grabowski, G.A., Shihabuddin, L.S., Schlossmacher, M.G., 2011. Acid β-glucosidase mutants linked to Gaucher disease, Parkinson disease, and Lewy body dementia alter α-synuclein processing. Ann. Neurol. 69, 940–953. https://doi.org/10.1002/ana.22400

Damier, P., Hirsch, E.C., Zhang, P., Agid, Y., Javoy-Agid, F., 1993. Glutathione peroxidase, glial cells and Parkinson's disease. Neuroscience 52, 1–6. https://doi.org/10.1016/0306-4522(93)90175-f

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287. https://doi.org/10.1038/ng.3656

Dauer, W., Przedborski, S., 2003. Parkinson's disease: mechanisms and models. Neuron 39, 889–909. https://doi.org/10.1016/s0896-6273(03)00568-3

Dawson, T.M., Dawson, V.L., 2003. Molecular pathways of neurodegeneration in Parkinson's disease. Science 302, 819–822. https://doi.org/10.1126/science.1087753

Del Rey, N.L.-G., Quiroga-Varela, A., Garbayo, E., Carballo-Carbajal, I., Fernández-Santiago, R., Monje, M.H.G., Trigo-Damas, I., Blanco-Prieto, M.J., Blesa, J., 2018. Advances in Parkinson's Disease: 200 Years Later. Front. Neuroanat. 12, 113. https://doi.org/10.3389/fnana.2018.00113

Deng, H.-X., Shi, Y., Yang, Y., Ahmeti, K.B., Miller, N., Huang, C., Cheng, L., Zhai, H., Deng, S., Nuytemans, K., Corbett, N.J., Kim, M.J., Deng, H., Tang, B., Yang, Z., Xu, Y., Chan, P., Huang, B., Gao, X.-P., Song, Z., Liu, Z., Fecto, F., Siddique, N., Foroud, T., Jankovic, J., Ghetti, B., Nicholson, D.A., Krainc, D., Melen, O., Vance, J.M., Pericak-Vance, M.A., Ma, Y.-C., Rajput, A.H., Siddique, T., 2016. Identification of TMEM230 mutations in familial Parkinson's disease. Nat. Genet. 48, 733–739. https://doi.org/10.1038/ng.3589

Dias, V., Junn, E., Mouradian, M.M., 2013. The role of oxidative stress in Parkinson's disease. J. Park. Dis. 3, 461–491. https://doi.org/10.3233/JPD-130230

Dickson, D.W., 2018. Neuropathology of Parkinson disease. Parkinsonism Relat. Disord. 46 Suppl 1, S30–S33. https://doi.org/10.1016/j.parkreldis.2017.07.033

Do, J., McKinney, C., Sharma, P., Sidransky, E., 2019. Glucocerebrosidase and its relevance to Parkinson disease. Mol. Neurodegener. 14, 36. https://doi.org/10.1186/s13024-019-0336-2

Dorsey, E.R., Sherer, T., Okun, M.S., Bloem, B.R., 2018. The Emerging Evidence of the Parkinson Pandemic. J. Park. Dis. 8, S3–S8. https://doi.org/10.3233/JPD-181474

Duran, R., Mencacci, N.E., Angeli, A.V., Shoai, M., Deas, E., Houlden, H., Mehta, A., Hughes, D., Cox, T.M., Deegan, P., Schapira, A.H., Lees, A.J., Limousin, P., Jarman, P.R., Bhatia, K.P., Wood, N.W., Hardy, J., Foltynie, T., 2013. The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 28, 232–236. https://doi.org/10.1002/mds.25248

Dvir, H., Harel, M., McCarthy, A.A., Toker, L., Silman, I., Futerman, A.H., Sussman, J.L., 2003. X-ray structure of human acid-β-glucosidase, the defective enzyme in Gaucher disease. EMBO Rep. 4, 704–709. https://doi.org/10.1038/sj.embor.embor873

Eichler, E.E., 2019. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. N. Engl. J. Med. 381, 64–74. https://doi.org/10.1056/NEJMra1809315

Fernandes, H.J.R., Hartfield, E.M., Christian, H.C., Emmanoulidou, E., Zheng, Y., Booth, H., Bogetofte, H., Lang, C., Ryan, B.J., Sardi, S.P., Badger, J., Vowles, J., Evetts, S., Tofaris, G.K., Vekrellis, K., Talbot, K., Hu, M.T., James, W., Cowley, S.A., Wade-Martins, R., 2016. ER Stress and Autophagic Perturbations Lead to Elevated Extracellular α-Synuclein in GBA-N370S Parkinson's iPSC-Derived Dopamine Neurons. Stem Cell Rep. 6, 342–356. https://doi.org/10.1016/j.stemcr.2016.01.013

Fernández-Santiago, R., Sharma, M., 2022. What have we learned from genome-wide association studies (GWAS) in Parkinson's disease? Ageing Res. Rev. 79, 101648. https://doi.org/10.1016/j.arr.2022.101648

Fiesel, F.C., Ando, M., Hudec, R., Hill, A.R., Castanedes-Casey, M., Caulfield, T.R., Moussaud-Lamodière, E.L., Stankowski, J.N., Bauer, P.O., Lorenzo-Betancor, O., Ferrer, I., Arbelo, J.M., Siuda, J., Chen, L., Dawson, V.L., Dawson, T.M., Wszolek, Z.K., Ross, O.A., Dickson, D.W., Springer, W., 2015. (Patho-)physiological relevance of PINK1-dependent ubiquitin phosphorylation. EMBO Rep. 16, 1114–1130. https://doi.org/10.15252/embr.201540514

Frank-Cannon, T.C., Tran, T., Ruhn, K.A., Martinez, T.N., Hong, J., Marvin, M., Hartley, M., Treviño, I., O'Brien, D.E., Casey, B., Goldberg, M.S., Tansey, M.G., 2008. Parkin deficiency increases vulnerability to inflammation-related nigral degeneration. J. Neurosci. Off. J. Soc. Neurosci. 28, 10825–10834. https://doi.org/10.1523/JNEUROSCI.3001-08.2008

Galpern, W.R., Lang, A.E., 2006. Interface between tauopathies and synucleinopathies: a tale of two proteins. Ann. Neurol. 59, 449–458. https://doi.org/10.1002/ana.20819

Galvagnion, C., Marlet, F.R., Cerri, S., Schapira, A.H.V., Blandini, F., Di Monte, D.A., 2022. Sphingolipid changes in Parkinson L444P GBA mutation fibroblasts promote α-synuclein aggregation. Brain J. Neurol. 145, 1038–1051. https://doi.org/10.1093/brain/awab371

Gan, B., Hu, J., Jiang, S., Liu, Y., Sahin, E., Zhuang, L., Fletcher-Sananikone, E., Colla, S., Wang, Y.A., Chin, L., Depinho, R.A., 2010. Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells. Nature 468, 701–704. https://doi.org/10.1038/nature09595

Gan-Or, Z., Amshalom, I., Kilarski, L.L., Bar-Shira, A., Gana-Weisz, M., Mirelman, A., Marder, K., Bressman, S., Giladi, N., Orr-Urtreger, A., 2015. Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology 84, 880–887. https://doi.org/10.1212/WNL.0000000000001315

Gan-Or, Z., Giladi, N., Rozovski, U., Shifrin, C., Rosner, S., Gurevich, T., Bar-Shira, A., Orr-Urtreger, A., 2008. Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. Neurology 70, 2277–2283. https://doi.org/10.1212/01.wnl.0000304039.11891.29

Gao, V., Briano, J.A., Komer, L.E., Burré, J., 2023. Functional and Pathological Effects of α-Synuclein on Synaptic SNARE Complexes. J. Mol. Biol. 435, 167714. https://doi.org/10.1016/j.jmb.2022.167714

Gegg, M.E., Burke, D., Heales, S.J.R., Cooper, J.M., Hardy, J., Wood, N.W., Schapira, A.H.V., 2012. Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains. Ann. Neurol. 72, 455–463. https://doi.org/10.1002/ana.23614

Geisler, S., Holmström, K.M., Skujat, D., Fiesel, F.C., Rothfuss, O.C., Kahle, P.J., Springer, W., 2010. PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. Nat. Cell Biol. 12, 119–131. https://doi.org/10.1038/ncb2012

Gillardon, F., Schmid, R., Draheim, H., 2012. Parkinson's disease-linked leucine-rich repeat kinase 2(R1441G) mutation increases proinflammatory cytokine release from activated primary microglial cells and resultant neurotoxicity. Neuroscience 208, 41–48. https://doi.org/10.1016/j.neuroscience.2012.02.001

Gilsbach, B.K., Kortholt, A., 2014. Structural biology of the LRRK2 GTPase and kinase domains: implications for regulation. Front. Mol. Neurosci. 7, 32. https://doi.org/10.3389/fnmol.2014.00032

Gloeckner, C.J., Kinkl, N., Schumacher, A., Braun, R.J., O'Neill, E., Meitinger, T., Kolch, W., Prokisch, H., Ueffing, M., 2006. The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. Hum. Mol. Genet. 15, 223–232. https://doi.org/10.1093/hmg/ddi439

Goedert, M., Spillantini, M.G., Del Tredici, K., Braak, H., 2013. 100 years of Lewy pathology. Nat. Rev. Neurol. 9, 13–24. https://doi.org/10.1038/nrneurol.2012.242

Goh, G., Choi, M., 2012. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. Genomics Inform. 10, 214–219. https://doi.org/10.5808/GI.2012.10.4.214

Goker-Alpan, O., Giasson, B.I., Eblan, M.J., Nguyen, J., Hurtig, H.I., Lee, V.M.-Y., Trojanowski, J.Q., Sidransky, E., 2006. Glucocerebrosidase mutations are an important risk factor for Lewy body disorders. Neurology 67, 908–910. https://doi.org/10.1212/01.wnl.0000230215.41296.18

Goker-Alpan, O., Schiffmann, R., LaMarca, M.E., Nussbaum, R.L., McInerney-Leo, A., Sidransky, E., 2004. Parkinsonism among Gaucher disease carriers. J. Med. Genet. 41, 937–940. https://doi.org/10.1136/jmg.2004.024455

Goker-Alpan, O., Stubblefield, B.K., Giasson, B.I., Sidransky, E., 2010. Glucocerebrosidase is present in α-synuclein inclusions in Lewy body disorders. Acta Neuropathol. (Berl.) 120, 641–649. https://doi.org/10.1007/s00401-010-0741-7

Goldman, J.G., Postuma, R., 2014. Premotor and nonmotor features of Parkinson's disease. Curr. Opin. Neurol. 27, 434–441. https://doi.org/10.1097/WCO.0000000000000112

Grabowski, G.A., 2008. Phenotype, diagnosis, and treatment of Gaucher's disease. Lancet Lond. Engl. 372, 1263–1271. https://doi.org/10.1016/S0140-6736(08)61522-6

Grabowski, G.A., Horowitz, M., 1997. Gaucher's disease: molecular, genetic and enzymological aspects. Baillieres Clin. Haematol. 10, 635–656. https://doi.org/10.1016/s0950-3536(97)80032-7

Greenfield, J.G., Bosanquet, F.D., 1953. The brain-stem lesions in Parkinsonism. J. Neurol. Neurosurg. Psychiatry 16, 213–226. https://doi.org/10.1136/jnnp.16.4.213

Grenn, F.P., Kim, J.J., Makarious, M.B., Iwaki, H., Illarionova, A., Brolin, K., Kluss, J.H., Schumacher-Schuh, A.F., Leonard, H., Faghri, F., Billingsley, K., Krohn, L., Hall, A., Diez-Fairen, M., Periñán, M.T., Foo, J.N., Sandor, C., Webber, C., Fiske, B.K., Gibbs, J.R., Nalls, M.A., Singleton, A.B., Bandres-Ciga, S., Reed, X., Blauwendraat, C., 2020a. The Parkinson's Disease Genome-Wide Association Study Locus Browser. Mov. Disord. 35, 2056–2067. https://doi.org/10.1002/mds.28197

Grenn, F.P., Kim, J.J., Makarious, M.B., Iwaki, H., Illarionova, A., Brolin, K., Kluss, J.H., Schumacher-Schuh, A.F., Leonard, H., Faghri, F., Billingsley, K., Krohn, L., Hall, A., Diez-Fairen, M., Periñán, M.T., Foo, J.N., Sandor, C., Webber, C., Fiske, B.K., Gibbs, J.R., Nalls, M.A., Singleton, A.B., Bandres-Ciga, S., Reed, X., Blauwendraat, C., International Parkinson's Disease Genomics Consortium (IPDGC), 2020b. The Parkinson's Disease Genome-Wide Association Study Locus Browser. Mov. Disord. Off. J. Mov. Disord. Soc. 35, 2056–2067. https://doi.org/10.1002/mds.28197

Gudmundsson, S., Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Genome Aggregation Database Consortium, Neale, B.M., Daly, M.J., MacArthur, D.G., 2021. Addendum: The

mutational constraint spectrum quantified from variation in 141,456 humans. Nature 597, E3–E4. https://doi.org/10.1038/s41586-021-03758-y

Guo, J.-F., Zhang, L., Li, K., Mei, J.-P., Xue, J., Chen, J., Tang, X., Shen, L., Jiang, H., Chen, C., Guo, H., Wu, X.-L., Sun, S.-L., Xu, Q., Sun, Q.-Y., Chan, P., Shang, H.-F., Wang, T., Zhao, G.-H., Liu, J.-Y., Xie, X.-F., Jiang, Y.-Q., Liu, Z.-H., Zhao, Y.-W., Zhu, Z.-B., Li, J., Hu, Z.-M., Yan, X.-X., Fang, X.-D., Wang, G.-H., Zhang, F.-Y., Xia, K., Liu, C.-Y., Zhu, X.-W., Yue, Z.-Y., Li, S.C., Cai, H.-B., Zhang, Z.-H., Duan, R.-H., Tang, B.-S., 2018. Coding mutations in NUS1 contribute to Parkinson's disease. Proc. Natl. Acad. Sci. U. S. A. 115, 11567–11572. https://doi.org/10.1073/pnas.1809969115

Guo, Y., Sun, Y., Song, Z., Zheng, W., Xiong, W., Yang, Y., Yuan, L., Deng, H., 2021. Genetic Analysis and Literature Review of SNCA Variants in Parkinson's Disease. Front. Aging Neurosci. 13, 648151. https://doi.org/10.3389/fnagi.2021.648151

Haaxma, C.A., Bloem, B.R., Borm, G.F., Oyen, W.J.G., Leenders, K.L., Eshuis, S., Booij, J., Dluzen, D.E., Horstink, M.W.I.M., 2007. Gender differences in Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 78, 819–824. https://doi.org/10.1136/jnnp.2006.103788

Halperin, A., Elstein, D., Zimran, A., 2006. Increased incidence of Parkinson disease among relatives of patients with Gaucher disease. Blood Cells. Mol. Dis. 36, 426–428. https://doi.org/10.1016/j.bcmd.2006.02.004

Hamza, T.H., Zabetian, C.P., Tenesa, A., Laederach, A., Montimurro, J., Yearout, D., Kay, D.M., Doheny, K.F., Paschall, J., Pugh, E., Kusel, V.I., Collura, R., Roberts, J., Griffith, A., Samii, A., Scott, W.K., Nutt, J., Factor, S.A., Payami, H., 2010. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat. Genet. 42, 781–785. https://doi.org/10.1038/ng.642

Hanss, Z., Larsen, S.B., Antony, P., Mencke, P., Massart, F., Jarazo, J., Schwamborn, J.C., Barbuti, P.A., Mellick, G.D., Krüger, R., 2021. Mitochondrial and Clearance Impairment in p.D620N VPS35 Patient-Derived Neurons. Mov. Disord. Off. J. Mov. Disord. Soc. 36, 704–715. https://doi.org/10.1002/mds.28365

Hardy, J., Lewis, P., Revesz, T., Lees, A., Paisan-Ruiz, C., 2009. The genetics of Parkinson's syndromes: a critical review. Curr. Opin. Genet. Dev. 19, 254–265. https://doi.org/10.1016/j.gde.2009.03.008

Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., Ferreira, J.J., Tolosa, E., Kay, D.M., Klein, C., Williams, D.R., Marras, C., Lang, A.E., Wszolek, Z.K., Berciano, J., Schapira, A.H.V., Lynch, T., Bhatia, K.P., Gasser, T., Lees, A.J., Wood, N.W., International LRRK2 Consortium, 2008. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol. 7, 583–590. https://doi.org/10.1016/S1474-4422(08)70117-0

Hiatt, S.M., Lawlor, J.M.J., Handley, L.H., Ramaker, R.C., Rogers, B.B., Partridge, E.C., Boston, L.B., Williams, M., Plott, C.B., Jenkins, J., Gray, D.E., Holt, J.M., Bowling, K.M., Bebin, E.M., Grimwood, J., Schmutz, J., Cooper, G.M., 2021. Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. HGG Adv. 2, 100023. https://doi.org/10.1016/j.xhgg.2021.100023

Hipp, G., Vaillant, M., Diederich, N.J., Roomp, K., Satagopam, V.P., Banda, P., Sandt, E., Mommaerts, K., Schmitz, S.K., Longhino, L., Schweicher, A., Hanff, A.-M., Nicolai, B., Kolber, P., Reiter, D., Pavelka, L., Binck, S., Pauly, C., Geffers, L., Betsou, F., Gantenbein, M., Klucken, J., Gasser, T., Hu, M.T., Balling, R., Krüger, R., 2018. The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. Front. Aging Neurosci. 10, 326. https://doi.org/10.3389/fnagi.2018.00326

Hirsch, E.C., Hunot, S., 2009. Neuroinflammation in Parkinson's disease: a target for neuroprotection? Lancet Neurol. 8, 382–397. https://doi.org/10.1016/S1474-4422(09)70062-6

Hirschhorn, J.N., Daly, M.J., 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95–108. https://doi.org/10.1038/nrg1521

Hong, C.T., Chan, L., Bai, C.-H., 2020. The Effect of Caffeine on the Risk and Progression of Parkinson's Disease: A Meta-Analysis. Nutrients 12, 1860. https://doi.org/10.3390/nu12061860

Horowitz, M., Elstein, D., Zimran, A., Goker-Alpan, O., 2016. New Directions in Gaucher Disease. Hum. Mutat. 37, 1121–1136. https://doi.org/10.1002/humu.23056

Horowitz, M., Wilder, S., Horowitz, Z., Reiner, O., Gelbart, T., Beutler, E., 1989. The human glucocerebrosidase gene and pseudogene: Structure and evolution. Genomics 4, 87–96. https://doi.org/10.1016/0888-7543(89)90319-4

Houlden, H., Singleton, A.B., 2012. The genetics and neuropathology of Parkinson's disease. Acta Neuropathol. (Berl.) 124, 325–338. https://doi.org/10.1007/s00401-012-1013-5

Hruska, K.S., LaMarca, M.E., Scott, C.R., Sidransky, E., 2008. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). Hum. Mutat. 29, 567–583. https://doi.org/10.1002/humu.20676

Huh, Y.E., Usnich, T., Scherzer, C.R., Klein, C., Chung, S.J., 2023. GBA Variants and Parkinson's Disease: Paving the Way for Targeted Therapy. J. Mov. Disord. https://doi.org/10.14802/jmd.23023

Ibanez, L., Dube, U., Saef, B., Budde, J., Black, K., Medvedeva, A., Del-Aguila, J.L., Davis, A.A., Perlmutter, J.S., Harari, O., Benitez, B.A., Cruchaga, C., 2017. Parkinson disease polygenic risk score is associated with Parkinson disease status and age at onset but not with alpha-synuclein cerebrospinal fluid levels. BMC Neurol. 17, 198. https://doi.org/10.1186/s12883-017-0978-z

Imamura, K., Hishikawa, N., Sawada, M., Nagatsu, T., Yoshida, M., Hashizume, Y., 2003. Distribution of major histocompatibility complex class II-positive microglia and cytokine profile of Parkinson's disease brains. Acta Neuropathol. (Berl.) 106, 518–526. https://doi.org/10.1007/s00401-003-0766-2

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J., Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C., Hsieh, C.-L., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittemore, A.S., Sieh, W., 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. 99, 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Isik, S., Yeman Kiyak, B., Akbayir, R., Seyhali, R., Arpaci, T., 2023. Microglia Mediated Neuroinflammation in Parkinson's Disease. Cells 12, 1012. https://doi.org/10.3390/cells12071012

Iwaki, H., Blauwendraat, C., Leonard, H.L., Liu, G., Maple-Grødem, J., Corvol, J.-C., Pihlstrøm, L., van Nimwegen, M., Hutten, S.J., Nguyen, K.-D.H., Rick, J., Eberly, S., Faghri, F., Auinger, P., Scott, K.M., Wijeyekoon, R., Van Deerlin, V.M., Hernandez, D.G., Day-Williams, A.G., Brice, A., Alves, G., Noyce, A.J., Tysnes, O.-B., Evans, J.R., Breen, D.P., Estrada, K., Wegel, C.E., Danjou, F., Simon, D.K., Ravina, B., Toft, M., Heutink, P., Bloem, B.R., Weintraub, D., Barker, R.A., Williams-Gray, C.H., van de Warrenburg, B.P., Van Hilten, J.J., Scherzer, C.R., Singleton, A.B., Nalls, M.A., 2019. Genetic risk of Parkinson disease and progression:: An analysis of 13 longitudinal cohorts. Neurol. Genet. 5, e348. https://doi.org/10.1212/NXG.0000000000000348

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., Chow, E.D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S.J., Farh, K.K.-H., 2019. Predicting Splicing from Primary Sequence with Deep Learning. Cell 176, 535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H.E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M., 2018. Nanopore sequencing and

assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36, 338–345. https://doi.org/10.1038/nbt.4060

Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. Erratum to: The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 17, 256. https://doi.org/10.1186/s13059-016-1122-x

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.-C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., Bras, J.M., Schymick, J.C., Hernandez, D.G., Traynor, B.J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H.M., Hardy, J.A., Rosenberg, N.A., Singleton, A.B., 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451, 998–1003. https://doi.org/10.1038/nature06742

Jankovic, J., Aguilar, L.G., 2008. Current approaches to the treatment of Parkinson's disease. Neuropsychiatr. Dis. Treat. 4, 743–757. https://doi.org/10.2147/ndt.s2006

Jesús, S., Huertas, I., Bernal-Bernal, I., Bonilla-Toribio, M., Cáceres-Redondo, M.T., Vargas-González, L., Gómez-Llamas, M., Carrillo, F., Calderón, E., Carballo, M., Gómez-Garre, P., Mir, P., 2016. GBA Variants Influence Motor and Non-Motor Features of Parkinson's Disease. PloS One 11, e0167749. https://doi.org/10.1371/journal.pone.0167749

Joselin, A.P., Hewitt, S.J., Callaghan, S.M., Kim, R.H., Chung, Y.-H., Mak, T.W., Shen, J., Slack, R.S., Park, D.S., 2012. ROS-dependent regulation of Parkin and DJ-1 localization during oxidative stress in neurons. Hum. Mol. Genet. 21, 4888–4903. https://doi.org/10.1093/hmg/dds325

Jović, M., Kean, M.J., Szentpetery, Z., Polevoy, G., Gingras, A.-C., Brill, J.A., Balla, T., 2012. Two phosphatidylinositol 4-kinases control lysosomal delivery of the Gaucher disease enzyme, β-glucocerebrosidase. Mol. Biol. Cell 23, 1533–1545. https://doi.org/10.1091/mbc.E11-06-0553

Kaakkola, S., 2000. Clinical pharmacology, therapeutic use and potential of COMT inhibitors in Parkinson's disease. Drugs 59, 1233–1250. https://doi.org/10.2165/00003495-200059060-00004

Kalia, L.V., Lang, A.E., 2015. Parkinson's disease. Lancet Lond. Engl. 386, 896–912. https://doi.org/10.1016/S0140-6736(14)61393-3

Kamp, F., Exner, N., Lutz, A.K., Wender, N., Hegermann, J., Brunner, B., Nuscher, B., Bartels, T., Giese, A., Beyer, K., Eimer, S., Winklhofer, K.F., Haass, C., 2010. Inhibition of mitochondrial fusion by α-synuclein is rescued by PINK1, Parkin and DJ-1. EMBO J. 29, 3571–3589. https://doi.org/10.1038/emboj.2010.223

Kasten, M., Hartmann, C., Hampf, J., Schaake, S., Westenberger, A., Vollstedt, E.-J., Balck, A., Domingo, A., Vulinovic, F., Dulovic, M., Zorn, I., Madoev, H., Zehnle, H., Lembeck, C.M., Schawe, L., Reginold, J., Huang, J., König, I.R., Bertram, L., Marras, C., Lohmann, K., Lill, C.M., Klein, C., 2018. Genotype-Phenotype Relations for the Parkinson's Disease Genes Parkin, PINK1, DJ1: MDSGene Systematic Review. Mov. Disord. Off. J. Mov. Disord. Soc. 33, 730–741. https://doi.org/10.1002/mds.27352

Kenborg, L., Lassen, C.F., Ritz, B., Andersen, K.K., Christensen, J., Schernhammer, E.S., Hansen, J., Wermuth, L., Rod, N.H., Olsen, J.H., 2015. Lifestyle, family history, and risk of idiopathic Parkinson disease: a large Danish case-control study. Am. J. Epidemiol. 181, 808–816. https://doi.org/10.1093/aje/kwu332

Khoo, T.K., Yarnall, A.J., Duncan, G.W., Coleman, S., O'Brien, J.T., Brooks, D.J., Barker, R.A., Burn, D.J., 2013. The spectrum of nonmotor symptoms in early Parkinson disease. Neurology 80, 276–281. https://doi.org/10.1212/WNL.0b013e31827deb74

Kilarski, L.L., Pearson, J.P., Newsway, V., Majounie, E., Knipe, M.D.W., Misbahuddin, A., Chinnery, P.F., Burn, D.J., Clarke, C.E., Marion, M.-H., Lewthwaite, A.J., Nicholl, D.J., Wood, N.W., Morrison, K.E., Williams-Gray, C.H., Evans, J.R., Sawcer, S.J., Barker, R.A., Wickremaratchi, M.M., Ben-Shlomo, Y., Williams, N.M., Morris, H.R., 2012. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 27, 1522–1529. https://doi.org/10.1002/mds.25132

Kim, Jong-hyeon, Choi, D., Jeong, H., Kim, Jun, Kim, D.W., Choi, S.Y., Park, S.-M., Suh, Y.H., Jou, I., Joe, E.-H., 2013. DJ-1 facilitates the interaction between STAT1 and its phosphatase, SHP-1, in brain microglia and astrocytes: A novel anti-inflammatory function of DJ-1. Neurobiol. Dis. 60, 1–10. https://doi.org/10.1016/j.nbd.2013.08.007

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., Shimizu, N., 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392, 605–608. https://doi.org/10.1038/33416

Klein, C., Westenberger, A., 2012. Genetics of Parkinson's disease. Cold Spring Harb. Perspect. Med. 2, a008888. https://doi.org/10.1101/cshperspect.a008888

Krüger, R., Kuhn, W., Müller, T., Woitalla, D., Graeber, M., Kösel, S., Przuntek, H., Epplen, J.T., Schöls, L., Riess, O., 1998. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat. Genet. 18, 106–108. https://doi.org/10.1038/ng0298-106

Kumar, K.R., Djarmati-Westenberger, A., Grünewald, A., 2011. Genetics of Parkinson's disease. Semin. Neurol. 31, 433–440. https://doi.org/10.1055/s-0031-1299782

Kuo, M.-C., Liu, S.C.-H., Hsu, Y.-F., Wu, R.-M., 2021. The role of noncoding RNAs in Parkinson's disease: biomarkers and associations with pathogenic pathways. J. Biomed. Sci. 28, 78. https://doi.org/10.1186/s12929-021-00775-x

La Cognata, V., Morello, G., D'Agata, V., Cavallaro, S., 2017. Copy number variability in Parkinson's disease: assembling the puzzle through a systems biology approach. Hum. Genet. 136, 13–37. https://doi.org/10.1007/s00439-016-1749-4

LaFramboise, T., 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. 37, 4181–4193. https://doi.org/10.1093/nar/gkp552

Lai, B.C.L., Marion, S.A., Teschke, K., Tsui, J.K.C., 2002. Occupational and environmental risk factors for Parkinson's disease. Parkinsonism Relat. Disord. 8, 297–309. https://doi.org/10.1016/s1353-8020(01)00054-2

Lamb, R., Rohrer, J.D., Lees, A.J., Morris, H.R., 2016. Progressive Supranuclear Palsy and Corticobasal Degeneration: Pathophysiology and Treatment Options. Curr. Treat. Options Neurol. 18, 42. https://doi.org/10.1007/s11940-016-0422-5

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., Maglott, D.R., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980-985. https://doi.org/10.1093/nar/gkt1113

Langston, J.W., Ballard, P., Tetrud, J.W., Irwin, I., 1983. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. Science 219, 979–980. https://doi.org/10.1126/science.6823561

Latham, T., Grabowski, G.A., Theophilus, B.D., Smith, F.I., 1990. Complex alleles of the acid beta-glucosidase gene in Gaucher disease. Am. J. Hum. Genet. 47, 79–86.

Lesage, S., Anheim, M., Condroyer, C., Pollak, P., Durif, F., Dupuits, C., Viallet, F., Lohmann, E., Corvol, J.-C., Honoré, A., Rivaud, S., Vidailhet, M., Dürr, A., Brice, A., French Parkinson's Disease Genetics Study Group, 2011. Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. Hum. Mol. Genet. 20, 202–210. https://doi.org/10.1093/hmg/ddq454

Lesage, S., Brice, A., 2009. Parkinson's disease: from monogenic forms to genetic susceptibility factors. Hum. Mol. Genet. 18, R48-59. https://doi.org/10.1093/hmg/ddp012

Lesage, S., Dürr, A., Tazir, M., Lohmann, E., Leutenegger, A.-L., Janin, S., Pollak, P., Brice, A., French Parkinson's Disease Genetics Study Group, 2006. LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. N. Engl. J. Med. 354, 422–423. https://doi.org/10.1056/NEJMc055540

Li, H., Ham, A., Ma, T.C., Kuo, S.-H., Kanter, E., Kim, D., Ko, H.S., Quan, Y., Sardi, S.P., Li, A., Arancio, O., Kang, U.J., Sulzer, D., Tang, G., 2019. Mitochondrial dysfunction and mitophagy defect triggered by heterozygous GBA mutations. Autophagy 15, 113–130. https://doi.org/10.1080/15548627.2018.1509818

Li, J.-Q., Tan, L., Yu, J.-T., 2014. The role of the LRRK2 gene in Parkinsonism. Mol. Neurodegener. 9, 47. https://doi.org/10.1186/1750-1326-9-47

Liao, J., Wu, C.-X., Burlak, C., Zhang, S., Sahm, H., Wang, M., Zhang, Z.-Y., Vogel, K.W., Federici, M., Riddle, S.M., Nichols, R.J., Liu, D., Cookson, M.R., Stone, T.A., Hoang, Q.Q., 2014. Parkinson disease-associated mutation R1441H in LRRK2 prolongs the "active state" of its GTPase domain. Proc. Natl. Acad. Sci. U. S. A. 111, 4055–4060. https://doi.org/10.1073/pnas.1323285111

Liu, X., Cheng, R., Ye, X., Verbitsky, M., Kisselev, S., Mejia-Santana, H., Louis, E., Cote, L., Andrews, H., Waters, C., Ford, B., Fahn, S., Marder, K., Lee, J., Clark, L., 2013. Increased Rate of Sporadic and Recurrent Rare Genic Copy Number Variants in Parkinson's Disease Among Ashkenazi Jews. Mol. Genet. Genomic Med. 1, 142–154. https://doi.org/10.1002/mgg3.18

Lock, E.A., Zhang, J., Checkoway, H., 2013. Solvents and Parkinson disease: a systematic review of toxicological and epidemiological evidence. Toxicol. Appl. Pharmacol. 266, 345–355. https://doi.org/10.1016/j.taap.2012.11.016

Logsdon, G.A., Vollger, M.R., Eichler, E.E., 2020. Long-read human genome sequencing and its applications. Nat. Rev. Genet. 21, 597–614. https://doi.org/10.1038/s41576-020-0236-x

Lohmann, E., Welter, M.-L., Fraix, V., Krack, P., Lesage, S., Laine, S., Tanguy, M.-L., Houeto, J.-L., Mesnage, V., Pollak, P., Durr, A., Agid, Y., Brice, A., French Parkinson's Disease Genetics study group, 2008. Are parkin patients particularly suited for deep-brain stimulation? Mov. Disord. Off. J. Mov. Disord. Soc. 23, 740–743. https://doi.org/10.1002/mds.21903

Lubbe, S.J., Bustos, B.I., Hu, J., Krainc, D., Joseph, T., Hehir, J., Tan, M., Zhang, W., Escott-Price, V., Williams, N.M., Blauwendraat, C., Singleton, A.B., Morris, H.R., 2021. Assessing the relationship between monoallelic PRKN mutations and Parkinson's risk. Hum. Mol. Genet. 30, 78–86. https://doi.org/10.1093/hmg/ddaa273

Lücking, C.B., Dürr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B.S., Meco, G., Denèfle, P., Wood, N.W., Agid, Y., Brice, A., French Parkinson's Disease Genetics Study Group, European Consortium on Genetic Susceptibility in Parkinson's Disease, 2000. Association between early-onset Parkinson's disease and mutations in the parkin gene. N. Engl. J. Med. 342, 1560–1567. https://doi.org/10.1056/NEJM200005253422103

Lunde, K.A., Chung, J., Dalen, I., Pedersen, K.F., Linder, J., Domellöf, M.E., Elgh, E., Macleod, A.D., Tzoulis, C., Larsen, J.P., Tysnes, O.-B., Forsgren, L., Counsell, C.E., Alves, G., Maple-Grødem, J., 2018. Association of glucocerebrosidase polymorphisms and mutations with dementia in incident Parkinson's disease. Alzheimers Dement. J. Alzheimers Assoc. 14, 1293–1301. https://doi.org/10.1016/j.jalz.2018.04.006

MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42, D986-992. https://doi.org/10.1093/nar/gkt958

Magistrelli, L., Contaldi, E., Comi, C., 2021. The Impact of SNCA Variations and Its Product Alpha-Synuclein on Non-Motor Features of Parkinson's Disease. Life 11, 804. https://doi.org/10.3390/life11080804

Mahlknecht, P., Seppi, K., Poewe, W., 2015. The Concept of Prodromal Parkinson's Disease. J. Park. Dis. 5, 681–697. https://doi.org/10.3233/JPD-150685

Makarious, M.B., Diez-Fairen, M., Krohn, L., Blauwendraat, C., Bandres-Ciga, S., Ding, J., Pihlstrøm, L., Houlden, H., Scholz, S.W., Gan-Or, Z., 2019. ARSA variants in α-synucleinopathies. Brain 142, e70. https://doi.org/10.1093/brain/awz340

Malekkou, A., Sevastou, I., Mavrikiou, G., Georgiou, T., Vilageliu, L., Moraitou, M., Michelakakis, H., Prokopiou, C., Drousiotou, A., 2020. A novel mutation deep within intron 7 of the GBA gene causes Gaucher disease. Mol. Genet. Genomic Med. 8, e1090. https://doi.org/10.1002/mgg3.1090

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. https://doi.org/10.1038/nature08494

Maraganore, D.M., de Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V.K., Frazer, K.A., Cox, D.R., Ballinger, D.G., 2005. High-resolution whole-genome association study of Parkinson disease. Am. J. Hum. Genet. 77, 685–693. https://doi.org/10.1086/496902

Marian, A.J., 2020. Clinical Interpretation and Management of Genetic Variants. JACC Basic Transl. Sci. 5, 1029–1042. https://doi.org/10.1016/j.jacbts.2020.05.013

Marras, C., Lang, A., van de Warrenburg, B.P., Sue, C.M., Tabrizi, S.J., Bertram, L., Mercimek-Mahmutoglu, S., Ebrahimi-Fakhari, D., Warner, T.T., Durr, A., Assmann, B., Lohmann, K., Kostic, V., Klein, C., 2016. Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force. Mov. Disord. Off. J. Mov. Disord. Soc. 31, 436–457. https://doi.org/10.1002/mds.26527

Martin, I., Kim, J.W., Dawson, V.L., Dawson, T.M., 2014. LRRK2 pathobiology in Parkinson's disease. J. Neurochem. 131, 554–565. https://doi.org/10.1111/jnc.12949

Mata, I.F., Samii, A., Schneer, S.H., Roberts, J.W., Griffith, A., Leis, B.C., Schellenberg, G.D., Sidransky, E., Bird, T.D., Leverenz, J.B., Tsuang, D., Zabetian, C.P., 2008. Glucocerebrosidase Gene Mutations. Arch. Neurol. 65, 379–382. https://doi.org/10.1001/archneurol.2007.68

Matsuda, N., Sato, S., Shiba, K., Okatsu, K., Saisho, K., Gautier, C.A., Sou, Y.-S., Saiki, S., Kawajiri, S., Sato, F., Kimura, M., Komatsu, M., Hattori, N., Tanaka, K., 2010. PINK1 stabilized by mitochondrial depolarization recruits Parkin to damaged mitochondria and activates latent Parkin for mitophagy. J. Cell Biol. 189, 211–221. https://doi.org/10.1083/jcb.200910140

Mazzulli, J.R., Xu, Y.-H., Sun, Y., Knight, A.L., McLean, P.J., Caldwell, G.A., Sidransky, E., Grabowski, G.A., Krainc, D., 2011. Gaucher disease glucocerebrosidase and α-synuclein form a bidirectional pathogenic loop in synucleinopathies. Cell 146, 37–52. https://doi.org/10.1016/j.cell.2011.06.001

Mazzulli, J.R., Zunke, F., Tsunemi, T., Toker, N.J., Jeon, S., Burbulla, L.F., Patnaik, S., Sidransky, E., Marugan, J.J., Sue, C.M., Krainc, D., 2016. Activation of β-Glucocerebrosidase Reduces Pathological α-Synuclein and Restores Lysosomal Function in Parkinson's Patient Midbrain Neurons. J. Neurosci. 36, 7693–7706. https://doi.org/10.1523/JNEUROSCI.0628-16.2016

McColgan, P., Tabrizi, S.J., 2018. Huntington's disease: a clinical review. Eur. J. Neurol. 25, 24–34. https://doi.org/10.1111/ene.13413

McGeer, P.L., Itagaki, S., Boyes, B.E., McGeer, E.G., 1988. Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains. Neurology 38, 1285–1291. https://doi.org/10.1212/wnl.38.8.1285

McNeill, A., Magalhaes, J., Shen, C., Chau, K.-Y., Hughes, D., Mehta, A., Foltynie, T., Cooper, J.M., Abramov, A.Y., Gegg, M., Schapira, A.H.V., 2014. Ambroxol improves lysosomal biochemistry in glucocerebrosidase mutation-linked Parkinson disease cells. Brain J. Neurol. 137, 1481–1495. https://doi.org/10.1093/brain/awu020

Meixner, A., Boldt, K., Van Troys, M., Askenazi, M., Gloeckner, C.J., Bauer, M., Marto, J.A., Ampe, C., Kinkl, N., Ueffing, M., 2011. A QUICK screen for Lrrk2 interaction partners--leucine-rich repeat kinase 2 is involved in actin cytoskeleton dynamics. Mol. Cell. Proteomics MCP 10, M110.001172. https://doi.org/10.1074/mcp.M110.001172

Merker, J.D., Wenger, A.M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., Waggott, D., Utiramerur, S., Hou, Y., Smith, K.S., Montgomery, S.B., Wheeler, M., Buchan, J.G., Lambert, C.C., Eng, K.S., Hickey, L., Korlach, J., Ford, J., Ashley, E.A., 2018. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet. Med. Off. J. Am. Coll. Med. Genet. 20, 159–163. https://doi.org/10.1038/gim.2017.86

Michel, P.P., Hirsch, E.C., Hunot, S., 2016. Understanding Dopaminergic Cell Death Pathways in Parkinson Disease. Neuron 90, 675–691. https://doi.org/10.1016/j.neuron.2016.03.038

Migdalska-Richards, A., Schapira, A.H.V., 2016. The relationship between glucocerebrosidase mutations and Parkinson disease. J. Neurochem. 139 Suppl 1, 77–90. https://doi.org/10.1111/jnc.13385

Miller, D.E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K.M., Lewis, A.P., Fuerte, E.P.A., Paschal, C.R., Walsh, T., Thies, J., Bennett, J.T., Glass, I., Dipple, K.M., Patterson, K., Bonkowski, E.S., Nelson, Z., Squire, A., Sikes, M., Beckman, E., Bennett, R.L., Earl,

D., Lee, W., Allikmets, R., Perlman, S.J., Chow, P., Hing, A.V., Wenger, T.L., Adam, M.P., Sun, A., Lam, C., Chang, I., Zou, X., Austin, S.L., Huggins, E., Safi, A., Iyengar, A.K., Reddy, T.E., Majoros, W.H., Allen, A.S., Crawford, G.E., Kishnani, P.S., King, M.-C., Cherry, T., Chong, J.X., Bamshad, M.J., Nickerson, D.A., Mefford, H.C., Doherty, D., Eichler, E.E., 2021. Targeted long-read sequencing identifies missing disease-causing variation. Am. J. Hum. Genet. 108, 1436–1449. https://doi.org/10.1016/j.ajhg.2021.06.006

Mitsui, J., Mizuta, I., Toyoda, A., Ashida, R., Takahashi, Y., Goto, J., Fukuda, Y., Date, H., Iwata, A., Yamamoto, M., Hattori, N., Murata, M., Toda, T., Tsuji, S., 2009. Mutations for Gaucher disease confer high susceptibility to Parkinson disease. Arch. Neurol. 66, 571–576. https://doi.org/10.1001/archneurol.2009.72

Montfort, M., Chabás, A., Vilageliu, L., Grinberg, D., 2004. Functional analysis of 13 GBA mutant alleles identified in Gaucher disease patients: Pathogenic changes and "modifier" polymorphisms. Hum. Mutat. 23, 567–575. https://doi.org/10.1002/humu.20043

Mukadam, A.S., Seaman, M.N.J., 2015. Retromer-mediated endosomal protein sorting: The role of unstructured domains. FEBS Lett. 589, 2620–2626. https://doi.org/10.1016/j.febslet.2015.05.052

Mullaney, J.M., Mills, R.E., Pittard, W.S., Devine, S.E., 2010. Small insertions and deletions (INDELs) in human genomes. Hum. Mol. Genet. 19, R131-136. https://doi.org/10.1093/hmg/ddq400

Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., Bras, J., Young, E., Coelln, R. von, Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrom, L., Siitonen, A., Iwaki, H., Leonard, H., Faghri, F., Gibbs, J.R., Hernandez, D.G., Scholz, S.W., Botia, J.A., Martinez, M., Corvol, J.-C., Lesage, S., Jankovic, J., Shulman, L.M., Team, T. 23andMe R., Consortium, S.G. of P.D. (SGPD), Sutherland, M., Tienari, P., Majamaa, K., Toft, M., Andreassen, O.A., Bangale, T., Brice, A., Yang, J., Gan-Or, Z., Gasser, T., Heutink, P., Shulman, J.M., Wood, N., Hinds, D.A., Hardy, J.A., Morris, H.R., Gratten, J., Visscher, P.M., Graham, R.R., Singleton, A.B., Consortium, for the I.P.D.G., 2019a. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. https://doi.org/10.1101/388165

Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., Bras, J., Young, E., von Coelln, R., Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrøm, L., Siitonen, A., Iwaki, H., Leonard, H., Faghri, F., Gibbs, J.R., Hernandez, D.G., Scholz, S.W., Botia, J.A., Martinez, M., Corvol, J.-C., Lesage, S., Jankovic, J., Shulman, L.M., Sutherland, M., Tienari, P., Majamaa, K., Toft, M., Andreassen, O.A., Bangale, T., Brice, A., Yang, J., Gan-Or, Z., Gasser, T., Heutink, P., Shulman, J.M., Wood, N.W., Hinds, D.A., Hardy, J.A., Morris, H.R., Gratten, J., Visscher, P.M., Graham, R.R., Singleton, A.B., 23andMe Research Team, System Genomics of Parkinson's Disease Consortium, International Parkinson's Disease Genomics Consortium, 2019b. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 18, 1091–1102. https://doi.org/10.1016/S1474-4422(19)30320-5

Nalls, M.A., Bras, J., Hernandez, D.G., Keller, M.F., Majounie, E., Renton, A.E., Saad, M., Jansen, I., Guerreiro, R., Lubbe, S., Plagnol, V., Gibbs, J.R., Schulte, C., Pankratz, N., Sutherland, M., Bertram, L., Lill, C.M., DeStefano, A.L., Faroud, T., Eriksson, N., Tung, J.Y., Edsall, C., Nichols, N., Brooks, J., Arepalli, S., Pliner, H., Letson, C., Heutink, P., Martinez, M., Gasser, T., Traynor, B.J., Wood, N., Hardy, J., Singleton, A.B., International Parkinson's Disease Genomics Consortium (IPDGC), Parkinson's Disease meta-analysis consortium, 2015. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol. Aging 36, 1605.e7–12. https://doi.org/10.1016/j.neurobiolaging.2014.07.028

Nalls, M.A., Duran, R., Lopez, G., Kurzawa-Akanbi, M., McKeith, I.G., Chinnery, P.F., Morris, C.M., Theuns, J., Crosiers, D., Cras, P., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Mann, D.M.A., Snowden, J., Pickering-Brown, S., Halliwell, N., Davidson, Y., Gibbons, L., Harris, J., Sheerin, U.-M., Bras, J., Hardy, J., Clark, L., Marder, K., Honig, L.S., Berg,

D., Maetzler, W., Brockmann, K., Gasser, T., Novellino, F., Quattrone, A., Annesi, G., De Marco, E.V., Rogaeva, E., Masellis, M., Black, S.E., Bilbao, J.M., Foroud, T., Ghetti, B., Nichols, W.C., Pankratz, N., Halliday, G., Lesage, S., Klebe, S., Durr, A., Duyckaerts, C., Brice, A., Giasson, B.I., Trojanowski, J.Q., Hurtig, H.I., Tayebi, N., Landazabal, C., Knight, M.A., Keller, M., Singleton, A.B., Wolfsberg, T.G., Sidransky, E., 2013. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. JAMA Neurol. 70, 727–735. https://doi.org/10.1001/jamaneurol.2013.1925

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., International Parkinson's Disease Genomics Consortium (IPDGC), Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI), 23andMe, GenePD, NeuroGenetics Research Consortium (NGRC), Hussman Institute of Human Genomics (HIHG), Ashkenazi Jewish Dataset Investigator, Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE), North American Brain Expression Consortium (NABEC), United Kingdom Brain Expression Consortium (UKBEC), Greek Parkinson's Disease Consortium, Alzheimer Genetic Analysis Group, Ikram, M.A., Ioannidis, J.P.A., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., Singleton, A.B., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. 46, 989–993. https://doi.org/10.1038/ng.3043

Neudorfer, O., Giladi, N., Elstein, D., Abrahamov, A., Turezkite, T., Aghai, E., Reches, A., Bembi, B., Zimran, A., 1996. Occurrence of Parkinson's syndrome in type I Gaucher disease. QJM Mon. J. Assoc. Physicians 89, 691–694. https://doi.org/10.1093/qjmed/89.9.691

Neumann, J., Bras, J., Deas, E., O'Sullivan, S.S., Parkkinen, L., Lachmann, R.H., Li, A., Holton, J., Guerreiro, R., Paudel, R., Segarane, B., Singleton, A., Lees, A., Hardy, J., Houlden, H., Revesz, T., Wood, N.W., 2009. Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. Brain J. Neurol. 132, 1783–1794. https://doi.org/10.1093/brain/awp044

Nguyen, A.P.T., Moore, D.J., 2017. Understanding the GTPase Activity of LRRK2: Regulation, Function, and Neurotoxicity. Adv. Neurobiol. 14, 71–88. https://doi.org/10.1007/978-3-319-49969-7_4

Nica, A.C., Dermitzakis, E.T., 2013. Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368, 20120362. https://doi.org/10.1098/rstb.2012.0362

Nichols, W.C., Pankratz, N., Marek, D.K., Pauciulo, M.W., Elsaesser, V.E., Halter, C.A., Rudolph, A., Wojcieszek, J., Pfeiffer, R.F., Foroud, T., Parkinson Study Group-PROGENI Investigators, 2009. Mutations in GBA are associated with familial Parkinson disease susceptibility and age at onset. Neurology 72, 310–316. https://doi.org/10.1212/01.wnl.0000327823.81237.d1

Oeda, T., Umemura, A., Mori, Y., Tomita, S., Kohsaka, M., Park, K., Inoue, K., Fujimura, H., Hasegawa, H., Sugiyama, H., Sawada, H., 2015. Impact of glucocerebrosidase mutations on motor and nonmotor complications in Parkinson's disease. Neurobiol. Aging 36, 3306–3313. https://doi.org/10.1016/j.neurobiolaging.2015.08.027

Ohnmacht, J., May, P., Sinkkonen, L., Krüger, R., 2020. Missing heritability in Parkinson's disease: the emerging role of non-coding genetic variation. J. Neural Transm. 127, 729–748. https://doi.org/10.1007/s00702-020-02184-0

O'Regan, G., deSouza, R.-M., Balestrino, R., Schapira, A.H., 2017. Glucocerebrosidase Mutations in Parkinson Disease. J. Park. Dis. 7, 411–422. https://doi.org/10.3233/JPD-171092

Osellame, L.D., Rahim, A.A., Hargreaves, I.P., Gegg, M.E., Richard-Londt, A., Brandner, S., Waddington, S.N., Schapira, A.H.V., Duchen, M.R., 2013. Mitochondria and quality control defects in a mouse model of Gaucher disease--links to Parkinson's disease. Cell Metab. 17, 941–953. https://doi.org/10.1016/j.cmet.2013.04.014

Owolabi, L., 2013. Progressive supranuclear palsy misdiagnosed as Parkinson's disease: a case report and review of literature. Ann. Med. Health Sci. Res. 3, S44-47. https://doi.org/10.4103/2141-9248.121221

Ozelius, L.J., Senthil, G., Saunders-Pullman, R., Ohmann, E., Deligtisch, A., Tagliati, M., Hunt, A.L., Klein, C., Henick, B., Hailpern, S.M., Lipton, R.B., Soto-Valencia, J., Risch, N., Bressman, S.B., 2006. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. N. Engl. J. Med. 354, 424–425. https://doi.org/10.1056/NEJMc055509

Pachchek, S., 2023. Accurate long-read sequencing identified GBA variants as a major risk factor in the Luxembourgish cohort.

Pal, G., Mangone, G., Hill, E.J., Ouyang, B., Liu, Y., Lythe, V., Ehrlich, D., Saunders-Pullman, R., Shanker, V., Bressman, S., Alcalay, R.N., Garcia, P., Marder, K.S., Aasly, J., Mouradian, M.M., Link, S., Rosenbaum, M., Anderson, S., Bernard, B., Wilson, R., Stebbins, G., Nichols, W.C., Welter, M.-L., Sani, S., Afshari, M., Verhagen, L., de Bie, R.M.A., Foltynie, T., Hall, D., Corvol, J.-C., Goetz, C.G., 2022. Parkinson Disease and Subthalamic Nucleus Deep Brain Stimulation: Cognitive Effects in GBA Mutation Carriers. Ann. Neurol. 91, 424–435. https://doi.org/10.1002/ana.26302

Pan, H., Liu, Z., Ma, J., Li, Y., Zhao, Y., Zhou, Xiaoxia, Xiang, Y., Wang, Y., Zhou, Xun, He, R., Xie, Y., Zhou, Q., Yuan, K., Xu, Q., Sun, Q., Wang, J., Yan, X., Zhang, H., Wang, Chunyu, Lei, L., Liu, W., Wang, X., Ding, X., Wang, T., Xue, Z., Zhang, Zhentao, Chen, L., Wang, Q., Liu, Y., Tang, J., Zhang, X., Peng, S., Wang, Chaodong, Ding, J., Liu, C., Wang, L., Chen, H., Shen, L., Jiang, H., Wu, X., Tan, H., Luo, D., Xiao, S., Chen, X., Tan, J., Hu, Z., Chen, C., Xia, K., Zhang, Zhuohua, Foo, J.N., Blauwendraat, C., Nalls, M.A., Singleton, A.B., Liu, J., Chan, P., Zheng, H., Li, J., Guo, J., Yang, J., Tang, B., Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China (PD-MDCNC), 2023. Genome-wide association study using whole-genome sequencing identifies risk loci for Parkinson's disease in Chinese population. NPJ Park. Dis. 9, 22. https://doi.org/10.1038/s41531-023-00456-6

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40, 1413–1415. https://doi.org/10.1038/ng.259

Pang, S.Y.-Y., Lo, R.C.N., Ho, P.W.-L., Liu, H.-F., Chang, E.E.S., Leung, C.-T., Malki, Y., Choi, Z.Y.-K., Wong, W.Y., Kung, M.H.-W., Ramsden, D.B., Ho, S.-L., 2022. LRRK2, GBA and their interaction in the regulation of autophagy: implications on therapeutics in Parkinson's disease. Transl. Neurodegener. 11, 5. https://doi.org/10.1186/s40035-022-00281-6

Pankratz, N., Dumitriu, A., Hetrick, K.N., Sun, M., Latourelle, J.C., Wilk, J.B., Halter, C., Doheny, K.F., Gusella, J.F., Nichols, W.C., Myers, R.H., Foroud, T., DeStefano, A.L., PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories, 2011. Copy number variation in familial Parkinson disease. PloS One 6, e20988. https://doi.org/10.1371/journal.pone.0020988

Pankratz, N., Wilk, J.B., Latourelle, J.C., DeStefano, A.L., Halter, C., Pugh, E.W., Doheny, K.F., Gusella, J.F., Nichols, W.C., Foroud, T., Myers, R.H., PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories, 2009. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum. Genet. 124, 593–605. https://doi.org/10.1007/s00439-008-0582-9

Papapetropoulos, S., Adi, N., Ellul, J., Argyriou, A.A., Chroni, E., 2007. A prospective study of familial versus sporadic Parkinson's disease. Neurodegener. Dis. 4, 424–427. https://doi.org/10.1159/000107702

Parihar, M.S., Parihar, A., Fujita, M., Hashimoto, M., Ghafourifar, P., 2009. Alpha-synuclein overexpression and aggregation exacerbates impairment of mitochondrial functions by augmenting oxidative stress in human neuroblastoma cells. Int. J. Biochem. Cell Biol. 41, 2015–2024. https://doi.org/10.1016/j.biocel.2009.05.008

Parker, W.D., Boyson, S.J., Parks, J.K., 1989. Abnormalities of the electron transport chain in idiopathic Parkinson's disease. Ann. Neurol. 26, 719–723. https://doi.org/10.1002/ana.410260606

Parkinson, J., 2002. An essay on the shaking palsy. 1817. J. Neuropsychiatry Clin. Neurosci. 14, 223–236; discussion 222. https://doi.org/10.1176/jnp.14.2.223

Pastor, P., Ezquerra, M., Muñoz, E., Martí, M.J., Blesa, R., Tolosa, E., Oliva, R., 2000. Significant association between the tau gene A0/A0 genotype and Parkinson's disease. Ann. Neurol. 47, 242–245.

Pastores, G.M., Hughes, D.A., 1993. Gaucher Disease, in: Adam, M.P., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A. (Eds.), GeneReviews®. University of Washington, Seattle, Seattle (WA).

Pavelka, L., Rauschenberger, A., Landoulsi, Z., Pachchek, S., May, P., Glaab, E., Krüger, R., NCER-PD Consortium, 2022. Age at onset as stratifier in idiopathic Parkinson's disease - effect of ageing and polygenic risk score on clinical phenotypes. NPJ Park. Dis. 8, 102. https://doi.org/10.1038/s41531-022-00342-7

Petrucci, S., Ginevrino, M., Trezzi, I., Monfrini, E., Ricciardi, L., Albanese, A., Avenali, M., Barone, P., Bentivoglio, A.R., Bonifati, V., Bove, F., Bonanni, L., Brusa, L., Cereda, C., Cossu, G., Criscuolo, C., Dati, G., De Rosa, A., Eleopra, R., Fabbrini, G., Fadda, L., Garbellini, M., Minafra, B., Onofrj, M., Pacchetti, C., Palmieri, I., Pellecchia, M.T., Petracca, M., Picillo, M., Pisani, A., Vallelunga, A., Zangaglia, R., Di Fonzo, A., Morgante, F., Valente, E.M., ITA-GENE-PD Study Group, 2020. GBA-Related Parkinson's Disease: Dissection of Genotype-Phenotype Correlates in a Large Italian Cohort. Mov. Disord. Off. J. Mov. Disord. Soc. 35, 2106–2111. https://doi.org/10.1002/mds.28195

Pihlstrøm, L., Morset, K.R., Grimstad, E., Vitelli, V., Toft, M., 2016. A cumulative genetic risk score predicts progression in Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 31, 487–490. https://doi.org/10.1002/mds.26505

Poewe, W., 2008. Non-motor symptoms in Parkinson's disease. Eur. J. Neurol. 15 Suppl 1, 14–20. https://doi.org/10.1111/j.1468-1331.2008.02056.x

Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E., Halliday, G., Goetz, C.G., Gasser, T., Dubois, B., Chan, P., Bloem, B.R., Adler, C.H., Deuschl, G., 2015. MDS clinical diagnostic criteria for Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 30, 1591–1601. https://doi.org/10.1002/mds.26424

Poulopoulos, M., Levy, O.A., Alcalay, R.N., 2012. The neuropathology of genetic Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 27, 831–842. https://doi.org/10.1002/mds.24962

Pringsheim, T., Jette, N., Frolkis, A., Steeves, T.D.L., 2014. The prevalence of Parkinson's disease: a systematic review and meta-analysis. Mov. Disord. Off. J. Mov. Disord. Soc. 29, 1583–1590. https://doi.org/10.1002/mds.25945

Quadri, M., Mandemakers, W., Grochowska, M.M., Masius, R., Geut, H., Fabrizio, E., Breedveld, G.J., Kuipers, D., Minneboo, M., Vergouw, L.J.M., Carreras Mascaro, A., Yonova-Doing, E., Simons, E., Zhao, T., Di Fonzo, A.B., Chang, H.-C., Parchi, P., Melis, M., Correia Guedes, L., Criscuolo, C., Thomas, A., Brouwer, R.W.W., Heijsman, D., Ingrassia, A.M.T., Calandra Buonaura, G., Rood, J.P., Capellari, S., Rozemuller, A.J., Sarchioto, M., Fen Chien, H., Vanacore, N., Olgiati, S., Wu-Chou, Y.-H., Yeh, T.-H., Boon, A.J.W., Hoogers, S.E., Ghazvini, M., IJpma, A.S., van IJcken, W.F.J., Onofrj, M., Barone, P., Nicholl, D.J., Puschmann, A., De Mari, M., Kievit, A.J., Barbosa, E., De Michele, G., Majoor-Krakauer, D., van Swieten, J.C., de Jong, F.J., Ferreira, J.J., Cossu, G., Lu, C.-S., Meco, G., Cortelli, P., van de Berg, W.D.J., Bonifati, V., International Parkinsonism Genetics Network, 2018. LRP10 genetic variants in familial Parkinson's disease and dementia with Lewy bodies: a genome-wide linkage and sequencing study. Lancet Neurol. 17, 597–608. https://doi.org/10.1016/S1474-4422(18)30179-0

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rang, F.J., Kloosterman, W.P., de Ridder, J., 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 19, 90. https://doi.org/10.1186/s13059-018-1462-9

Raninga, P.V., Di Trapani, G., Tonissen, K.F., 2017. The Multifaceted Roles of DJ-1 as an Antioxidant. Adv. Exp. Med. Biol. 1037, 67–87. https://doi.org/10.1007/978-981-10-6583-5_6

Reczek, D., Schwake, M., Schröder, J., Hughes, H., Blanz, J., Jin, X., Brondyk, W., Van Patten, S., Edmunds, T., Saftig, P., 2007. LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of beta-glucocerebrosidase. Cell 131, 770–783. https://doi.org/10.1016/j.cell.2007.10.018

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, Junjun, Zerjal, T., Zhang, Jane, Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. Nature 444, 444–454. https://doi.org/10.1038/nature05329

Revel-Vilk, S., Szer, J., Mehta, A., Zimran, A., 2018. How we manage Gaucher Disease in the era of choices. Br. J. Haematol. 182, 467–480. https://doi.org/10.1111/bjh.15402

Riboldi, G.M., Di Fonzo, A.B., 2019. GBA, Gaucher Disease, and Parkinson's Disease: From Genetic to Clinic to New Therapeutic Approaches. Cells 8, 364. https://doi.org/10.3390/cells8040364

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. Off. J. Am. Coll. Med. Genet. 17, 405–424. https://doi.org/10.1038/gim.2015.30

Richter, F., Fleming, S.M., Watson, M., Lemesre, V., Pellegrino, L., Ranes, B., Zhu, C., Mortazavi, F., Mulligan, C.K., Sioshansi, P.C., Hean, S., De La Rosa, K., Khanna, R., Flanagan, J., Lockhart, D.J., Wustman, B.A., Clark, S.W., Chesselet, M.-F., 2014. A GCase chaperone improves motor function in a mouse model of synucleinopathy. Neurother. J. Am. Soc. Exp. Neurother. 11, 840–856. https://doi.org/10.1007/s13311-014-0294-x

Robak, L.A., Jansen, I.E., van Rooij, J., Uitterlinden, A.G., Kraaij, R., Jankovic, J., International Parkinson's Disease Genomics Consortium (IPDGC), Heutink, P., Shulman, J.M., 2017. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. Brain J. Neurol. 140, 3191–3203. https://doi.org/10.1093/brain/awx285

Roberts, R.J., Carneiro, M.O., Schatz, M.C., 2013. The advantages of SMRT sequencing. Genome Biol. 14, 405. https://doi.org/10.1186/gb-2013-14-6-405

Rosenbloom, B., Balwani, M., Bronstein, J.M., Kolodny, E., Sathe, S., Gwosdow, A.R., Taylor, J.S., Cole, J.A., Zimran, A., Weinreb, N.J., 2011. The incidence of Parkinsonism in patients with type 1 Gaucher disease: data from the ICGG Gaucher Registry. Blood Cells. Mol. Dis. 46, 95–102. https://doi.org/10.1016/j.bcmd.2010.10.006

Saiki, M., Baker, A., Williams-Gray, C.H., Foltynie, T., Goodman, R.S., Taylor, C.J., Compston, D.A.S., Barker, R.A., Sawcer, S.J., Goris, A., 2010. Association of the human leucocyte antigen region with susceptibility to Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 81, 890–891. https://doi.org/10.1136/jnnp.2008.162883

Sala, G., Marinig, D., Arosio, A., Ferrarese, C., 2016. Role of Chaperone-Mediated Autophagy Dysfunctions in the Pathogenesis of Parkinson's Disease. Front. Mol. Neurosci. 9, 157. https://doi.org/10.3389/fnmol.2016.00157

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sardi, S.P., Clarke, J., Viel, C., Chan, M., Tamsett, T.J., Treleaven, C.M., Bu, J., Sweet, L., Passini, M.A., Dodge, J.C., Yu, W.H., Sidman, R.L., Cheng, S.H., Shihabuddin, L.S., 2013. Augmenting CNS glucocerebrosidase activity as a therapeutic strategy for parkinsonism and other Gaucher-related synucleinopathies. Proc. Natl. Acad. Sci. 110, 3537–3542. https://doi.org/10.1073/pnas.1220464110

Sarihan, E.I., Pérez-Palma, E., Niestroj, L.-M., Loesch, D., Inca-Martinez, M., Horimoto, A.R.V.R., Cornejo-Olivas, M., Torres, L., Mazzetti, P., Cosentino, C., Sarapura-Castro, E., Rivera-Valdivia, A., Dieguez, E., Raggio, V., Lescano, A., Tumas, V., Borges, V., Ferraz, H.B., Rieder, C.R., Schumacher-Schuh, A.F., Santos-Lobato, B.L., Velez-Pardo, C., Jimenez-Del-Rio, M., Lopera, F., Moreno, S., Chana-Cuevas, P., Fernandez, W., Arboleda, G., Arboleda, H., Arboleda-Bustos, C.E., Yearout, D., Zabetian, C.P., Thornton, T.A., O'Connor, T.D., Lal, D., Mata, I.F., Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD)‡, 2021. Genome-Wide Analysis of Copy Number Variation in Latin American Parkinson's Disease Patients. Mov. Disord. Off. J. Mov. Disord. Soc. 36, 434–441. https://doi.org/10.1002/mds.28353

Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., Tsunoda, T., Watanabe, M., Takeda, A., Tomiyama, H., Nakashima, K., Hasegawa, K., Obata, F., Yoshikawa, T., Kawakami, H., Sakoda, S., Yamamoto, M., Hattori, N., Murata, M., Nakamura, Y., Toda, T., 2009. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat. Genet. 41, 1303–1307. https://doi.org/10.1038/ng.485

Schapira, A.H., Cooper, J.M., Dexter, D., Clark, J.B., Jenner, P., Marsden, C.D., 1990. Mitochondrial complex I deficiency in Parkinson's disease. J. Neurochem. 54, 823–827. https://doi.org/10.1111/j.1471-4159.1990.tb02325.x

Schmitz, M., Alfalah, M., Aerts, J.M.F.G., Naim, H.Y., Zimmer, K.-P., 2005. Impaired trafficking of mutants of lysosomal glucocerebrosidase in Gaucher's disease. Int. J. Biochem. Cell Biol. 37, 2310–2320. https://doi.org/10.1016/j.biocel.2005.05.008

Schöndorf, D.C., Aureli, M., McAllister, F.E., Hindley, C.J., Mayer, F., Schmid, B., Sardi, S.P., Valsecchi, M., Hoffmann, S., Schwarz, L.K., Hedrich, U., Berg, D., Shihabuddin, L.S., Hu, J., Pruszak, J., Gygi, S.P., Sonnino, S., Gasser, T., Deleidi, M., 2014. iPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. Nat. Commun. 5, 4028. https://doi.org/10.1038/ncomms5028

Schuepbach, W.M.M., Rau, J., Knudsen, K., Volkmann, J., Krack, P., Timmermann, L., Hälbig, T.D., Hesekamp, H., Navarro, S.M., Meier, N., Falk, D., Mehdorn, M., Paschen, S., Maarouf, M., Barbe, M.T., Fink, G.R., Kupsch, A., Gruber, D., Schneider, G.-H., Seigneuret, E., Kistner, A., Chaynes, P., Ory-Magne, F., Brefel Courbon, C., Vesper, J., Schnitzler, A., Wojtecki, L., Houeto, J.-L., Bataille, B., Maltête, D., Damier, P., Raoul, S., Sixel-Doering, F., Hellwig, D., Gharabaghi, A., Krüger, R., Pinsker, M.O., Amtage, F., Régis, J.-M., Witjas, T., Thobois, S., Mertens, P., Kloss, M., Hartmann, A., Oertel, W.H., Post, B., Speelman, H., Agid, Y., Schade-Brittinger, C., Deuschl, G., EARLYSTIM Study Group, 2013. Neurostimulation for Parkinson's disease with early motor complications. N. Engl. J. Med. 368, 610–622. https://doi.org/10.1056/NEJMoa1205158

Schüle, B., McFarland, K.N., Lee, K., Tsai, Y.-C., Nguyen, K.-D., Sun, C., Liu, M., Byrne, C., Gopi, R., Huang, N., Langston, J.W., Clark, T., Gil, F.J.J., Ashizawa, T., 2017. Parkinson's disease associated with pure ATXN10 repeat expansion. NPJ Park. Dis. 3, 27. https://doi.org/10.1038/s41531-017-0029-x

Setó-Salvia, N., Pagonabarraga, J., Houlden, H., Pascual-Sedano, B., Dols-Icardo, O., Tucci, A., Paisán-Ruiz, C., Campolongo, A., Antón-Aguirre, S., Martín, I., Muñoz, L., Bufill, E., Vilageliu, L., Grinberg, D., Cozar, M., Blesa, R., Lleó, A., Hardy, J., Kulisevsky, J., Clarimón, J., 2012. Glucocerebrosidase mutations confer a greater risk of dementia during Parkinson's disease course. Mov. Disord. Off. J. Mov. Disord. Soc. 27, 393–399. https://doi.org/10.1002/mds.24045

Shemesh, E., Deroma, L., Bembi, B., Deegan, P., Hollak, C., Weinreb, N.J., Cox, T.M., 2015. Enzyme replacement and substrate reduction therapy for Gaucher disease. Cochrane Database Syst. Rev. 2015, CD010324. https://doi.org/10.1002/14651858.CD010324.pub2

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311. https://doi.org/10.1093/nar/29.1.308

Shimura, H., Hattori, N., Kubo, S. i, Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., Suzuki, T., 2000. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat. Genet. 25, 302–305. https://doi.org/10.1038/77060

Siddiqui, I.J., Pervaiz, N., Abbasi, A.A., 2016. The Parkinson Disease gene SNCA: Evolutionary and structural insights with pathological implication. Sci. Rep. 6, 24475. https://doi.org/10.1038/srep24475

Sidransky, E., Lopez, G., 2012. The link between the GBA gene and parkinsonism. Lancet Neurol. 11, 986–998. https://doi.org/10.1016/S1474-4422(12)70190-4

Sidransky, E., Nalls, M.A., Aasly, J.O., Aharon-Peretz, J., Annesi, G., Barbosa, E.R., Bar-Shira, A., Berg, D., Bras, J., Brice, A., Chen, C.-M., Clark, L.N., Condroyer, C., De Marco, E.V., Dürr, A., Eblan, M.J., Fahn, S., Farrer, M.J., Fung, H.-C., Gan-Or, Z., Gasser, T., Gershoni-Baruch, R., Giladi, N., Griffith, A., Gurevich, T., Januario, C., Kropp, P., Lang, A.E., Lee-Chen, G.-J., Lesage, S., Marder, K., Mata, I.F., Mirelman, A., Mitsui, J., Mizuta, I., Nicoletti, G., Oliveira, C., Ottman, R., Orr-Urtreger, A., Pereira, L.V., Quattrone, A., Rogaeva, E., Rolfs, A., Rosenbaum, H., Rozenberg, R., Samii, A., Samaddar, T., Schulte, C., Sharma, M., Singleton, A., Spitz, M., Tan, E.-K., Tayebi, N., Toda, T., Troiano, A.R., Tsuji, S., Wittstock, M., Wolfsberg, T.G., Wu, Y.-R., Zabetian, C.P., Zhao, Y., Ziegler, S.G., 2009. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N. Engl. J. Med. 361, 1651–1661. https://doi.org/10.1056/NEJMoa0901281

Simón-Sánchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., Krüger, R., Federoff, M., Klein, C., Goate, A., Perlmutter, J., Bonin, M., Nalls, M.A., Illig, T., Gieger, C., Houlden, H., Steffens, M., Okun, M.S., Racette, B.A., Cookson, M.R., Foote, K.D., Fernandez, H.H., Traynor, B.J., Schreiber, S., Arepalli, S., Zonozi, R., Gwinn, K., van der Brug, M., Lopez, G., Chanock, S.J., Schatzkin, A., Park, Y., Hollenbeck, A., Gao, J., Huang, X., Wood, N.W., Lorenz, D., Deuschl, G., Chen, H., Riess, O., Hardy, J.A., Singleton, A.B., Gasser, T., 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. 41, 1308–1312. https://doi.org/10.1038/ng.487

Smith, L., Schapira, A.H.V., 2022. GBA Variants and Parkinson Disease: Mechanisms and Treatments. Cells 11, 1261. https://doi.org/10.3390/cells11081261

Sorge, J.A., West, C., Kuhl, W., Treger, L., Beutler, E., 1987. The human glucocerebrosidase gene has two functional ATG initiator codons. Am. J. Hum. Genet. 41, 1016–1024.

Spillantini, M.G., Goedert, M., 2018. Neurodegeneration and the ordered assembly of α-synuclein. Cell Tissue Res. 373, 137–148. https://doi.org/10.1007/s00441-017-2706-9

Spillantini, M.G., Schmidt, M.L., Lee, V.M., Trojanowski, J.Q., Jakes, R., Goedert, M., 1997. Alpha-synuclein in Lewy bodies. Nature 388, 839–840. https://doi.org/10.1038/42166

Stefanis, L., 2012. α-Synuclein in Parkinson's disease. Cold Spring Harb. Perspect. Med. 2, a009399. https://doi.org/10.1101/cshperspect.a009399

Stirnemann, J., Belmatoug, N., Camou, F., Serratrice, C., Froissart, R., Caillaud, C., Levade, T., Astudillo, L., Serratrice, J., Brassier, A., Rose, C., Billette de Villemeur, T., Berger, M.G., 2017. A Review of Gaucher Disease Pathophysiology, Clinical Presentation and Treatments. Int. J. Mol. Sci. 18, 441. https://doi.org/10.3390/ijms18020441

Stoker, T.B., Torsney, K.M., Barker, R.A., 2018. Pathological Mechanisms and Clinical Aspects of GBA1 Mutation-Associated Parkinson's Disease, in: Stoker, Thomas B., Greenland, J.C. (Eds.), Parkinson's Disease: Pathogenesis and Clinical Aspects. Codon Publications, Brisbane (AU).

Stuppia, L., Antonucci, I., Palka, G., Gatta, V., 2012. Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. Int. J. Mol. Sci. 13, 3245–3276. https://doi.org/10.3390/ijms13033245

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. 20, 467–484. https://doi.org/10.1038/s41576-019-0127-1

Tansey, M.G., Goldberg, M.S., 2010. Neuroinflammation in Parkinson's disease: its role in neuronal death and implications for therapeutic intervention. Neurobiol. Dis. 37, 510–518. https://doi.org/10.1016/j.nbd.2009.11.004

Tayebi, N., Stubblefield, B.K., Park, J.K., Orvisky, E., Walker, J.M., LaMarca, M.E., Sidransky, E., 2003. Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease. Am. J. Hum. Genet. 72, 519–534. https://doi.org/10.1086/367850

Thomas, K.J., McCoy, M.K., Blackinton, J., Beilina, A., van der Brug, M., Sandebring, A., Miller, D., Maric, D., Cedazo-Minguez, A., Cookson, M.R., 2011. DJ-1 acts in parallel to the PINK1/parkin pathway to control mitochondrial function and autophagy. Hum. Mol. Genet. 20, 40–50. https://doi.org/10.1093/hmg/ddq430

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192. https://doi.org/10.1093/bib/bbs017

Tofaris, G.K., Layfield, R., Spillantini, M.G., 2001. alpha-synuclein metabolism and aggregation is linked to ubiquitin-independent degradation by the proteasome. FEBS Lett. 509, 22–26. https://doi.org/10.1016/s0014-5793(01)03115-5

Toffoli, M., Chen, X., Sedlazeck, F.J., Lee, C.-Y., Mullin, S., Higgins, A., Koletsi, S., Garcia-Segura, M.E., Sammler, E., Scholz, S.W., Schapira, A.H.V., Eberle, M.A., Proukakis, C., 2022. Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene. Commun. Biol. 5, 670. https://doi.org/10.1038/s42003-022-03610-7

Toft, M., Ross, O.A., 2010. Copy number variation in Parkinson's disease. Genome Med. 2, 62. https://doi.org/10.1186/gm183

Tolosa, E., Wenning, G., Poewe, W., 2006. The diagnosis of Parkinson's disease. Lancet Neurol. 5, 75–86. https://doi.org/10.1016/S1474-4422(05)70285-4

Tong, S.-Y., Fan, K., Zhou, Z.-W., Liu, L.-Y., Zhang, S.-Q., Fu, Y., Wang, G.-Z., Zhu, Y., Yu, Y.-C., 2022. mvPPT: A Highly Efficient and Sensitive Pathogenicity Prediction Tool for Missense Variants. Genomics Proteomics Bioinformatics S1672-0229(22)00091–2. https://doi.org/10.1016/j.gpb.2022.07.005

Torkamani, A., Wineinger, N.E., Topol, E.J., 2018. The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. 19, 581–590. https://doi.org/10.1038/s41576-018-0018-x

Trinh, J., Lohmann, K., Baumann, H., Balck, A., Borsche, M., Brüggemann, N., Dure, L., Dean, M., Volkmann, J., Tunc, S., Prasuhn, J., Pawlack, H., Imhoff, S., Lill, C.M., Kasten, M., Bauer, P., Rolfs, A., International Parkinson's Disease Genomics Consortium (IPDGC), Klein, C., 2019. Utility and implications of exome sequencing in early-onset Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 34, 133–137. https://doi.org/10.1002/mds.27559

Tysnes, O.-B., Storstein, A., 2017. Epidemiology of Parkinson's disease. J. Neural Transm. Vienna Austria 1996 124, 901–905. https://doi.org/10.1007/s00702-017-1686-y

Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M.K., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G., Albanese, A., Nussbaum, R., González-Maldonado, R., Deller, T., Salvi, S., Cortelli, P., Gilks, W.P., Latchman, D.S., Harvey, R.J., Dallapiccola, B., Auburger, G., Wood, N.W., 2004. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304, 1158–1160. https://doi.org/10.1126/science.1096284

Vellingiri, B., Suriyanarayanan, A., Abraham, K.S., Venkatesan, D., Iyer, M., Raj, N., Gopalakrishnan, A.V., 2022. Influence of heavy metals in Parkinson's disease: an overview. J. Neurol. 269, 5798–5811. https://doi.org/10.1007/s00415-022-11282-w

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vogiatzi, T., Xilouri, M., Vekrellis, K., Stefanis, L., 2008. Wild type alpha-synuclein is degraded by chaperone-mediated autophagy and macroautophagy in neuronal cells. J. Biol. Chem. 283, 23542–23556. https://doi.org/10.1074/jbc.M801992200

Volkmann, J., 2004. Deep brain stimulation for the treatment of Parkinson's disease. J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc. 21, 6–17. https://doi.org/10.1097/00004691-200401000-00003

Wang, C., Zhou, C., Guo, T., Huang, P., Xu, X., Zhang, M., 2022. Association between cigarette smoking and Parkinson's disease: a neuroimaging study. Ther. Adv. Neurol. Disord. 15, 17562864221092566. https://doi.org/10.1177/17562864221092566

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

Wang, Y., Liu, L., Xiong, J., Zhang, X., Chen, Z., Yu, L., Chen, C., Huang, J., Zhang, Z., Mohmed, A.A., Lin, Z., Xiong, N., Wang, T., 2012. Glucocerebrosidase L444P mutation confers genetic risk for Parkinson's disease in central China. Behav. Brain Funct. BBF 8, 57. https://doi.org/10.1186/1744-9081-8-57

Webb, J.L., Ravikumar, B., Atkins, J., Skepper, J.N., Rubinsztein, D.C., 2003. Alpha-Synuclein is degraded by both autophagy and the proteasome. J. Biol. Chem. 278, 25009–25013. https://doi.org/10.1074/jbc.M300227200

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. 37, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

West, A.B., Moore, D.J., Biskup, S., Bugayenko, A., Smith, W.W., Ross, C.A., Dawson, V.L., Dawson, T.M., 2005. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. Proc. Natl. Acad. Sci. U. S. A. 102, 16842–16847. https://doi.org/10.1073/pnas.0507360102

Williams, E.T., Chen, X., Moore, D.J., 2017. VPS35, the Retromer Complex and Parkinson's Disease. J. Park. Dis. 7, 219–233. https://doi.org/10.3233/JPD-161020

Winder-Rhodes, S.E., Evans, J.R., Ban, M., Mason, S.L., Williams-Gray, C.H., Foltynie, T., Duran, R., Mencacci, N.E., Sawcer, S.J., Barker, R.A., 2013. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. Brain J. Neurol. 136, 392–399. https://doi.org/10.1093/brain/aws318

Winfield, S.L., Tayebi, N., Martin, B.M., Ginns, E.I., Sidransky, E., 1997. Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: implications for Gaucher disease. Genome Res. 7, 1020–1026. https://doi.org/10.1101/gr.7.10.1020

Xia, N., Cabin, D.E., Fang, F., Reijo Pera, R.A., 2022. Parkinson's Disease: Overview of Transcription Factor Regulation, Genetics, and Cellular and Animal Models. Front. Neurosci. 16, 894620. https://doi.org/10.3389/fnins.2022.894620

Xue, A., Wu, Yang, Zhu, Z., Zhang, F., Kemper, K.E., Zheng, Z., Yengo, L., Lloyd-Jones, L.R., Sidorenko, J., Wu, Yeda, eQTLGen Consortium, McRae, A.F., Visscher, P.M., Zeng, J., Yang, J., 2018. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat. Commun. 9, 2941. https://doi.org/10.1038/s41467-018-04951-w

Yoon, E., Ahmed, S., Li, R., Bandres-Ciga, S., Blauwendraat, C., Dustin, I., Scholz, S., Hallett, M., Ehrlich, D., 2023. Association of polygenic risk score with response to deep brain stimulation in Parkinson's disease. BMC Neurol. 23, 143. https://doi.org/10.1186/s12883-023-03188-5

Zampieri, S., Cattarossi, S., Bembi, B., Dardis, A., 2017. GBA Analysis in Next-Generation Era: Pitfalls, Challenges, and Possible Solutions. J. Mol. Diagn. JMD 19, 733–741. https://doi.org/10.1016/j.jmoldx.2017.05.005

Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., Rossle, S.C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., Hotzy, C., Reinthaler, E., Harutyunyan, A., Kralovics, R., Peters, A., Zimprich, F., Brücke, T., Poewe, W., Auff, E., Trenkwalder, C., Rost, B., Ransmayr, G., Winkelmann, J., Meitinger, T., Strom, T.M., 2011. A mutation in VPS35, encoding a subunit of the retromer complex,

causes late-onset Parkinson disease. Am. J. Hum. Genet. 89, 168–175. https://doi.org/10.1016/j.ajhg.2011.06.008

Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B., Stoessl, A.J., Pfeiffer, R.F., Patenge, N., Carbajal, I.C., Vieregge, P., Asmus, F., Müller-Myhsok, B., Dickson, D.W., Meitinger, T., Strom, T.M., Wszolek, Z.K., Gasser, T., 2004. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44, 601–607. https://doi.org/10.1016/j.neuron.2004.11.005

Appendices

https://doi.org/10.1093/brain/awab446 BRAIN 2022: 145; 964–978 | 964

BRAIN
ORIGINAL ARTICLE

Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state

Semra Smajić,[1,†] Cesar A. Prada-Medina,[2,†] Zied Landoulsi,[1] Jenny Ghelfi,[1] Sylvie Delcambre,[1] Carola Dietrich,[2] Javier Jarazo,[1,3] Jana Henck,[2] Saranya Balachandran,[4] Sinthuja Pachchek,[1] Christopher M. Morris,[5] Paul Antony,[1] Bernd Timmermann,[2] Sascha Sauer,[6] Sandro L. Pereira,[1] Jens C. Schwamborn,[1,3] Patrick May,[1] Anne Grünewald[1,7,‡] and Malte Spielmann[2,4,8,‡]

[†,‡]These authors contributed equally to this work.

See Lin and Narendra (https://doi.org/10.1093/brain/awac071) for a scientific commentary on this article.

Idiopathic Parkinson's disease is characterized by a progressive loss of dopaminergic neurons, but the exact disease aetiology remains largely unknown. To date, Parkinson's disease research has mainly focused on nigral dopaminergic neurons, although recent studies suggest disease-related changes also in non-neuronal cells and in midbrain regions beyond the substantia nigra. While there is some evidence for glial involvement in Parkinson's disease, the molecular mechanisms remain poorly understood. The aim of this study was to characterize the contribution of all cell types of the midbrain to Parkinson's disease pathology by single-nuclei RNA sequencing and to assess the cell type-specific risk for Parkinson's disease using the latest genome-wide association study.

We profiled >41 000 single-nuclei transcriptomes of post-mortem midbrain from six idiopathic Parkinson's disease patients and five age-/sex-matched controls. To validate our findings in a spatial context, we utilized immunolabelling of the same tissues. Moreover, we analysed Parkinson's disease-associated risk enrichment in genes with cell type-specific expression patterns. We discovered a neuronal cell cluster characterized by *CADPS2* overexpression and low *TH* levels, which was exclusively present in idiopathic Parkinson's disease midbrains. Validation analyses in laser-microdissected neurons suggest that this cluster represents dysfunctional dopaminergic neurons. With regard to glial cells, we observed an increase in nigral microglia in Parkinson's disease patients. Moreover, nigral idiopathic Parkinson's disease microglia were more amoeboid, indicating an activated state. We also discovered a reduction in idiopathic Parkinson's disease oligodendrocyte numbers with the remaining cells being characterized by a stress-induced upregulation of *S100B*. Parkinson's disease risk variants were associated with glia- and neuron-specific gene expression patterns in idiopathic Parkinson's disease cases. Furthermore, astrocytes and microglia presented idiopathic Parkinson's disease-specific cell proliferation and dysregulation of genes related to unfolded protein response and cytokine signalling. While reactive patient astrocytes showed *CD44* overexpression, idiopathic Parkinson's disease microglia revealed a pro-inflammatory trajectory characterized by elevated levels of *IL1B*, *GPNMB* and *HSP90AA1*.

Taken together, we generated the first single-nuclei RNA sequencing dataset from the idiopathic Parkinson's disease midbrain, which highlights a disease-specific neuronal cell cluster as well as 'pan-glial' activation as a central mechanism in the pathology of the movement disorder. This finding warrants further research into inflammatory signalling and immunomodulatory treatments in Parkinson's disease.

Received June 22, 2021. Revised September 21, 2021. Accepted November 18, 2021. Advance access publication December 17, 2021

1 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg
2 Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany
3 OrganoTherapeutics SARL-S, L-4362 Esch-sur-Alzette, Luxembourg
4 Institute of Human Genetics, Kiel University, D-42118 Kiel, Germany
5 Newcastle Brain Tissue Resource, Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, NE1 7RU Newcastle upon Tyne, UK
6 Max-Delbrück-Centrum für Molekulare Medizin, Genomics Group, D-13125 Berlin, Germany
7 Institute of Neurogenetics, University of Lübeck, D-23562 Lübeck, Germany
8 Institute of Human Genetics, University of Lübeck, D-23562 Lübeck, Germany

Correspondence to: Malte Spielmann
Institute of Human Genetics
University of Lübeck
Ratzeburger Allee 160
23562 Lübeck, Germany
E-mail: malte.spielmann@uksh.de

Correspondence may also be addressed to: Anne Grünewald
Luxembourg Centre for Systems Biomedicine
University of Luxembourg
6 avenue du Swing
L-4367 Belvaux, Luxembourg
E-mail: anne.gruenewald@uni.lu

Keywords: Parkinson's disease; midbrain substantia nigra; single-cell sequencing; microglia; neuroinflammation

Abbreviations: DaNs = dopaminergic neurons; OPCs = oligodendrocyte precursor cells; SN = substantia nigra; SNP = single nucleotide polymorphism; snRNA-seq = single-nucleus RNA sequencing; UMAP = uniform manifold approximation and projection

Introduction

Parkinson's disease is a neurological disorder that is commonly characterized by a progressive loss of neuromelanin-containing dopaminergic neurons (DaNs) in the substantia nigra (SN).[1,2] Age, genetic and environmental factors contribute to Parkinson's disease pathogenesis, but disease pathology and aetiology remain mostly unknown.[3] Approximately 95% of Parkinson's disease patients do not harbour an interpretable genetic cause; therefore, they are classified as idiopathic Parkinson's disease.[4]

So far Parkinson's disease research has mainly focused on nigral dopaminergic neurons. By contrast, recent studies suggest disease-related changes also in non-neuronal cells and in brain regions beyond the SN. For instance, PET of drug-naive Parkinson's disease patients revealed microglial activation of the entire brain.[5] This finding is supported by histological analyses in post-mortem Parkinson's disease tissue, which indicate microglial activation in the nigra but also in the putamen, hippocampus, and cortex.[6] Reactive microglia can trigger the induction of neurotoxic reactive astrocytes,[7] which, in turn, interfere with oligodendrocyte survival.[8] Accordingly, glial pathology is suspected to drive neuroinflammatory processes throughout the brain, which contribute to neuronal demise in Parkinson's disease.

The current understanding of neuron-glia cellular perturbations in Parkinson's disease relies largely on experimental models that lack adequate representation of the disease complexity. For instance, toxin-induced animal models capture neither the nature of the human brain nor the multifactorial aspect of the disease.[9] Also, induced pluripotent stem cell (iPSC) models derived from idiopathic Parkinson's disease patients lack the complex cellular composition and dynamics found in a human brain. Several transcriptomic studies using human post-mortem midbrain tissue have investigated the transcriptional programs disrupted in idiopathic Parkinson's disease. However, these studies used either bulk RNA-seq approaches on midbrain tissue or on laser capture-microdissected dopaminergic neurons, thereby failing to disentangle cell-type-specific contributions to the disease pathology.[10] The recent development of single-cell sequencing technologies offers the possibility to overcome these challenges. In particular, transcriptional profiling of single cells (scRNA-seq) or nuclei (snRNA-seq) has proved itself to be an effective strategy to obtain a global view of disease-associated changes at an unprecedented resolution.[11] Moreover, this single-cell approach can be linked to known disease-specific genetic variants to reveal disease trait association in specific cell types.

To address the above-described knowledge gaps and technical limitations, we performed snRNA-seq of post-mortem adult human midbrain tissue of idiopathic Parkinson's disease patients and age-matched control subjects. Using this approach, we obtained an unbiased and global view of the cell type composition as well as the transcriptional programmes disrupted in idiopathic Parkinson's disease glia and neurons at single-cell resolution.

Materials and methods

Human brain tissue cryosectioning

Frozen human post-mortem midbrain tissue sections and the associated clinical and neuropathological data were supplied by the Parkinson's UK Brain Bank and the Newcastle Brain Tissue Resource. According to the neuropathological procedure, after removing the brainstem and cerebellum, the brain hemispheres were divided down the midline, with the hemi-midbrain associated

with each hemisphere. The left hemi-midbrain was removed with a transverse section by taking a line from just behind the mammillary body through the superior colliculus. This midbrain block was then snap-frozen at −120°C and cryosectioned at ∼15 μm thickness in the transverse plane. The resulting sections were stored at −80°C.

Patients and control subjects gave written informed consent with the brain banks, which, together with the ethics review panel of the University of Luxembourg, approved the study.

Sample preparation for nuclei isolation

Six to eight sections were combined from one individual for nuclei isolation. Nuclei were isolated by adapting the published 10X Genomics® protocol for 'Isolation of Nuclei for Single Cell RNA Sequencing'. In brief, the tissue was lysed in a chilled lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0,1% Nonidet™ P40). Then, the suspension was filtered and nuclei were pelleted by centrifugation. Nuclei-pellets were then washed in a 'nuclei wash and resuspension buffer' [1× PBS (phosphate-buffered saline), 1% BSA (bovine serum albumin), 0.2 U/μl RNase inhibitor], filtered, and pelleted again. Nuclei-pellets were suspended in the DAPI solution (1.5 μM DAPI in 1× PBS) and incubated for 5 min prior to FACS sorting. After dissociation, single DAPI-positive nuclei were filtered by size and granularity using a FACSDiva Cell Sorter (BD Biosciences) to minimize the amount of cell debris in the suspension. The resulting nuclei were inspected under the microscope. Only those that appeared intact were considered when adjusting the nuclei concentration prior to loading of the sequencer.

Library preparation and sequencing

Sorted nuclei were processed using the Chromium Next GEM Single Cell 3′ Kit v3.1 to generate the cDNA libraries. The quality of cDNA was assessed using the Agilent 2100 Bioanalyzer System. Sequencing was performed on Illumina NovaSeq 6000-S2.

Transcript quantification and filtering

FASTQ files were generated from the raw base call (BCL) outputs with the Cell Ranger (10× Genomics) *mkfastq* pipeline v.3.0. From this, we obtained a gene-barcode UMI count matrix per sample using the Cell Ranger (10× Genomics) count pipeline v.3.0 using default parameters. The Cell Ranger count pipeline only considers exon-mapping reads during UMI-counting. Also, single-nuclei sequencing readouts are enriched in intronic regions. To account for this, we used the Cell Ranger recommended variation of the human reference transcriptome (hg38), where introns are annotated as exons. The CellRanger pipeline predicted 51 929 barcodes to represent intact single nuclei across all samples, from which 10 494 were filtered out. We retained barcodes with >1500 UMIs and >1000 genes, as well as <10% of mitochondrial-encoded (mtDNA) and <10% of ribosomal gene counts. We only kept genes that were detected in at least three barcodes. Next, we removed ribosomal and mtDNA-encoded genes from the count matrix. We then used Scrublet[12] to identify potential multiplet-barcodes, and only kept barcodes with an estimated doublet score <0.15 for downstream analysis.

Normalization, sample integration and cell clustering

To identify the major cell types comprising the human midbrain, we combined the samples in a single embedding following the Seurat v3[13] CCA integration workflow. First, each sample was normalized using the SCTransform approach.[14] Cell-cycle phase assignment was performed based on this normalized expression matrix. We used the Seurat *CellCycleScoring* function and the Seurat v3 reference genes for the S and G2/M cell-cycle phases. To determine the inter-sample anchors for integration, we used the *FindIntegrationAnchors* Seurat function with the top 4000 consistent highly variable genes across the samples, identified with the *SelectIntegrationFeatures* function. We then used the *IntegrateData* Seurat function to obtain a combined and centred expression matrix. Principal component analysis was carried out on this centred expression matrix. The top 25 principal components were used to build a shared nearest neighbour (SNN) cell graph, which was then clustered using the Louvain algorithm (resolution = 1.5) implemented in the Seurat *FindClusters* function. The top 25 principal components were embedded onto two dimensions using the Uniform Manifold Approximation and Projection (UMAP) algorithm with the number of neighbours set to 30 and a minimum distance set to 0.3 following Seurat3 default implementation.[15] We identified marker genes for each cluster by using the ROC method of the Seurat *FindAllMarkers* function. The top marker genes were used to assign cell type annotations manually for each cell cluster. We compared the cell types by correlating their pseudo bulk profiles. The resulting gene-cell type matrix was normalized (transcript per million) and log2 transformed. The Pearson correlation estimates among the normalized cell type profiles were used as the input distance matrix for hierarchical clustering.

Machine learning cross-validation of cell-type annotation

To quantitatively validate the cell-cluster definition and annotation, we implemented a stratified cross-validation machine learning approach. Briefly, we removed the sample effects on the combined UMI count dataset using Harmony.[16] For normalization, we used the loess transform[17] to fit a smooth curve between mean and variance using the log-transformed data. We then scaled the data with the fitted mean and standard deviation (SD). The identified marker genes (Supplementary Table 3) were selected as the features of the model. We considered each cell type's median cell type to subsample the dataset as a few cell types (DaNs and CADPS2^{high}) were under-represented. To ensure similar label composition in the training and test sets, we split the data using scikit-learn[18] *StratifiedKFold* with 70% of the data as training and 30% as test dataset 5-fold cross-validation. We performed dimensionality reduction with *truncatedSVD* to 30 components. These 30 components were classified based on scikit-multilearn's ensemble classification,[19] which uses Louvain-based clustering[20,21] and a random forest classification to account for the clustered and sparse nature of the snRNA-seq data. The predicted cell types were then compared to the manually curated cell label assignments using a confusion matrix.

Differential cell-type composition

We estimated the differential cell type composition by comparing the UMAP embeddings and the cell type proportions between the idiopathic Parkinson's disease and control samples. We considered the two-dimensional kernel cell density of the idiopathic Parkinson's disease and control cells independently on the first two UMAP components using the *kde2d* function (bins = 100) implemented in the *MASS* R package.[22] The idiopathic Parkinson's

disease log2 differential UMAP density was calculated. Also, for each cell type, we compared the proportion of cells per sample between the idiopathic Parkinson's disease patients and control individuals. We assessed this difference with the Student's *t*-test implemented in the *t.test* function of the R stats package.[23] Furthermore, we used the beta-regression model to estimate the contribution of the sample clinical features [e.g. condition, post-mortem interval (PMI), age] on the cell proportion variation. We modelled the cell type proportion using the *betareg* R package.[24]

Sub-clustering, trajectory reconstruction, and differential gene expression in three glial cell types

We subset cell-type-specific UMI raw counts. To identify the glial subpopulations, we integrated cells from different samples following the Seurat3 reciprocal principal component analysis based protocol considering the top 1000 highly variable genes for astrocytes and oligodendrocytes and the top 500 highly variable genes for microglia. Then we used the unsupervised and network-based Louvain clustering approach based on the top 25 principal components of the integrated datasets. Marker genes were defined as described before. We reconstructed the cellular activation trajectories following the *monocle3* approach. Briefly, cells from different samples were integrated, and factor size normalized. The sample effect was removed using the Mutual Nearest Neighbor method.[25] Then the highly variable genes defined before were embedded in the first 25 principal components used for dimensionality reduction and trajectory inference using the DDR algorithm implemented in the *learn_graph* function of the *monocle3* R package.[11] Pseudotime ordering was done in a supervised manner by rooting the trajectory in the graph node that maximizes the distance to the known activated cell subpopulation. We identified cell type-specific perturbed genes in idiopathic Parkinson's disease using the Quasi-Poisson generalized linear model implemented in the *fit_models* function of the *monocle3* R package.[11] Idiopathic Parkinson's disease differential expression coefficient with $q < 0.05$ were considered as differentially expressed genes. Highly variable genes associated with the cell trajectories were identified using the spatial correlation analysis Moran's I approach implemented in the *graph_test* function of the *monocle3* R package.[11] Functional enrichment analysis of the differentially perturbed genes was done using Enrichr.[26]

CADPS2 expression validation in dopaminergic neurons

Midbrain tissues on PEN slides were fixed in ice-cold 75% ethanol for 3 min, then in 99% ethanol for 1 min and then air-dried for 5–10 min prior to dissection.[27] From the SN of each sample, 150 neuromelanin-positive neurons were captured with laser-microdissection using the PALM Microbeam (Zeiss) in 20 µl nuclease-free water with 0.2 U/µl RNase inhibitor (Roche). RNA was extracted with the NucleoSpin RNA XS purification kit (Macherey-Nagel) according to the manufacturer's protocol. The reverse-transcription into cDNA was performed with SuperScript III Reverse Transcriptase (ThermoFisher).

The *CADPS2* expression was quantified by means of digital PCR (dPCR) using the QuantStudio™ 3D Digital PCR System (Applied Biosystems). Samples were prepared following the manufacturer's instructions using SYBR™ Green (S5763, Life Technologies) and QuantStudio™ 3D digital Master Mix v2 (A26359, Life Technologies). Primer sequences for CADPS2 are: forward 3′-AAACTCTGTGCCCTGGATGG-5′ and reverse 3′-GACAACACGCC TTCCAACAC-5′. Primer sequences for Actin are: forward 3′-CGA GGACTTTGATTGCACATTGTT-5′ and reverse 3′-TGGGGTGGCTTTT AGGATGG-5′. Samples were loaded on a QuantStudio™ 3D digital PCR Chip v2 using the QuantStudio™ 3D Digital PCR Chip loader. The PCR was then performed on the ProFlex™ 2X Flat PCR System using the following parameters: 95°C for 5 min, 45 cycles of 95°C for 10 s, and 60°C for 10 s, 72°C for 10 s. The chips were read using the QuantStudio™ 3D Digital PCR instrument and the data were analysed using the QuantStudio™ 3D AnalysisSuite, version 3.1.6-PRC-build18.

Multi-fluorescence immunolabelling of the tissue

Paraffin-embedded PFA-fixed midbrain sections were deparaffinized by incubation at 60°C for 30 min. This was followed by washing with Histoclear (2×5 min) and ethanol gradient series (100%, 100%, 95%, 70% vol/vol, 5 min each), and finally in distilled water for 10 min. Antigen retrieval was performed in 1 mM EDTA, pH = 8, in a pressure cooker for 40 min. Next, the slides were washed in distilled water and 1% TBST and blocked with 10% NGS in 1% TBST for 1 h. The sections were then incubated in the primary antibody (anti-tyrosine hydroxylase MAB318, 1:100, Millipore; anti-myelin PLP, 1:100, Abcam; anti-IBA1 019-19741, 1:500, FUJIFILM Wako; anti-GFAP ab7260, 1:100, Abcam), diluted to a working concentration in 1% NGS in 0.1% TBST, at 4°C overnight. This was followed by washing 3×5 min in 1% TBST. Then, the midbrain sections were incubated with a secondary antibody (goat anti-mouse IgG1 Alexa Fluor 647, A21240; goat anti-mouse IgG2a Alexa Fluor 546, A-21133; goat anti-rabbit IgG Alexa Fluor 488, A-27034), diluted to a working concentration of 1:100 in 1% NGS in 0.1% TBST, for 1 h. Sections were rewashed 3×5 min in 1% TBST and incubated in Sudan black solution for 2 min. This was followed by three washes in 1% TBST and mounting in ProLong Gold mounting medium.

Automated image analysis

Immunofluorescence images of human post-mortem midbrain sections were acquired with Carl Zeiss Axio Observer Inverted Microscope Z1 with 20× objective and analysed in MATLAB (Version 2019B, Mathworks). Automated in-house developed image analysis algorithms segmented the fluorescent cell areas (neurons, astrocytes, microglia, oligodendrocyte) extracting features such as area and perimeter. The segmentation of dopaminergic neurons was computed by convolving the raw TH channel with a Gaussian filter. TH-positive cell areas were detected by setting a pixel threshold followed by *bwareaopen* to remove small connected components to generate a TH area mask. The neuromelanin mask was computed by identifying areas below the selected pixel threshold and subtracting the small connected components with *bwareaopen*. The segmentation of astrocytes and microglia was calculated by selecting a pixel threshold, followed by an *imfill* filter to generate the cell area masks for GFAP or IBA1, respectively. Further, the skeleton of the IBA1 mask was generated with a thinning function to identify the branching of the mask. Because of the massive oligodendrocyte population, we generated the mask by selecting a pixel threshold to identify all the PLP1 positive areas without segmentation. The mean area of each individual was calculated, and the groups were compared with an unpaired *t*-test. The results were visualized with ggplot2 in R 4.0.0.

disease log2 differential UMAP density was calculated. Also, for each cell type, we compared the proportion of cells per sample between the idiopathic Parkinson's disease patients and control individuals. We assessed this difference with the Student's *t*-test implemented in the *t.test* function of the R stats package.[23] Furthermore, we used the beta-regression model to estimate the contribution of the sample clinical features [e.g. condition, post-mortem interval (PMI), age] on the cell proportion variation. We modelled the cell type proportion using the *betareg* R package.[24]

Sub-clustering, trajectory reconstruction, and differential gene expression in three glial cell types

We subset cell-type-specific UMI raw counts. To identify the glial subpopulations, we integrated cells from different samples following the Seurat3 reciprocal principal component analysis based protocol considering the top 1000 highly variable genes for astrocytes and oligodendrocytes and the top 500 highly variable genes for microglia. Then we used the unsupervised and network-based Louvain clustering approach based on the top 25 principal components of the integrated datasets. Marker genes were defined as described before. We reconstructed the cellular activation trajectories following the *monocle3* approach. Briefly, cells from different samples were integrated, and factor size normalized. The sample effect was removed using the Mutual Nearest Neighbor method.[25] Then the highly variable genes defined before were embedded in the first 25 principal components used for dimensionality reduction and trajectory inference using the DDR algorithm implemented in the *learn_graph* function of the *monocle3* R package.[11] Pseudotime ordering was done in a supervised manner by rooting the trajectory in the graph node that maximizes the distance to the known activated cell subpopulation. We identified cell type-specific perturbed genes in idiopathic Parkinson's disease using the Quasi-Poisson generalized linear model implemented in the *fit_models* function of the *monocle3* R package.[11] Idiopathic Parkinson's disease differential expression coefficient with $q < 0.05$ were considered as differentially expressed genes. Highly variable genes associated with the cell trajectories were identified using the spatial correlation analysis Moran's I approach implemented in the *graph_test* function of the *monocle3* R package.[11] Functional enrichment analysis of the differentially perturbed genes was done using Enrichr.[26]

CADPS2 expression validation in dopaminergic neurons

Midbrain tissues on PEN slides were fixed in ice-cold 75% ethanol for 3 min, then in 99% ethanol for 1 min and then air-dried for 5–10 min prior to dissection.[27] From the SN of each sample, 150 neuromelanin-positive neurons were captured with laser-microdissection using the PALM Microbeam (Zeiss) in 20 µl nuclease-free water with 0.2 U/µl RNase inhibitor (Roche). RNA was extracted with the NucleoSpin RNA XS purification kit (Macherey-Nagel) according to the manufacturer's protocol. The reverse-transcription into cDNA was performed with SuperScript III Reverse Transcriptase (ThermoFisher).

The *CADPS2* expression was quantified by means of digital PCR (dPCR) using the QuantStudio™ 3D Digital PCR System (Applied Biosystems). Samples were prepared following the manufacturer's instructions using SYBR™ Green (S5763, Life Technologies) and QuantStudio™ 3D digital Master Mix v2 (A26359, Life Technologies). Primer sequences for CADPS2 are: forward 3′-AAACTCTGTGCCCTGGATGG-5′ and reverse 3′-GACAACACGCC TTCCAACAC-5′. Primer sequences for Actin are: forward 3′-CGA GGACTTTGATTGCACATTGTT-5′ and reverse 3′-TGGGGTGGCTTTT AGGATGG-5′. Samples were loaded on a QuantStudio™ 3D digital PCR Chip v2 using the QuantStudio™ 3D Digital PCR Chip loader. The PCR was then performed on the ProFlex™ 2X Flat PCR System using the following parameters: 95°C for 5 min, 45 cycles of 95°C for 10 s, and 60°C for 10 s, 72°C for 10 s. The chips were read using the QuantStudio™ 3D Digital PCR instrument and the data were analysed using the QuantStudio™ 3D AnalysisSuite, version 3.1.6-PRC-build18.

Multi-fluorescence immunolabelling of the tissue

Paraffin-embedded PFA-fixed midbrain sections were deparaffinized by incubation at 60°C for 30 min. This was followed by washing with Histoclear (2×5 min) and ethanol gradient series (100%, 100%, 95%, 70% vol/vol, 5 min each), and finally in distilled water for 10 min. Antigen retrieval was performed in 1 mM EDTA, pH = 8, in a pressure cooker for 40 min. Next, the slides were washed in distilled water and 1% TBST and blocked with 10% NGS in 1% TBST for 1 h. The sections were then incubated in the primary antibody (anti-tyrosine hydroxylase MAB318, 1:100, Millipore; anti-myelin PLP, 1:100, Abcam; anti-IBA1 019-19741, 1:500, FUJIFILM Wako; anti-GFAP ab7260, 1:100, Abcam), diluted to a working concentration in 1% NGS in 0.1% TBST, at 4°C overnight. This was followed by washing 3×5 min in 1% TBST. Then, the midbrain sections were incubated with a secondary antibody (goat anti-mouse IgG1 Alexa Fluor 647, A21240; goat anti-mouse IgG2a Alexa Fluor 546, A-21133; goat anti-rabbit IgG Alexa Fluor 488, A-27034), diluted to a working concentration of 1:100 in 1% NGS in 0.1% TBST, for 1 h. Sections were rewashed 3×5 min in 1% TBST and incubated in Sudan black solution for 2 min. This was followed by three washes in 1% TBST and mounting in ProLong Gold mounting medium.

Automated image analysis

Immunofluorescence images of human post-mortem midbrain sections were acquired with Carl Zeiss Axio Observer Inverted Microscope Z1 with 20× objective and analysed in MATLAB (Version 2019B, Mathworks). Automated in-house developed image analysis algorithms segmented the fluorescent cell areas (neurons, astrocytes, microglia, oligodendrocyte) extracting features such as area and perimeter. The segmentation of dopaminergic neurons was computed by convolving the raw TH channel with a Gaussian filter. TH-positive cell areas were detected by setting a pixel threshold followed by *bwareaopen* to remove small connected components to generate a TH area mask. The neuromelanin mask was computed by identifying areas below the selected pixel threshold and subtracting the small connected components with *bwareaopen*. The segmentation of astrocytes and microglia was calculated by selecting a pixel threshold, followed by an *imfill* filter to generate the cell area masks for GFAP or IBA1, respectively. Further, the skeleton of the IBA1 mask was generated with a thinning function to identify the branching of the mask. Because of the massive oligodendrocyte population, we generated the mask by selecting a pixel threshold to identify all the PLP1 positive areas without segmentation. The mean area of each individual was calculated, and the groups were compared with an unpaired *t*-test. The results were visualized with ggplot2 in R 4.0.0.

Genotyping of Parkinson's disease cases using NeuroChip

DNA samples from all idiopathic Parkinson's disease cases underwent genotyping at the Institute of Human Genetics at the Helmholtz Zentrum München using the Illumina (San Diego, CA) NeuroChip.[28] Standard genotype data quality control (QC) steps were carried out.[29] Single nucleotide polymorphism (SNP) imputation was carried out on our NeuroChip data using the Michigan Imputation Server[30] to produce a final list of common (minor allele frequency ≥1%) variants for further analyses. Imputed SNP positions were based on Genome Reference Consortium Human 37/human genome version 19 (GRCh37/hg19). All cases were screened for disease-associated variants in known major Parkinson's disease genes (*SNCA*, *LRRK2*, *DJ-1*, *PRKN*, *GBA*, *PINK1*, *ATP13A2*, *VPS35*, *MAPT*, *DCTN1*, *DNAJC6*, *SYNJ1*, *VPS13C* and *MAPT*) covered by the NeuroChip.

Cell type association with genetic risk of Parkinson's disease

Association analysis of cell type-specific expressed genes with genetic risk of Parkinson's disease was performed using Multi-marker Analysis of GenoMic Annotation (MAGMA) v1.08, to identify idiopathic Parkinson's disease-relevant cell types in the midbrain. MAGMA is a gene-set enrichment analysis method that tests the joint association of all risk SNPs in a gene with the phenotype while accounting for the linkage disequilibrium (LD) structure between SNPs.[31] In our study, the SNPs and their *P*-values were taken from the summary statistics of the Parkinson's disease genome-wide association study (GWAS) from Nalls *et al.*[32] (excluding 23andMe). The publicly available European subset of 1000 Genomes Phase 3 was used as a reference panel to estimate LD between SNPs. MAGMA analysis consists of three steps. First, the annotation step, where SNPs were mapped to genes using the NCBI GRCh37 build (annotation release 105). Gene boundaries were defined as the transcribed region of each gene. An extended window of 35 kb upstream and 10 kb downstream of each gene was added to the gene boundaries.[33] Second, the gene analysis step computes gene-wise *P*-values based on SNP GWAS *P*-values. The third step is the competitive gene-set analysis implemented as a linear regression model on a gene-set data matrix. The gene-sets used here are the differentially expressed genes in every cell type or the cell-type-specific expressed genes [filtered for false discovery rate (FDR)-corrected *P*-values < 0.05, percentage of cells of the cluster where the expression was detected >0.5, and logFC > 0.25]. MAGMA gene-set analysis provides association results for every gene-set and for every gene in the gene-sets. The association of a gene with a cell type is quantified as a z-score. Z-scores will be close to zero if a gene is not differentially expressed, while high positive z-scores indicate most differentially expressed genes.

Data availability

Raw snRNA-seq data for the 11 samples presented in this study are available in the Gene Expression Omnibus (GEO) with accession number GSE157783. Imaging data are available upon request.

Results

We sampled adult human post-mortem midbrain tissue from five idiopathic Parkinson's disease cases, for which pathology reports described a severe neuronal loss in the SN without a family history of Parkinson's disease (Supplementary Table 1). We confirmed the idiopathic nature by SNP-Chip profiling of 179 467 known variants associated with neurological diseases, including Parkinson's disease,[34] which did not reveal a genetic aetiology (Supplementary Table 2). We sampled six control midbrains to match the idiopathic Parkinson's disease patient characteristics. The average age of idiopathic Parkinson's disease patients and control individuals were ∼77 [standard error of the mean (SEM = 3)] and ∼81 (SEM = 4) years, respectively, and both groups had similar post-mortem intervals (idiopathic Parkinson's disease ∼22 and controls ∼16 h) (Supplementary Table 1).

We sequenced single nuclei from frozen ventral sections of human post-mortem midbrains (Fig. 1A) and obtained ∼2000–6000 high-quality nuclei per sample with an average of ∼7600 transcripts and ∼2700 genes per nucleus after filtering out poorly sequenced nuclei and potential doublets (Fig. 1B). This dataset comprised 22 433 and 19 002 single nuclei from control individuals and patients with idiopathic Parkinson's disease, respectively (Fig. 1C).

We embedded the 41 435 nuclei transcriptomes into two dimensions using the UMAP algorithm. We found that the overall cluster structure was mostly driven by cell-type identity and inter-sample variability (Supplementary Fig. 1A and B). Of note, patient and control cells gathered together within the major cell clusters (Supplementary Fig. 1B). To account for this inter-individual variation during the cell-type identification, we followed the Seurat3 sample integration protocol (see 'Materials and methods' section) (Fig. 1D). Using this corrected principal component analysis embedding and the unsupervised, network-based Louvain clustering approach, we found that the human midbrain comprised 12 major cell types (Fig. 1D and E and Supplementary Fig. 1C).

The studied human midbrain tissue was composed of glial, neuronal, and vascular cells (Fig. 1D and E). We annotated most cell clusters by manually comparing well-known marker genes in the literature and the identified marker genes of each unsupervised cell cluster (Supplementary Fig. 1C–I and Supplementary Table 3). Oligodendrocytes, the most abundant cell type in the midbrain (Fig. 1F), were characterized by the expression of *MOBP*.[35] Oligodendrocyte precursor cells (OPCs) highly express *VCAN*.[36] Expression of *AQP4* was characteristic for astrocytes[37] and *FOXJ1* for ependymal cells (Fig. 1E–G and Supplementary Fig. 1C–I).[38] Also, immune and vascular cells displayed a highly specific expression of well-known marker genes; *CD74* in microglia,[39] *CLDN5* in endothelial cells,[40,41] and *GFRB* in pericytes[42] (Fig. 1E–G and Supplementary Fig. 1C). Regarding neuronal cells, we identified four cell types: excitatory (*SLC17A6*),[43] inhibitory (*GAD2*),[43] GABAergic (*GAD2*/*GRIK1*)[44,45] and, dopaminergic neurons (*TH*) (Fig. 1E–G and Supplementary Fig. 1C–I).[46]

A closer look at the number of profiled nuclei indicated that DaNs only comprised 0.18% of the total cell count limiting the comparison between the idiopathic Parkinson's disease and control DaNs (Fig. 1F). Therefore, we performed quantitative immunofluorescence imaging analysis of TH- or neuromelanin-positive cells in idiopathic Parkinson's disease and control tissues and confirmed a significant reduction in TH- or neuromelanin-positive nigral DaNs in the idiopathic Parkinson's disease midbrains compared to controls (Supplementary Fig. 2A–D).

Interestingly, we also found a neuronal cluster of 120 cells, which we could not annotate initially based on known marker genes, that was characterized by high expression of *CADPS2* (CADPS2^{high} cells) (Fig. 1E–F and H–I and Supplementary Fig. 1C and I). These cells almost exclusively originated from idiopathic

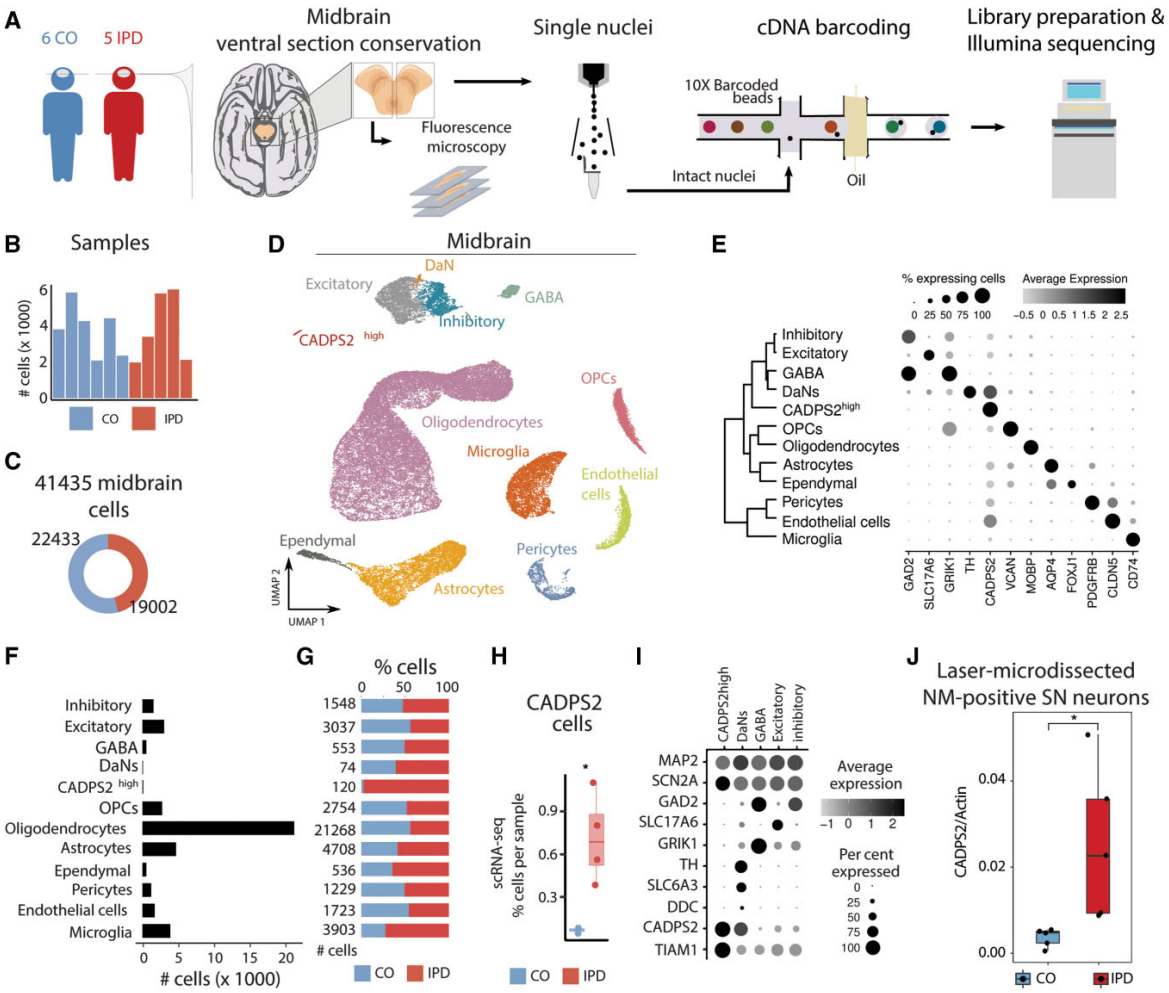


Figure 1 Cell type composition of human midbrain. (**A**) The experimental approach to midbrain tissue processing and nuclei extraction. Nuclei suspensions were processed with the 10× Genomics platform and sequenced with an Illumina sequencer. (**B**) Contribution of nuclei from idiopathic Parkinson's disease (IPD) patients or controls to each cell type. (**C**) The number of high-quality nuclei per sample. Overall, the population consists of 19 002 nuclei from idiopathic Parkinson's disease patients and 22 433 nuclei from controls. (**D**) UMAP embedding of the 41 435 human midbrain nuclei. Cells are coloured by cell type. (**E**) Cell type transcriptome similarity and representative marker genes. CADPS2^{high} cells cluster together with the neuronal cells. (**F**) The number of profiled nuclei per cell type. (**G**) The proportion of idiopathic Parkinson's disease and control profiled cells per cell type. (**H**) CADPS2^{high} cell proportion per sample, (*t*-test $P = 0.02$). (**I**) CADPS2^{high} cells are neurons. They express *MAP2*, *SCN2A*, and *TIAM1*, but have low levels of *TH*. (**J**) Digital PCR reveals significantly higher expression of *CADPS2* in neuromelanin-positive nigral neurons ($n = 150$ per person) dissected from idiopathic Parkinson's disease midbrain sections compared to those isolated from control tissue (*t*-test $P = 0.027$).

Parkinson's disease patients (idiopathic Parkinson's disease, 98.4%; control, 1.6%) (Fig. 1G). Quantitative assessment of the cell annotation assignment validated our cell-type annotation (Supplementary Fig. 1D). With regard to neuronal markers, these cells show a similar profile to DaNs, except for low *TH* abundance (Fig. 1I). Moreover, CADPS2^{high} cells express even higher levels of *TIAM1* than DaNs (Fig. 1I). TIAM1 has been identified as a regulator of the Wnt/Dvl/Rac1 pathway, which controls midbrain DaN differentiation.[47,48] Thus, we hypothesized that CADPS2^{high} cells might constitute degenerating DaNs.

In order to test this hypothesis, we applied laser-capture microdissection (LCM) to frozen midbrain sections from additional idiopathic Parkinson's disease patients and aged control subjects. Our 'validation cohort' included C1 and IPD4 from the original 'snRNA-seq cohort' as well as sections from four previously unstudied cases and four new controls (Supplementary Table 1). From each individual, we isolated 150 neuromelanin-containing neurons that were subjected to *CADPS2* gene expression analysis by means of digital PCR, whereby *Beta-actin* served as a house-keeping gene. This experiment indicated significantly higher *CADPS2*:*Beta-actin* ratios in idiopathic Parkinson's disease compared to control DaNs with neuromelanin deposits (Fig. 1J), suggesting that CADPS2^{high} cells are indeed of dopaminergic origin.

Next, we aimed to reveal cell-type composition changes of the midbrain associated with idiopathic Parkinson's disease and followed three approaches. We compared idiopathic Parkinson's disease and control cell density distributions in the 2D UMAP representation (Fig. 2A and B) and analysed the idiopathic

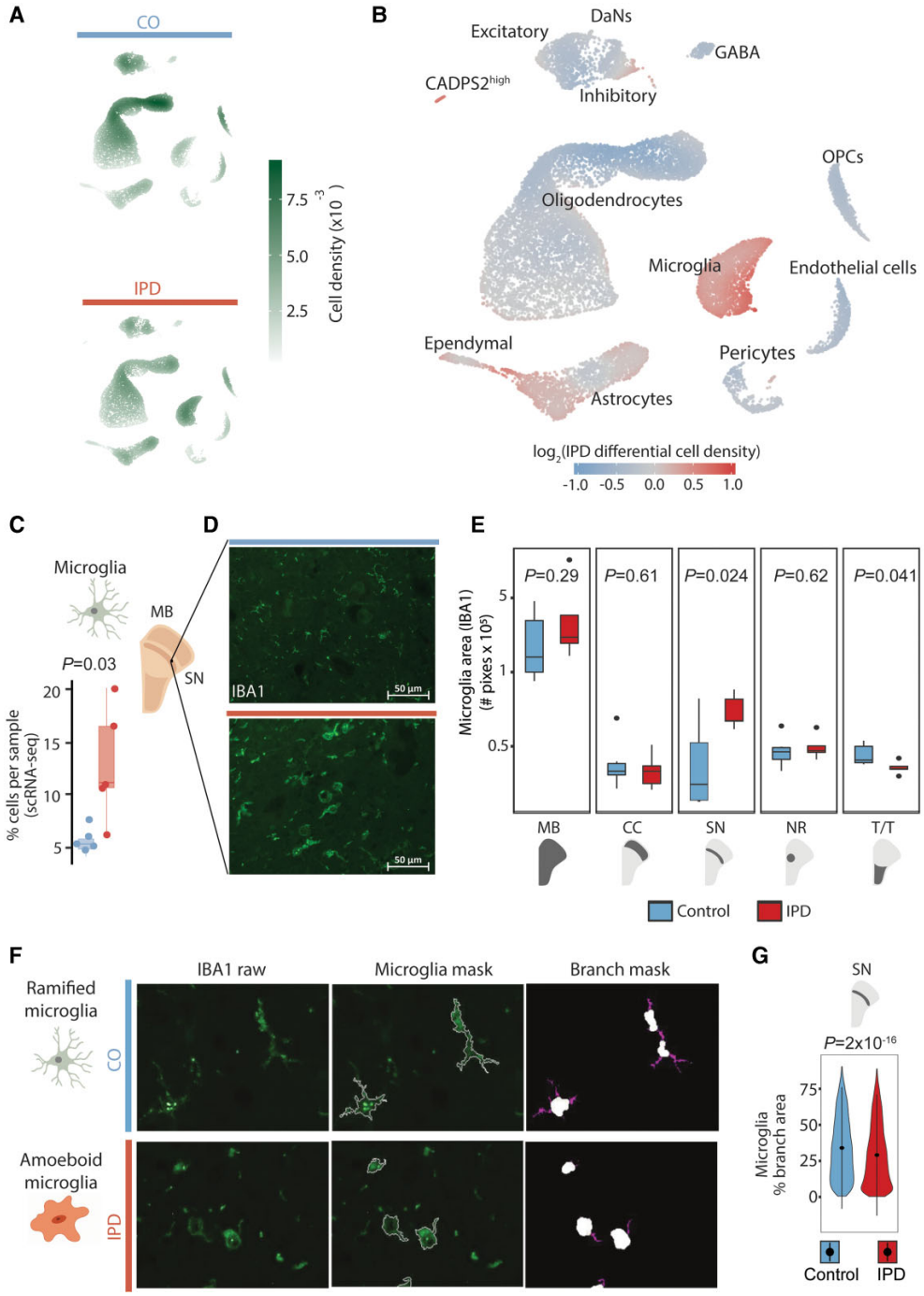


Figure 2 Idiopathic Parkinson's disease (IPD) midbrain is characterized by an increase in microglia. Differential cell type composition in idiopathic Parkinson's disease patients compared to age-matched control subjects. (**A**) Two-dimensional cell density in the first UMAP embeddings of the human midbrain for idiopathic Parkinson's disease patients and control subjects independently. (**B**) Differential 2D cell density in idiopathic Parkinson's disease midbrain. Idiopathic Parkinson's disease midbrain has a larger population of microglia and astrocytes than control midbrain tissue. (**C**) Microglia cell proportion per sample. Idiopathic Parkinson's disease patients display a higher proportion of microglia cells (*t*-test $P = 0.03$). (**D**) IBA1 immunofluorescence in idiopathic Parkinson's disease and control ventral midbrain sections. (**E**) IBA1-positive areas in the entire midbrain and individual regions of 11 individuals. The Parkinson's disease-associated increase of microglia is the most significant in the SN (*t*-test $P = 0.024$). (**F**) Microglia morphology analysis. (**G**) An idiopathic Parkinson's disease-associated reduction of microglia branching indicates less ramified microglia in the SN (*t*-test $P = 2 \times 10^{-16}$), which implies increased cell reactivity. MB = midbrain; PD = Parkinson's disease; NR = nucleus ruber; TT = tectum/tegmentum; CC = crus cerebri. IPD: red bar; control: blue bar; scale bar = 50 µm.

Parkinson's disease and control distributions of the cell type proportions per sample. Altogether, these results revealed an increase in the fraction of microglia and astrocytes and a decreased fraction of oligodendrocytes in idiopathic Parkinson's disease midbrains compared to controls (Fig. 2C, Supplementary Fig. 3A and D and Supplementary Table 6). To validate these results with an independent approach, we examined paraformaldehyde-fixed paraffin-embedded sections from the right hemi-midbrain of the same 11 individuals by performing multi-labelling immunofluorescence analysis.[49] First, we confirmed the increased fraction of microglia in idiopathic Parkinson's disease midbrains by labelling it with an antibody against the marker protein IBA1 (Fig. 2D). Automated image analysis demonstrated an increase in IBA1-positive areas in idiopathic Parkinson's disease midbrain tissue compared to control samples (Fig. 2E and Supplementary Table 7). This microglia increase was the most significant in the SN compared to other midbrain regions (Fig. 2E and Supplementary Table 7). Further image analysis of the microglia cellular shape in the SN of age- and sex-matched idiopathic Parkinson's disease and control cases (Fig. 2F) revealed an idiopathic Parkinson's disease-related decrease in microglial branching, indicating cellular activation (Fig. 2G and Supplementary Table 7).[50] Second, we validated the increased fraction of astrocytes and a decreased fraction of oligodendrocytes in idiopathic Parkinson's disease midbrains. We labelled astrocytes and oligodendrocytes with antibodies against their marker proteins GFAP and PLP1, respectively (Supplementary Fig. 3B and E). We observed a trend towards a higher abundance of GFAP-positive areas throughout all the regions in idiopathic Parkinson's disease midbrain tissue compared to control subjects (Supplementary Fig. 3C, Supplementary Table 7). Moreover, we detected a reduction of PLP1-positive areas in the idiopathic Parkinson's disease midbrain sections compared to controls with the highest significance in the SN (Supplementary Fig. 3F). In contrast, the other midbrain cell types, OPCs, pericytes, ependymal, excitatory, inhibitory, and GABAergic cells, did not display significant deviations associated with idiopathic Parkinson's disease (Supplementary Fig. 4 and Supplementary Table 6).

We also investigated how other clinical characteristics, in addition to the disease status (condition), affect the midbrain cellular composition. For this, we modelled the percentage of each cell type per sample as a function of age, post-mortem interval, and condition. We used beta-regression modelling to estimate the

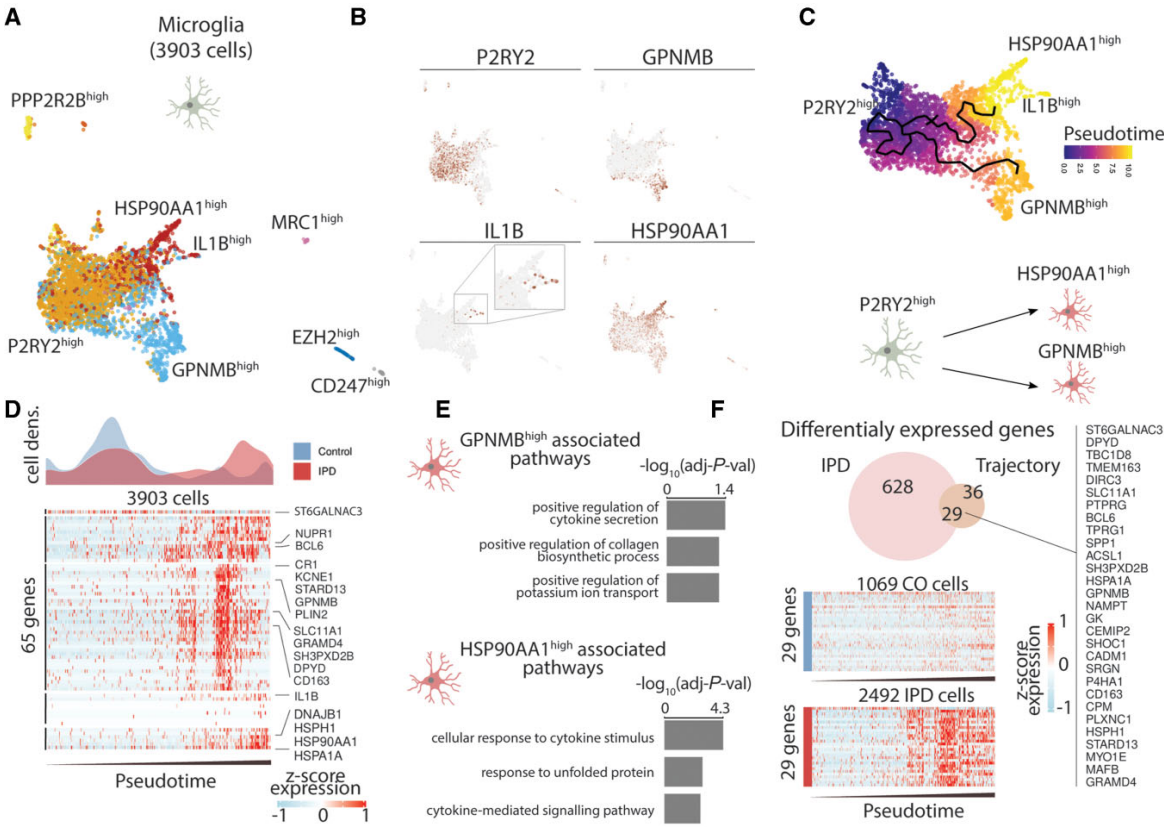


Figure 3 Trajectory reconstruction reveals microglia differential activation in idiopathic Parkinson's disease (IPD). (**A**) Microglia subpopulations labelled with a representative marker gene. (**B**) Expression of *P2RY12*, *GPNMB*, *IL1B* and *HSP90AA1* along the ~4000 microglia cells. These genes are characteristic of the major three microglia subpopulations. (**C**). Trajectory reconstruction and pseudotime representation based on the *P2RY1*^{high}, *GPNMB*^{high}, and *HSP90AA1B*^{high} subpopulations. This reveals a two-branches activation trajectory. (**D**) Differential cell-density distribution along pseudotime for idiopathic Parkinson's disease and control samples. Also, the expression of 65 genes, whose expression is associated with the microglia activation trajectory. Z-score normalized expression is presented for each gene over ~4000 microglia cells organized by their pseudotime. (**E**) Gene ontology (GO) molecular function enrichment of the genes associated with the *GPNMB* and *HSP90AA1* activation trajectories. (**F**) Twenty-nine idiopathic Parkinson's disease differentially expressed genes intersect with the differentially expressed genes along the microglia activation trajectory.

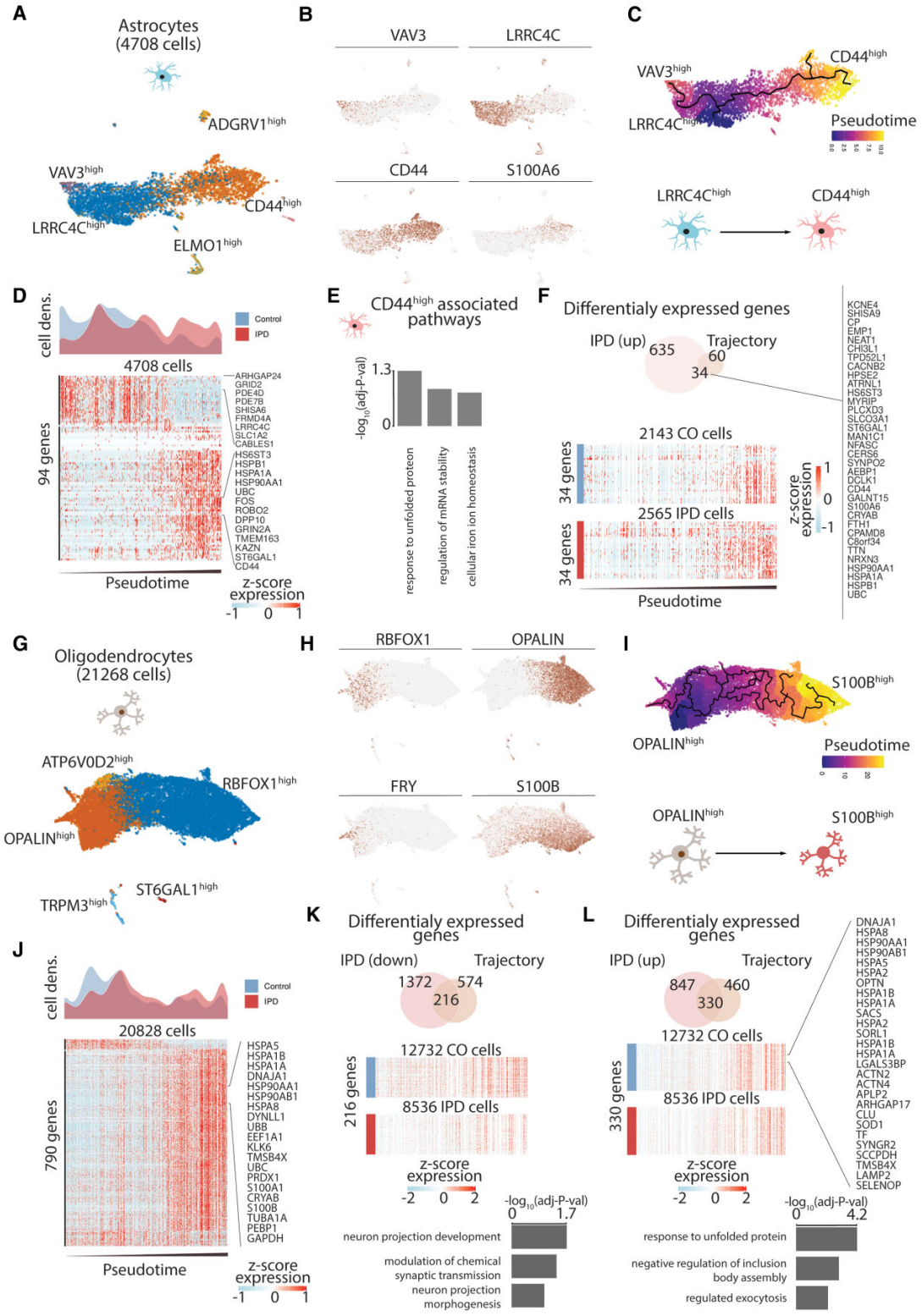


Figure 4 Trajectory reconstruction reveals astrocyte differential activation, loss of myelinating oligodendrocytes, and differential activation in idiopathic Parkinson's disease (IPD).

Continued

coefficients of these clinical features (Supplementary Fig. 5). Condition (idiopathic Parkinson's disease) appeared to be the sample characteristics with the highest impact on the midbrain cellular composition. For instance, the most significant coefficients were the loss of DaNs and the gain of CADPS2^{high} cells associated with idiopathic Parkinson's disease (Supplementary Fig. 5).

To reveal the transcriptional programmes and pathways associated with the increased fraction of microglia and astrocytes in idiopathic Parkinson's disease, we subclustered these cell types to identify glial subpopulations and reconstruct their activation trajectories. We identified seven microglia subpopulations characterized by the expression of a few marker genes (Fig. 3A). The three biggest subpopulations are defined by the high expression of *P2RY12*, *GPNMB*, and *HSP90AA1* (Fig. 3B). Given that these three subpopulations conform to a continuum in the UMAP projection and both *GPNMB* and *HSP90AA1* are microglia activation markers, we estimated a cell trajectory structure comprising these major subpopulations (Fig. 3C). We then organized cells along this trajectory (pseudotime), starting from the trajectory node that maximizes the distance to the *GPNMB* and *HSP70AA1* trajectory branch tips (Fig. 3C). This microglia activation trajectory spans from *P2RY12*^{high} cells towards two activation branches, one containing *GPNMB*^{high} cells and another with cells highly expressing *HSP90AA1* or *IL1B* (Fig. 3C). Idiopathic Parkinson's disease cells differentially distribute along the microglia activation trajectory being enriched towards the activated state compared (Fig. 3D and Supplementary Fig. 6). While P2RY12 is highly abundant in the resting microglia,[51] GPNMB,[52] HSP90,[53] and IL-1β[54] are involved in the inflammatory response and have previously been linked to neurodegeneration[53,55,56] supporting the notion that the idiopathic Parkinson's disease-specific upregulation of *GPNMB* and *HSP90AA1* are markers of microglial activation. To further characterize the molecular phenotype of these two activated microglia states, we identified genes whose expression was associated with the activation trajectory and functionally enriched them to gene-ontology molecular functions (Fig. 3D and E). This analysis revealed that these subpopulations are enriched in cytokine secretion and the stress response to unfolded protein pathways (Fig. 3E). Next, we identified the genes whose expression was differentially upregulated in idiopathic Parkinson's disease microglia across the activation trajectory. We intersected the upregulated genes in idiopathic Parkinson's disease and the activation-trajectory associated genes in microglia and identified 29 genes linked with the differential activation of the midbrain microglia in idiopathic Parkinson's disease (Fig. 3F and Supplementary Table 8), several of which have previously been associated with idiopathic Parkinson's disease.[57–59]

We also characterized the astrocyte and oligodendrocyte subpopulations, reconstructed their activation trajectories, and identified gene signatures associated with idiopathic Parkinson's disease differential activation (Fig. 4). First, we identified five astrocyte subpopulations characterized by high expression of *VAV3*, *LRRC4C*, *ELMO1*, *ADGRV1* and *CD44* (Fig. 4A and B). We recovered the astrocyte activation trajectory based on the main cell types comprising *VAV3*^{high}, *LRRC4C*^{high}, and *CD44*/*S100A6*^{high} subpopulations (Fig. 4C). Given that *CD44* expression implicates reactive astrogliosis,[60] we ordered cells on the activation trajectory by setting the root in the trajectory graph-node that maximizes the distance from the *CD44*^{high} branch end. These results implied an astrocyte activation transition from *LRRC4C*^{high} to *CD44*^{high} subpopulations (Fig. 4C). Indeed, idiopathic Parkinson's disease astrocytes were highly enriched at the end of the astrogliosis trajectory compared to control astrocytes (Fig. 4D and Supplementary Fig. 6). We further characterized the molecular phenotype of the *CD44*^{high} astrocyte activated state by enriching GO molecular functions to the highly upregulated genes across the astrocytes activation trajectory (Fig. 4D and E). The *CD44*^{high} subpopulation was related to the unfolded protein response (UPR) pathway, which has recently been linked to a specific astrocyte reactivity state that is detrimental to the survival of neurons (Fig. 4E).[61] Next, we calculated idiopathic Parkinson's disease-differentially upregulated genes, which were also highly expressed towards the end of the astrogliosis trajectory (Fig. 4F), and identified 34 genes associated with idiopathic Parkinson's disease differential astrogliosis (Fig. 4F and Supplementary Table 8). These genes include several heat-shock proteins that have previously been shown to co-localize with α-synuclein deposits in the human brain.[62] Similarly, we investigated the oligodendrocyte diversity and reconstructed its differentiation trajectory (Fig. 4G–L). We identified five subpopulations characterized by the expression of *ATP6V02*, *OPALIN*, *TRPM3*, *ST6GAL1*, and *RBFOX1* (Fig. 4G and H). The inferred trajectory based on subpopulations recovered differentiation trajectory spanning from *FRY*/*OPALIN*^{high} cells towards *RBFOX1*/*S100B*^{high} cells (Fig. 4H and I). *OPALIN* (also denominated as *Tmem*) is a marker of myelinating oligodendrocytes,[63] while *S100B* has been associated with glial stress response in Parkinson's disease post-mortem midbrain.[64] When comparing idiopathic Parkinson's disease oligodendrocyte density across this trajectory, we found a reduced fraction of myelinating *OPALIN*^{high} cells compared to controls (Fig. 4J and Supplementary Fig. 6). An overlay of the idiopathic Parkinson's disease-differentially expressed genes and of such genes defining the oligodendrocyte trajectory identified 216 and 330 downregulated and upregulated genes across the trajectory. Downregulated genes are associated with neuronal maintaining pathways, while upregulated genes are related to the response to unfolded protein pathways (Fig. 4K and L and Supplementary Table 8). We also investigated the idiopathic Parkinson's disease differential expression

Figure 4 Continued
(**A**) Astroglial subpopulations are named based on characteristic marker genes. (**B**) *VAV3*, *LRRC4C*, *CD44*, and *S100A6* expression across the ∼4700 astrocytes. (**C**) Inferred cell trajectory and pseudotime for the major astrocyte subpopulations, *VAV3*^{high}, *LRRC4C*^{high} and *CD44*^{high} cells. (**D**) Idiopathic Parkinson's disease and control differential cell-density distribution over pseudotime and the expression of the 94 genes highly associated with the astrogliosis trajectory in the ∼4700 astrocytes organized by pseudotime. (**E**) GO molecular function pathway enrichment of the upregulated genes in the *CD44*^{high} activated branch. (**F**) The 34 intersected genes between the upregulated genes in idiopathic Parkinson's disease and across the astrocyte activation trajectory. (**G**) Oligodendrocyte subpopulations are named based on representative marker genes. (**H**) Expression of *OPALIN*, *RBFOX1*, *FRY* and *S100B* in the ∼21 000 oligodendrocytes. (**I**) Inferred cell trajectory and pseudotime ordering of the major oligodendrocytes subpopulations, *OPALIN*^{high}, *ATP6V0D2*^{high}, and *S100B*^{high} cells. (**J**) Idiopathic Parkinson's disease and control differential cell density over pseudotime. Expression levels of 790 highly variable genes across the oligodendrocyte trajectory. Expression is presented for ∼21 000 oligodendrocytes organized by their pseudotime. (**K** and **L**) The intersection of idiopathic Parkinson's disease differentially expressed and trajectory-associated genes. Also, the GO molecular enrichment of the intersected genes is presented. (**K**) Two hundred and sixteen idiopathic Parkinson's disease downregulated genes across the trajectory are associated with pathways important for neuron projection and synaptic transmission. (**L**) Three hundred and thirty genes are idiopathic Parkinson's disease upregulated along the oligodendrocyte trajectory. These genes are mainly associated with the unfolded protein response.

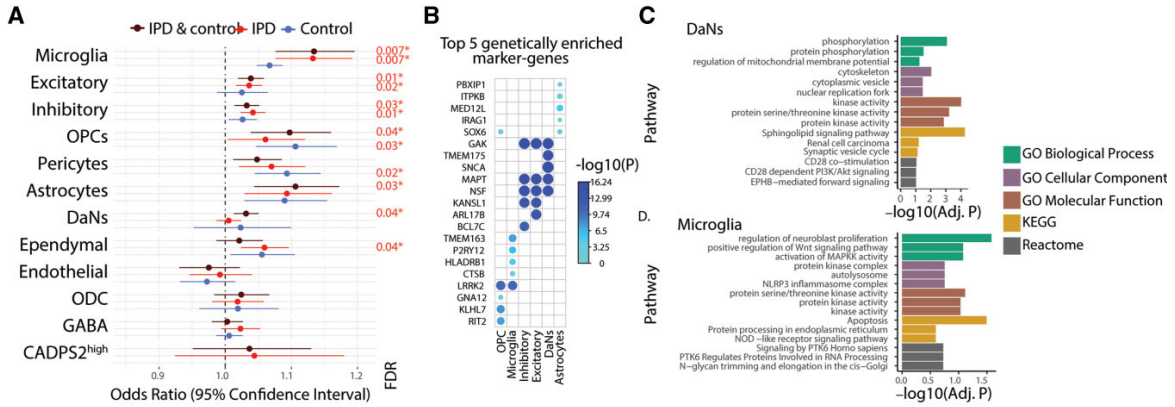


Figure 5 Idiopathic Parkinson's disease (IPD)-associated genetic variants enriched in microglia and neuron-specific genes. (**A**) Forest plots of the odds ratio (OR) and 95% confidence intervals for the association between the Parkinson's disease-associated variants and the marker genes of the midbrain cell types from idiopathic Parkinson's disease patients, control subjects and both conditions. This approach describes the enrichment of Parkinson's disease risk variants, taken from the latest Parkinson's disease GWAS, in genes with cell-type-specific patterns in order to identify idiopathic Parkinson's disease relevant cell types in the midbrain. Only significant association *P*-values were shown (*$P < 0.05$). (**B**) Top five enriched genes in six midbrain cell types. The association of a gene with a cell type is quantified and the most responsible genes for the genetic variant enrichment observed in (**A**) were shown. The *P*-values of genes association are colour coded from light to dark blue and the size of circles is inversely proportional to *P*-values. (**C** and **D**) Gene Ontology terms (GO) and molecular pathways (KEGG, Reactome) associated respectively with the DaNs and microglia marker genes responsible for the Parkinson's disease variant enrichment.

across all cell types in the human midbrain and evidenced that the unfolded protein pathways are also upregulated in the OPCs, and the vascular cells (Supplementary Table 8).

To gain cellular mechanistic insights into how the Parkinson's disease-associated genetic variants could affect the midbrain physiology, we evaluated the enrichment of midbrain cell-type marker genes with the Parkinson's disease-associated genetic variants. We found that Parkinson's disease risk variants are significantly associated with microglia, neurons, astrocytes, and OPCs (Fig. 5A and Supplementary Table 4). Having access to both control and idiopathic Parkinson's disease tissue, we tested whether these associations depend on the disease context. After analysing each condition separately, we found that the Parkinson's disease risk variants associate differently with patients and controls (Fig. 5A). Considering idiopathic Parkinson's disease samples alone, microglia and neurons remain significantly associated with Parkinson's disease risk variants (Fig. 5A and Supplementary Table 4). By contrast, in control subjects, disease variants are associated with pericytes and OPCs (Fig. 5A and Supplementary Table 4). These results show that the link between Parkinson's disease genetic risk and cell type is highly influenced by the disease status. When analysing the samples separately (Fig. 5A), the association of DaNs to risk variants is weaker, presumably due to the low number of DaNs. Therefore, we utilized the entire dataset (controls and idiopathic Parkinson's disease cases) for further analyses (Fig. 5B–D).

We prioritized the cell-type-specific and Parkinson's disease risk-associated genes based on their enrichment contribution for each cell type (Fig. 5B). We found that *LRRK2* showed the highest association with microglia and OPCs, and *SNCA* was the most prominent Parkinson's disease-associated gene in DaNs (Fig. 5B and Supplementary Table 5). These findings are in line with previous reports of Parkinson's disease-associated mutations in α-synuclein promoting Lewy body formation in DaNs[65] and with studies suggesting a role for *LRRK2* mutations in the activation of microglia in Parkinson's disease.[66] Lastly, we investigated which pathways are associated with the Parkinson's disease variant enrichment in DaNs and microglial differentially expressed genes (Fig. 5C and D). Among the key hits from GO, KEGG, and Reactome, we identified terms such as 'phosphorylation' and 'kinase activity' in DaNs and 'NLRP3 inflammasome complex' in microglia (Fig. 5D). In particular, the latter finding further supports a role for inflammatory signalling in Parkinson's disease.

Discussion

This study provides the first single-cell atlas of the human midbrain from idiopathic Parkinson's disease patients and age-matched control subjects to the best of our knowledge. Rather than exclusively focusing on nigral DaNs, the most studied cell type in Parkinson's disease, we aimed to characterize cell- and disease-specific molecular signatures associated with idiopathic Parkinson's disease in the entire midbrain. In addition, we associated Parkinson's disease risk variants to specific midbrain cell types in idiopathic Parkinson's disease patients and control subjects.

Our key observations include an increment in the astrocytes and microglia midbrain fractions, which coincided with a reduction of oligodendrocyte fraction in the idiopathic Parkinson's disease midbrain. Immunofluorescence analysis and pseudotime trajectory reconstructions revealed glial activation in idiopathic Parkinson's disease—a finding that was further supported by Parkinson's disease variant enrichment in the idiopathic Parkinson's disease microglia. Finally, we discovered a small CADPS2-positive neuronal cell cluster in idiopathic Parkinson's disease midbrain tissue, which warrants further investigations in a larger sample set.

When assessing DaNs in our snRNA-seq data, we did not observe a significant loss in idiopathic Parkinson's disease tissue. The low abundance of DaNs likely hampered this compared to other cell types in the midbrain. However, the automated image analysis of immunofluorescence-labelled idiopathic Parkinson's disease and control midbrain sections confirmed a significant loss of TH-positive DaNs and neuromelanin aggregates. This result was in line with the neuropathological reports, which described

severe DaN degeneration in all idiopathic Parkinson's disease patient samples (*cf.* Supplementary Table 1). Thus, technical limitations may have caused the under-representation of DaNs in the transcriptomic data. First, we used 15-µm thick midbrain slices, which are in the size range of the rather large DaN nuclei (10–20 µM). Hence, a considerable proportion of nuclei may not have remained intact during the sectioning process—a prerequisite for high-quality snRNA-seq results. Second, rather than sampling only SN cells, we extracted nuclei from the entire midbrain. This may have led to an under-representation of nigral neurons in our dataset. Despite these constraints, when combining the latest GWAS[32] with our snRNA-seq results from idiopathic Parkinson's disease and control midbrain sections, we observed an enrichment of Parkinson's disease variants in DaNs. Pathway analyses of differentially expressed DaN marker genes with Parkinson's disease variant enrichment identified processes such as 'mitochondrial function' and 'kinase activity' that have previously been associated with Parkinson's disease.

In addition, we identified a disease-specific cell type, consisting of only 120 cells, characterized by its transcriptional similarity to midbrain DaNs but with low *TH* levels and high *CADPS2* expression. CADPS2 has previously been linked to catecholamine uptake and genetic Parkinson's disease.[67–69] In addition, elevated levels of *TIAM1*, which is involved in Wnt/Dvl/Rac1 signalling,[47,48] made us wonder whether CADPS2^{high} cells constitute degenerating DaNs that have lost their dopaminergic identity. Aberrant dopamine function in metabolically impaired but viable neurons in the SN has previously been observed in Parkinson's disease post-mortem tissue.[70] Corroborating our hypothesis, *CADPS2* quantification in neuromelanin-positive DaNs isolated from midbrain tissue of two scRNA-sequenced samples (C1 and IPD4) and four additional idiopathic Parkinson's disease patients and four new control subjects revealed higher levels of *CADPS2* in the former cells. However, further mechanistic studies beyond the scope of the manuscript will be needed to uncover the physiological cause and consequence of *CADPS2* upregulation in DaNs.

In our dataset, glia made up ~80% of all sequenced cells, enabling an in-depth analysis of their contribution to the pathogenesis of idiopathic Parkinson's disease. We identified a disease-specific upregulation of microglia, which mediates the innate immune defence in the brain. During microgliosis, microglia amplify, undergo morphological changes, and secrete cytokines, which can further engage surrounding microglia and astrocytes.[7,71] Suggestive of an activated state, we detected fewer ramified microglia in idiopathic Parkinson's disease post-mortem SN tissue using a quantitative immunofluorescence approach. Moreover, we identified a significant Parkinson's disease risk variant enrichment in microglia, showing the strongest association with the Parkinson's disease gene *LRRK2*. The kinase LRRK2 is most abundant in immune cells and may contribute to inflammasome formation via the phosphorylation of Rab GTPases.[72] In line with this finding, Parkinson's disease risk variant enrichment analysis in microglial differentially expressed genes highlighted the kinase activity and NLRP3 inflammasome pathways. By inferring the activation trajectories of the microglial subpopulations, we observed an increase in cells from resting into an activated state. Interestingly, our finding of *GPNMB* upregulation in activated microglia is supported by recent results in Alzheimer's disease brains. Reactive patient microglia, which presented an amoeboid shape, also showed higher GPNMB protein levels in this immunohistochemistry study.[55] Moreover, pathway analyses in the activated cell populations identified cytokine signalling and, likely upstream of this, induction of the UPR pathway in the microglia. We also found chaperones and heat-shock proteins to be overexpressed along the disease trajectory, which, when they are released from the cell, can act as damage-associated molecular patterns (DAMPs) that trigger an immune reaction.[73]

Astrocytes can equally act as immune effector cells in the brain by releasing proinflammatory cytokines.[74] When modelling astroglial activation trajectories, we detected reactive astrogliosis specifically in idiopathic Parkinson's disease patient cells.[60] As for microglia, pathway analysis along the trajectory identified the UPR pathway, which has recently been described to influence the astrocytic secretome.[61] Neurotrophic factors released from reactive astrocytes were shown to accelerate neuronal demise[61]—a disease mechanism that has not gained much attention in Parkinson's disease research so far. Besides neurons, reactive astrocytes can also affect oligodendrocyte function and survival.[7]

Accordingly, our snRNA-seq data also showed a trend towards decreased oligodendrocyte numbers in idiopathic Parkinson's disease midbrain tissue. Immunofluorescence analyses suggest that this reduction is the most profound in the SN. In the white matter, oligodendrocytes generate myelin sheets, which provide insulation of axons and ensure saltatory conduction.[75] However, since Parkinson's disease has long been considered a 'grey matter' disease, oligodendrocytes only recently gained attention in the field. A single-cell study[76] in nigral tissue from controls showed that common genetic Parkinson's disease risk variants are associated with oligodendrocyte-specific expression. Another study on the entire mouse nervous system also reported an association with oligodendrocytes.[77] By contrast, we did not observe an enrichment of Parkinson's disease risk variants in oligodendrocytes from control or idiopathic Parkinson's disease tissue. This may be explained by the fact that our data are based on nuclei from the entire midbrain, possibly masking nigra-specific genetic effects. However, a closer look into trajectory inference analysis in oligodendrocytes revealed a transition from high *OPALIN* to high *S100B* expression subpopulations. S100B was shown to control the maturation process of oligodendrocytes[78] and has previously been linked to neurodegeneration.[79] *S100B* overexpression in response to cytokine injections mediates dystrophic neurite formation in an Alzheimer rat model.[79] Accordingly, the oligodendrocyte-specific upregulation of *S100B* observed in the idiopathic Parkinson's disease midbrains may be the result of enhanced cytokine release from microglia and astrocytes. These results further implicate glial cells in the propagation of neuroinflammatory and neurodegenerative processes in idiopathic Parkinson's disease.

In summary, our study reinforces the relevance of neuroinflammation in idiopathic Parkinson's disease. Applying snRNA-seq for the first time to post-mortem midbrain tissue from patients and matched control subjects, we identified a disease-specific upregulation of microglia and astrocytes as well as a loss of oligodendrocytes. In addition, we discovered a small neuronal cell population that was almost exclusively identified in idiopathic Parkinson's disease midbrain tissue, likely representing degenerating DaNs. Disease trajectory analyses in the glial cell populations identified stress in response to misfolded proteins as the major trigger of inflammatory signalling in idiopathic Parkinson's disease, extending from microglia via astrocytes to oligodendrocytes. Further strengthening this finding, Parkinson's disease risk variants were specifically enriched in microglia from idiopathic Parkinson's disease patients.

Our study also has several limitations. Due to the precious nature of post-mortem brain tissue, our results are based on snRNA-seq of sections from 11 individuals and dPCR analysis of laser-capture microdissected DaNs from eight additional

individuals. Therefore, single-cell RNA analyses in independent cohorts will be needed to validate the key findings from our study. Moreover, additional *in vitro* and *in vivo* experiments are necessary to explore the role of CADPS2 in the pathogenesis of Parkinson's disease and to elucidate how the observed glial interplay perpetuates or induces DaN demise.

Despite these challenges, our unique human single-cell dataset provides the basis for new research approaches investigating the role of the different midbrain cell types in idiopathic Parkinson's disease and for translational programmes that aim to develop immunomodulatory Parkinson's disease therapies.

Acknowledgements

We thank the patients and their families for their generous participation.

Funding

S.Sm. and S.P. received funding from the Luxembourg National Research Fund (FNR) within the PARK-QC DTU (PRIDE17/12244779/PARK-QC). A.G. is supported by the FNR within the framework of the ATTRACT (Model-idiopathic Parkinson's disease, FNR9631103) career development program, and were generously funded by a donation from Le Foyer Assurances Luxembourg. Moreover, A.G., P.M. and Z.L. obtained FNR funding as part of the National Centre of Excellence in Research on Parkinson's disease (NCER-Parkinson's disease, FNR11264123). In addition, A.G. and P.M. received funding within the framework of the FNR CORE grant 'MiRisk-Parkinson's disease' (C17/BM/11676395) and FNR INTER grant 'ProtectMove' (INTER/DFG/19/14429377). M.S. is a DZHK principal investigator and is supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (SP1532/3-1, SP1532/4-1, and SP1532/5-1) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR 01GM1925). C.M.M. is supported by grants to the Newcastle Brain Tissue Resource from UK MRC (MR/L016451/1), the Alzheimer's Society and Alzheimer's Research Trust through the Brains for Dementia Research Initiative, and from National Institutes for Health Research Biomedical Research Centre Newcastle. The Parkinson's UK Brain Bank is funded by Parkinson's UK, a charity registered in England and Wales (258197) and in Scotland (SC037554).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Grünewald A, Kumar KR, Sue CM. New insights into the complex role of mitochondria in Parkinson's disease. *Prog Neurobiol*. 2019;177:73–93.
2. Smolders S, Van Broeckhoven C. Genetic perspective on the synergistic connection between vesicular transport, lysosomal and mitochondrial pathways associated with Parkinson's disease pathogenesis. *Acta Neuropathol Commun*. 2020;8:63.
3. Inamdar NN, Arulmozhi DK, Tandon A, Bodhankar SL. Parkinson's disease: genetics and beyond. *Curr Neuropharmacol*. 2007;5:99–113.
4. Obeso JA, Stamelou M, Goetz CG, *et al*. Past, present, and future of Parkinson's disease: a special essay on the 200th Anniversary of the Shaking Palsy. *Mov Disord*. 2017;32:1264–1310.
5. Ouchi Y, Yagi S, Yokokura M, Sakamoto M. Neuroinflammation in the living brain of Parkinson's disease. *Parkinsonism Relat Disord*. 2009;15(Suppl 3):S200–S204.
6. Imamura K, Hishikawa N, Sawada M, Nagatsu T, Yoshida M, Hashizume Y. Distribution of major histocompatibility complex class II-positive microglia and cytokine profile of Parkinson's disease brains. *Acta Neuropathol*. 2003;106:518–526.
7. Liddelow SA, Guttenplan KA, Clarke L, *et al*. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*. 2017;541:481–487.
8. Kıray H, Lindsay SL, Hosseinzadeh S, Barnett SC. The multifaceted role of astrocytes in regulating myelination. *Exp Neurol*. 2016;283:541–549.
9. Blesa J, Przedborski S. Parkinson's disease: animal models and dopaminergic cell vulnerability. *Front Neuroanat*. 2014;8:155.
10. Borrageiro G, Haylett W, Seedat S, Kuivaniemi H, Bardien S. A review of genome-wide transcriptomics studies in Parkinson's disease. *Eur J Neurosci*. 2018;47:1–16.
11. Cao J, Spielmann M, Qiu X, *et al*. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566:496–502.
12. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Systems*. 2019;8:281–291.e9.
13. Stuart T, Butler A, Hoffman P, *et al*. Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–1902.e21.
14. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:296.
15. McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426
16. Korsunsky I, Millard N, Fan J, *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–1296.
17. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc Stat Assoc*. 1979;74:829–836.
18. Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–2830.
19. Szymański P, Kajdanowicz T, Kersting K. How is a data-driven approach better than random choice in label space division for multi-label classification? *Entropy*. 2016;18:282.
20. Hagberg A, Swart P, S Chult, D. Exploring network structure, dynamics, and function using networkx. https://www.osti.gov/biblio/960616-exploring-network-structure-dynamics-function-using-networkx
21. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech*. 2008;2008:P10008.
22. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Springer; 2002. https://doi.org/10.1007/978-0-387-21706-2
23. R Development Core Team. R: A language and environment for statistical computing (ISBN 3-900051-07-0). https://www.scienceopen.com/document?vid=300a2dc0-3207-4383-818c-51eb0f49f561
24. Cribari-Neto F, Zeileis A. Beta regression in R. *J Stat Softw*. 2010;34:1–24.
25. Haghverdi L, Lun ATL, Morgan MD, Marioni, JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*. 2018;36:421–427.

26. Chen EY, Tan CM, Kou Y, *et al*. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform*. 2013;14:128.
27. Clément-Ziza M, Munnich A, Lyonnet S, Jaubert F, Besmond C. Stabilization of RNA during laser capture microdissection by performing experiments under argon atmosphere or using ethanol as a solvent in staining solutions. *RNA*. 2008;14:2698–2704.
28. Maher MP, Pine J, Wright J, Tai YC. The neurochip: a new multielectrode device for stimulating and recording from cultured neurons. *J Neurosci Methods*. 1999;87:45–56.
29. Reed E, Nunez S, Kulp D, Qian J, Reilly M, Foulkes AS A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med*. 2015;34:3769–3792.
30. Das S, Forer L, Schönherr S, *et al*. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–1287.
31. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015; 11:e1004219.
32. Nalls MA, Blauwendraat C, Vallerga CL, *et al*. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18:1091–1102.
33. Reynolds RH, Botía J, Nalls MA, *et al*. Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. *NPJ Parkinsons Dis*. 2019;5:6.
34. Blauwendraat C, Faghri F, Pihlstrom L, *et al*. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging*. 2017;57:247.e9–247.e13.
35. Mitkus SN, Hyde TM, Vakkalanka R, *et al*. Expression of oligodendrocyte-associated genes in dorsolateral prefrontal cortex of patients with schizophrenia. *Schizophr Res*. 2008;98:129–138.
36. van Bruggen D, Agirre E, Castelo-Branco G. Single-cell transcriptomic analysis of oligodendrocyte lineage cells. *Curr Opin Neurobiol*. 2017;47:168–175.
37. Ikeshima-Kataoka H. Neuroimmunological Implications of AQP4 in Astrocytes. *Int J Mol Sci*. 2016;17:1306.
38. Shah PT, Stratton JA., Stykel MG, *et al*. Single-cell transcriptomics and fate mapping of ependymal cells reveals an absence of neural stem cell function. *Cell*. 2018;173:1045–1057.e9.
39. Hwang IK, Park JH, Lee T-K, *et al*. CD74-immunoreactive activated M1 microglia are shown late in the gerbil hippocampal CA1 region following transient cerebral ischemia. *Mol Med Rep*. 2017;15:4148–4154.
40. Maher TJ, Ren Y, Li Q, *et al*. ATP-binding cassette transporter Abcg2 lineage contributes to the cardiac vasculature after oxidative stress. *Am J Physiol Heart Circ Physiol*. 2014;306:H1610–H1618.
41. Jang AS, Concel VJ, Bein K, *et al*. Endothelial dysfunction and claudin 5 regulation during acrolein-induced lung injury. *Am J Respir Cell Mol Biol*. 2011;44:483–490.
42. Bell RD, Winkler EA, Sagare AP, *et al*. Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron*. 2010;68:409–427.
43. Kodama T, Guerrero S, Shin M, Moghadam S, Faulstich M, du Lac S. Neuronal classification and marker gene identification via single-cell expression profiling of brainstem vestibular neurons subserving cerebellar learning. *J Neurosci*. 2012;32:7819–7831.
44. Merrill CB, Friend LN, Newton ST, Hopkins ZH, Edwards JG. Ventral tegmental area dopamine and GABA neurons: Physiological properties and expression of mRNA for endocannabinoid biosynthetic elements. *Sci Rep*. 2015;5:16176.
45. Wu L-J, Ko SW, Toyoda H, *et al*. Increased anxiety-like behavior and enhanced synaptic efficacy in the amygdala of GluR5 knockout mice. *PLoS One*. 2007;2:e167.
46. Thompson L, Barraud P, Andersson E, Kirik D, Björklund A. Identification of dopaminergic neurons of nigral and ventral tegmental area subtypes in grafts of fetal ventral mesencephalon based on cell morphology, protein expression, and efferent projections. *J Neurosci*. 2005;25:6467–6477.
47. Čajánek L, Ganji RS, Henriques-Oliveira C, *et al*. Tiam1 regulates the Wnt/Dvl/Rac1 signaling pathway and the differentiation of midbrain dopaminergic neurons. *Mol Cell Biol*. 2013;33:59–70.
48. Arenas E. Wnt signaling in midbrain dopaminergic neuron development and regenerative medicine for Parkinson's disease. *J Mol Cell Biol*. 2014;6:42–53.
49. Grünewald A, Lax NZ, Rocha MC, *et al*. Quantitative quadruple-label immunofluorescence of mitochondrial and cytoplasmic proteins in single neurons from human midbrain tissue. *J Neurosci Methods*. 2014;232:143–149.
50. Torres-Platas SG, Comeau S, Rachalski A, *et al*. Morphometric characterization of microglial phenotypes in human cerebral cortex. *J Neuroinflammation*. 2014;11:12.
51. Walker DG, Tang TM, Mendsaikhan A, *et al*. Patterns of expression of purinergic receptor p2ry12, a putative marker for non-activated microglia, in aged and Alzheimer's disease brains. *Int J Mol Sci*. 2020;21:678.
52. van der Poel M, Ulas T, Mizee MR, *et al*. Transcriptional profiling of human microglia reveals grey-white matter heterogeneity and multiple sclerosis-associated changes. *Nat Commun*. 2019; 10:1139.
53. Kakimura J-I, Kitamura Y, Takata K, *et al*. Microglial activation and amyloid-β clearance induced by exogenous heat-shock proteins. *FASEB J*. 2002;16:601–603.
54. Burm SM, Zuiderwijk-Sick EA, 't Jong AEJ, *et al*. Inflammasome-induced IL-1β secretion in microglia is characterized by delayed kinetics and is only partially dependent on inflammatory caspases. *J Neurosci*. 2015;35:678–687.
55. Hüttenrauch M, Ogorek I, Klafki H, *et al*. Glycoprotein NMB: a novel Alzheimer's disease associated marker expressed in a subset of activated microglia. *Acta Neuropathol Commun*. 2018;6:108.
56. Parajuli B, Sonobe Y, Horiuchi H, Takeuchi H, Mizuno T, Suzumura A. Oligomeric amyloid β induces IL-1β processing via production of ROS: implication in Alzheimer's disease. *Cell Death Dis*. 2013;4:e975.
57. Nalls MA, Pankratz N, Lill CM, *et al*. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46:989–993.
58. Wu K-C, Liou H-H, Kao Y-H, Lee C-Y, Lin C-J. The critical role of Nramp1 in degrading α-synuclein oligomers in microglia under iron overload condition. *Neurobiol Dis*. 2017;104:61–72.
59. Li S, Liu H, Bian S, Sha X, Li Y, Wang Y. The accelerated aging model reveals critical mechanisms of late-onset Parkinson's disease. *BioData Min*. 2020;13, 4.
60. Bradford BM, Wijaya CAW, Mabbott NA. Discrimination of prion strain targeting in the central nervous system via reactive astrocyte heterogeneity in CD44 expression. *Front Cell Neurosci*. 2019; 13, 411.
61. Smith HL, Freeman OJ, Butcher AJ, *et al*. Astrocyte unfolded protein response induces a specific reactivity state that causes non-cell-autonomous neuronal degeneration. *Neuron*. 2020; 105:855–866.e5.
62. Muchowski PJ, Wacker JL. Modulation of neurodegeneration by molecular chaperones. *Nat Rev Neurosci*. 2005;6:11–22.
63. Jiang W, Yang W, Yang W, *et al*. Identification of Tmem10 as a novel late-stage oligodendrocytes marker for detecting hypomyelination. *Int J Biol Sci*. 2013;10:33–42.
64. Sathe K, Maetzler W, Lang JD, *et al*. S100B is increased in Parkinson's disease and ablation protects against MPTP-induced

toxicity through the RAGE and TNF-α pathway. *Brain*. 2012;135: 3336–3347.
65. Spillantini MG, Schmidt ML, Lee VMY, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. *Nature*. 1997;388: 839–840.
66. Schapansky J, Nardozzi JD, LaVoie MJ. The complex relationships between microglia, alpha-synuclein, and LRRK2 in Parkinson's disease. *Neuroscience*. 2015;302:74–88.
67. Tsuda L, Lim Y-M. Regulatory system for the G1-arrest during neuronal development in *Drosophila*. *Dev Growth Differ*. 2014; 56:358–367.
68. Brunk I, Blex C, Speidel D, Brose N, Ahnert-Hilger G. Ca2+-dependent activator proteins of secretion promote vesicular monoamine uptake. *J Biol Chem*. 2009;284: 1050–1056.
69. Reinhardt P, Schmid B, Burbulla LF, *et al*. Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell*. 2013;12:354–367.
70. Kordower JH, Olanow CW, Dodiya HB, *et al*. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. *Brain*. 2013;136:2419–2431.
71. Joers V, Tansey MG, Mulas G, Carta AR. Microglial phenotypes in Parkinson's disease and animal models of the disease. *Prog Neurobiol*. 2017;155:57–75.
72. Alessi DR, Sammler E. LRRK2 kinase in Parkinson's disease. *Science*. 2018;360:36–37.
73. Smith JA. Regulation of cytokine production by the unfolded protein response; implications for infection and autoimmunity. *Front Immunol*. 2018;9:422.
74. Dong Y, Benveniste EN. Immune function of astrocytes. *Glia*. 2001;36:180–190.
75. Hamanaka G, Ohtomo R, Takase H, Lok J, Arai K. White-matter repair: Interaction between oligodendrocytes and the neurovascular unit. *Brain Circ*. 2018;4:118–123.
76. Agarwal D, Sandor C, Volpato V, *et al*. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat Commun*. 2020;11:4183.
77. Bryois J, Skene NG, Hansen TF, *et al*. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet*. 2020;52:482–493.
78. Deloulme JC, Raponi E, Gentil BJ, *et al*. Nuclear expression of S100B in oligodendrocyte progenitor cells correlates with differentiation toward the oligodendroglial lineage and modulates oligodendrocytes maturation. *Mol Cell Neurosci*. 2004;27: 453–465.
79. Sheng JG, Ito K, Skinner RD, *et al*. In vivo and in vitro evidence supporting a role for the inflammatory cytokine interleukin-1 as a driving force in Alzheimer pathogenesis. *Neurobiol Aging*. 1996;17:761–766.

npj | parkinson's disease

www.nature.com/npjparkd

ARTICLE OPEN

Check for updates

Age at onset as stratifier in idiopathic Parkinson's disease – effect of ageing and polygenic risk score on clinical phenotypes

L. Pavelka[1,2✉], A. Rauschenberger[3], Z. Landoulsi[4], S. Pachchek[1,4], P. May[4], E. Glaab[3], R. Krüger[1,2,5✉] and on behalf of the NCER-PD Consortium*

Several phenotypic differences observed in Parkinson's disease (PD) patients have been linked to age at onset (AAO). We endeavoured to find out whether these differences are due to the ageing process itself by using a combined dataset of idiopathic PD ($n = 430$) and healthy controls (HC; $n = 556$) excluding carriers of known PD-linked genetic mutations in both groups. We found several significant effects of AAO on motor and non-motor symptoms in PD, but when comparing the effects of age on these symptoms with HC (using age at assessment, AAA), only positive associations of AAA with burden of motor symptoms and cognitive impairment were significantly different between PD vs HC. Furthermore, we explored a potential effect of polygenic risk score (PRS) on clinical phenotype and identified a significant inverse correlation of AAO and PRS in PD. No significant association between PRS and severity of clinical symptoms was found. We conclude that the observed non-motor phenotypic differences in PD based on AAO are largely driven by the ageing process itself and not by a specific profile of neurodegeneration linked to AAO in the idiopathic PD patients.

npj Parkinson's Disease (2022)8:102 ; https://doi.org/10.1038/s41531-022-00342-7

INTRODUCTION

Although considered as one disease entity, Parkinson's disease (PD) displays substantial clinical heterogeneity with various phenotypes that translate into different combinations of both motor and non-motor symptoms. To address this heterogeneity, the age at onset (AAO) has been suggested as a key indicator associated with the clinical profile and progression of PD[1–3]. Previous studies with cross-sectional design have identified later AAO to be related with a stronger motor as well as non-motor impairment suggesting that late AAO is associated with higher progression rate of motor symptoms and cognitive decline. Conversely, early onset PD has been reported to show a specific disease profile with higher rate of motor complications such as early dyskinesia and dystonia[4–6]. Furthermore, both prospective[7] and retrospective studies with autopsy-proven PD[8] have shown similar findings, but given the heterogeneity of the study designs and various cut-offs used for categorising AAO, the reproducibility of the findings is limited. Despite reporting multiple AAO-related phenotypic differences, no study so far has endeavoured to integrate the effect of the physiological ageing process. Therefore, the associations between AAO and severity of PD phenotypes require further analysis.

Apart from AAO, the concept of polygenic risk scores (PRS) in sporadic forms of PD has recently been established to assess the complex genetic architecture of PD beyond known rare familial forms of PD with Mendelian inheritance of mutations in disease-causing genes[9]. Even though PRS were reported to be significantly negatively correlated with AAO[10], potential effects of PRS on the disease severity and the phenotypic profile have not yet been explored in detail.

Previous studies focusing on the role of AAO in PD were limited by (i) not addressing the concomitant effect of the physiological ageing process on the clinical phenotype by modelling age-related effects in a healthy control group, (ii) including relatively small numbers of PD patients from highly specific subgroups (e.g. drug naïve), (iii) using different AAO cut-offs across the studies and (iv) lacking a detailed genetic profiling of the study sample to exclude individuals with monogenic forms and variants presenting a genetic risk factor for developing PD. Therefore, our study addresses these issues by combining a mono-centric idiopathic PD dataset and healthy control group (HC) with detailed genetic data with the aim (i) to investigate the effect of AAO on clinical phenotype in idiopathic PD, (ii) to separate the PD-related ageing effect from the natural ageing effect and finally (iii) to explore the effect of the genetic background reflected by PRS on the disease severity in idiopathic PD.

RESULTS

Effect of AAO on clinical outcomes in PD

Several traits in PD phenotypic profiles were found in association with AAO. An overview of clinical outcomes, sociodemographic characteristics and comorbidities among participants of the Luxembourg Parkinson's Study is shown in Tables 1 and 2. As expected, the PD group comprised more males than females (67% vs. 33%) with mean AAO of 61.8 ± 12.0 years and mean disease duration since diagnosis of 5.5 ± 5.5 years. The mean age at assessment (AAA) was 67.3 ± 11.0 years. To investigate the effects of AAO on the clinical outcomes, a multiple regression analysis adjusting for disease duration was performed with results shown in Fig. 1. The overall motor disease severity as reflected by modified H&Y, MDS-UPDRS III, frequency of falls and gait disorder were all significantly positively associated with AAO. With regard

[1]Clinical and Experimental Neuroscience, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg. [2]Parkinson's Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg, Luxembourg. [3]Biomedical Data Science Group, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg. [4]Bioinformatics Core, Luxembourg Centre for Systems Biomedicine (LCSB), Esch-sur-Alzette, Luxembourg. [5]Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg. *A list of authors and their affiliations appears at the end of the paper. ✉email: lukas.pavelka@uni.lu; rejko.krueger@uni.lu

Published in partnership with the Parkinson's Foundation

Table 1. Overview of sociodemographic characteristics of study dataset including comorbidities and polygenic risk score with *p* values from Mann–Whitney *U* test for numerical variables and Fisher's exact test for binary variables.

	HC $n = 556$			PD = 430			
Demographic, PRS and comorbidities	Mean or YES in %	SD or NO/YES	n.a.	Mean or YES in %	SD or NO/YES	n.a.	*p* value
Gender (male)*	56%	243/313	0	67%	142/288	0	7.8e−04″
Age at onset (years)	–	–	556	61.84	11.99	0	–
Age at assessment (years)	59.61	11.78	0	67.30	11.04	0	6.8e−23″
Disease duration since diagnosis (years)	–	–	556	5.49	5.54	0	–
Years of education	14.27	3.88	5	13.09	4.10	0	5.4e−06″
Family history of parkinsonism*	26%	408/146	2	25%	324/106	0	5.6e−01
Family history of dementia*	32%	373/178	5	24%	325/103	2	5.4e−03′
Polygenic risk score for PD	−0.21	0.91	6	0.16	0.94	6	7.1e−09″
De novo*	–	0/0	556	8%	395/35	0	–
Treatment with DBS*	0%	556/0	0	5%	410/20	0	4.8e−08″
History or presence of RLS*	6%	520/36	0	9%	392/38	0	1.8e−01
Diabetes (type not specified)*	6%	523/33	0	10%	385/45	0	1.2e−02′
Arterial hypertension*	33%	375/181	0	44%	239/191	0	1.6e−04″
Cardiovascular disease*	9%	504/52	0	21%	340/90	0	3.8e−07″
Hypercholesterolemia*	38%	347/209	0	42%	248/182	0	1.5e−01
History of stroke*	3%	539/17	0	5%	410/20	0	2.4e−01

Single and double ticks indicate significance at the 5% level and the Bonferroni-adjusted 5% level respectively. The binary variables are annotated by asterisk. *n.a.* corresponds to total number of missing values per variable, *PD* Parkinson's disease, *PRS* Polygenic risk score, *HC* Healthy controls, *DBS* Deep brain stimulation, *SD* Standard deviation.

to the motor complications of PD, no significant association of AAO was found with total hours of dyskinesia/day, dystonia/day, nor OFF time/day, however, a significant negative association of AAO with the MDS-UPDRS IV total score was identified. Additionally, SCOPA-AUT total score and Starkstein Apathy scale had significant positive associations with AAO indicating that patients with higher AAO experience more non-motor symptoms including urinary incontinence. Cognition as reflected by the MoCA score was significantly negatively associated with AAO showing higher impairment in patients with an older AAO. Similarly, AAO was significantly negatively associated with olfactory dysfunction. All other putative associations were not significantly associated with AAO as shown in Fig. 1.

Analysing the difference in ageing effect in PD vs HC

When investigating the effects of AAA and AAO on the clinical phenotypes of PD, all associations were found to be comparable in both models (cf. Table 3). The reason is the strong correlation between AAA and AAO (statistically significant Kendall's tau $\rho = 0.73$, see Supplementary Fig. 1). To investigate an effect of physiological ageing on the PD phenotypes, we also included the HC group into the regression models. When investigating the ageing-associated effects in PD, we determined a significant positive association in PD between AAA and H&Y, MDS-UPDRS III, frequency of falls and urine incontinence, SCOPA-AUT, Starkstein Apathy Scale as well as significant negative association between AAA and MoCA and Sniffin' Stick test (cf. Table 3). Similarly in the HC group, we found a significant positive association between AAA and MDS-UPDRS III, SCOPA-AUT, Starkstein Apathy Scale, frequency of urine incontinence and gait disorder as well as significant negative association between AAA and MoCA and Sniffin' Stick test as demonstrated in Table 4. Surprisingly, after comparing the ageing effect between PD vs HC (i.e. comparing effect of AAA on the clinical variables; see Table 5, column AAA:status), the only significant differences between PD and HC were found for H&Y, MDS-UPDRS III, MDS-UPDRS IV and MoCA indicating that the concomitant ageing process might be the main determinant of the non-motor PD phenotypic differences when studying the isolated effect of age in PD.

Correlation between AAO and PRS and its effect on severity of the PD phenotype

Using a polygenic risk score defined by the imputed genotypic data from the Luxembourg Parkinson's Study and the summary statistics of 90 single nucleotide polymorphisms (SNP) that were previously identified to be genome-wide significantly associated with PD risk, we identified a significant negative correlation between PRS and AAO as shown in Fig. 2. However, neither Kendall's tau correlation test for continuous variables nor Mann–Whitney *U* test for binary variables estimating the effect of PRS on clinical outcomes nor multiple regression models including PRS adjusted for AAA and disease duration showed effects of PRS on the severity of the clinical phenotype as demonstrated in Tables 6 and 7 respectively.

DISCUSSION

The presented cross-sectional analysis of PD patients and HC at the baseline clinical visit uses data from one of the largest ongoing observational studies, focusing on PD with demographic and clinical parameters corresponding closely to other recently published large PD datasets[11–13]. In our study, we have identified several significant associations of different PD-associated motor and non-motor symptoms with AAO using a comprehensive set of clinical assessments. This is in line with previous cross-sectional, retrospective and prospective studies suggesting that later onset PD is associated with a more rapid progression rate of motor symptoms[4,11,14,15]. Conversely, comparing to the Cardiff community-based PD longitudinal cohort[16] and the longitudinal study at the Movement Disorders Clinic Saskatchewan[4], both demonstrating higher frequency of dyskinesia, motor fluctuations and dystonia in the younger onset groups vs. older onset groups,

Table 2. Overview of dataset with clinical variables in healthy control group (HC) and Parkinson's disease patients (PD) with *p* values from Mann–Whitney *U* test for numerical variables and Fisher's exact test for binary variables.

	HC $n = 556$			PD = 430			
Clinical symptoms and scales	Mean or YES in %	SD or NO/YES	n.a.	Mean or YES in %	SD or NO/YES	n.a.	*p* value
H&Y	0.00	0.00	2	2.24	0.81	2	1.5e−196″
MDS-UPDRS III	3.45	4.76	6	34.70	17.02	9	3.1e−150″
MDS-UPDRS II	1.21	2.37	6	11.69	8.32	8	6.2e−126″
LEDD (g/day)	0.0035	0.037	0	0.53	0.42	0	7.5e−160″
Gait disorder*	2%	546/10	0	57%	185/245	0	6.9e−97″
Repetitive falls*	1%	552/4	0	18%	351/79	0	1.1e−25″
MDS-UPDRS IV	0.00	0.00	4	1.88	3.52	5	1.4e−43″
Dyskinesia/day (hours)	0.00	0.00	0	0.69	2.73	1	1.2e−21″
OFF time/day (hours)	0.00	0.00	0	0.53	1.44	2	3.3e−34″
Dystonia/day (hours)	0.00	0.00	0	0.048	0.22	2	6.8e−12″
Dyskinesia*	0%	556/0	0	13%	375/55	0	1.9e−21″
Motor fluctuations*	0%	556/0	0	17%	357/73	0	1.1e−28″
Freezing of gait*	0%	556/0	0	23%	331/99	0	1.6e−39″
MoCA	27.03	2.55	3	24.28	4.41	8	4.4e−28″
Sniffin' stick test	12.86	2.39	2	8.03	3.41	13	1.6e−94″
PDQ-39	10.31	13.20	16	39.69	26.31	39	1.3e−80″
SCOPA-AUT	7.34	5.81	16	14.82	8.02	21	6.9e−53″
MDS-UPDRS I	4.58	4.42	9	10.47	7.04	10	9.0e−51″
BDI-I	5.29	5.03	15	9.97	7.11	23	2.3e−30″
Starkstein Apathy Scale	9.41	4.71	16	13.93	5.70	26	8.2e−35″
PDSS	122.81	19.61	13	105.17	23.85	28	2.3e−34″
Probable RBD*	8%	496/42	18	24%	305/95	30	1.2e−11″
Excessive daily sleepiness*	3%	541/15	0	30%	299/131	0	2.4e−36″
Insomnia*	8%	514/42	0	24%	327/103	0	9.1e−13″
Hallucinations*	0%	554/2	0	17%	357/73	0	1.2e−25″
Impulse Control Disorder*	0%	555/1	0	9%	392/38	0	1.8e−13″
Orthostatic hypotension*	6%	525/31	0	27%	312/118	0	8.2e−22″
Dysphagia*	1%	552/4	0	25%	323/107	0	5.6e−37″
Constipation*	5%	528/28	0	42%	250/180	0	3.6e−47″
Urinary Incontinence*	5%	530/26	0	32%	293/137	0	3.6e−31″

Single and double ticks indicate significance at the 5% level and the Bonferroni-adjusted 5 % level respectively.
Clinical symptoms and scales are described in Supplementary Material. The binary variables are annotated by asterisk.
n.a. (not acquired) corresponds to total number of individuals with missing value, *SD* Standard deviation.

we could not identify such associations with AAO. Only an overall burden of motor complications reflected by MDS-UPDRS IV score was significantly negatively associated with AAO in our study. The significant positive association of olfactory dysfunction and significant negative association of cognitive performance with AAO observed in our study correlate with previous findings[17,18] and in terms of cognitive impairment it might point to a decreased ability of senescent brain to cope with the pathological neurodegenerative process known as cognitive resilience[19]. Additionally, another large multi-centric study using the Quebec Parkinson Network (QPN) dataset of over 1000 PD individuals showed comparable results with a positive association between late-onset PD and higher motor burden reflected by H&Y, higher cognitive decline and higher frequency of falls, but differed on significantly higher frequency of constipation and hallucinations late-onset PD (defined as AAO > 50 years) compared to early onset PD[11]. However, most scales applied in QPN differ from our study and different categorical approaches were used in QPN both for AAO and disease duration, influencing the comparability of results. To summarise our results, the earlier AAO, patients experience a lower level of motor impairment, lower cognitive impairment and less global autonomic dysfunction, apathy and olfactory deficit, but present with more motor complications even after adjusting for disease duration as a main determinant of disease severity.

These phenotypic differences observed in PD based on different AAO were previously not clearly separated from the physiological ageing process and challenged the concept that phenotypic differences are related specifically to the age at which the disease first manifests. This intriguing aspect evolves from the inherent close correlation between the main co-variates (AAA, AAO and disease duration) and thus raises a major methodological concern in most of the cross-sectional studies when aiming at determining the effect of all three co-variates on the clinical outcomes in a single model as discussed by Johnson et al. 2002[20]. Therefore, we tried to disentangle the effect of ageing on the clinical phenotype in the cross-sectional setting by determining the ageing effect in individuals with and without PD. Surprisingly, the effect of ageing (AAA) on clinical outcomes in PD vs HC differed significantly only in motor disease severity (H&Y, MDS-UPDRS III), motor complications (MDS-UPDRS IV) and cognitive performance. These results suggest

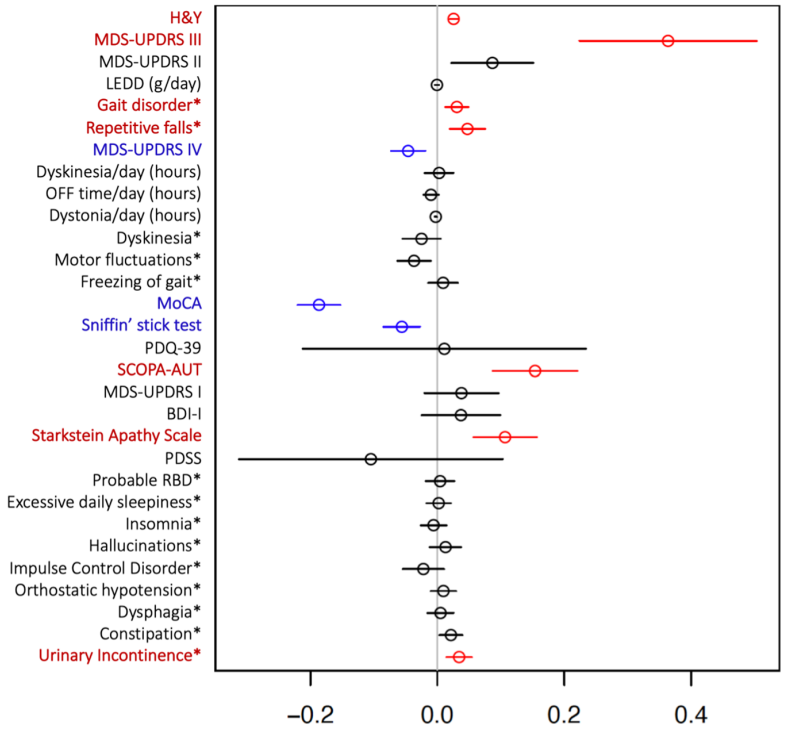


Fig. 1 Forrest plot with estimated coefficients and corresponding confidence intervals (±1.96 × standard error) for AAO, from linear/logistic regression of numerical/binary outcome on disease duration and AAO. The colour blue indicates significant negative effects of AAO on the clinical outcome, and the colour red indicates significant positive effects at the Bonferroni-adjusted 5% level. The binary variables are annotated by asterisk. Clinical symptoms and scales are described in Supplementary Material.

that the majority of the observed significant non-motor phenotypic differences in PD should be attributed rather to the physiological ageing process itself than age-specific dynamics of PD.

When considering the effect and role of AAO and age in classification of the respective PD phenotypes, potential underlying genetic determinants need to be considered. It is well known that rare disease-causing mutations in monogenic PD (e.g. in PARKIN, PINK1, SNCA or GBA[21–23]) have an effect on both AAO and PD phenotype. However, until now only few studies have explored the cumulative effects of common genetic variants with small effect sizes (as defined by PRS) on the clinical phenotype[24]. Here our results are in line with several recent studies observing no significant association between PRS and cognitive decline, severity of motor symptoms[25] or ICD[26] in contrast to other longitudinal prospective study[27]. It is worth noting that our statistical models included individuals without any known PD causing monogenic mutation or genetic risk variant (i.e. PD-associated variants in the GBA gene). Nevertheless, the significance of the PRS effect on clinical outcomes did not change in the models including PD-associated mutation or genetic variant carriers. Together with the significant negative correlation between AAO and PRS (cf. Fig. 2), our findings suggest that PRS may increase the risk to develop PD but might not have an effect on the severity of the disease phenotype. This observation is in favour of the hypothesis that initiation of the disease on one hand and the disease progression rate on the other might be driven by distinct factors.

Besides the mentioned strengths of our study design, several limitations need to be considered. First, the cross-sectional design does not allow for the identification of causal relations between AAO and clinical phenotypes. Second, we cannot consider the Luxembourg Parkinson's Study as community-based by design, although some clinical indicators (such as mean AAO and male-to-female ratio) correspond closely to several community-based studies[28–31]. Third, we observe a relatively high frequency of positive family history of parkinsonism in the HC group (26% vs. 25% in PD) as well as high frequency of a family history of dementia in HC (32% vs 24%). We assume that there are two principal reasons why we observe increased frequencies of neurodegenerative diseases in HC group: (i) HC with personal experience with parkinsonism and/or dementia in their family are more aware to support research and (ii) family members of study participants are more inclined to participate in the study. To address these points and eliminate a potential bias, we excluded 1st, 2nd and 3rd degree relatives from our statistical models.

In summary, our study sought to overcome limitations identified in previous studies on the role of AAO in PD by (i) including substantially higher number of PD patients and HCs in the model accounting for the independent effect of ageing, (ii) our study being based on monocentric data collection and including PD patients of all disease stages regardless of the cognitive status, (iii) investigating an idiopathic dataset of PD and PD-related mutation free HC, (iv) refuting the categorisation bias by a priori arbitrary AAO grouping, and finally (v) exploring the effect of PRS on severity of the PD phenotype in a large genotyped sample.

METHODS

Study population

All subjects were recruited from March 2015 until 10th December 2020 in the frame of the nation-wide monocentric observational longitudinal Luxembourg Parkinson's Study. The diagnosis of PD was based on

Table 3. Multiple regression of clinical outcomes on age at onset (AAO), age at assessment (AAA) and disease duration for Parkinson's disease group.

Clinical symptoms and scales	Intercept	Disease duration	AAA	Intercept	Disease duration	AAO	Intercept	AAA	AAO
H&Y	0.23	**0.05″**	**0.03″**	0.23	**0.07″**	**0.03″**	0.23	**0.07″**	**−0.05″**
MDS-UPDRS III	5.95	**0.76″**	**0.36″**	6.04	**1.13″**	**0.36″**	5.98	**1.13″**	**−0.76″**
MDS-UPDRS II	2.38	**0.63″**	0.09′	2.39	**0.72″**	0.09′	2.41	**0.72″**	**−0.63″**
LEDD (g/day)	0.37	**0.04″**	0.00	0.37	**0.03″**	0.00	0.38	**0.03″**	**−0.04″**
Gait disorder*	−2.23	**0.08″**	0.03′	−2.23	**0.12″**	0.03′	−2.22	**0.12″**	**−0.08″**
Repetitive falls*	−5.79	**0.14″**	**0.05″**	−5.74	**0.19″**	**0.05″**	−5.80	**0.19″**	**−0.15″**
MDS-UPDRS IV	3.43	**0.28″**	−0.05′	3.46	**0.24″**	−0.05′	3.43	**0.24″**	**−0.28″**
Dyskinesia/day (hours)	−0.16	**0.12″**	0.00	−0.16	**0.12″**	0.00	−0.15	**0.12″**	**−0.12″**
OFF time/day (hours)	0.87	**0.06″**	−0.01	0.87	**0.05″**	−0.01	0.87	**0.05″**	**−0.06″**
Dystonia/day (hours)	0.15	**0.01″**	0.00′	0.16	**0.01″**	0.00′	0.15	**0.01″**	**−0.01″**
Dyskinesia*	−1.57	**0.18″**	−0.02	−1.55	**0.15″**	−0.02	−1.56	**0.15″**	**−0.18″**
Motor fluctuations*	−0.35	**0.18″**	−0.04′	−0.39	**0.14″**	−0.04′	−0.33	**0.14″**	**−0.18″**
Freezing of gait*	−2.82	**0.16″**	0.01	−2.87	**0.17″**	0.01	−2.79	**0.17″**	**−0.16″**
MoCA	37.04	−0.05	**−0.19″**	37.10	**−0.24″**	**−0.19″**	37.03	**−0.23″**	**0.05**
Sniffin' stick test	12.40	**−0.11″**	**−0.06″**	12.41	**−0.17″**	**−0.06″**	12.40	**−0.17″**	**0.11″**
PDQ-39	29.72	**1.78″**	0.01	29.46	**1.79″**	0.01	29.87	**1.77″**	**−1.76″**
SCOPA-AUT	2.20	**0.43″**	**0.15″**	2.16	**0.58″**	**0.15″**	2.23	**0.58″**	**−0.42″**
MDS-UPDRS I	6.05	**0.35″**	0.04	5.96	**0.39″**	0.04	6.08	**0.38″**	**−0.35″**
BDI-I	6.14	**0.24″**	0.04	6.19	**0.28″**	0.04	6.15	**0.28″**	**−0.24″**
Starkstein Apathy Scale	6.75	−0.01	**0.11″**	6.81	0.10	**0.11″**	6.74	0.10	0.00
PDSS	117.20	**−0.98″**	−0.10	117.45	**−1.08″**	−0.10	117.11	**−1.06″**	**0.96″**
Probable RBD*	−2.13	**0.11″**	0.00	−2.14	**0.12″**	0.00	−2.12	**0.11″**	**−0.11″**
Excessive daily sleepiness*	−1.31	**0.07″**	0.00	−1.37	**0.07″**	0.00	−1.29	0.07′	**−0.06″**
Insomnia*	−1.00	0.04′	−0.01	−1.01	0.04	−0.01	−1.00	0.04	**−0.04′**
Hallucinations*	−3.08	**0.09″**	0.01	−3.06	**0.10″**	0.01	−3.08	**0.11″**	**−0.09″**
Impulse Control Disorder*	−1.62	**0.11″**	−0.02	−1.64	0.09′	−0.02	−1.60	0.09′	**−0.11″**
Orthostatic hypotension*	−1.88	0.05′	0.01	−1.92	0.06′	0.01	−1.87	0.06′	−0.05′
Dysphagia*	−1.77	0.06′	0.00	−1.80	0.06′	0.01	−1.76	**0.06′**	−0.06′
Constipation*	−2.14	**0.07″**	0.02′	−2.19	**0.10″**	0.02′	−2.13	**0.09″**	**−0.07″**
Urinary Incontinence*	−3.40	0.04′	**0.04″**	−3.36	**0.08″**	**0.03″**	−3.41	**0.08″**	−0.05′

Regression coefficients for different outcomes (rows) from three equivalent models with each two out of three features (columns). Single and double ticks indicate significance at the 5% level and the Bonferroni-adjusted 5% level respectively. The bold indicates significant effect where minus value indicates negative significant effect and positive value positive significant effect respectively. The binary variables are annotated by asterisk. Clinical symptoms and scales are described in Supplementary Material.

UKPDSBB diagnostic criteria[32]. The initial visit dataset of 430 PD patients and 556 HC genetically screened by both NeuroChip and PacBio were analysed after exclusion of 6 PD and 39 HC individuals for 1st, 2nd and 3rd degree relationships and after exclusion of 53 PD carriers and 27 HC carriers of pathogenic PD-associated variants. The overall study design, inclusion and exclusion workflow are illustrated in Fig. 3.

All participants taking part in Luxembourg Parkinson's Study agreed and signed a written informed consent. The study has been approved by the National Ethics Board (CNER Ref: 201407/13). The patients with PD were included regardless of the disease duration, cognitive status, age or disease stage. The HC were partially recruited from the pool of independent observational studies in Luxembourg (ORISCAV-LUX study; EHES-LUX) or were recruited from Luxembourg or the surrounding area of Greater Region based on individual interest not meeting any of the exclusion criteria (presence of a neurodegenerative disorder, active cancer; age under 18 and pregnant women)[33].

Clinical assessment and data. A description of the design of the Luxembourg Parkinson's Study was previously published[33]. Sociodemographic characteristics and clinical outcomes validated for PD were chosen from the basic clinical assessment battery and listed in Tables 1 and 2. Validated self-administered questionnaires and scales for PD were used. All patients have been evaluated in medication ON state and where applicable, in deep brain stimulation ON state. AAO is defined as age at diagnosis of PD. The clinical symptoms as scales are defined in detail in the Supplementary material.

Missing data statement. The absolute number of missing data per variable are shown in Tables 1 and 2. Given the low proportions of missing values in the outcome variables and 0% of missing values in the co-variates (AAA, AAO and disease duration), we used a pairwise deletion for all statistical models.

Genotyping and quality-control analyses. DNA samples were genotyped using the NeuroChip array (v.1.0 and v1.1; Illumina, San Diego, CA) that was specifically designed to integrate rare and common neurodegenerative disease-related variants[34]. Quality-control (QC) analysis was performed as follows: samples with call rates < 95% and whose genetically determined sex deviated from reported sex in clinical data were excluded from the analysis, and the filtered variants were checked for cryptic relatedness and excess of heterozygosity. Samples exhibiting excess heterozygosity (F statistic > 0.2) and first-degree relatedness were excluded. Once sample QC was completed, SNPs with Hardy–Weinberg equilibrium P value < 1E−6, and missingness rates >5% were excluded. All samples except for twelve from all individuals entering the analysis after exclusion of the 1st, 2nd and 3rd degree relatives and presence of PD-linked mutation and genetic risk factors passed the QC (424 PD and 550 HC). The data were then imputed using the Haplotype Reference Consortium r1.1 2016 and the Michigan Imputation Server and filtered for

Table 4. Simple regression of clinical outcomes with healthy controls.

Clinical symptoms and scales	Intercept	AAA
H&Y	0.00	0.00
MDS-UPDRS III	−3.70	**0.12″**
MDS-UPDRS II	−0.21	0.02′
LEDD (g/day)	−0.02	0.00′
Gait disorder*	−12.56	**0.13″**
Repetitive falls*	−5.14	0.00
MDS-UPDRS IV	0.00	0.00
Dyskinesia/day (hours)	0.00	0.00
OFF time/day (hours)	0.00	0.00
Dystonia/day (hours)	0.00	0.00
Dyskinesia*	−26.57	0.00
Motor fluctuations*	−26.57	0.00
Freezing of gait*	−26.57	0.00
MoCA	29.84	**−0.05″**
Sniffin' stick test	15.44	**−0.04″**
PDQ-39	10.68	−0.01
SCOPA-AUT	2.53	**0.08″**
MDS-UPDRS I	3.12	0.02
BDI-I	4.06	0.02
Starkstein Apathy Scale	5.11	**0.07″**
PDSS	130.37	−0.13
Probable RBD*	−1.58	−0.02
Excessive daily sleepiness*	−6.53	0.05
Insomnia*	−2.00	−0.01
Hallucinations*	−3.31	−0.04
Impulse Control Disorder*	−4.32	−0.04
Orthostatic hypotension*	−3.92	0.02
Dysphagia*	−7.28	0.04
Constipation*	−2.37	−0.01
Urinary Incontinence*	−8.43	**0.08″**

Regression coefficients are shown from linear regression of numerical outcome and from logistic regression of binary outcome on age at assessment (AAA). Single and double ticks indicate significance at the 5% level and the Bonferroni adjusted 5% level, the bold indicates significant effect where minus value indicates negative significant effect and positive value positive significant effect respectively. The binary variables are annotated by asterisk. Clinical symptoms and scales are described in Supplementary Material.

imputation quality (RSQ > 0.8)[35]. Genetic analysis and QC was done using PLINK v1.9. Additionally, all samples underwent targeted sequencing of the GBA locus using single-molecule sequencing on a Sequel II sequencer from Pacific BioScience[36]. Variants were called with DeepVariant 1.0[37]. PD causing rare variants were defined by the ClinVar classification 'pathogenic/likely-pathogenic'. All PD causing variants (listed in Supplementary material) identified by any method were Sanger validated and all samples with a validated PD causing variant were excluded from further analysis.

Polygenic risk score (PRS). We generated PRSs with PRSice-2 under default settings. PRSs for each individual were calculated using the imputed genotype data from Luxembourg Parkinson's Study as a target sample. The base GWAS data used to determine PRS for PD was the summary statistics of the 90 SNPs that were previously found to be genome-wide significantly associated with PD risk[38]. The criteria for linkage disequilibrium (LD) clumping of SNPs were pairwise LD r2 < 0.1 within the 250 kb window. Briefly, PRSs were calculated by summing the weighted effects of GWAS PD risk genetic variants present in the target samples, with a possible proxy of $R^2 > 0.9$, meeting p value thresholds ranging from 5e−08 to 0.5. The values of PRS were Z-normalised.

Table 5. Multiple regression model with PD and HC investigating the difference in effect of ageing in HC (AAA) and in PD (AAA:status) adjusted for disease duration.

Clinical symptoms and scales	Intercept	AAA	Status	Disease duration	AAA:Status
H&Y	0.00	0.00	0.23	**0.05″**	**0.03″**
MDS-UPDRS III	−3.70	0.12′	9.65	**0.76″**	**0.24″**
MDS-UPDRS II	−0.21	0.02	2.59	**0.63″**	0.06′
LEDD (g/day)	−0.02	0.00	0.39	**0.04″**	0.00
Gait disorder*	−12.56	0.13′	10.34	**0.08″**	−0.10′
Repetitive falls*	−5.14	0.00	−0.66	**0.14″**	0.04
MDS-UPDRS IV	0.00	0.00	3.43	**0.28″**	**−0.05″**
Dyskinesia/day (hours)	0.00	0.00	−0.16	**0.12″**	0.00
OFF time/day (hours)	0.00	0.00	0.87	**0.06″**	−0.01
Dystonia/day (hours)	0.00	0.00	0.15	**0.01″**	0.00′
Dyskinesia*	−20.57	0.00	18.99	**0.18″**	−0.02
Motor fluctuations*	−20.57	0.00	20.22	**0.18″**	−0.04
Freezing of gait*	−20.57	0.00	17.75	**0.16″**	0.01
MoCA	29.84	**−0.05″**	7.20	−0.05	**−0.14″**
Sniffin' stick test	15.44	**−0.04″**	−3.03	**−0.11″**	−0.01
PDQ-39	10.68	−0.01	19.04	**1.78″**	0.01
SCOPA-AUT	2.53	0.08′	−0.33	**0.43″**	0.07
MDS-UPDRS I	3.12	0.02	2.94	**0.35″**	0.01
BDI-I	4.06	0.02	2.09	**0.24″**	0.02
Starkstein Apathy Scale	5.11	**0.07″**	1.64	−0.01	0.04
PDSS	130.37	−0.13	−13.17	**−0.98″**	0.03
Probable RBD*	−1.58	−0.02	−0.54	**0.11″**	0.02
Excessive daily sleepiness*	−6.53	0.05	5.22	**0.07″**	−0.05
Insomnia*	−2.00	−0.01	0.99	0.04′	0.00
Hallucinations*	−3.31	−0.04	0.23	**0.09″**	0.05
Impulse Control Disorder*	−4.32	−0.04	2.70	**0.11″**	0.01
Orthostatic hypotension*	−3.92	0.02	2.04	0.05′	−0.01
Dysphagia*	−7.28	0.04	5.51	0.06′	−0.03
Constipation*	−2.37	−0.01	0.23	**0.07″**	0.03
Urinary Incontinence*	−8.43	**0.08″**	5.03	0.04′	−0.05′

Regression coefficients are shown for different outcomes (rows). Status takes the value 0 for HC and 1 for PD, the AAA:status is the interaction term of AAA and being PD (status = 1). Single and double ticks indicate significance at the 5% level and the Bonferroni-adjusted 5% level respectively, the bold indicates significant effect where minus value indicates negative significant effect and positive value positive significant effect respectively. The column AAA:status indicates whether the effect of AAA on clinical outcomes differs between PD and HC. The binary variables are annotated by asterisk. Clinical symptoms and scales are described in Supplementary Material.

Statistical analysis

Firstly, we performed an intergroup comparison (PD vs HC) of socio-demographic and clinical characteristics as well as polygenic risk score and comorbidities with the Mann−Whitney U test for numerical variables and

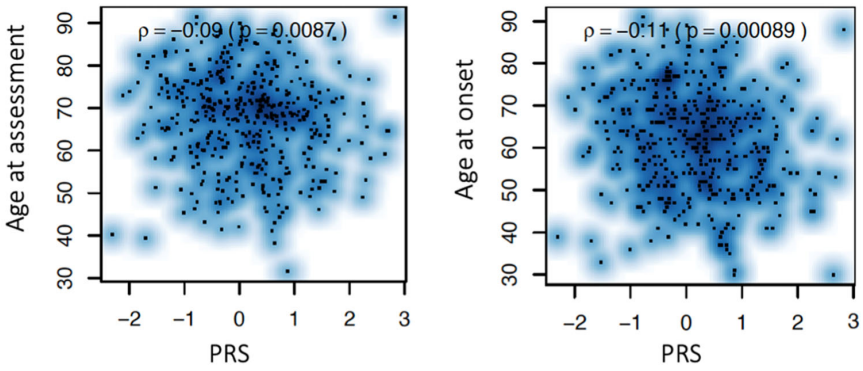


Fig. 2 Pairwise association between age at onset (AAO), age at assessment (AAA) (y-axis) and polygenic risk score (PRS) (x-axis) with Kendall correlation coefficient. Significant inverse association was determined between AAO and PRS and AAA and PRS indicating the younger the AAO of PD, the higher cummulative burden of small effect size variants (represented by PRS).

Table 6. Kendall correlation coefficient between clinical outcome (row) and polygenic risk score (PRS) for healthy controls (HC) (left) and Parkinson's disease patients (PD) (right), with annotation by bold indicating significant effect where minus value indicates negative significant effect and positive value positive significant effect respectively (Kendall correlation test).

Clinical symptoms and scales	HC	PD
H&Y	0.0272	0.0272
MDS-UPDRS III	−0.0088	−0.0341
MDS-UPDRS II	−0.0242	0.0058
LEDD (g/day)	0.0091	−0.0147
Gait disorder*	0.4070	−0.0177
Repetitive falls*	0.6209	0.1690
MDS-UPDRS IV	0.0625	0.0625
Dyskinesia/day (hours)	0.0576	0.0576
OFF time/day (hours)	0.0100	0.0100
Dystonia/day (hours)	0.0491	0.0491
Dyskinesia*	–	0.1426
Motor fluctuations*	–	0.0864
Freezing of gait*	–	0.0084
MoCA	0.0542	−0.0493′
Sniffin' stick test	0.1012	0.0576
PDQ-39	−0.0519	0.0386
SCOPA-AUT	−0.0113	0.0150
MDS-UPDRS I	−0.0556	0.0105
BDI-I	−0.0155	0.0290
Starkstein Apathy Scale	−0.0425	−0.0068
PDSS	−0.0128	−0.0332
Probable RBD*	−0.1142	−0.0703
Excessive daily sleepiness*	0.1638	−0.1261
Insomnia*	−0.0356	0.1387
Hallucinations*	−0.9806	−0.1225
Impulse Control Disorder*	−2.1177	−0.0872
Orthostatic hypotension*	0.0839	−0.1331
Dysphagia*	−0.6401	−0.2650′
Constipation*	0.2701	−0.0740
Urinary Incontinence*	0.1011	0.0434

Single and double ticks indicate significance at the 5% level and the Bonferroni-adjusted 5% level. The binary variables are annotated by asterisk. Clinical symptoms and scales are described in Supplementary Material.

Table 7. Multiple regression model with coefficients shown from linear regression of numerical outcome and from logistic regression of binary outcome on disease duration, age at assessment (AAA) and polygenic risk score (PRS) in PD group.

Clinical symptoms and scales	Intercept	Disease duration	AAA	PRS
H&Y	0.17	**0.05″**	**0.03″**	−0.02
MDS-UPDRS III	4.44	**0.76″**	**0.39″**	−0.57
MDS-UPDRS II	1.81	**0.63″**	0.10′	−0.47
LEDD (g/day)	0.32	**0.04″**	0.00	0.01
Gait disorder*	−2.29	**0.09″**	0.03′	−0.04
Repetitive falls*	−5.91	**0.14″**	0.05′	0.15
MDS-UPDRS IV	3.37	**0.28″**	−0.04′	0.03
Dyskinesia/day (hours)	−0.36	**0.12″**	0.01	0.04
OFF time/day (hours)	0.87	**0.06″**	−0.01	−0.04
Dystonia/day (hours)	0.15	**0.01″**	0.00′	0.01
Dyskinesia*	−1.62	**0.18″**	−0.02	−0.05
Motor fluctuations*	−0.12	**0.18″**	−0.04′	−0.12
Freezing of gait*	−2.66	**0.17″**	0.01	−0.13
MoCA	37.25	−0.07′	**−0.19″**	0.40′
Sniffin' stick test	12.64	**−0.11″**	**−0.06″**	−0.22
PDQ-39	30.96	**1.86″**	−0.01	−2.98′
SCOPA-AUT	2.74	**0.47″**	**0.15″**	−0.98′
MDS-UPDRS I	6.29	**0.36″**	0.03	−0.84′
BDI-I	6.81	**0.27″**	0.03	−0.76′
Starkstein Apathy Scale	6.99	0.00	**0.10″**	−0.18
PDSS	116.69	**−1.00″**	−0.10	0.12
Probable RBD*	−2.07	**0.12″**	0.00	−0.21
Excessive daily sleepiness*	−1.23	**0.07″**	0.00	−0.20
Insomnia*	−1.04	0.04′	−0.01	0.11
Hallucinations*	−2.95	**0.10″**	0.01	−0.23
Impulse Control Disorder*	−1.37	**0.11″**	−0.03	−0.25
Orthostatic hypotension*	−1.64	0.06′	0.01	−0.20
Dysphagia*	−1.77	**0.07″**	0.00	−0.37′
Constipation*	−2.07	**0.08″**	0.02′	−0.13
Urinary Incontinence*	−3.67	0.05′	**0.04″**	0.06

Single and double ticks indicate significance at the 5% level and the Bonferroni adjusted 5% level, the bold indicates significant effect where minus value indicates negative significant effect and positive value positive significant effect respectively. The binary variables are annotated by asterisk. Clinical symptoms and scales are described in Supplementary Material.

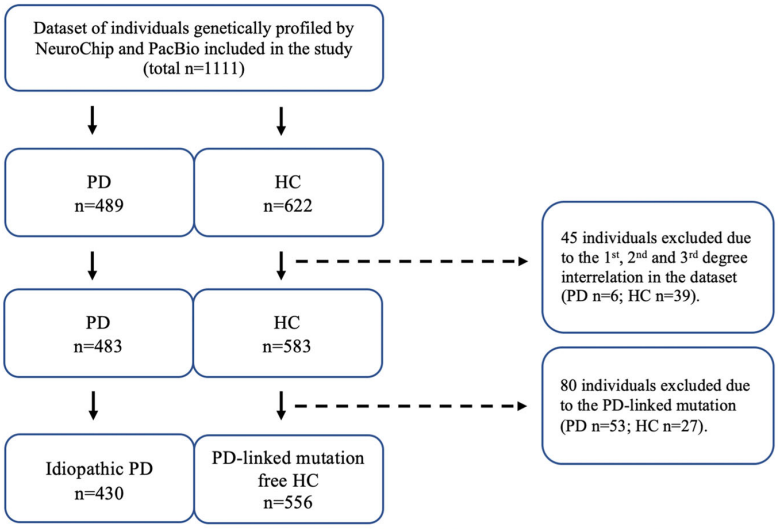


Fig. 3 Description of the study design and study dataset. PD individuals with Parkinson's disease, HC healthy control.

Fisher's exact test for binary variables (Tables 1 and 2). Secondly, we used multiple regression models (linear and logistic) to identify effects of AAO (as a numerical variable) on numerical or binary clinical outcomes accounting for disease duration (Fig. 1). Subsequently, we performed a multiple regression model for both HC and PD (Table 5) to examine whether the effect of ageing (AAA) on clinical outcomes differs between HC and PD adjusted for disease duration. For this, we included the main effects of the continuous variable AAA and the binary variable status (HC: status = 0, PD: status = 1), their interaction effect (HC: status*AAA = 0, PD: status*AAA > 0), and the main effect of the continuous variable disease duration (HC: duration = 0, PD: duration > 0). To investigate the role of PRS in PD, a pairwise association analysis with Kendall's tau correlation test between PRS and AAO and AAA was performed (Fig. 2). Furthermore, we performed a Kendall correlation test between PRS and clinical outcome for PD and HC respectively (Table 6). As a last step, we employed a multiple regression model including PRS adjusting for AAA and disease duration, to investigate the effect of PRS on the clinical phenotype in PD (Table 7). At all instances, the significance at the 5% level and the Bonferroni-adjusted 5% level was set.

DATA AVAILABILITY

The dataset for this manuscript is not publicly available as it is linked to the Luxembourg Parkinson's Study and its internal regulations. Any requests for accessing the dataset can be directed to *request.ncer-pd@uni.lu*.

CODE AVAILABILITY

The code for the statistical models is available at: https://doi.org/10.17881/hr67-ba06.

Received: 30 January 2022; Accepted: 30 May 2022;
Published online: 09 August 2022

REFERENCES

1. Elizabeth, Qian & Yue, Huang Subtyping of Parkinson's disease - where are we up to? *Aging Dis.* **10**, 1130–1139 (2019).
2. Rodríguez-Violante, M., Cervantes-Arriaga, A., Fahn, S. & Tolosa, E. Two-hundred years later: Is Parkinson's disease a single defined entity? *Rev. Invest Clin.* **69**, 308–313 (2017).
3. Krüger, R. et al. Classification of advanced stages of Parkinson's disease: translation into stratified treatments. *J. Neural Transm.* **124**, 1015–1027 (2017).
4. Pagano, G., Ferrara, N., Brooks, D. J. & Pavese, N. Age at onset and Parkinson disease phenotype. *Neurology* **86**, 1400–1407 (2016).
5. Jellinger, K. A. Parkinson disease with old-age onset. *Arch. Neurol.* **60**, 1814–1815 (2003).
6. Zhou, M. Z. et al. The association between non-motor symptoms in Parkinson's disease and age at onset. *Clin. Neurol. Neurosurg.* **115**, 2103–2107 (2013).
7. Ferguson, L. W., Rajput, A. H. & Rajput, A. Early-onset vs. Late-onset Parkinson's disease: a clinical-pathological study. *Can. J. Neurol. Sci.* **43**, 113–119 (2016).
8. Jellinger, K. A. Very old onset parkinsonism: a clinical-pathological study. *Parkinsonism Relat. Disord.* **57**, 39–43 (2018).
9. Ohnmacht, J., May, P., Sinkkonen, L. & Krüger, R. Missing heritability in Parkinson's disease: the emerging role of non-coding genetic variation. *J. Neural Transm.* **127**, 729–748 (2020).
10. Escott-Price, V., International Parkinson's Disease Genomics Consortium & Nalls, M. A. et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann. Neurol.* **77**, 582–591 (2015).
11. Gan-Or, Z. et al. The Quebec Parkinson network: a researcher-patient matching platform and multimodal biorepository. *J. Parkinsons Dis.* **10**, 301–313 (2020).
12. Santos García, D. et al. COPPADIS-2015 (COhort of Patients with PArkinson's DIsease in Spain, 2015): an ongoing global Parkinson's disease project about disease progression with more than 1000 subjects included. Results from the baseline evaluation. *Eur. J. Neurol.* **26**, 1399–1407 (2019).
13. Marek, K. et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann. Clin. Transl. Neurol.* **5**, 1460–1477 (2018).
14. Diederich, N. J., Moore, C. G., Leurgans, S. E., Chmura, T. A. & Goetz, C. G. Parkinson disease with old-age onset: a comparative study with subjects with middle-age onset. *Arch. Neurol.* **60**, 529–533 (2003).
15. Wickremaratchi, M. M. et al. The motor phenotype of Parkinson's disease in relation to age at onset. *Mov. Disord.* **26**, 457–463 (2011).
16. Mehanna, R., Moore, S., Hou, J. G., Sarwar, A. I. & Lai, E. C. Comparing clinical features of young onset, middle onset and late onset Parkinson's disease. *Parkinsonism Relat. Disord.* **20**, 530–534 (2014).
17. Chase, B. A. & Markopoulou, K. Olfactory dysfunction in familial and sporadic Parkinson's Disease. *Front Neurol.* **11**, 447 (2020).
18. Kim, R., Shin, J. H., Park, S., Kim, H. J. & Jeon, B. Longitudinal evolution of non-motor symptoms according to age at onset in early Parkinson's disease. *J. Neurol. Sci.* **418**, 117157 (2020).
19. Pettigrew, C. & Soldan, A. Defining cognitive reserve and implications for cognitive aging. *Curr. Neurol. Neurosci. Rep.* **19**, 1 (2019).
20. Johnson, S. B. & Meltzer, L. J. Disentangling the effects of current age, onset age, and disease duration: parent and child attitudes toward diabetes as an exemplar. *J. Pediatr. Psychol.* **27**, 77–86 (2002).
21. Blauwendraat, C. et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia [published correction appears in Brain]. *Brain* **143**, 234–248 (2020).
22. Yahalom, G. et al. Age at onset of Parkinson's disease among Ashkenazi Jewish patients: contribution of environmental factors, LRRK2 p.G2019S and GBA p.N370S mutations. *J. Parkinsons Dis.* **10**, 1123–1132 (2020).
23. Hernandez, D. G., Reed, X. & Singleton, A. B. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J. Neurochem.* **139**, 59–74 (2016).
24. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

25. Liu, G. et al. Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease. *Nat. Genet* **53**, 787–793 (2021).
26. Ihle, J. et al. DIGPD study group; Parkinson's disease polygenic risk score is not associated with impulse control disorders: a longitudinal study. *Parkinsonism Relat. Disord.* **75**, 30–33 (2020).
27. Paul, K. C., Schulz, J., Bronstein, J. M., Lill, C. M. & Ritz, B. R. Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. *JAMA Neurol.* **75**, 360–366 (2018).
28. Alves, G. et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *J. Neurol. Neurosurg. Psychiatry* **80**, 851–857 (2009).
29. Granieri, E. et al. Parkinson's disease in Ferrara, Italy, 1967 through 1987. *Arch. Neurol.* **48**, 854–857 (1991).
30. Winter, Y. et al. Incidence of Parkinson's disease and atypical parkinsonism: Russian population-based study. *Mov. Disord.* **25**, 349–356 (2010).
31. Savica, R., Grossardt, B. R., Rocca, W. A. & Bower, J. H. Parkinson disease with and without Dementia: a prevalence study and future projections. *Mov. Disord.* **33**, 537–543 (2018).
32. Litvan, I. et al. SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders. *Mov. Disord.* **18**, 467–486 (2003).
33. Hipp, G. et al. The Luxembourg Parkinson's study: a comprehensive approach for stratification and early diagnosis. *Front Aging Neurosci.* **10**, 326 (2018).
34. Blauwendraat, C. et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol. Aging* **57**, 247.e9–247.e13 (2017).
35. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet* **48**, 1284–1287 (2016).
36. Korlach, Jonas et al. Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol.* **472**, 431–455 (2010).
37. Poplin, R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).
38. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).

ACKNOWLEDGEMENTS

We would like to give special thanks to all participants in the Luxembourg Parkinson's Study. Furthermore, we acknowledge the joint effort of the NCER-PD consortium members generally contributing to the Luxembourg Parkinson's Study as listed below. The work presented here was supported within the framework of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) funded by the Luxembourg National Research Fund (FNR/NCER13/BM/11264123), the PEARL program (FNR/P13/6682797 to R.K.), MotaSYN (12719684 to R.K.), MAMaSyn (to R.K.), MiRisk-PD (C17/BM/11676395 to R.K., E.G., P.M.), the FNR/DFG Core INTER (ProtectMove, FNR11250962 to P.M.), and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to R.K. and S.P.), FNR PD-Strat (INTER/11651464 to EG), DIGIPD (ERAPERMED 2020-314 to EG).

AUTHOR CONTRIBUTIONS

L.P.: Conceived, organised, and executed the research project; co-executed the statistical analysis and interpretation of results; wrote the manuscript; substantially participated in data collection, data exportation and data curation. R.K.: Conceived, organised, and co-executed the research project; participated in interpretation of results; critically revised the manuscript. A.R.: Executed the statistical analysis and interpretation, critically revised the manuscript, substantially participated in data curation. E.G.: Co-executed the research project; participated in interpretation of results; critically revised the manuscript. P.M.: Executed the genetic analysis; participated in interpretation of results; critically revised the manuscript. S.P.: Contributed genetic data; co-executed the research project; critically revised the manuscript. Z.L.: Executed the genetic analysis; participated in interpretation of results; critically revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41531-022-00342-7.

Correspondence and requests for materials should be addressed to L. Pavelka or R. Krüger.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ON BEHALF OF THE NCER-PD CONSORTIUM

Geeta Acharya[6], Gloria Aguayo[6], Myriam Alexandre[6], Muhammad Ali[7], Dominic Allen[6], Wim Ammerlann[6], Rudi Balling[7], Michele Bassis[7], Katy Beaumont[6], Regina Becker[7], Camille Bellora[6], Guy Berchem[8], Daniela Berg[9], Alexandre Bisdorff[10], Kathrin Brockmann[9], Jessica Calmes[6], Lorieza Castillo[6], Gessica Contesotto[6], Nico Diederich[8], Rene Dondelinger[10], Daniela Esteves[6], Guy Fagherazzi[6], Jean-Yves Ferrand[6], Manon Gantenbein[6], Thomas Gasser[9], Piotr Gawron[7], Soumyabrata Ghosh[7], Enrico Glaab[7], Clarissa Gomes[7], Elisa Gómez De Lope[7], Nikolai Goncharenko[6], Jérôme Graas[6], Mariella Graziano[11], Valentin Groues[7], Anne Grünewald[7], Wei Gu[7], Gaël Hammot[6], Anne-Marie Hanff[6], Linda Hansen[7,8], Maxime Hansen[7,8], Michael Heneka[7], Estelle Henry[6], Sylvia Herbrink[12], Eve Herenne[6], Sascha Herzinger[7], Michael Heymann[6], Michele Hu[13], Alexander Hundt[6], Nadine Jacoby[14], Jacek Jaroslaw Lebioda[7], Yohan Jaroz[7], Quentin Klopfenstein[7], Rejko Krüger[6,7,8], Pauline Lambert[6], Zied Landoulsi[7], Roseline Lentz[15], Inga Liepelt[9], Robert Liszka[16], Laura Longhino[8], Victoria Lorentz[6], Paula Cristina Lupu[6], Clare Mackay[17], Walter Maetzler[18], Katrin Marcus[19], Guilherme Marques[6], Tainá Marques[7], Patrick May[7], Deborah Mcintyre[6], Chouaib Mediouni[6], Francoise Meisch[7], Myriam Menster[6], Maura Minelli[6], Michel Mittelbronn[7,20], Brit Mollenhauer[21], Kathleen Mommaerts[6], Carlos Moreno[7], Serge Moudio[7], Friedrich Mühlschlegel[20], Romain Nati[8], Ulf Nehrbass[6], Sarah Nickels[7], Beatrice Nicolai[8], Jean-Paul Nicolay[22], Wolfgang Oertel[23], Marek Ostaszewski[7], Sinthuja Pachchek[7], Claire Pauly[7,8], Laure Pauly[7], Lukas Pavelka[7,8], Magali Perquin[6], Roslina Ramos Lima[6], Armin Rauschenberger[7], Rajesh Rawal[7], Dheeraj Reddy Bobbili[7], Eduardo Rosales[6], Isabel Rosety[7], Kirsten Rump[7], Estelle Sandt[6], Venkata Satagopam[7], Marc Schlesser[8], Margaux Schmitt[6], Sabine Schmitz[7], Reinhard Schneider[7], Jens Schwamborn[7], Amir Sharify[6], Ekaterina Soboleva[7], Kate Sokolowska[6], Olivier Terwindt[7,8], Hermann Thien[6], Elodie Thiry[8], Rebecca Ting Jiin Loo[7], Christophe Trefois[7], Johanna Trouet[6], Olena Tsurkalenko[6], Michel Vaillant[6], Mesele Valenti[6], Liliana Vilas Boas[8], Maharshi Vyas[7], Richard Wade-Martins[24] and Paul Wilmes[7]

[6]Luxembourg Institute of Health, Strassen, Luxembourg. [7]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [8]Centre Hospitalier de Luxembourg, Strassen, Luxembourg. [9]Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Tübingen, Germany. [10]Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg. [11]Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg. [12]Centre Hospitalier du Nord, Ettelbrück, Luxembourg. [13]Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [14]Private practice, Ettelbruck, Luxembourg. [15]Parkinson Luxembourg Association, Leudelange, Luxembourg. [16]Westpfalz-Klinikum GmbH, Kaiserslautern, Germany. [17]Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK. [18]Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany. [19]Ruhr-University of Bochum, Bochum, Germany. [20]Laboratoire National de Santé, Dudelange, Luxembourg. [21]Paracelsus-Elena-Klinik, Kassel, Germany. [22]Private practice, Luxembourg-City, Luxembourg. [23]Department of Neurology Philipps, University Marburg, Marburg, Germany. [24]Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK.

Journal of Parkinson's Disease 12 (2022) 2561–2573
DOI 10.3233/JPD-223511
IOS Press

Research Report

Body-First Subtype of Parkinson's Disease with Probable REM-Sleep Behavior Disorder Is Associated with Non-Motor Dominant Phenotype

Lukas Pavelka[a,b,*], Armin Rauschenberger[c], Zied Landoulsi[d], Sinthuja Pachchek[d,e], Taina Marques[e], Clarissa P.C. Gomes[e], Enrico Glaab[c], Patrick May[d] and Rejko Krüger[e,a,b,*] on behalf of the NCER-PD Consortium

[a]*Parkinson's Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg, Luxembourg*
[b]*Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg*
[c]*Biomedical Data Science Group, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg*
[d]*Bioinformatics Core, Luxembourg Centre for Systems Biomedicine (LCSB), Esch-sur-Alzette, Luxembourg*
[e]*Translational Neuroscience, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg*

Accepted 25 September 2022
Pre-press 14 October 2022

Abstract.
Background: The hypothesis of *body-first* vs. *brain-first* subtype of PD has been proposed with REM-Sleep behavior disorder (RBD) defining the former. The body-first PD presumes an involvement of the brainstem in the pathogenic process with higher burden of autonomic dysfunction.
Objective: To identify distinctive clinical subtypes of idiopathic Parkinson's disease (iPD) in line with the formerly proposed concept of *body-first* vs. *brain-first* subtypes in PD, we analyzed the presence of probable RBD (pRBD), sex, and the *APOE* ε4 carrier status as potential sub-group stratifiers.
Methods: A total of 400 iPD patients were included in the cross-sectional analysis from the baseline dataset with a completed RBD Screening Questionnaire (RBDSQ) for classifying as pRBD by using the cut-off RBDSQ ≥ 6. Multiple regression models were applied to explore (i) the effect of pRBD on clinical outcomes adjusted for disease duration and age, (ii) the effect of sex on pRBD, and (iii) the association of *APOE* ε4 and pRBD.
Results: iPD-pRBD was significantly associated with autonomic dysfunction (SCOPA-AUT), level of depressive symptoms (BDI-I), MDS-UPDRS I, hallucinations, and constipation, whereas significantly negatively associated with quality of life (PDQ-39) and sleep (PDSS). No significant association between sex and pRBD or *APOE* ε4 and pRBD in iPD was found nor did we determine a significant effect of *APOE* ε4 on the PD phenotype.

*Correspondence to: Lukas Pavelka, MD, Physician scientist, National Centre for Excellence in Research in Parkinson's disease. Tel.: +352 621621578; E-mail: lukas.pavelka@lih.lu and Prof. Rejko Krüger, MD, Coordinator of Luxembourg National Centre for Excellence in Research in Parkinson's disease. Tel.: +352 26 970 458; E-mail: rejko.krueger@uni.lu.

Conclusion: We identified an RBD-specific PD endophenotype, characterized by predominant autonomic dysfunction, hallucinations, and depression, corroborating the concept of a distinctive *body-first* subtype of PD. We did not observe a significant association between *APOE* ε4 and pRBD suggesting both factors having an independent effect on cognitive decline in iPD.

Keywords: Idiopathic Parkinson's disease, probable REM-Sleep behavior disorder, RBDSQ, non-motor symptoms, *APOE*, stratification

INTRODUCTION

The phenotypic heterogeneity of Parkinson's disease (PD) has been a challenge for both clinicians and researchers for decades. Several efforts were made to identify an underlying pattern explaining this heterogeneity by subtyping PD patients. They can be grouped into two distinct methods. The first approach uses a single clinical or genetic metric determining the clinical phenotype, such as age at onset, sex, motor phenotype, or being a carrier of the PD-causing rare genetic mutations. The second approach has been using hypothesis-free data-driven models identifying phenotypic clusters in PD based on clinical symptoms, but this approach failed reproducibility checks, possibly due to a limited methodological overlap between the studies and a wide variety of clinical metrics entering the models [1]. Interestingly, both approaches systematically reported REM-sleep behavior disorder (RBD) as a relevant clinical variable. Not only is RBD currently known as the most robust prodromal marker of future pheno-conversion to the alpha-synucleinopathies (i.e., PD, dementia with Lewy bodies or multiple system atrophy) [2], but it was suggested that RBD is associated with more rapid progression of motor symptoms, a higher burden of non-motor symptoms and lower quality of life [3–5].

RBD received increasing attention in the last years, with several cross-sectional and longitudinal studies investigating the association between RBD and the clinical phenotype of PD. On the one hand, we observe an overall consensus regarding a non-motor dominant profile of PD with higher autonomic dysfunction and more rapid cognitive decline. On the other hand, prior studies have reported contradictory findings on the effect of comorbid RBD on motor progression in PD [5–8]. Moreover, genetic risk factors and PD-causing rare mutations with a substantial effect on the clinical phenotype were rarely systematically addressed in the context of concomitant RBD and PD and their effect on the severity of the clinical phenotype. Recently, the *APOE* epsilon4 (*APOE* ε4) genotype has been linked to faster cognitive decline and motor progression in PD [9], although studies on the role of *APOE* ε4 and clinical progression of PD remain controversial [10, 11]. Whether an additive or multiplicative potentiation effect of RBD and *APOE* ε4 on cognitive decline in PD exists has not been adequately addressed so far. Currently, no association of the *APOE* ε4 carriers status with idiopathic RBD has been observed [12, 13], but a potential role of the *APOE* ε4 genotype as a modifier of the clinical phenotype of PD with RBD has not yet been explored.

RBD has been suggested to represent a key element in distinguishing body-first from brain-first subtype of PD, a concept recently proposed to explain the phenotypic differences and variability of dynamics in PD and supported by several clinical and imaging studies [14, 15]. It has been proposed that the body-first subtype of PD starts in the peripheral nervous system with spreading of neurodegeneration via brainstem thus associated with RBD, higher burden of autonomic dysfunction and higher rate of cognitive decline [16].

In order to test the hypothesis of body-first subtype of PD with comorbid pRBD, we used a large baseline visit dataset from the Luxembourg Parkinson's Study, a monocentric longitudinal observational study with a previously described recruitment design [17]. In our study, we primarily aimed to determine the effect of pRBD on clinical outcomes in idiopathic PD (iPD) by excluding known PD-linked rare mutations or genetic risk variant carriers. Next, we investigated potential confounding effects of sex and the *APOE* ε4 carrier status as potential stratifiers of iPD.

MATERIALS AND METHODS

Study population

The data used in this study were acquired from participants recruited in the frame of the nationwide monocentric observational longitudinal Luxembourg Parkinson's Study [17]. The diagnosis of PD relied on the UK Parkinson's Disease Society Brain Bank

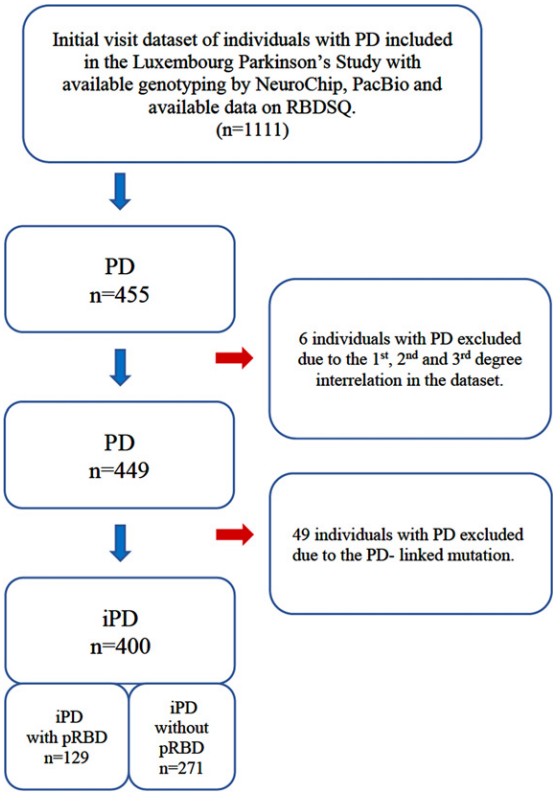


Fig. 1. Description of the study design and study dataset. PD, individuals with Parkinson's disease; iPD, idiopathic Parkinson's disease; pRBD, probable REM-sleep behavior disorder; RBDSQ, REM sleep behavior disorder screening questionnaire.

(UKPDSBB) diagnostic criteria [18]. All participants were genotyped for disease-causing mutations and PD-associated risk variants using both NeuroChip® and PacBio sequencing. Available data on RBDSQ were analyzed after excluding six PD patients for 1st, 2nd, and 3rd degree relationships and after excluding 49 PD patients carrying PD-associated mutations. The overall study design, inclusion, and exclusion workflow are illustrated in Fig. 1. Though the diagnostic gold standard of RBD remains polysomnography (PSG) [19], the accessibility of the sleep laboratory and performing PSG on a large scale is problematic due to the sleep laboratory capacities and costs. We therefore applied a classification of probable RBD (pRBD) by REM-sleep behavior disorder screening questionnaire (RBDSQ) as used in several previous studies [20–24]. The group assignment of pRBD in iPD individuals uses the criterion RBDSQ ≥ 6 to optimize the specificity and sensitivity for pRBD in line with the Oxford Discovery Study [24].

All participants taking part in the Luxembourg Parkinson's Study agreed and signed a written informed consent. The study has been approved by the National Research Ethics Committee (CNER Ref: 201407/13).

Clinical assessment and data

The design and recruitment of the Luxembourg Parkinson's Study were previously published in detail [17]. Sociodemographic characteristics and clinical outcomes validated for PD were chosen from the basic clinical assessment battery and are listed in Tables 1 and 2. All patients have been evaluated in medication ON state and, where applicable, in deep brain stimulation ON state. The clinical symptoms as scales are defined in detail in the Supplementary Material.

Missing data statement

The absolute number of missing data per variable is described in Tables 1 and 2. Given the low proportions of missing values in the dataset, we used a pairwise deletion for all statistical models.

Genotyping and quality-control analyses

The methods for genotyping in our dataset have been described previously [25]. PD causing rare variants were defined by the ClinVar classification as "pathogenic/likely pathogenic". All PD-causing variants (listed in the Supplementary Material) identified by any method were Sanger validated, and all samples with a validated PD-causing variant were excluded from further analysis with a list of excluded variants described in the Supplementary Material.

APOE genotyping

APOE genotypes were called for all individuals from two SNPs investigated by NeuroChip array (rs429358, rs7412) that distinguish the $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$ alleles classifying the respective *APOE* carriers. The NeuroChip provides high accuracy of 98.1% for genotyping of *APOE* $\varepsilon 4$ [26], and the approach was aligned with other large studies [27].

Statistical analysis

Mann-Whitney's *U* test was used for numerical variables and Fisher's exact test for binary variables in

Table 1
Descriptive and comparative statistics of demographic data and frequency of *APOE* ε4 genotype in PD individuals with (right) and without (left) probable REM-sleep behavior disorder (pRBD). For intergroup comparisons, *p*-values are shown from Mann-Whitney *U* test for numerical variables and Fisher's exact test for binary variables. Binary variables are annotated by asterisk. Results are shown as mean and standard deviation (SD) for numerical variables, number of zeros ('NO') and ones ('YES') for binary variables and percentage of YES, and number of missing values (NA). Single and double ticks indicate significance at the 5% level, and the Bonferroni-adjusted 5% level. Age at onset was calculated based on the year of the PD diagnosis. PD, Parkinson's disease

	PD non-pRBD ($n = 271$)			PD pRBD ($n = 129$)			p
	Mean or YES in %	SD or NO/YES	NA	Mean or YES in %	SD or NO/YES	NA	
Disease duration since diagnosis (y)	4.20	4.55	0	7.86	6.36	0	8.2e-11″
Age at assessment (y)	66.19	11.29	0	68.31	9.85	0	1.2e-01
Age at onset (y)	62.01	11.64	0	60.48	11.98	0	2.5e-01
Sex (male)*	65%	96/175	0	74%	34/95	0	8.6e-02
APOE (ε2/ ε4; ε3/ ε4;ε4/ε4)*	21%	213/58	0	26%	95/34	0	3.1e-01
Years of education	13.29	4.12	0	12.99	3.90	0	6.7e-01
Total languages spoken	2.86	1.06	0	2.89	1.04	0	8.0e-01

Table 2
Descriptive and comparative statistics of clinical outcomes for iPD group with and without probable REM-sleep behavior disorder (pRBD). Results are shown as mean and standard deviation (SD) for numerical variables, number of zeros ('NO') and ones ('YES') for binary variables and percentage of YES, and number of missing values (NA). For intergroup comparisons, *p*-values are shown from Mann-Whitney *U* test for numerical variables and Fisher's exact test for binary variables. Binary variables are annotated by asterisk. Single and double ticks indicate significance at the 5% level, and the Bonferroni-adjusted 5% level. All clinical outcomes are defined and described in the Supplementary Material

	PD non-pRBD ($n = 271$)			PD pRBD ($n = 129$)			p
	Mean or YES in %	SD or NO/YES	NA	Mean or YES in %	SD or NO/YES	NA	
H&Y	2.12	0.78	2	2.37	0.75	0	1.2e-04″
MDS-UPDRS III	32.00	16.11	5	38.02	16.76	2	4.5e-04″
MDS-UPDRS II	9.79	7.45	3	14.50	8.64	3	1.0e-07″
LEDD (g/day)	0.45	0.38	0	0.68	0.41	0	2.8e-08″
Gait disorder*	48%	141/130	0	71%	37/92	0	1.0e-05″
Repetitive falls*	11%	240/31	0	29%	91/38	0	1.7e-05″
MDS-UPDRS IV	1.37	3.01	2	2.75	3.98	3	5.2e-05″
Dyskinesia/day (hours)	0.47	2.29	0	1.21	3.57	1	9.3e-05″
OFF time/day (hours)	0.40	1.41	0	0.72	1.38	2	3.2e-04″
Dystonia/day (hours)	0.027	0.15	1	0.088	0.31	1	7.3e-03′
Dyskinesia*	9%	246/25	0	20%	103/26	0	3.5e-03′
Motor fluctuations*	11%	241/30	0	27%	94/35	0	8.1e-05″
Freezing of gait*	16%	227/44	0	34%	85/44	0	9.4e-05″
MoCA	24.85	3.93	5	24.02	4.45	2	6.9e-02
Sniffin' stick test	8.52	3.34	7	7.50	3.27	3	1.0e-02′
PDQ-39	33.65	23.88	12	52.23	27.05	6	7.2e-11″
SCOPA-AUT	12.59	6.97	2	19.59	8.11	0	6.7e-15″
MDS-UPDRS I	8.54	5.78	6	13.62	7.36	4	5.1e-12″
BDI-I	8.79	6.65	7	12.62	7.33	3	6.2e-08″
Starkstein Apathy Scale	13.46	5.31	4	14.67	6.24	3	1.2e-01
PDSS	111.40	21.55	4	92.64	23.05	3	2.3e-13″
Probable RBD*	0%	271/0	0	100%	0/129	0	1.4e-108″
Excessive daily sleepiness*	23%	208/63	0	41%	76/53	0	3.8e-04″
Insomnia*	24%	205/66	0	21%	102/27	0	5.3e-01
Hallucinations*	9%	247/24	0	29%	91/38	0	4.8e-07″
Impulse Control Disorder*	6%	255/16	0	16%	108/21	0	1.4e-03′
Orthostatic hypotension*	23%	210/61	0	36%	82/47	0	3.9e-03′
Dysphagia*	20%	218/53	0	33%	87/42	0	5.6e-03′
Constipation*	31%	187/84	0	63%	48/81	0	2.8e-09″
Urinary Incontinence*	27%	197/74	0	39%	79/50	0	2.8e-02′

intergroup comparison analyses (iPD pRBD vs. iPD non-pRBD; male sex iPD vs. female sex iPD). Multiple linear and logistic regression models were applied to investigate the effect of pRBD on clinical outcomes in iPD, adjusted for age at assessment (AAA) and disease duration. To investigate the potential effect of the *APOE* genotype on clinical outcomes, we pooled the heterozygotes ($\varepsilon2/\varepsilon4$; $\varepsilon3/\varepsilon4$) and homozygotes ($\varepsilon4/\varepsilon4$), allowing us to quantify a potential association between *APOE* $\varepsilon4$ genotype and pRBD in iPD. Furthermore, we applied regression of clinical symptoms in PD on *APOE* $\varepsilon4$, AAA and disease duration. For all analyses, we assessed significance at the 5% level and the Bonferroni-adjusted 5% level.

RESULTS

Frequency of pRBD and effect of pRBD on clinical outcomes in iPD

According to the RBDSQ classification of pRBD, we observed a relative pRBD frequency of 32.3% in the iPD group (129 iPD pRBD out of 400). The demographic characteristics of iPD pRBD ($n = 129$) and iPD non-pRBD patients ($n = 271$) are shown in Tables 1 and 2. We investigated the effect of pRBD on the clinical outcomes adjusted for AAA and disease duration.

As key results, we observed a significant positive association between iPD pRBD (as opposed to iPD non-pRBD) and burden of non-motor symptoms, i.e., autonomic dysfunction (SCOPA-AUT) and frequency of constipation; MDS-UPDRS I, burden of depression symptoms assessed by BDI-I, frequency of hallucinations and PDQ-39, showing lower quality of life in iPD pRBD, as demonstrated in Fig. 2. Furthermore, a significant negative association was determined between iPD pRBD and the Parkinson's Disease Sleep Scale (PDSS), indicating lower quality of sleep in the group of iPD pRBD vs. iPD non-pRBD. Other considered clinical outcomes showed no significant associations after multiple testing correction.

APOE genotype and iPD pRBD

We found no significant association between pooled heterozygote and homozygote *APOE* $\varepsilon4$ carriers and iPD with pRBD. Additionally, no significant association was observed between *APOE* $\varepsilon4$ and the clinical outcomes of iPD with pRBD vs. iPD non-pRBD adjusted for AAA and disease duration, as shown in Fig. 3.

Effect of sex on frequency of pRBD and other clinical outcomes in iPD

Clinical and demographic characteristics and outcomes of sex-stratified iPD are shown in Table 3. We did not observe a significant effect of male sex on the frequency of pRBD in iPD. Interestingly, from all the putative variables, only olfactory performance (measured by Sniffin' Stick test) was significantly negatively, and FOG significantly positively associated with male sex in PD after adjustment for AAA and disease duration (see Fig. 4).

Effect of education and number of spoken languages on cognitive performance

We analyzed a potential confounding effect of the years of education (YoE) and the total languages spoken (TLS) on cognitive performance in our dataset. As shown in the Supplementary Table 1, only YoE (not TLS) had a significant positive effect on Montreal Cognitive Assessment (MoCA) in a multiple regression model adjusted for AAA and disease duration.

DISCUSSION

The results of our study support the classification of RBD as a distinctive characteristic of the body-first subtype by identifying a significant association of iPD pRBD with the non-motor dominant disease profile, a result that matched remarkably well with the majority of previous studies [4–8]. It favors the concept of pathological process beginning in the peripheral nervous system with further centripetal spreading of alpha-synuclein in a subgroup of PD patients and hence the associated neurodegeneration causing a significantly higher autonomic dysfunction, higher depression burden as well as hallucinations through dysregulation of dopaminergic and noradrenergic system in the brainstem. Although we assessed RBD via a screening questionnaire, our results were consistent with a prior study using PSG-proven RBD, which indicated an association of a non-motor dominant phenotype in PD with PSG-proven RBD [4]. However, we observed only a trend in the negative effect of pRBD on global cognitive performance in PD, which did not correspond to several cross-

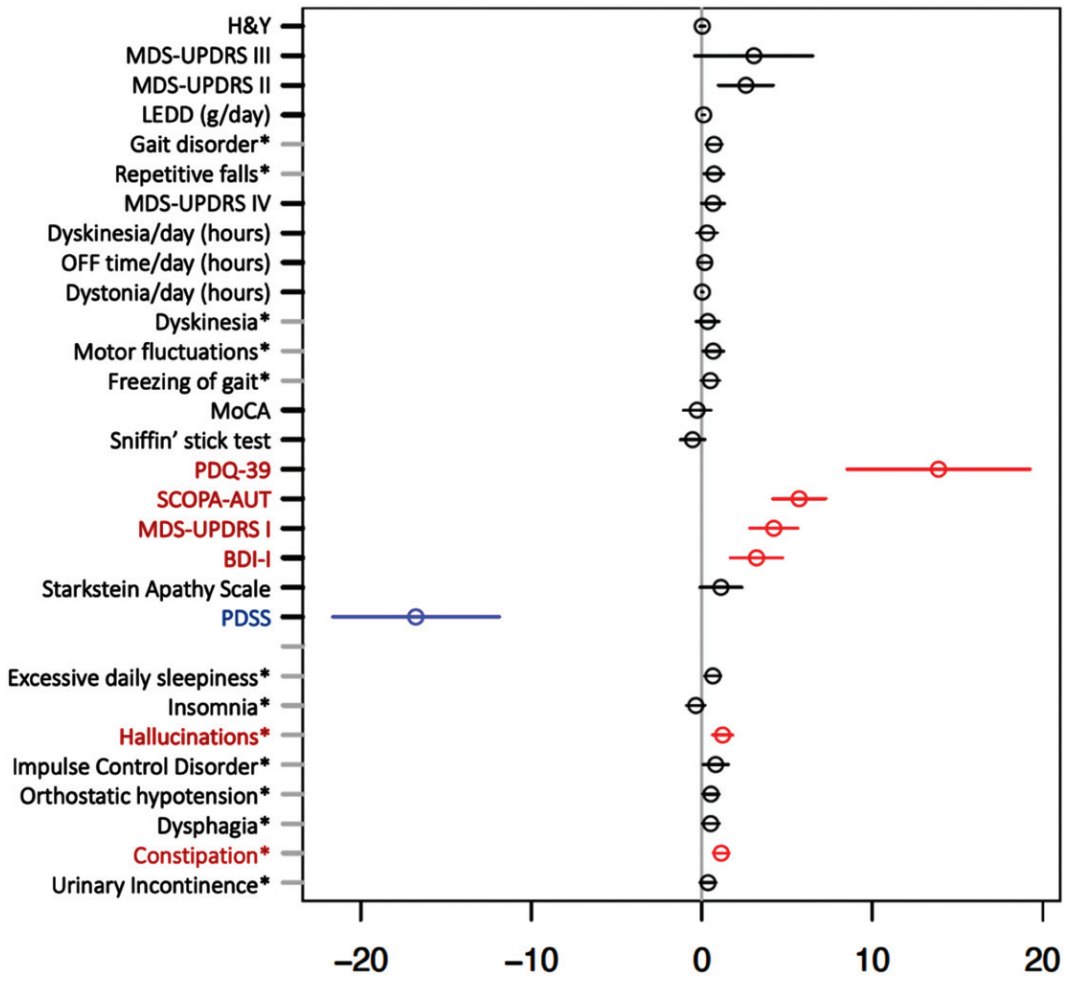


Fig. 2. Multiple regression model for investigating effect of probable REM-Sleep behavior disorder on clinical outcomes in idiopathic Parkinson's disease adjusted for age at assessment and disease duration. Forrest plot with estimated coefficients and corresponding confidence intervals (±1.96 x standard error) for pRBD, from linear/logistic regression of numerical/binary outcome on disease duration, age at assessment (AAA) and pRBD (binary outcomes are annotated by asterisk). The color blue indicates significant negative effects of pRBD on the clinical outcome, and the color red indicates significant positive effects at the Bonferroni-adjusted 5% level. Clinical symptoms and scales are described in the Supplementary Material.

sectional and longitudinal studies [8]. To assess a potential independent variable influencing cognitive performance, we identified a protective effect of YoE on cognitive decline in the overall PD group, but we did not identify a significant difference in pRBD PD vs. non-pRBD PD in terms of YoE or TLS. Therefore, we did not consider these two factors (YoE and TLS) as confounding factors for the effect of pRBD on cognitive performance assessed by MoCA in our dataset. Moreover, the *APOE* ε4 genotype, known to exacerbate beta amyloid pathology in Alzheimer's disease, has been suggested to play a role in accelerated cognitive decline in PD [27, 28]. As RBD was associated with a higher rate of cognitive decline and dementia in previous studies, we explored a potential association between pRBD and *APOE* ε4 carrier status. However, no significant association between the two was observed in our study. This would argue for an independent effect of pRBD and *APOE* ε4 status without a synergistic effect on cognitive decline in iPD. Therefore, we conclude that *APOE* ε4 genotype might not play a role as a stratifier in body-first vs. brain-first concept. It is important to stress that we excluded a potential effect of PD-linked genetic mutations and genetic risk factors for PD, which may have contributed to confounding effects on clinical phenotype in other studies, as in the case of highly prevalent mutations in the GBA gene [29].

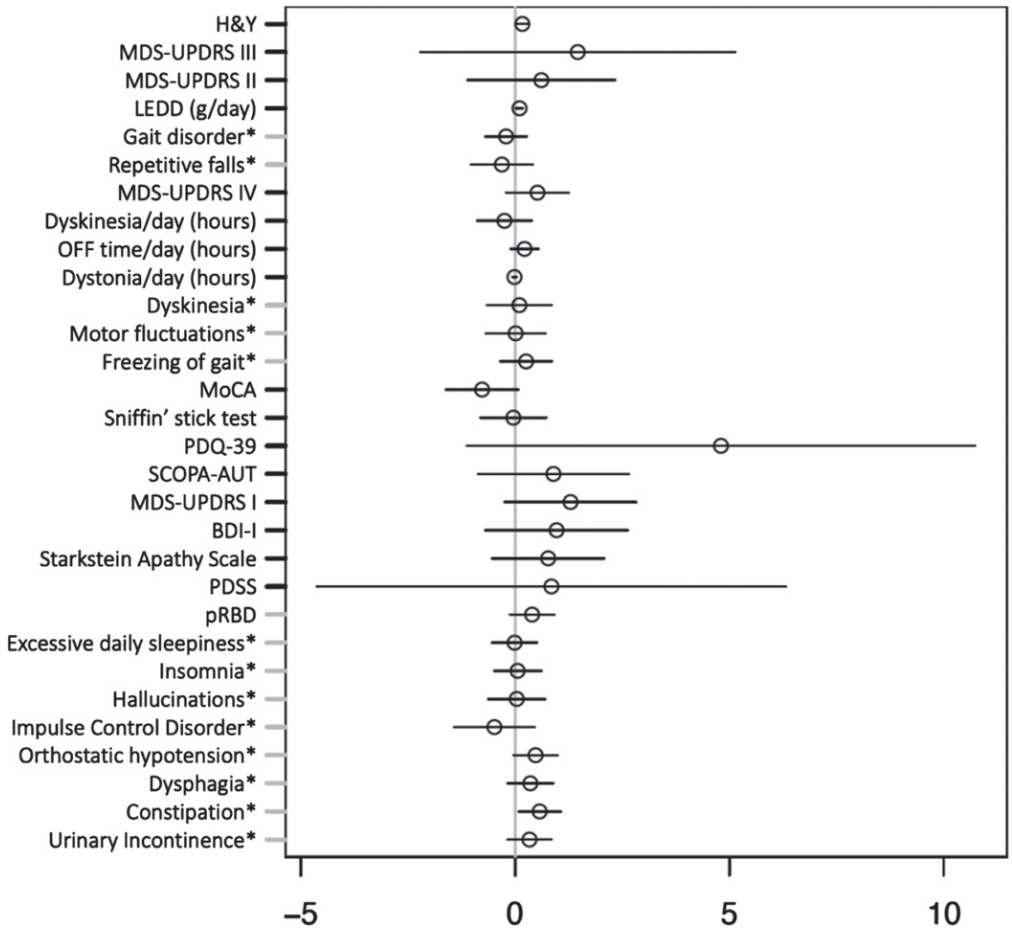


Fig. 3. Multiple regression model investigating effect of *APOE* ε4 carrier status on clinical outcomes in idiopathic Parkinson's disease adjusted for age at assessment and disease duration. Forrest plot with estimated coefficients and corresponding confidence intervals (±1.96 x standard error) for *APOE* ε4 genotype, from linear/logistic regression of numerical/binary outcome on disease duration, age at assessment (AAA), and *APOE* (binary outcomes are annotated by asterisk). The color blue indicates significant negative effects of *APOE* ε4 genotype on the clinical outcome, and the color red indicates significant positive effects at the Bonferroni-adjusted 5% level. Clinical symptoms and scales are described in the Supplementary Material.

Our investigation of potential sex-related differences in iPD phenotype did not reveal a significant association between pRBD and male sex, as suggested by several prior studies using either a similar screening questionnaire approach or PSG [30–32]. This adds to the open debate about whether there are significant differences in the prevalence of RBD in males vs. females. We would like to point out that the higher frequency of RBD in males was observed in studies using the dataset of individuals referred primarily to sleep laboratories which may cause a referral bias, given the fact that males are reported to have more violent RBD symptoms and are therefore more likely to be referred for PSG [33–36].

Next, we studied the potential confounding effects of sex on other motor and non-motor symptoms. We observed a higher frequency of males vs. females in the overall PD group (67.5% vs. 32.5%), in line with the results from recently published large cohort studies [37–39]. Interestingly, we found only olfactory dysfunction and FOG to be positively associated with males, while other putative motor and non-motor outcomes showed no significant associations with sex after multiple testing correction. These findings might indicate that sex does not play a substantial role in defining the phenotype of iPD and thus do not account for the phenotypic differences associated with pRBD.

Our study displays several specific strengths: (i) a large dataset was analyzed relative to previous studies; (ii) PD cases were genetically stratified by NeuroChip and targeted sequencing of GBA, avoid-

Table 3
Descriptive statistics for sex stratified iPD. Results are shown as mean and standard deviation (SD) for numerical variables, number of zeros ('NO') and ones ('YES') for binary variables and percentage of YES, and number of missing values (NA). The last column shows *p*-values from Mann-Whitney *U* test for numerical variables and Fisher's exact test for binary variables. Binary variables are annotated by asterisk. Single and double ticks indicate significance at the 5% level and the Bonferroni-adjusted 5% level. Age at onset was calculated based on the year of the PD diagnosis

	PD female ($n = 130$)			PD male ($n = 270$)			
	Mean or YES in %	SD or NO/YES	NA	Mean or YES in %	SD or NO/YES	NA	*p*
Disease duration since diagnosis (y)	5.44	5.53	0	5.35	5.46	0	8.1e-01
Age at assessment (y)	66.71	10.74	0	66.95	10.97	0	9.1e-01
Age at onset (y)	61.30	11.03	0	61.62	12.11	0	8.6e-01
H&Y	2.21	0.84	1	2.20	0.75	1	9.3e-01
MDS-UPDRS III	33.49	18.03	2	34.17	15.81	5	4.3e-01
MDS-UPDRS II	11.09	8.38	2	11.40	8.04	4	4.7e-01
LEDD (g/day)	0.47	0.36	0	0.55	0.42	0	1.2e-01
Gait disorder*	50%	65/65	0	58%	113/157	0	1.3e-01
Repetitive falls*	20%	104/26	0	16%	227/43	0	3.2e-01
MDS-UPDRS IV	1.90	3.61	4	1.77	3.30	1	9.3e-01
Dyskinesia/day (h)	0.87	3.22	1	0.63	2.55	0	8.6e-01
OFF time/day (h)	0.55	1.91	2	0.48	1.10	0	7.3e-01
Dystonia/day (h)	0.035	0.17	2	0.052	0.24	0	1.0e-01
Dyskinesia*	12%	115/15	0	13%	234/36	0	7.5e-01
Motor fluctuations*	11%	116/14	0	19%	219/51	0	4.3e-02′
Freezing of gait*	13%	113/17	0	26%	199/71	0	2.9e-03′
MoCA	24.92	3.84	3	24.41	4.24	4	3.4e-01
Sniffin' stick test	9.10	3.26	4	7.76	3.30	6	2.2e-04″
PDQ-39	43.28	26.38	8	37.92	26.26	10	4.0e-02′
SCOPA-AUT	14.92	8.01	2	14.83	8.08	0	1.0e+00
MDS-UPDRS I	10.22	6.32	3	10.14	6.96	7	5.5e-01
BDI-I	11.20	7.75	4	9.47	6.71	6	2.9e-02′
Starkstein Apathy Scale	13.84	5.77	6	13.86	5.60	1	9.7e-01
PDSS	102.64	25.08	4	106.68	22.94	3	1.3e-01
Probable RBD*	26%	96/34	0	35%	175/95	0	8.6e-02
Excessive daily sleepiness*	20%	104/26	0	33%	180/90	0	6.7e-03′
Insomnia*	27%	95/35	0	21%	212/58	0	2.6e-01
Hallucinations*	16%	109/21	0	15%	229/41	0	8.8e-01
Impulse Control Disorder*	7%	121/9	0	10%	242/28	0	3.6e-01
Orthostatic hypotension*	27%	95/35	0	27%	197/73	0	1.0e+00
Dysphagia*	26%	96/34	0	23%	209/61	0	4.5e-01
Constipation*	40%	78/52	0	42%	157/113	0	7.5e-01
Urinary Incontinence*	30%	91/39	0	31%	185/85	0	8.2e-01

ing a potential confounding by PD-causing mutations that are known to significantly influence the clinical phenotype; (iii) the study design included all disease stages of PD regardless of the cognitive status, and (iv) a monocentric data collection assured the consistency of the dataset.

Conversely, some limitations of our study should also be noted: We investigated the research questions using a cross-sectional analysis, and further studies on longitudinal data are still warranted. Additionally, RBD was not assessed by gold standard PSG but by a more accessible method using a screening questionnaire, potentially including in part false positive patients for RBD with another sleep pathology. Furthermore, the presence of hallucinations might be wrongly considered by the patients to classify as RBD symptoms. Nevertheless, the association of RBD in PD with hallucinations has been widely reported in the literature [40–42], thus we do not consider the significant positive association of pRBD with hallucinations in our dataset as a potential mis-classifier of pRBD vs. non-pRBD. Finally, we did not have complementary data on the time relation between pRBD and PD, i.e., describing whether pRBD preceded PD or evolved during the clinical phase of PD.

However, the overall concordance of the results on the association of pRBD in PD with a non-motor dominant phenotype indicates that applying RBDSQ may provide a useful tool for patient stratification in future studies and clinical trials. It might

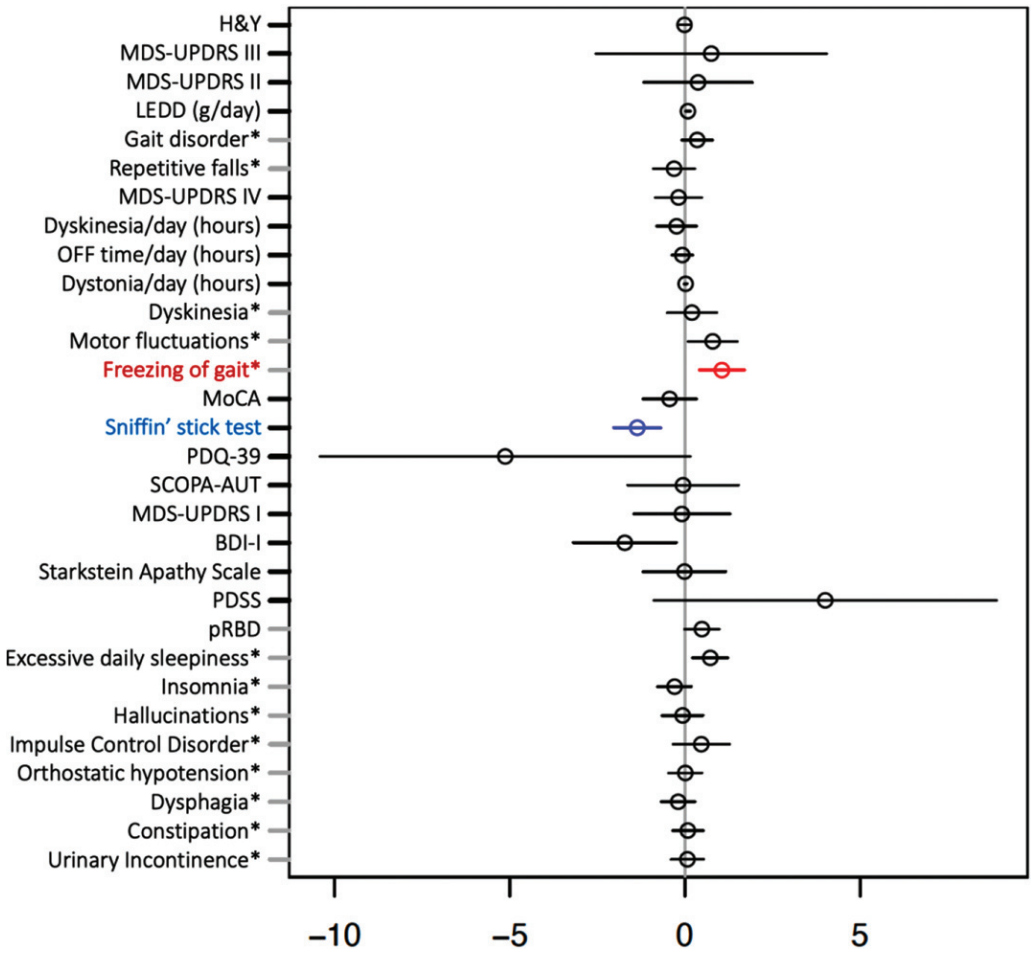


Fig. 4. Multiple regression model investigating effect of sex on clinical outcomes in idiopathic Parkinson's disease adjusted for age at assessment and disease duration. Forrest plot with estimated coefficients and corresponding confidence intervals (±1.96 x Standard error) for sex, from linear/logistic regression of numerical/binary outcome on disease duration, AAA, and sex (binary outcomes are annotated by asterisk). The color blue indicates significant negative effects of male vs. female sex on the clinical outcome, and the color red indicates significant positive effects at the Bonferroni-adjusted 5% level. Clinical symptoms and scales are described in the Supplementary Material.

prove to be a clinically relevant mean to screen for pRBD during the regular follow-up of PD patients in order to personalize and adapt the therapy and its potential secondary effects by the treating physicians. Finally, this study adds to the prior body of evidence that PD subtyping, in general, may serve the patient by providing treatment-relevant phenotype-genotype stratifications as a tool for future clinical trials.

ACKNOWLEDGMENTS

We would like to give special thanks to all participants in the Luxembourg Parkinson's Study. Furthermore, we acknowledge the joint effort of the NCER-PD consortium members generally contributing to the Luxembourg Parkinson's Study as listed below:

Geeta ACHARYA[2], Gloria AGUAYO[2], Myriam ALEXANDRE[2], Muhammad ALI[1], Wim AMMERLANN[2], Rudi BALLING[1], Michele BASSIS[1], Katy BEAUMONT[2], Regina BECKER[1], Camille BELLORA[2], Guy BERCHEM[3], Daniela BERG[11], Alexandre BISDORFF[5], Kathrin BROCKMANN[11], Jessica CALMES[2], Lorieza CASTILLO[2], Gessica CONTESOTTO[2], Giuseppe ARENA[1], Nico DIEDERICH[3], Rene DONDELINGER[5], Daniela ESTEVES[2], Guy FAGHERAZZI[2], Jean-Yves FERRAND[2], Manon GANTENBEIN[2], Thomas GASSER[11], Piotr GAWRON[1], Soumyabrata GHOSH[1], Marijus

GIRAITIS[2,3], Enrico GLAAB[1], Clarissa GOMES[1], Elisa GÓMEZ DE LOPE[1], Jérôme GRAAS[2], Mariella GRAZIANO[17], Valentin GROUES[1], Anne GRÜNEWALD[1], Wei GU[1], Gaël HAMMOT[2], Anne-Marie HANFF[2], Linda HANSEN[1,3], Maxime HANSEN[1,3], Michael HENEKA[1], Estelle HENRY[2], Sylvia HERBRINK[6], Sascha HERZINGER[1], Michael HEYMANN[2], Michele HU[8], Alexander HUNDT[2], Ivana PACCOUD[2], Nadine JACOBY[18], Jacek JAROSLAW LEBIODA[1], Yohan JAROZ[1], Quentin KLOPFENSTEIN[1], Jochen KLUCKEN[1,2,3], Rejko KRÜGER[1,2,3], Pauline LAMBERT[2], Zied LANDOULSI[1], Roseline LENTZ[7], Inga LIEPELT[11], Robert LISZKA[14], Laura LONGHINO[3], Victoria LORENTZ[2], Paula Cristina LUPU[2], Clare MACKAY[10], Walter MAETZLER[15], Katrin MARCUS[13], Guilherme MARQUES[2], Tainá MARQUES[1], Patricia MARTINS CONDE[1], Patrick MAY[1], Deborah MCINTYRE[2], Chouaib MEDIOUNI[2], Francoise MEISCH[1], Myriam MENSTER[2], Maura MINELLI[2], Michel MITTELBRONN[1,4], Brit MOLLENHAUER[12], Carlos MORENO[1], Friedrich MÜHLSCHLEGEL[4], Romain NATI[3], Ulf NEHRBASS[2], Sarah NICKELS[1], Beatrice NICOLAI[3], Jean-Paul NICOLAY[19], Fozia NOOR[2], Marek OSTASZEWSKI[1], Sinthuja PACHCHEK[1], Claire PAULY[1,3], Laure PAULY[1], Lukas PAVELKA[1,2,3], Magali PERQUIN[2], Rosalina RAMOS LIMA[2], Armin RAUSCHENBERGER[1], Rajesh RAWAL[1], Dheeraj REDDY BOBBILI[1], Eduardo ROSALES[2], Isabel ROSETY[1], Kirsten RUMP[1], Estelle SANDT[2], Stefano SAPIENZA[1], Venkata SATAGOPAM[1], Margaux SCHMITT[2], Sabine SCHMITZ[1], Reinhard SCHNEIDER[1], Jens SCHWAMBORN[1], Jean-Edouard SCHWEITZER[1], Amir SHARIFY[2], Ekaterina SOBOLEVA[1], Kate SOKOLOWSKA[2], Olivier TERWINDT[1,3], Hermann THIEN[2], Elodie THIRY[3], Rebecca TING JIIN LOO[1], Christophe TREFOIS[1], Johanna TROUET[2], Olena TSURKALENKO[2], Michel VAILLANT[2], Mesele VALENTI[2], Sijmen VAN SCHAGEN[1], Liliana VILAS BOAS[3], Maharshi VYAS[1], Richard WADE-MARTINS[9], Paul WILMES[1], Evi WOLLSCHEID-LENGELING[1], Gelani ZELIMKHANOV[3]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]Luxembourg Institute of Health, Strassen, Luxembourg

[3]Centre Hospitalier de Luxembourg, Strassen, Luxembourg

[4]Laboratoire National de Santé, Dudelange, Luxembourg

[5]Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg

[6]Centre Hospitalier du Nord, Ettelbrück, Luxembourg

[7]Parkinson Luxembourg Association, Leudelange, Luxembourg

[8]Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[9]Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK

[10]Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK

[11]Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Tübingen, Germany

[12]Paracelsus-Elena-Klinik, Kassel, Germany

[13]Ruhr-University of Bochum, Bochum, Germany

[14]Westpfalz-Klinikum GmbH, Kaiserslautern, Germany

[15]Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany

[16]Department of Neurology Philipps, University Marburg, Marburg, Germany

[17]Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg

[18]Private practice, Ettelbruck, Luxembourg

[19]Private practice, Luxembourg, Luxembourg

FUNDING SOURCES

The National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) is funded by the Luxembourg National Research Fund (FNR/NCER13/BM/11264123), the PEARL program (FNR/P13/6682797 to RK), FNR grant projects PD-Strat (INTER/11651464 to EG), DIGIPD (ERAPERMED 2020-314 to EG); MotaSYN (12719684 to RK), MAMaSyn (to RK), MiRisk-PD (C17/BM/11676395 to RK, EG, PM), the FNR/DFG Core INTER (ProtectMove, FNR11250962 to PM), and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to RK, SP).

CONFLICT OF INTEREST

All authors have no conflict of interest to report.

DATA AND CODE AVAILABILITY STATEMENT

The dataset for this manuscript is not publicly available as it is linked to the Luxembourg Parkinson's Study and its internal regulations. Any requests for accessing the dataset can be directed to request.ncer-pd@uni.lu. The code for the statistical models is available at: https://doi.org/10.17881/sw04-1w80.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JPD-223511.

REFERENCES

[1] Berg D, Borghammer P, Fereshtehnejad SM, Heinzel S, Horsager J, Schaeffer E, Postuma RB (2021) Prodromal Parkinson disease subtypes – key to understanding heterogeneity. *Nat Rev Neurol* **17**, 349-361.

[2] Postuma RB, Iranzo A, Hu M, Högl B, Boeve BF, Manni R, Oertel WH, Arnulf I, Ferini-Strambi L, Puligheddu M, Antelmi E, Cochen De Cock V, Arnaldi D, Mollenhauer B, Videnovic A, Sonka K, Jung KY, Kunz D, Dauvilliers Y, Provini F, Lewis SJ, Buskova J, Pavlova M, Heidbreder A, Montplaisir JY, Santamaria J, Barber TR, Stefani A, St Louis EK, Terzaghi M, Janzen A, Leu-Semenescu S, Plazzi G, Nobili F, Sixel-Doering F, Dusek P, Bes F, Cortelli P, Ehgoetz Martens K, Gagnon JF, Gaig C, Zucconi M, Trenkwalder C, Gan-Or Z, Lo C, Rolinski M, Mahlknecht P, Holzknecht E, Boeve AR, Teigen LN, Toscano G, Mayer G, Morbelli S, Dawson B, Pelletier A (2019) Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain* **142**, 744-759.

[3] Roguski A, Rayment D, Whone AL, Jones MW, Rolinski M (2020) A neurologist's guide to REM sleep behavior disorder. *Front Neurol* **11**, 610.

[4] Neikrug AB, Avanzino JA, Liu L, Maglione JE, Natarajan L, Corey-Bloom J, Palmer BW, Loredo JS, Ancoli-Israel S (2014) Parkinson's disease and REM sleep behavior disorder result in increased non-motor symptoms. *Sleep Med* **15**, 959-966.

[5] Postuma RB, Bertrand JA, Montplaisir J, Desjardins C, Vendette M, Rios Romenets S, Panisset M, Gagnon JF (2012) Rapid eye movement sleep behavior disorder and risk of dementia in Parkinson's disease: a prospective study. *Mov Disord* **27**, 720-726.

[6] Barber TR, Muhammed K, Drew D, Lawton M, Crabbe M, Rolinski M, Quinnell T, Zaiwalla Z, Ben-Shlomo Y, Husain M, Hu MTM (2018) Apathy in rapid eye movement sleep behaviour disorder is common and under-recognized. *Eur J Neurol* **25**, 469-e32.

[7] Assogna F, Liguori C, Cravello L, Macchiusi L, Belli C, Placidi F, Pierantozzi M, Stefani A, Mercuri B, Izzi F, Caltagirone C, Mercuri NB, Pontieri FE, Spalletta G, Pellicano C (2021) Cognitive and neuropsychiatric profiles in idiopathic rapid eye movement sleep behavior disorder and Parkinson's disease. *J Pers Med* **11**, 51.

[8] Mao J, Huang X, Yu J, Chen L, Huang Y, Tang B, Guo J (2020) Association between REM sleep behavior disorder and cognitive dysfunctions in Parkinson's disease: a systematic review and meta-analysis of observational studies. *Front Neurol* **11**, 577874.

[9] Pu JL, Jin CY, Wang ZX, Fang Y, Li YL, Xue NJ, Zheng R, Lin ZH, Yan YQ, Si XL, Chen Y, Liu Y, Song Z, Yan YP, Tian J, Yin XZ, Zhang BR (2022) Apolipoprotein E genotype contributes to motor progression in Parkinson's disease. *Mov Disord* **37**, 196-200.

[10] Mengel D, Dams J, Ziemek J, Becker J, Balzer-Geldsetzer M, Hilker R, Baudrexel S, Kalbe E, Schmidt N, Witt K, Liepelt-Scarfone I, Gräber S, Petrelli A, Neuser P, Schulte C, Linse K, Storch A, Wittchen HU, Riedel O, Mollenhauer B, Ebentheuer J, Trenkwalder C, Klockgether T, Spottke A, Wüllner U, Schulz JB, Reetz K, Heber IA, Ramirez A, Dodel R (2016) Apolipoprotein E ε4 does not affect cognitive performance in patients with Parkinson's disease. *Parkinsonism Relat Disord* **29**, 112-116.

[11] Federoff M, Jimenez-Rolando B, Nalls MA, Singleton AB (2012) A large study reveals no association between APOE and Parkinson's disease. *Neurobiol Dis* **46**, 389-392.

[12] Sunwoo JS, Byun JI, Jun JS, Kim TJ, Shin JW, Kim HJ, Jung KY (2021) Apolipoprotein E ε4 is not associated with cognitive impairment in patients with idiopathic REM sleep behavior disorder. *Parkinsonism Relat Disord* **92**, 13-14.

[13] Gan-Or Z, Montplaisir JY, Ross JP, Poirier J, Warby SC, Arnulf I, Strong S, Dauvilliers Y, Leblond CS, Hu MTM, Högl B, Stefani A, Monaca CC, De Cock VC, Boivin M, Ferini-Strambi L, Plazzi G, Antelmi E, Young P, Heidbreder A, Barber TR, Evetts SG, Rolinski M, Dion PA, Desautels A, Gagnon JF, Dupré N, Postuma RB, Rouleau GA (2017) The dementia-associated APOE ε4 allele is not associated with rapid eye movement sleep behavior disorder. *Neurobiol Aging* **49**, 218.e13-218.e15.

[14] Borghammer P, Van Den Berge N (2019) Brain-first versus gut-first Parkinson's disease: a hypothesis. *J Parkinsons Dis* **9**, S281-S295.

[15] Horsager J, Andersen KB, Knudsen K, Skjærbæk C, Fedorova TD, Okkels N, Schaeffer E, Bonkat SK, Geday J, Otto M, Sommerauer M, Danielsen EH, Bech E, Kraft J, Munk OL, Hansen SD, Pavese N, Göder R, Brooks DJ, Berg D, Borghammer P (2020) Brain-first versus body-first Parkinson's disease: a multimodal imaging case-control study. *Brain* **143**, 3077-3088.

[16] Horsager J, Knudsen K, Sommerauer M. Clinical and imaging evidence of brain-first and body-first Parkinson's disease (2022) Clinical and imaging evidence of brain-first and body-first Parkinson's disease. *Neurobiol Dis* **164**, 105626.

[17] Hipp G, Vaillant M, Diederich NJ, Roomp K, Satagopam VP, Banda P, Sandt E, Mommaerts K, Schmitz SK, Longh-

ino L, Schweicher A, Hanff AM, Nicolai B, Kolber P, Reiter D, Pavelka L, Binck S, Pauly C, Geffers L, Betsou F, Gantenbein M, Klucken J, Gasser T, Hu MT, Balling R, Krüger R (2018) The Luxembourg Parkinson's Study: a comprehensive approach for stratification and early diagnosis. *Front Aging Neurosci* **10**, 326.

[18] Litvan I, Bhatia KP, Burn DJ, Goetz CG, Lang AE, McKeith I, Quinn N, Sethi KD, Shults C, Wenning GK, Movement Disorders Society Scientific Issues Committee (2003) Movement Disorders Society Scientific Issues Committee report: SIC Task Force appraisal of clinical diagnostic criteria for Parkinsonian disorders. *Mov Disord* **18**, 467-486.

[19] Antelmi E, Lippolis M, Biscarini F, Tinazzi M, Plazzi G (2021) REM sleep behavior disorder: Mimics and variants. *Sleep Med Rev* **60**, 101515.

[20] Ashraf-Ganjouei A, Moradi K, Aarabi M, Abdolalizadeh A, Kazemi SZ, Kasaeian A, Vahabi Z (2021) The association between REM sleep behavior disorder and autonomic dysfunction in Parkinson's disease. *J Parkinsons Dis* **11**, 747-755.

[21] Barasa A, Wang J, Dewey RB Jr (2021) Probable REM sleep behavior disorder is a risk factor for symptom progression in Parkinson disease. *Front Neurol* **12**, 651157.

[22] Long K, Wan C, Xiang Y, Liu J, Xu Q, Sun Q, Wang Z, Tian Y, Fang L, Yang Y, Yan X, Tang B, Guo J (2020) Study on the clinical features of Parkinson's disease with probable rapid eye movement sleep behavior disorder. *Front Neurol* **11**, 979.

[23] Liu Y, Zhu XY, Zhang XJ, Kuo SH, Ondo WG, Wu YC (2017) Clinical features of Parkinson's disease with and without rapid eye movement sleep behavior disorder. *Trans Neurodegener* **6**, 35.

[24] Liu Y, Lawton MA, Lo C, Bowring F, Klein JC, Querejeta-Coma A, Scotton S, Welch J, Razzaque J, Barber T, Ben-Shlomo Y, Hu MT (2021) Longitudinal changes in Parkinson's disease symptoms with and without rapid eye movement sleep behavior disorder: The Oxford Discovery Cohort Study. *Mov Disord* **36**, 2821-2832.

[25] Pavelka L, Rauschenberger A, Landoulsi Z, Pachchek S, May P, Glaab E, Krüger R; NCER-PD Consortium (2022) Age at onset as stratifier in idiopathic Parkinson's disease – effect of ageing and polygenic risk score on clinical phenotypes. *NPJ Parkinsons Dis* **8**, 102.

[26] Blauwendraat C, Faghri F, Pihlstrom L, Geiger JT, Elbaz A, Lesage S, Corvol JC, May P, Nicolas A, Abramzon Y, Murphy NA, Gibbs JR, Ryten M, Ferrari R, Bras J, Guerreiro R, Williams J, Sims R, Lubbe S, Hernandez DG, Mok KY, Robak L, Campbell RH, Rogaeva E, Traynor BJ, Chia R, Chung SJ; International Parkinson's Disease Genomics Consortium (IPDGC), COURAGE-PD Consortium, Hardy JA, Brice A, Wood NW, Houlden H, Shulman JM, Morris HR, Gasser T, Krüger R, Heutink P, Sharma M, Simón-Sánchez J, Nalls MA, Singleton AB, Scholz SW (2017) NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging* **57**, 247.e9-247.e13.

[27] Tunold JA, Geut H, Rozemuller JMA, Henriksen SP, Toft M, van de Berg WDJ, Pihlstrøm L (2021) APOE and MAPT are associated with dementia in neuropathologically confirmed Parkinson's disease. *Front Neurol* **12**, 631145.

[28] Davis AA, Inman CE, Wargel ZM, Dube U, Freeberg BM, Galluppi A, Haines JN, Dhavale DD, Miller R, Choudhury FA, Sullivan PM, Cruchaga C, Perlmutter JS, Ulrich JD, Benitez BA, Kotzbauer PT, Holtzman DM (2020) APOE genotype regulates pathology and disease progression in synucleinopathy. *Sci Trans Med* **12**, eaay3069.

[29] Gan-Or Z, Liong C, Alcalay RN (2018) GBA-associated Parkinson's disease and other synucleinopathies. *Curr Neurol Neurosci Rep* **18**, 44.

[30] Sixel-Döring F, Trautmann E, Mollenhauer B, Trenkwalder C (2011) Associated factors for REM sleep behavior disorder in Parkinson disease. *Neurology* **77**, 1048-1054.

[31] Bjørnarå KA, Dietrichs E, Toft M (2013) REM sleep behavior disorder in Parkinson's disease–is there a gender difference? *Parkinsonism Relat Dis* **19**, 120-122.

[32] Baumann-Vogel H, Hor H, Poryazova R, Valko P, Werth E, Baumann CR (2020) REM sleep behavior in Parkinson disease: Frequent, particularly with higher age. *PloS One* **15**, e0243454.

[33] Wong JC, Li J, Pavlova M, Chen S, Wu A, Wu S, Gao X (2016) Risk factors for probable REM sleep behavior disorder: A community-based study. *Neurology* **86**, 1306-1312.

[34] Haba-Rubio J, Frauscher B, Marques-Vidal P, Toriel J, Tobback N, Andries D, Preisig M, Vollenweider P, Postuma R, Heinzer R (2018) Prevalence and determinants of rapid eye movement sleep behavior disorder in the general population. *Sleep* **41**, zsx197.

[35] Kang SH, Yoon IY, Lee SD, Han JW, Kim TH, Kim KW (2013) REM sleep behavior disorder in the Korean elderly population: prevalence and clinical characteristics. *Sleep* **36**, 1147-1152.

[36] Postuma RB, Montplaisir JY, Pelletier A, Dauvilliers Y, Oertel W, Iranzo A, Ferini-Strambi L, Arnulf I, Hogl B, Manni R, Miyamoto T, Mayer G, Stiasny-Kolster K, Puligheddu M, Ju Y, Jennum P, Sonka K, Santamaria J, Fantini ML, Zucconi M, Leu-Semenescu S, Frauscher B, Terzaghi M, Miyamoto M, Unger MM, Cochen De Cock V, Wolfson C (2012) Environmental risk factors for REM sleep behavior disorder: a multicenter case-control study. *Neurology* **79**, 428-434.

[37] Gan-Or Z, Rao T, Leveille E, Degroot C, Chouinard S, Cicchetti F, Dagher A, Das S, Desautels A, Drouin-Ouellet J, Durcan T, Gagnon JF, Genge A, Karamchandani J, Lafontaine AL, Sun SLW, Langlois M, Levesque M, Melmed C, Panisset M, Parent M, Poline JB, Postuma RB, Pourcher E, Rouleau GA, Sharp M, Monchi O, Dupré N, Fon EA (2020) The Quebec Parkinson Network: a researcher-patient matching platform and multimodal biorepository. *J Parkinsons Dis* **10**, 301-313.

[38] Marek K, Chowdhury S, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, Simuni T, Jennings D, Tanner CM, Trojanowski JQ, Shaw LM, Seibyl J, Schuff N, Singleton A, Kieburtz K, Toga AW, Mollenhauer B, Galasko D, Chahine LM, Weintraub D, Foroud T, Tosun-Turgut D, Poston K, Arnedo V, Frasier M, Sherer T; Parkinson's Progression Markers Initiative (2018) The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Ann Clin Transl Neurol* **5**, 1460-1477.

[39] Mollenhauer B, Trautmann E, Sixel-Döring F, Wicke T, Ebentheuer J, Schaumburg M, Lang E, Focke NK, Kumar KR, Lohmann K, Klein C, Schlossmacher MG, Kohnen R, Friede T, Trenkwalder C; DeNoPa Study Group (2013) Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology* **81**, 1226-1234.

[40] Onofrj M, Thomas A, D'Andreamatteo G, Iacono D, Luciano AL, Di Rollo A, Di Mascio R, Ballone E, Di Iorio

A (2002) Incidence of RBD and hallucination in patients affected by Parkinson's disease: 8-year follow-up. *Neurol Sci* **23**, S91-S94.

[41] Gjerstad MD, Boeve B, Wentzel-Larsen T, Aarsland D, Larsen JP (2008) Occurrence and clinical correlates of REM sleep behaviour disorder in patients with Parkinson's disease over time. *J Neurol Neurosurg Psychiatry* **79**, 387-391.

[42] Lenka A, Hegde S, Jhunjhunwala KR, Pal PK (2016) Interactions of visual hallucinations, rapid eye movement sleep behavior disorder and cognitive impairment in Parkinson's disease: A review. *Parkinsonism Relat Disord* **22**, 1-8.