UNIVERSITÄT ZU LÜBECK

From the Institute of Neurogenetics
of the University of Lübeck
Director: Prof. Dr. Christine Klein

“Gene-environment modifiers in Parkinson’s disease”

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck
from the Department of Natural Sciences

Submitted by

Carolin Gabbert
from Heide

Lübeck 2024

First referee: Prof. Dr. Joanne Trinh

Second referee: Prof. Dr. Lars Redecke

Date of oral examination: 11.10.2024

Approved for printing. Lübeck, 03.04.2025

Table of Contents

ABSTRACT

Parkinson's disease (PD) is a multifaceted, highly complex disease and the fastest-growing neurodegenerative disorder. It is characterized by great heterogeneity and can manifest due to a variety of causes. Although the majority of patients with PD are idiopathic, about 15% of patients can be attributed to monogenic forms and genetic risk factors of PD. Monogenic forms of PD can be caused by a single pathogenic variant in, e.g., *LRRK2,* while genetic risk factors, e.g., variants in *GBA1*, are not directly causative of PD but are more frequently found in patients with PD than healthy controls. Although PD patients with the same genotype often present with a similar phenotype, wide ranges in age at onset (AAO), clinical severity, and disease progression have been found. Additional modifiers such as environmental and lifestyle factors can contribute to the variable expressivity in PD. However, our understanding of these factors and possible gene-environment and gene-lifestyle interactions is limited thus far.

The hypothesis of this doctoral thesis is that genetic as well as environmental and lifestyle factors can separately and jointly modify the susceptibility, AAO, and clinical severity in patients with PD. To test this hypothesis, the first objective was to investigate the association of the lifestyle factors smoking, consumption of caffeine, and intake of aspirin with AAO in patients with PD. Based on these findings, the second objective focused on analyzing the association of these lifestyle factors with clinical severity in PD by investigating motor and non-motor symptoms cross-sectionally. In the third objective, variants in *GBA1* were investigated by employing and evaluating Oxford Nanopore sequencing, determining the frequency of *GBA1* variants in a Norwegian study group of patients with PD and healthy controls, and summarizing frequencies of *GBA1* variants across populations in a systematic literature search. In the last objective, the combined effect of genetics and environmental and lifestyle factors on AAO was assessed. For this, associations between the cumulative genetic burden, represented by a PD-specific polygenic score (PGS) and nuclear-encoded genes associated with mitochondrial function (MGS), with environmental and lifestyle factors on AAO were investigated in patients with idiopathic PD, *GBA1*-PD, and *LRRK2*-PD.

The investigations showed that smoking (smokers: median AAO=63.5 years, *IQR*=56.1-69.1; non-smokers: median AAO=60.8 years, *IQR*=53.7-66.7; $p<0.0001$), coffee drinking (coffee drinkers: median AAO=61.9 years, *IQR*=54.7-67.6; non-coffee drinkers: median AAO=59.4 years, *IQR*=52.1-65.6; $p<0.0001$), and aspirin intake (aspirin users: median AAO=64.0 years, *IQR*=57.9-69.0; aspirin non-user: median AAO=59.1 years, *IQR*=51.8-64.9; $p<0.0001$) are associated with a later AAO in PD. In contrast, aspirin users and smokers had a higher probability of experiencing motor- and non-motor symptoms compared to non-users.

Long-read Nanopore sequencing has been shown to detect variants in *GBA1* accurately. Rare variants within *GBA1* were more frequently found in patients with PD (17.1%) than in healthy

controls (8.4%). In particular, the odds of carrying a p.L483P or p. N409S variant were estimated to be more than four times higher in patients with PD than in controls (*OR*=4.11 [1.39, 12.12]).

The investigations of polygenic scores showed that there are additive effects between tobacco use (*β*=3.16, *SE*=0.50, $p=3\times10^{-10}$) and PGS (*β*=-1.11, *SE*=0.24, $p=4\times10^{-6}$) on AAO and between aspirin intake (*β*=7.62, *SE*=1.48, $p=9\times10^{-7}$) and PGS (*β*=-1.58, *SE*=0.64, *p*=0.0149) on AAO in patients with idiopathic PD. In addition, in patients with idiopathic PD, an interaction between MGS and smoking (*β*=1.32, 95% *CI*=0.32 to 2.32, *p*=0.010) on AAO was found, while in patients with *LRRK2*-PD, an interaction between MGS and caffeinated soda (*β*=-5.65, 95% *CI*=-9.37 to -1.94, *p*=0.003) was found.

In conclusion, these results highlight that environmental and lifestyle factors are essential modifiers of AAO and clinical severity and that there are genotype-specific differences in how genetic and lifestyle factors interact to modify AAO in PD.

ZUSAMMENFASSUNG

Morbus Parkinson (MP) ist eine vielschichtige, hochkomplexe Erkrankung und die am schnellsten wachsende neurodegenerative Störung. Sie zeichnet sich durch große Heterogenität aus und kann durch eine Vielzahl von Ursachen ausgelöst werden. Obwohl die Mehrheit der Patienten mit MP idiopathisch ist, können etwa 15 % der Patienten auf monogene Formen und genetische Risikofaktoren für MP zurückgeführt werden. Monogene Formen von MP können durch eine einzige pathogene Variante, z. B. in *LRRK2*, verursacht werden, während genetische Risikofaktoren, z.B. Varianten in *GBA1*, nicht direkt ursächlich für MP sind, aber bei Patienten häufiger auftreten als bei gesunden Kontrollpersonen. Obwohl Patienten mit MP desselben Genotyps oft auch einen ähnlichen Phänotyp aufweisen, wurden große Unterschiede im Alter bei Krankheitsbeginn (AAO), im klinischen Schweregrad und im Krankheitsverlauf festgestellt. Zusätzliche Faktoren wie Umwelt- und Lebensstilfaktoren können zur variablen Ausprägung von MP beitragen. Unser Verständnis dieser Faktoren und möglicher Gen-Umwelt sowie Gen-Lebensstil Wechselwirkungeist ist bisher jedoch begrenzt.
Die Hypothese dieser Doktorarbeit lautet, dass sowohl genetische Faktoren als auch Umwelt- und Lebensstilfaktoren die Anfälligkeit, das AAO und den klinischen Schweregrad bei Patienten mit MP separat und gemeinsam beeinflussen können. Um diese Hypothese zu überprüfen, wurde zunächst der Zusammenhang zwischen den Lebensstilfaktoren Rauchen, Koffeinkonsum und Aspirineinnahme mit dem AAO bei Patienten mit MP untersucht. Auf der Grundlage dieser Ergebnisse konzentrierte sich das zweite Ziel auf die Analyse des Zusammenhangs zwischen diesen Lebensstilfaktoren und dem klinischen Schweregrad von MP, indem motorische und nicht-motorische Symptome im Querschnitt untersucht wurden. Im dritten Ziel wurden Varianten in *GBA1* untersucht, indem die Oxford-Nanopore-Sequenzierung angewandt und bewertet, die Häufigkeit von *GBA1*-Varianten in einer norwegischen Studiengruppe von Patienten mit MP und gesunden Kontrollpersonen bestimmt und die Häufigkeit von *GBA1*-Varianten in verschiedenen Populationen in einer systematischen Literatursuche zusammengefasst wurde. Im letzten Ziel wurde die gemeinsame Wirkung von Genetik, Umwelt- und Lebensstilfaktoren auf das AAO untersucht. Zu diesem Zweck wurden die Zusammenhänge zwischen der kumulativen genetischen Last, dargestellt durch einen MP-spezifischen polygenen Score (PGS), und nuklear kodierten Genen, die mit der mitochondrialen Funktion assoziiert sind (MGS), mit Umwelt- und Lebensstilfaktoren auf das AAO bei Patienten mit idiopathischem MP, *GBA1*-MP und *LRRK2*-MP untersucht.
Die Untersuchungen zeigten, dass Rauchen (Raucher: medianes AAO = 63,5 Jahre, *IQR* = 56,1 - 69,1; Nichtraucher: medianes AAO = 60,8 Jahre, *IQR* = 53,7 - 66,7; $p < 0,0001$), Kaffeetrinken (Kaffeetrinker: medianes AAO = 61,9 Jahre, *IQR* = 54,7 – 67,6; Nicht-Kaffeetrinker: medianes AAO = 59,4 Jahre, *IQR* = 52,1- 6 5,6; $p < 0,0001$) und

Aspirineinnahme (Aspirin-Nutzer: medianes AAO = 64,0 Jahre, *IQR* = 57,9 - 69,0; Aspirin-Nicht-Nutzer: medianes AAO = 59,1 Jahre, *IQR* = 51,8 - 64,9; $p < 0{,}0001$) mit einem späteren AAO bei MP verbunden sind. Im Gegensatz dazu hatten Aspirin-Nutzer und Raucher eine höhere Wahrscheinlichkeit, motorische und nicht-motorische Symptome zu haben, als Nicht-Nutzer.

Es hat sich gezeigt, dass die Long-Read-Nanopore-Sequenzierung Varianten in *GBA1* präzise erkennen kann. Seltene Varianten in *GBA1* wurden bei Patienten mit MP häufiger gefunden (17,1 %) als bei gesunden Kontrollpersonen (8,4 %). Insbesondere wurde die Wahrscheinlichkeit eine p.L483P- oder p.N409S-Variante zu tragen, bei Patienten mit MP mehr als viermal so hoch eingeschätzt wie bei Kontrollpersonen (OR = 4,11 [1,39, 12,12]).

Die Untersuchungen der polygenen Scores zeigten, dass es additive Effekte zwischen Tabakkonsum ($\beta = 3{,}16$, $SE = 0{,}50$, $p = 3\times10^{-10}$) und dem PGS ($\beta = -1{,}11$, $SE = 0{,}24$, $p = 4\times10^{-6}$) auf das AAO und zwischen Aspirineinnahme ($\beta = 7{,}62$, $SE = 1{,}48$, $p = 9\times10^{-7}$) und dem PGS ($\beta = -1{,}58$, $SE = 0{,}64$, $p = 0{,}0149$) auf das AAO bei Patienten mit idiopathischem MP gibt. Darüber hinaus wurde bei Patienten mit idiopathischem MP eine Interaktion zwischen dem MGS und Rauchen ($\beta = 1{,}32$, 95% *CI* = 0,32 bis 2,32, $p = 0{,}010$) mit dem AAO gefunden, während bei Patienten mit *LRRK2*-MP eine Interaktion zwischen dem MGS und koffeinhaltigem Soda ($\beta = -5{,}65$, 95% *CI* = -9,37 bis -1,94, $p = 0{,}003$) gefunden wurde.

Zusammenfassend unterstreichen diese Ergebnisse, dass Umwelt- und Lebensstilfaktoren wesentliche Einflussfaktoren auf das AAO und den klinischen Schweregrad sind und dass es genotypspezifische Unterschiede in der Art und Weise gibt, wie genetische und Lebensstilfaktoren zusammenwirken, um das AAO bei Morbus Parkinson zu verändern.

1. INTRODUCTION

1.1 Parkinson's disease – a complex neurodegenerative disorder

Parkinson's disease (PD) is a fast-growing progressive neurodegenerative disorder affecting approximately 10 million patients worldwide (Ball et al., 2019). It has been estimated that by 2040, more than 17 million people might be affected, suggesting the term "Parkinson pandemic" (Albin and Grotewold, 2023; Bloem et al., 2021; Dorsey et al., 2018). PD is a multifaceted and highly complex disease that can present with a variety of genotypes and phenotypes, which complicates the analysis and treatment of patients with PD. Given the disease heterogeneity and to better categorize PD, several subtype classifications of PD have been proposed (Greenland et al., 2019; Outeiro et al., 2023). Although there is still no consensus on the most valuable PD subtype classification, three main strategies have been used to define PD: genetically, pathologically, and clinically.

1.1.1 Genetic variation in PD – Monogenic forms of PD and their pathways

The majority of patients with PD are idiopathic, with no known genetic cause for the disease. However, about 15% of patients can be explained by monogenic forms of PD or risk variants in *GBA1* (Lim and Klein, 2024). Well-established PD genes with autosomal dominant modes of inheritance include *SNCA*, *LRRK2*, and *VPS35*, and genes with autosomal recessive forms include *PRKN*, *PINK1*, and *DJ1* (Jia et al., 2022; Lange et al., 2022). Pathogenic variants in these genes can lead to mitochondrial, proteasomal, and lysosomal dysfunction. Pathogenic variants within the *SNCA* gene, including p.Ala53Thr as the first pathogenic variant of PD identified (Polymeropoulos et al., 1997), and duplications and triplications of *SNCA* have been shown to stimulate misfolding and aggregation of α-synuclein, resulting in Lewy bodies (Balestrino and Schapira, 2020; Jia et al., 2022; Serratos et al., 2022). In addition, variants in *SNCA* have been implicated in impaired neuronal activity, reduced mitochondrial respiration, and altered lipid metabolism in dopaminergic neurons (Barbuti et al., 2021). The p.Gly2019Ser variant in *LRRK2* is the most common genetic cause of PD, with a prevalence of about 1% in the PD population and an even higher frequency in North African Arab Berber (39%) and Ashkenazi Jewish (18%) populations (Lesage et al., 2006; Ozelius et al., 2006). Vairants in *LRRK2* interfere with autophagy and have been reported to impair α-synuclein degradation, contributing to the increased accumulation (Ho et al., 2020). Other possible disease mechanisms include the impairment of mitochondrial function. Variants in *PRKN* and *PINK1* likely disturb PINK1/parkin-mediated ubiquitin-dependent mitophagy, which is responsible for the degradation of damaged mitochondria (Harper et al., 2018; Li et al., 2023).

In addition to the established PD genes, new genes with a possible link to PD are regularly investigated and accordingly classified (Lange et al., 2022; Marras et al., 2016). *CHCHD2* has recently been confirmed as a gene of hereditary parkinsonism with a classical form of PD (Lange et al. 2022). Additionally, a promising new PD gene candidate is *RAB32* (Gustavsson et al., 2024). The pathogenic variant RAB32 p.Ser71Arg has recently been reported as a novel genetic risk factor for PD with reduced penetrance (Gustavsson et al., 2024). Genetic testing may, therefore, help to characterize patients with PD.

1.1.2 Pathology and pathophysiology of PD

Neuropathologically, PD is characterized by dopaminergic neuronal loss in the substantia nigra and the presence of Lewy neurites and Lewy bodies, the major component of which is α-synuclein (Braak et al., 2003). The Braak *et al.* staging suggested that the pathology may spread in a predictable fashion as the disease progresses (Braak et al., 2003). It has been described that in the first stages of PD, the pathology often starts in the medulla and the olfactory bulb, leading to the appearance of non-motor features such as rapid eye movement sleep behavior disorder, hyposmia, constipation, urinary dysfunction, orthostatic hypotension, excessive daytime sleepiness, and depression (Armstrong and Okun, 2020; Berg et al., 2015). Pathology spreads as the disease progresses to the substantia pars compacta and other midbrain and basal forebrain structures (Armstrong and Okun, 2020). With the increasing loss of dopaminergic cells, the classical PD motor signs develop in the majority of patients, which include tremor, bradykinesia, rigidity, and imbalance. In later stages, cognitive impairment and hallucinations might additionally occur, together with an overall deterioration in health.

However, while this pathology might apply to most PD cases, not all patients with PD can be categorized based on these stages. There are reported cases of patients with PD who did not present with Lewy bodies or Lewy neurites (Jackson et al., 2024; Johansen et al., 2018). In contrast, cases with pathology graded at Braak stages 4-6 at postmortem with no evidence of neurological signs have been documented (Burke et al., 2008). Furthermore, synucleinopathies are not unique to PD but have also been implicated in other neurodegenerative diseases, e.g., Dementia with Lewy bodies (DLB), Multiple systems atrophy (MSA), Pure autonomic failure (PAF) and REM sleep behavior disorder (RBD) (Calabresi et al., 2023). Similarly, the aggregation and deposition of tau protein, which has been observed in about half of PD brains (Zhang et al., 2018), has been found in other tauopathies such as Alzheimer's disease (AD), progressive supranuclear palsy (PSP), and argyrophilic grain disease (AGD) (Pan et al., 2021).

Thus, alternative models that may explain the differences in pathological findings need to be considered. It has been suggested that α-synuclein pathology arises in a single location from where it spreads, which can be characterized by two different subtypes (Borghammer, 2023).

The first one is the clinical body-first subtype, which develops from a gut-first pathology and is characterized by prodromal autonomic symptoms and REM sleep behavior disorder. The second one is the body-first subtype resulting from an olfactory-first pathology and is associated with fewer non-motor symptoms before PD diagnosis but is at an increased risk of dementia (Borghammer, 2023). However, additional investigations are needed to test different aspects of these models.

1.1.3 Clinical heterogeneity

Another critical aspect demonstrating the considerable heterogeneity in PD is the variable age at onset (AAO). Age is the biggest risk factor for developing PD, showing an increasing prevalence of PD with age. The average AAO in PD is about 59 years (Grover et al., 2022), however, the range in AAO can span decades across different phenotypes in PD. An AAO under 50 years of age is considered early onset, while an AAO over 50 years of age is regarded as late onset. The incidence of late-onset PD is five to tenfold higher than that of early-onset PD (Poewe et al., 2017). The underlying mechanisms of a variable AAO remain largely unknown but suggest a major role of neuroinflammation, immune dysfunction, and oxidative stress (Bottigliengo et al., 2022; Pajares et al., 2020; Tansey et al., 2022). However, the extensive range in AAO is certainly affected by a variety of factors, including both genetic modifiers as well as environmental and lifestyle factors. It has further been suggested that the AAO can be linked to disease progression. Patients with late-onset PD have been described to have more frequently presented with motor- and non-motor symptoms after a disease duration of five years compared to patients with early-onset PD, indicating a faster disease progression (De Carolis et al., 2023). Motor and non-motor signs and symptoms cover a broad spectrum in PD, with not all symptoms present in every patient. For example, although the cardinal parkinsonism manifestation includes bradykinesia, rigidity, and rest tremor (Postuma et al., 2015), up to 20% of patients with PD do not have a tremor (Bloem et al., 2021). Two subtypes have been described based on the occurrence of motor signs: a tremor-dominant form and a nontremor-dominant phenotype. Patients with a predominant tremor form of PD have been observed to have a milder phenotype, to slower develop gait and postural problems, and to be less likely to develop cognitive impairment (Alves et al., 2006; Greenland et al., 2019; Jankovic et al., 1990; Williams-Gray et al., 2013). In contrast, patients who present with predominant bradykinesia and rigidity are more likely to experience a decline in mobility and postural stability and to develop early dementia (Greenland et al., 2019; Jankovic et al., 1990; Williams-Gray et al., 2007; Williams-Gray et al., 2013). In addition, patients with early postural instability and gait impairment are more likely to progress faster and to have more severe concurrent cognitive impairment (Greenland et al., 2019; Jankovic et al., 1990; Lewis et al.,

2005; Williams-Gray et al., 2013). However, there is an overlap between these phenotypes and patients can present with a combination of these motor- and non-motor signs and symptoms (Greenland et al., 2019).

Until today, there has been no cure for PD, however, pharmacological treatments help with motor- and non-motor signs and symptoms. These treatments are primarily dopamine-based, e.g., levodopa preparations, dopamine agonists, and monoamine oxidase-B inhibitors, and have been helpful for most patients with PD (Fox et al., 2018). However, to fully understand the variable clinical phenotypes in patients with PD and ultimately improve the treatment of patients, additional factors modifying AAO and clinical severity need to be thoroughly investigated.

1.2 Genetic data analysis – Investigation of genetic risk factors of PD

1.2.1 Brief review – The history of sequencing technologies

In genetic data analysis, sequencing technologies play an immense role, with a constant effort to develop faster and more accurate techniques. It started with the development of the first-generation DNA sequencing approaches, i.e., Sanger 'chain-termination' sequencing (Sanger et al., 1977). Sanger sequencing uses dideoxynucleotides (ddNTPs), which lack the 3′ hydroxyl group that is required for the extension of DNA chains, resulting in the production of DNA strands of each possible length. The sequence can be detected through radiolabeled ddNTPs that are part of the DNA extension mix (Heather and Chain, 2016). The efforts of automating DNA sequencing led to whole-genome shotgun sequencing approaches of the human genome, which represent the peak of the first-generation sequencing era (Venter et al., 2001).

Larger-scale dideoxy sequencing efforts led to the development of the pyrosequencing technique, which had the advantage of observation in real time. Pyrosequencing was later introduced as the first major success in next-generation sequencing (NGS) technology (Nyren and Lundin, 1985; Ronaghi et al., 1998). The NGS platforms all share a common feature: massive parallel sequencing of DNA molecules that are spatially separated in a flow cell and are performed by repeated cycles of polymerase-mediated nucleotide extensions or iterative cycles of oligonucleotide ligation (Voelkerding et al., 2009). The ability to sequence complete human genomes with NGS at comparably low costs increased the international effort to sequence 1000 genomes over the decade (1000 Genomes Project Consortium et al., 2015). The 1000 Genomes Project was founded that used high-density SNP microarrays for genotyping to discover genetic variants and estimate individual genotypes and haplotypes (1000 Genomes Project Consortium et al., 2010; 1000 Genomes Project Consortium et al., 2012).

In comparison to the short-read NGS technologies, third-generation sequencing, with Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) as the most prominent representatives, can achieve read length upwards of 10 kb (Hu et al., 2021). PacBio uses the single molecule real-time (SMRT) sequencing method (Ling et al., 2023). This technology uses a circular DNA molecular template, which comprises double-stranded DNA inserts and single-stranded hairpin adapters at both ends (Ardui et al., 2018). Each hairpin loop contains a binding site for the sequencing primers. Once the SMRTbell is assembled, it is bound by DNA polymerase and loaded onto an SMRT cell, a disposable chip with an array of 150,000 holes called zero-mode waveguides (ZMWs) (Rhoads and Au, 2015). During DNA replication, fluorescently labeled nucleotides bind to the growing complementary strand and emit fluorescent signals when excited by a laser, which are then analyzed to determine the sequence of each nucleotide. In contrast, ONT sequencing is based on the passage of a single-stranded nucleic acid through a nanoscale protein pore that is embedded in an electrically resistant polymer membrane (Figure 1) (Hu et al., 2021; Wang et al., 2021).

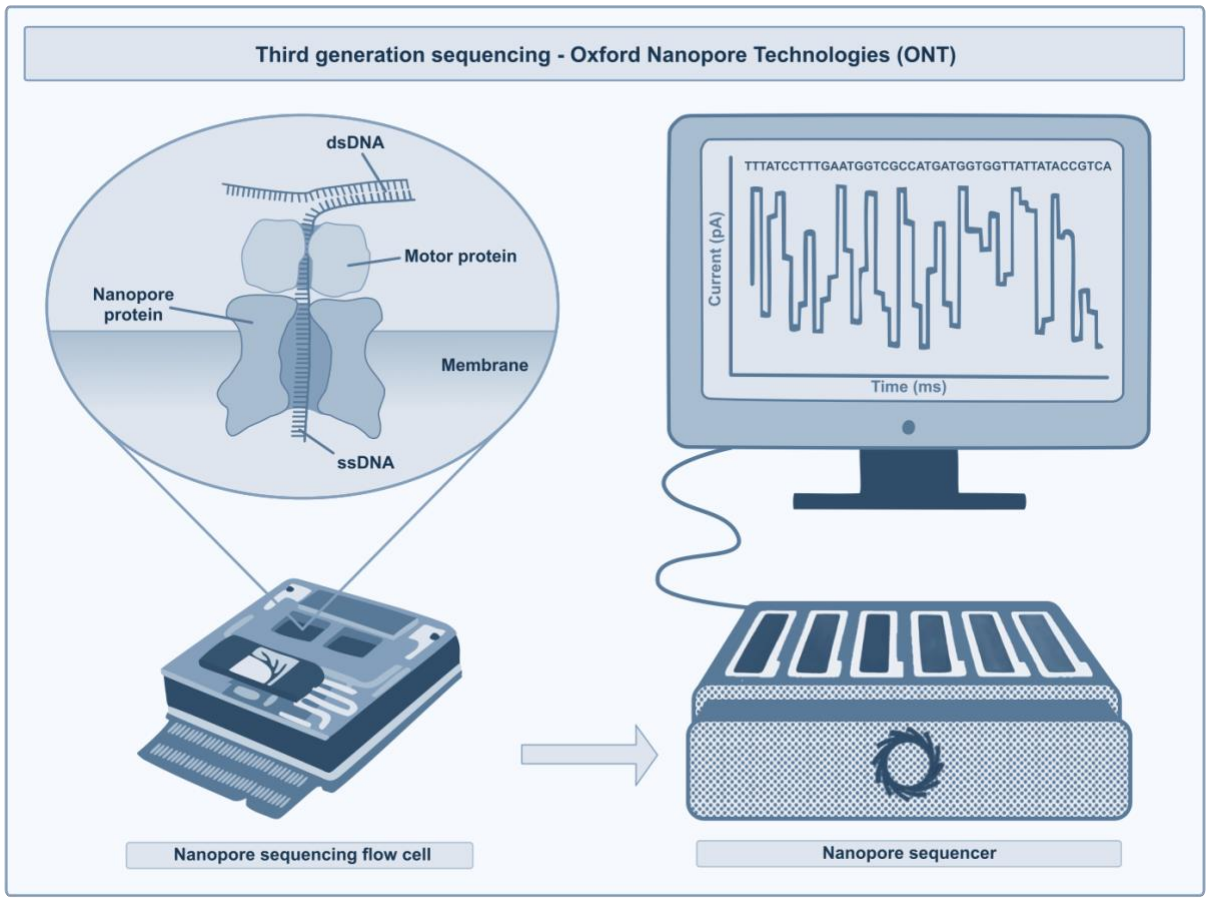


Figure 1: Overview of the Nanopore sequencing technology. The left side shows the structure of a Nanopore of an Oxford Nanopore Technologies (ONT) flow cell and the right side depicts the Nanopore sequencer with the bioinformatical analysis of the change current while the DNA strand passes through the pore (basecalling).

A constant voltage is applied to the pore to produce an ionic current through the nanopore that allows DNA molecules to pass, controlled by a motor protein. The passage of each nucleotide through the pore results in a characteristic disruption in ionic current, which is later decoded using computational algorithms, allowing real-time sequencing of single molecules. The different sequencing devices, i.e., MinION, GridION, and PromethION, differ in their flow cell capacity, allowing the generation of 50-300 Gb of long-read data. Long-read sequencing increases the possibilities of genetic data analysis vastly compared to short-read sequencing, enabling the study of structural variants, haplotypic phasing, and the assessment of DNA modifications across the genome with high precision.

1.2.2 Genome-wide association studies and genetic risk variants of PD

In the last two decades, there has been an increased interest in investigating genetic risk factors and their effect on the human body. Genome-wide association studies (GWAS) are an important research tool for identifying associations between single nucleotide polymorphisms (SNP) and phenotypic traits (Marees et al., 2018). The identification of trait-associated SNPs allows new insights into the biological mechanisms and pathways underlying these phenotypes and has been successfully used to reveal novel disease susceptibility genes (Marees et al., 2018; Tam et al., 2019). This way, GWAS have been used to identify genetic modifiers of phenotypic traits such as AAO (Laabs et al., 2021; Li et al., 2021). GWAS are mainly performed using data obtained from SNP arrays. However, genotype imputation is commonly done to increase the number of variants tested for association, which enables the analysis of low-frequency and rare variants that were not initially genotyped (Tam et al., 2019). Several reference panels are available, including large and ethnically diverse ones, e.g., the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015), the International HapMap Project (International HapMap, 2003), and the Haplotype Reference Consortium (HRC) (McCarthy et al., 2016). However, there are also limitations to genome-wide approaches, such as the need of high numbers of patients and controls for sufficient power to account for the multiple testing, usually setting a threshold of $p=5\times10^{-8}$ for associations to be considered significant, and the fact that GWAS do not necessarily identify causal variants (Tam et al., 2019).

Besides causative monogenic variants, which are rare variants with a large effect size, common variants have been more frequently found in patients with PD than healthy controls. Although these variants are not directly disease-causing, they may increase the risk for PD. In addition to the pathogenic rare variants, common variants in *SNCA* and *LRRK2* have been identified that alter PD risk. The variants rs356220, rs356219, rs6532194, rs356182 in *SNCA* and rs34778348, rs76904798, and rs1491942 in *LRRK2* showed significant associations with

PD in GWAS (Chang et al., 2017; Lill et al., 2012; Nalls et al., 2014; U. K. Parkinson's Disease Consortium et al., 2011). Two more variants in *LRRK2*, p.Gly2385Arg and p.Arg1628Pro, are further associated with a higher risk of PD in Asian populations, while they are rarely found among European populations (Billingsley et al., 2018; Di Fonzo et al., 2006; Farrer et al., 2007; Kim et al., 2010; Tan et al., 2007; Wang et al., 2012). Other variants found to be associated with PD risk include, e.g., rs2414739 in *VPS13C*, rs12637471 in *MCCC1*, rs3793947 in *DLG2*, rs7077361 in *ITGA8*, rs34016896 in *NMD3*, rs11158026 in *GCH1*, rs11724635 and rs4538475 in *BST1*, and rs7215239 and rs17649553 in *MAPT* (Chang et al., 2017; Lill et al., 2012; Nalls et al., 2014; U. K. Parkinson's Disease Consortium et al., 2011). However, while these variants were found to increase the risk for PD in GWAS, associations might be population-specific. In a large meta-GWAS of individuals of African and African admixed ancestry, rs3115534 in *GBA1* was associated with PD risk and AAO, which was not found in European populations (Rizig et al., 2023). The variant rs61204179 in *HEATR6* has further been reported as a Chinese-specific PD risk variant, which was rarely found in populations of European ancestry (Pan et al., 2023). In addition, GWAS only explain a fraction of the missing heritability, underlining the importance of further investigating gene-environment and gene-gene interactions. The majority of complex traits, in particular PD, are most likely polygenic and, therefore, influenced by thousands of genetic variants, each having a small effect (Dudbridge, 2016).

1.2.3 Cumulative genetic burden – Polygenic scores for PD

The polygenic scoring method uses GWAS datasets to estimate the effect size of each SNP and to select SNPs with a major effect. The weighted sum of risk alleles at the selected SNPs is used in the study dataset to calculate the polygenic score (PGS) (Dudbridge, 2016; International Schizophrenia et al., 2009). The largest meta-GWAS of PD to date, involving the analysis of 7.8 million SNPs in 37,700 cases, 18,600 proxy cases, and 1.4 million controls, identified 90 independent risk loci across 78 genomic regions (Nalls et al., 2019). In this study, a PGS comprising 1805 variants was estimated, which explained about 26% of PD heritability. A variety of polygenic scores have been calculated based on this and other GWAS to investigate the risk of developing PD, the penetrance of PD, associations with AAO, response to deep brain stimulation, or other phenotypic traits (Huang et al., 2024; Iwaki et al., 2020; Koch et al., 2021; Liu et al., 2022; Pan et al., 2023; Sia et al., 2021; Yoon et al., 2023). Especially PGS, which focus on PD-related pathways instead of the general risk of PD, can add new insight into disease mechanisms. These pathway-dependent PGS are based on genetic variants assigned to genes involved in biological pathways and mechanisms related to the disease development in PD, e.g., signal transduction mechanisms affecting protein

misfolding and aggregation, vesicular-mediated transport, and lysosomal or mitochondrial function (Bandres-Ciga et al., 2020). Lysosomal PGS have been validated to predict PD status and have further been associated with a faster progression of cognitive decline in PD (Dehestani et al., 2022; Tunold et al., 2024). Since respiratory chain, mitophagy, and mitochondrial biogenesis impairment are associated with PD (Grunewald et al., 2019), genes linked to mitochondrial function can be used to evaluate PD pathogenesis. Similar to lysosomal PGS, mitochondrial PGS have been shown to predict PD status and additionally correlate with an earlier AAO in patients with PD (Arena et al., 2024; Billingsley et al., 2019; Dehestani et al., 2022). Thus, the cumulative genetic burden may have an essential impact on the PD phenotype.

1.2.4 PD risk variants in GBA1

Variants in *GBA1* are considered the most common genetic risk factor of PD and occur in about 3-20% of patients, depending on the population (Blauwendraat et al., 2018; Blauwendraat et al., 2020; Gan-Or et al., 2015; Lesage et al., 2011; Skrahina et al., 2021). The *GBA1* gene is an 8.9 kb large gene located on chromosome 1, which comprises 11 exons and 10 introns (Figure 2). The nearby pseudogene *GBAP1*, which shares approximately 96% exonic sequence homology with *GBA1*, is located 6.9 kb downstream of *GBA1* (Toffoli et al., 2022). The existence of this pseudogene increases the risk of recombination between homologous regions (Leija-Salazar et al., 2019).

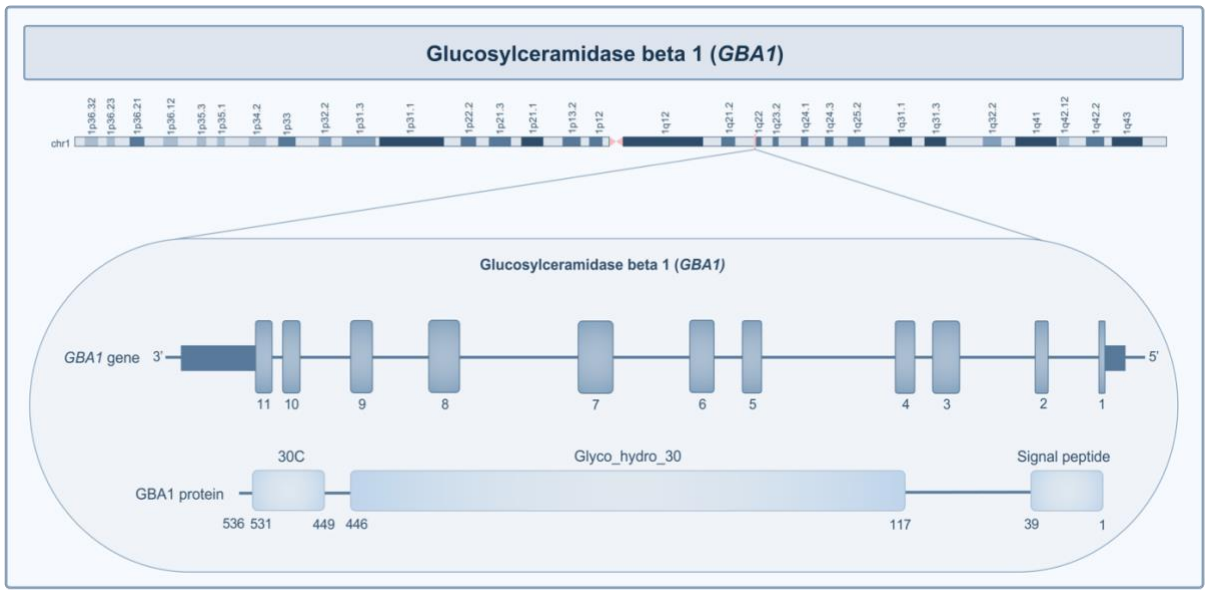


Figure 2 Schematic representation of the structure of the *GBA1* gene. The structure of the exons (dark gray boxes) and introns (gray lines) of the *GBA1* gene and its location within chromosome 1. The corresponding GBA1 protein is composed of a 39-residue signal peptide, the conserved catalytic domain Glyco_hydro_30, and Glyco_hydro_30C domain.

GBA1 encodes the lysosomal membrane enzyme glucocerebrosidase (GCase), which cleaves the beta-glucosidic linkage of glucosylceramide. It has been suggested that loss of GCase activity reduces the ability to degrade α-synuclein, encoded by *SNCA* (Blauwendraat et al., 2020).

Homozygous and compound heterozygous variants in *GBA1* are classically a cause for Gaucher's disease (GD), with almost 300 *GBA1* variants identified in patients with GD (Hruska et al., 2008; Vieira and Schapira, 2021). However, multiple heterozygous variants in *GBA1* have been further shown to increase the risk of PD. Two variants, p.Glu365Lys and p.Thr408Met, are solely associated with PD but do not cause GD (Greuel et al., 2020). Other common variants include p.Asn409Ser and p.Leu483Pro, with varying frequencies depending on the ethnicity and population (Blauwendraat et al., 2018; Leija-Salazar et al., 2019; Sidransky et al., 2009). Patients with *GBA1*-PD have a similar phenotype compared to idiopathic PD. However, variant carriers present with an earlier AAO and a more rapid progression of motor impairment and cognitive decline than patients without *GBA1* variants (Barkhuizen et al., 2016; Cilia et al., 2016; Sidransky et al., 2009).

1.3 Lifestyle and environmental factors as modifiers of PD

1.3.1 Historical background – first studies on environmental factors in PD

The characteristics of what is today known as Parkinson's disease were first described by James Parkinson more than 200 years ago (Parkinson, 2002). In his book, James Parkinson suggested that environmental influences might cause the disease. The impact of environmental and lifestyle factors on the risk of developing PD has already been investigated for several decades. In the late 1960s and early 1970s, an association between smoking and PD showed a decreased risk of death from PD for smokers and ex-smokers in comparison to non-smokers (Hammond, 1966; Kahn, 1966). In addition, patients with PD were less likely to have ever smoked and among smokers, patients with PD tend to smoke less than the controls (Kessler and Diamond, 1971). In 1983, a report of cases with a sudden onset of parkinsonism after using an illicit drug was published (Langston, 2017; Langston et al., 1983). The patients showed signs of generalized slowing and difficulty in moving with near total immobility, marked generalized increase in tone, a complete inability to speak intelligibly, marked diminution of blinking, constant drooling, and cogwheel rigidity in the upper extremities within a week after using the drug, which were immediately reversed after treatment with a combination of L-dopa and carbidopa. The substance was identified as 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a byproduct of the synthesis of 1-methyl-4-phenyl-4-propionoxypiperidine (MPPP). This lipophilic compound may pass the blood-brain barrier, resulting in a conversion into 1-methyl-4-phenylpyridinium (MPP+), which has been confirmed as the toxic metabolite

responsible for the Parkinsonian symptoms (Langston et al., 1984; Markey et al., 1984). An analog of MPP+, which differs by only one methyl group, is paraquat, one of the most widely used herbicides worldwide since 1962. Paraquat is highly poisonous, leading to heart, lung, and liver failure, as well as severe cerebral damage (Grant et al., 1980). It has further been presented with similar but reduced effects as MPTP, e.g., a decreased brain dopamine concentration (Barbeau et al., 1985) and specific features of PD (Betarbet et al., 2002). These findings show that the influence of environmental and lifestyle factors on PD pathogenesis, susceptibility, and progression may not be negligible.

1.3.2 Protective environmental and lifestyle factors in PD and their mechanism of action

Most environmental and lifestyle factors can be grouped into predominantly protective or predominantly risk factors with regard to PD susceptibility and AAO. The role of tobacco use and smoking as a protective factor for PD susceptibility and AAO has been frequently reported with a smaller percentage of smokers among PD cases compared to controls, a later AAO among smokers compared to non-smokers, and less prevalent motor symptoms in smokers than non-smokers (De Reuck et al., 2005; Gigante et al., 2017; Jacobs et al., 2020; Kessler and Diamond, 1971; McCulloch et al., 2008). These observations were further confirmed in large meta-studies (Grover et al., 2019; Larsson and Burgess, 2022; Noyce et al., 2012), which additionally reported that a genetic liability to smoking was associated with a decreased risk of PD. The underlying mechanism of tobacco use and smoking that leads to the protective effect has not been fully elucidated. However, there is broad consensus that the neuroprotective effect of tobacco can be mainly attributed to nicotine, a naturally occurring alkaloid from tobacco that has consistently shown beneficial pro-cognitive effects in cellular and animal models of PD (Parain et al., 2003; Ruan et al., 2023; Yang et al., 2019). Nicotine has a neuroprotective effect on dopaminergic neurons via the stimulation of central nicotinic cholinergic receptors, resulting in the release of dopamine in the brain and reducing the production of reactive oxygen species, which may protect against nigrostriatal damage and PD progression (Barreto et al., 2014; Benowitz, 2009). It is further involved in an anti-inflammatory mechanism mediated by decreasing microglial activation, leading to a significant reduction of tumor necrosis factor (TNF)-alpha mRNA expression and TNF-alpha release induced by LPS stimulation (Park et al., 2007). In mouse models, nicotine has been shown to decrease the electron leak at the site of respiratory chain complex I in mitochondria (Xie et al., 2005) and suppress astrocyte apoptosis via the mitochondrial pathway through the stimulation of α7-nicotinic acetylcholine receptors (Liu et al., 2015). Other protective lifestyle factors are the caffeinated beverages coffee and black tea, which have been consistently described as protective for patients with PD concerning a smaller risk for PD and a later AAO (Ahmed et al.,

2014; Benedetti et al., 2000; McCulloch et al., 2008; Noyce et al., 2012). Caffeine is an adenosine analog and acts as a nonselective adenosine receptor antagonist, which helps to block the adenosine receptors, protects against neurodegeneration, and lowers the risk of PD (Bata-Garcia et al., 2010; Hu et al., 2007; Kalda et al., 2006; Nehlig et al., 1992; Singh et al., 2010). Caffeine administration has also been reported to reduce LPS-induced microglia activation in three regions of the hippocampus in a dose-dependent manner (Brothers et al., 2010) and to halt LPS-induced neuroinflammation and synaptic dysfunction (Badshah et al., 2019). Similar to tobacco and coffee, the use of non-steroidal anti-inflammatory drugs (NSAIDs) has been associated with a decreased risk and a later AAO in patients with PD (Gagne and Power, 2010; Gao et al., 2011; Noyce et al., 2012; San Luciano et al., 2020). However, the evidence is rather limited compared to smoking and coffee (Manthripragada et al., 2011; Rees et al., 2011; Samii et al., 2009). In general, inflammation, especially in the context of aging, is a major factor in the pathogenesis and progression of PD. Excess low-level production of circulating inflammatory mediators, including the C-reactive protein (CRP), IL-6, and TNF from chronically stimulated innate and adaptive immune cells are characteristic of the inflammation (Tansey et al., 2022). Anti-inflammatory drugs might, therefore, be crucial in controlling and delaying the inflammatory mechanisms leading to neurodegeneration. Aspirin has been shown to increase ferroportin 1 (Fpn1) expression by inhibiting hepcidin expression via the IL-6/JAK/STAT3 pathway (Huang et al., 2018). The increased expression of Fpn1 induced an increase in iron release from the cells and a reduction in iron content inside of the cells, leading to a reduced risk for the cells to be injured by iron-mediated oxidative stress and free radical reaction (Huang et al., 2018). Aspirin and other NSAIDs have further been exhibited to diminish the decrease in dopamine uptake caused by glutamate, indicating preservation of neuronal integrity and protection against excitotoxicity (Casper et al., 2000). In addition, there is evidence that ibuprofen increases the relative number of dopaminergic neurons (Casper et al., 2000). Another lifestyle factor that can be classified as both protective and harmful, often depending on the dosage, is alcohol consumption (Bettiol et al., 2015; Eriksson et al., 2013; Jacobs et al., 2020; Jimenez-Jimenez et al., 2019; Liu et al., 2020; Zhang et al., 2014). Although low alcohol intake might be considered protective, heavy alcohol drinking can increase the risk for PD. Chronic heavy alcohol intake may lead to elevated glutamate-induced excitotoxicity, oxidative stress, and neuronal damage due to a decrease in dopamine levels and an increase in α-synuclein (Peng et al., 2020). In addition to environmental factors, the patient's lifestyle might influence the PD risk. Physical activity has been frequently shown to reduce the risk of PD and to improve symptoms by reducing the accumulation of α-synuclein protein, inflammation, and oxidative stress while enhancing nerve regeneration and mitochondrial function (Fan et al., 2020).

1.3.3 Environmental risk factors in PD and their pathways

In contrast to the protective environmental and lifestyle factors, several factors have been shown to increase the risk of PD (Figure 3). One of the best-studied environmental risk factors for PD is pesticides, with paraquat and rotenone as the most prominent representatives, which have been reported to increase the risk for PD up to 7-fold and lead to an earlier AAO in patients with PD (Elbaz et al., 2004; Fitzmaurice et al., 2014; Gamache et al., 2019; Ratner et al., 2014; Rosler et al., 2018; Tanner et al., 2011).

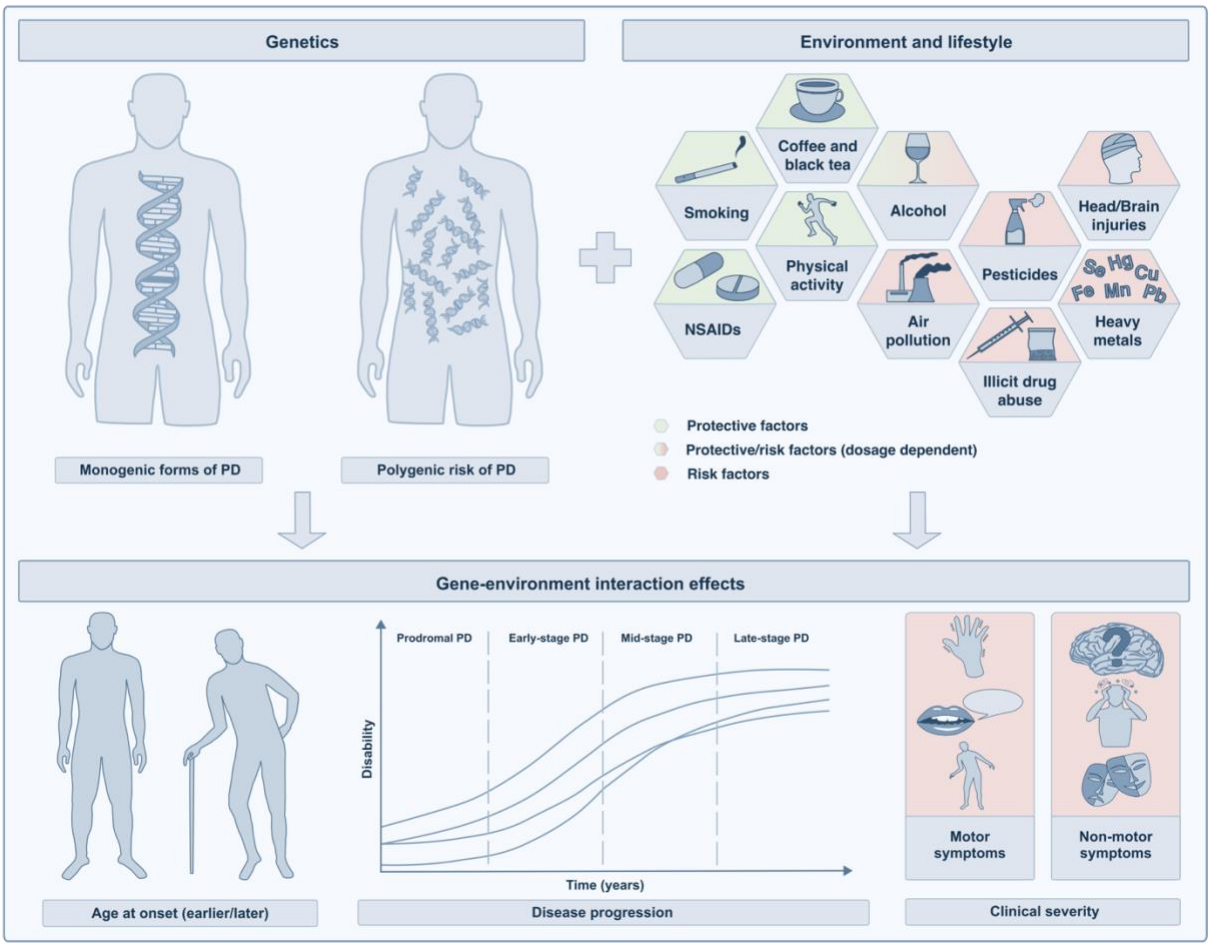


Figure 3: Overview of possible gene-environment interaction effects. The upper left panel represents genetic variations, the upper right panel environmental and lifestyle factors that may be protective or risk factors, and the bottom panel depicts possible effects of genetics and environmental and lifestyle factors including earlier or later age at onset, an altered disease progression and differences in clinical severity.

Paraquat, which belongs to the group of oxidative stressors, has been demonstrated to cause the degeneration of dopaminergic neurons in the substantia nigra due to the generation of reactive oxygen species, the decrease in antioxidant enzyme levels, neuroinflammation, mitochondrial dysfunction, and ER stress (See et al., 2024). This affects multiple molecular pathways, including mitogen-activated protein kinase (MAPK), phosphatidylinositol-4,5-

bisphosphate 3-kinase (PI3K)/AKT, mammalian target of rapamycin (mTOR), and Wnt/β-catenin signaling pathways, resulting in the initiation of apoptosis (See et al., 2024). Rotenone, which belongs to the group of mitochondrial respiratory chain complex I inhibitors, has been shown to increase oxidative stress, generate microtubule depolymerization, and induce α-synuclein aggregation, leading to neuronal death (Ibarra-Gutierrez et al., 2023; Ren et al., 2005; Shin and Chung, 2020). Similar effects as for pesticides have also been reported for heavy metals, with patients exposed to toxic metals showing an earlier AAO compared to patients with PD who were not exposed to metals (Gamache et al., 2019; Ratner et al., 2014). Heavy metals, e.g., cobalt, nickel, mercury, chromium, and thallium, can be included in α-synuclein aggregation, resulting in oxidative stress, DNA damage, impairment of mitochondrial function, and death of dopaminergic neurons (Vellingiri et al., 2022). As an additional comorbidity that increases the risk of PD, traumatic brain injuries are associated with a higher risk of developing PD (Jafari et al., 2013). Following an acute head injury, a neuroinflammatory response sets in involving the alteration of microglia, astrocytes, oligodendrocytes, and endothelial cells due to the release of pro- and anti-inflammatory cytokines, neurotrophic factor modulation, phagocytosis, and accumulation of pathological proteins, e.g., amyloid β peptide and hyperphosphorylated tau (Brett et al., 2022).

1.3.4 Gene-environment interactions in PD

Expanding our knowledge of how genetic and environmental factors interact can be beneficial in understanding PD pathogenesis and the various phenotypes of PD better. However, replications of reported gene-environment interactions are limited. There is evidence of several interactions between genes and tobacco use and smoking. For example, *SNCA REP1*, rs4240705 in *RXRA*, and rs1900586 in *SLC17A6* have all been reported to interact with smoking (Lee et al., 2018; McCulloch et al., 2008) and additionally showed that smoking was significantly less frequent in patients with PD than controls among variant carriers. A gene-smoking interaction was further found for *SV2C* with PD risk, however, no association between *SV2C* and the smoking habit was detected (Hill-Burns et al., 2013). An inverse association between *HLA-DRB1* and smoking has repeatedly been reported. An interaction was found between rsrs660895 in *HLA-DRB1* and smoking in patients with PD (Chuang et al., 2017), which has also been replicated (Domenighetti et al., 2022a). However, while a strong protective effect for PD was found for the combination of a positive history of smoking together with valine at amino acid position 11 in *HLA-DRB1* (Hollenbach et al., 2019), the opposite effect was reported in another study, which found an inverse association between *HLA-DRB1* in genetically predicted smoking initiation in PD only in the absence of valine at amino acid position 11 (Domenighetti et al., 2022a). It has further been suggested that nicotine may

interact with other genes, such as *SIRT6*, which is a member of the sirtuin family and promotes apoptosis in numerous cell types, leading to neurodegeneration (Nicholatos et al., 2018). Nicotine can inhibit SIRT6 activity and prevent neuron death for those suffering from PD and other neurodegenerative disorders. There is also a close association between the cholinergic and dopaminergic systems in the striatum (Zhou et al., 2002). Interactions with coffee have been revealed in interaction tests with APOE (McCulloch et al., 2008). Significant interactions have also been reported for *ADORA2A* rs5760423 with heavy vs. light coffee consumption in incident but not prevalent PD (Chuang et al., 2016). A large GWAIS has been performed that predicted that signals for PD-associated SNPs in *GRIN2A* were enhanced by SNP*coffee interaction (Hamza et al., 2011). However, these findings could not be replicated thus far (Ahmed et al., 2014; Hamza et al., 2014). Besides a potential effect on PD susceptibility, an additional study showed that the *GRIN2A* T allele was associated with more rapid clinical progression of PD among subjects taking creatine with high levels of caffeine intake, but not among those with low caffeine intake, although no significant association between *GRIN2A* rs4998386 genotype and caffeine intake was detected (Simon et al., 2017).

Many studies have previously investigated gene-environment interactions with pesticides that have frequently been reported to act neurotoxic and induce mitochondrial dysfunction. For example, an association of exposure to ALDH-inhibiting pesticides with an increased PD risk has been described, which was further exacerbated in subjects with genetic variation in *ALDH2* (Fitzmaurice et al., 2014). The variant rs1803274 in *BCHE*, which is not directly associated with PD, has been shown to interact with pesticide exposure, including insecticides but not herbicides (Rosler et al., 2018). In addition, in carriers of the minor allele of rs1803274, pesticide exposure significantly increased the risk of PD compared to individuals with the same genotype who were not exposed (Rosler et al., 2018). In a Taiwanese study group, the variants A340T in *PINK1* and V66M in *BDNF* both showed interactions with past pesticide exposure and influenced the occurrence of PD (Lin et al., 2011). More interactions have also been reported in ambient residential and workplace for *APEX1* rs1130409 and *OGG1* rs1052133 with pesticide exposure of the class of oxidative stressors, which increased the risk for PD for pesticide-exposed risk allele carriers (Sanders et al., 2017). Significantly different ORs among subjects with high ambient pesticide exposure at both residence and workplace were also reported for *SKP1* rs2284312, who either carried at least one T allele of *SKP1* rs2284312 or the CC genotype (Rhodes et al., 2013).

A major concern of reported gene-environment interactions is that these are often not replicated and might vary in different study groups, PD genotypes and phenotypes, and ethnic backgrounds. The interaction between *HLA-DRB1* and smoking has been replicated (Chuang et al., 2017; Domenighetti et al., 2022a), while the interaction between *GRIN2A* and coffee was not replicated (Ahmed et al., 2014; Hamza et al., 2011; Hamza et al., 2014). Most reported

gene-environment interactions are still in need of independent replication. In addition, while gene-environment interactions are most frequently investigated in the context of PD risk, associations with AAO in PD, motor signs, and non-motor symptoms may further be of interest to unravel this diverse disease. It is well known that not only patients with idiopathic PD but also with monogenic forms of PD may present with a wide range of AAO and varying clinical severity, which strengthens the role of environmental and lifestyle factors as essential modifiers of PD susceptibility and AAO. The relationship between lifestyle factors and AAO and clinical severity has been understudied in the past, especially in combination with the possible interactions with genetic factors. Thus, investigating these factors is of great importance in identifying the underlying pathogenic mechanisms and finally revealing new therapeutic strategies.

1.4 Public data resources

Publicly available data platforms are an essential research tool for the PD community and investigators. They allow access to large PD cohorts providing clinical, genetic, and environmental data. There is a variety of large consortia, data platforms, and networks that aim to investigate PD to improve the patient's quality of life. One of them is the Michael J. Fox Foundation for Parkinson's Research (MJFF), which is committed to accelerating PD treatments and ultimately finding a cure for PD (The Michael J. Fox Foundation, https://www.michaeljfox.org/our-promise). Therefore, MJFF funds several activities and projects to improve the lived experience of patients with PD, family members, and caregivers. One of their data platforms is the Fox Insight Data Exploration Network. Fox Insight is an online longitudinal study that contains data from almost 55,000 patients with PD and controls, with a targeted enrollment set to at least 125,000 individuals. It provides surveys on health and disease, motor and non-motor assessments, daily activities and exercises, quality of life, environmental and lifestyle questionnaires, and other factors relevant to the field of PD (The Michael J. Fox Foundation, https://www.michaeljfox.org/fox-insight). Cross-sectional and longitudinal assessments are further combined with genotyping data through 23andMe (Smolensky et al., 2020). Another MJFF-sponsored platform is the Parkinson's Progression Markers Initiative (PPMI), an international, multicenter study designed to identify PD progression biomarkers to improve understanding of disease etiology and to enhance the likelihood of success of PD modifying therapeutic trials (Parkinson Progression Marker, 2011). The study comprises patients with PD and healthy controls who have been followed longitudinally since 2010 for clinical, imaging, and biospecimen biomarker assessment using standardized data acquisition protocols. Biological samples include blood, cerebrospinal fluid, and urine that can be applied for by scientists separately from the data.

The Accelerating Medicines Partnership in Parkinson's Disease (AMP-PD) is a large program between the National Institute of Neurological Disorders and Stroke (NINDS), National Institute on Aging (NIA), the Food and Drug Administration (FDA), GSK, Pfizer, Sanofi, Bristol-Myers Squibb, Verily, MJFF, Aligning Science Across Parkinson's (ASAP) initiative and AbbVie (The Accelerating Medicines Partnership in Parkinson's Disease (AMP-PD), https://amp-pd.org/about). The foundation of AMP-PD is predicated on PD-related study data that is collected and harmonized to identify new biomarkers and develop new treatments for PD (The Accelerating Medicines Partnership in Parkinson's Disease (AMP-PD), https://amp-pd.org/unified-cohorts). Data is available through the collective efforts of the MJFF and National Institutes of Neurological Disorders and Stroke (NINDS) BioFIND study, Harvard Biomarkers Study (HBS), the NIA International Lewy Body Dementia Genetics Consortium Genome Sequencing in Lewy body dementia case-control cohort (LBD), the MJFF LRRK2 Cohort Consortium (LCC), the NINDS Parkinson's disease Biomarkers Program (PDBP), MJFF PPMI, the NINDS Study of Isradipine as a Disease Modifying Agent in Subjects With Early Parkinson Disease, Phase 3 (STEADY-PD3), and the Study of Urate Elevation in Parkinson's Disease, Phase 3 (SURE-PD3). The centrally harmonized AMP-PD data includes cross-sectional assessments, including clinical, transcriptomic, genomic, and proteomic data.
One last example is the Aligning Science Across Parkinson's (ASAP) initiative that aspires to better understand the underlying causes of PD (Aligning Science Across Parkinson's, https://parkinsonsroadmap.org/#). The Global Parkinson's Genetics Program (GP2) is a resource program of the ASAP initiative focused on improving our understanding of the genetic architecture of PD and making this knowledge globally relevant (Global Parkinson's Genetics Program, https://gp2.org/about-gp2/). GP2 works toward diversifying our understanding of PD by collecting samples and data from around the world, specifically with the goal of including groups traditionally underrepresented in genetics research. Thus far, the program has assembled more than 100 cohorts from around the world with the ultimate goal of collecting and genotyping more than 150,000 unique samples.

1.5 Hypothesis and Objectives

Hypothesis: Genetic and environmental and lifestyle factors can separately and jointly modify the susceptibility, AAO, and clinical severity in patients with Parkinson's disease.
This hypothesis can further be separated into four objectives (Figure 4):
Objective 1: To investigate the association of the lifestyle factors smoking, consumption of caffeine, and the use of aspirin with AAO in patients with PD.
Objective 2: To analyze the association of these lifestyle factors with motor and non-motor symptom progression cross-sectionally in patients with PD.

Objective 3: To characterize variants in *GBA1* in patients with PD and healthy controls by employing and evaluating Oxford Nanopore sequencing.

Objective 4: To examine the AAO association of polygenic scores with environmental and lifestyle factors.

A) Using a PD-specific PGS to assess the AAO association and the combined effect of coffee drinking, tobacco use, and aspirin intake in PD.

B) Using a mitochondrial polygenic score to investigate the impact of the MGS and the tobacco use, caffeine consumption, and pesticide exposure on the AAO in PD.

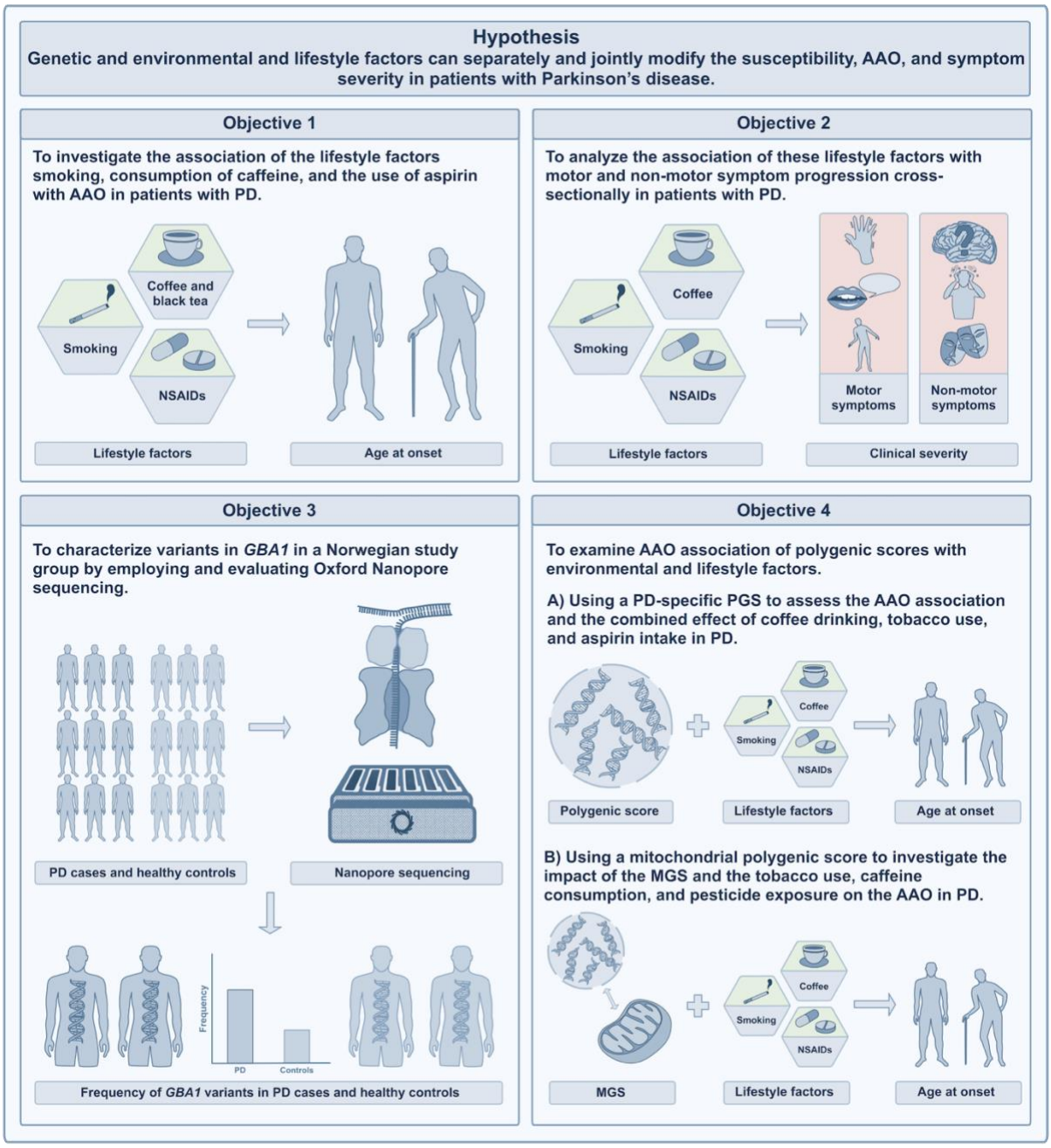


Figure 4: Overview of the hypothesis and the objectives of the thesis. The aims of objectives 1-4 are displayed in the four panels.

2. RESULTS WITH RELATED PUBLICATIONS

2.1 Objective 1: The association of smoking, coffee, and aspirin use with AAO in patients with iPD

PD onset is widely variable and can be influenced by both genetic and environmental factors. With regard to possible future treatments, the investigation of potential protective factors is of great interest to the research community. Association studies are a valuable assessment tool for investigating the relationship between lifestyle factors and AAO (Benedetti et al., 2000; Delamarre and Meissner, 2017; Gigante et al., 2018; Kandinov et al., 2009; Yahalom et al., 2020). This includes large study groups, in particular, as these provide sufficient statistical power. One of these is the Fox Insight study, an online health study that cross-sectionally assesses factors relevant to the field of PD (Smolensky et al., 2020).

In this study, the relationship between the environmental and lifestyle factors coffee, black tea, smoking, NSAIDs, and pesticides with AAO was investigated in patients with PD from the Fox Insight cohort. For this, non-parametric Mann-Whitney *U* tests were used to compare the median AAO between the different groups of lifestyle factor use, and non-parametric Spearman's correlations and multi-linear regression analysis were used to assess the relationship between variables. Although some of the investigated environmental and lifestyle factors did not show an association with AAO, i.e., black tea, ibuprofen, and pesticides in a working setting, three lifestyle factors showed a correlation with a later AAO when investigated binary (yes-no indication), in a dosage-dependent manner, and as duration of use in years: smoking, coffee drinking, and aspirin intake. Patients with PD who regularly drank coffee, smoked, or took aspirin had a 2- to 5-year later AAO compared to patients who did not consume these substances. In a sensitivity analysis, which additionally included the covariates age at examination (AAE), gender, and potential comorbidities (i.e., lung diseases for smoking and heart diseases, arthritis, back pain, and surgeries with anesthesia for aspirin intake) in the regression models, the protective association between coffee, smoking, and aspirin with AAO was further demonstrated. In a systematic literature review, associations between smoking, coffee, and aspirin with AAO were investigated. These included the assessment of 332 publications in total. After screening, 21 articles were found that investigated AAO associations with smoking, 11 articles described associations between coffee and AAO, and no articles assessed AAO associations with aspirin. The majority of studies showed that smoking and coffee have presented a protective effect on the AAO in PD. However, no other study has previously explored aspirin and AAO in a large PD cohort and found an association. Therefore, a replication analysis in patients with PD from the EPIPARK cohort, a German cohort of patients with idiopathic PD, was performed, showing comparable results as in Fox Insight with

an on average 6-year later AAO in patients with PD who take aspirin regularly compared to non-users.

The study covering this objective is titled “Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson’s disease”. For this, I downloaded all relevant data sheets from the Fox Data Exploration Network, e.g., the “Environmental Exposure Questionnaire” for all lifestyle factors, “General”, “About You”, “Your Current Health”, and “Health History”, and harmonized the data accordingly for all >35,000 participants. Using the environmental and lifestyle data questionnaires, I estimated consumers of lifestyle factors and calculated the duration of usage until the AAO by including all periods the patients consumed either substance. This did not only include the described environmental and lifestyle factors smoking, caffeinated beverages (i.e., coffee and black tea), and NSAIDs (i.e., aspirin, ibuprofen, and other NSAIDs), but also included the exposure to pesticides, which was not included in the final paper as the definition of pesticide exposure did not allow to assess the association with AAO sufficiently. For the replication of the aspirin analysis, I received data from the EPIPARK study. Beke Kolms, whom I supervised on this project during her Master’s thesis, and I prepared the lifestyle factor data and estimated the duration of lifestyle factor use. After some initial statistical overview performed by Prof. Inke R. König, I performed all multi-linear regression models used in the paper in SPSS and the Mann-Whitney *U* tests and Spearman’s correlations in GraphPad Prism. To compare our results to previous findings, I performed a systematic literature review on environmental factors influencing AAO in PD in PubMed. All articles were pre-sorted by Beke Kolms and then filtered and summarized by me. Finally, I wrote the first draft of the manuscript, including all tables, figures, the Supplementary Material, and the response letters in the subsequent peer-review process, supervised and revised by PD Dr. Joanne Trinh. The review process further included extensive changes to the manuscript that led to the separation into two independent publications, which I additionally performed.

Title:

Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson’s disease

Authors:

Carolin Gabbert[1], MSc, Prof. Inke R. König[2], Theresa Lüth[1], MSc, Beke Kolms[1], BSc, Meike Kasten[1,3], MD, Eva-Juliane Vollstedt[1], MD, Alexander Balck[1], MD, Anne Grünewald[1,4], PhD, Christine Klein[1], MD, Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
[2]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany
[3]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany
[4]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[*]Corresponding author

Published in *Journal of Neurology*, 2022, DOI: doi.org/10.1007/s00415-022-11041-x

Abstract:

Parkinson's disease (PD) is a progressive neurodegenerative disorder. Genetic modifiers, environmental factors, and gene-environment interactions have been found to modify PD risk and disease progression. The objective of this study was to evaluate the association of smoking, caffeine, and anti-inflammatory drugs with age at onset (AAO) in a large PD cohort. A total of 35,963 American patients with idiopathic PD (iPD) from the Fox Insight Study responded to health and lifestyle questionnaires. We compared the median AAO between different groups using the non-parametric Mann-Whitney *U* test. Non-parametric Spearman's correlation was used for correlation assessments and regression analysis was used to assess interaction between variables. We found that smoking ($p<0.0001$), coffee drinking ($p<0.0001$), and aspirin intake ($p<0.0001$) show an exploratory association with AAO in PD that was further supported by multivariate regression models. The association of aspirin with PD AAO was replicated in another cohort (EPIPARK) ($n=237$ patients with PD).

Introduction:

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by dopaminergic neuronal loss in the substantia nigra and the presence of Lewy Bodies (Bloem et al., 2021; Braak et al., 2003). It is the second-most common neurodegenerative disorder and the fastest-growing neurological disease, currently affecting over 7 million patients worldwide (Dorsey et al., 2018).

A phenomenon in PD is variable age at onset (AAO) that is considered a consequence of genetic and environmental factors. Tobacco use and smoking are already known protective factors for PD risk (Delamarre and Meissner, 2017; Gallo et al., 2019; Mappin-Kasirer et al., 2020). However, research specifically on AAO is not as extensive. Studies report that disease onset in patients with idiopathic or monogenic PD is later among smokers, dependent on the

dosage (De Reuck et al., 2005; Gigante et al., 2017; Kandinov et al., 2009; Luth et al., 2020; Martino et al., 2017; Wijeyekoon et al., 2017; Yahalom et al., 2020). The largest cross-sectional cohort was comprised of 715 PD patients, of whom 312 were smokers and 404 never smoked (Gallo et al., 2019). Likewise, caffeine consumption was associated with lower PD risk, with a dosage-dependent level of protection (Ascherio et al., 2001). In terms of AAO, there is evidence that the onset of PD among coffee drinkers is later compared to non-drinkers (Benedetti et al., 2000; Tan et al., 2007; Wijeyekoon et al., 2017), also indicating a dosage effect (Gigante et al., 2018; Yahalom et al., 2020). However, earlier studies report opposing effects of an earlier AAO with higher coffee intake (Kandinov et al., 2009). Non-steroidal anti-inflammatory drug (NSAID) intake has been found to be associated with a lower risk for PD (Wahner et al., 2007), supporting work that describe a role for neuro-inflammatory signaling in PD (Chen et al., 2005). NSAIDs (ibuprofen and aspirin) have been found to influence the penetrance of *LRRK2* (San Luciano et al., 2020). However, there are currently no studies published that investigate an association between aspirin and AAO in idiopathic PD (iPD).
Herein, we focused on lifestyle factors implicated in PD risk and investigated the association of smoking, the consumption of caffeine, and the use of aspirin on AAO in patients with iPD. We hypothesize these factors are associated with AAO in a large cohort of American iPD patients (*n*=35,963).

Methods:

Demographics

Our study is composed of 35,963 American patients with PD (Supplementary Table S1) from the Fox Insight Study (Supplementary text and Supplementary Figure S1). Due to the nature of the data collection and accessibility via an online data platform, some entries were highly unlikely or impossible; thus, we excluded PD patients with an AAO lower than 3 years. Most of the patients were White/Caucasian (89.9%) (Supplementary Table S1). PD patients had a mean age at examination (AAE) of 65.7±10.2 SD years (range 13.8-119.0 years) and a mean AAO of 60.4±11.0 SD years (range 5.1-115.4 years); 40.4% of PD patients were female. Patient recruitment for the Fox Insight Study has been previously described (Smolensky et al., 2020). Data from a separate replication cohort of German iPD patients (EPIPARK) were used to test novel associations (Kasten et al., 2013). In the EPIPARK cohort, PD patients had a mean AAE of 67.7±10.3 SD years (range 30.0-90.0 years) and a mean AAO of 54.8±13.2 SD years (range 13.0-81.0 years); 37.3% of PD patients were female. Participant questionnaires are described in detail in the Supplementary text.

Lifestyle factors

Patients were classified as tobacco users if they smoked more than 100 cigarettes in their lifetime or if they smoked at least one cigarette per day over a minimal period of 6 months or if they used smokeless tobacco at least once per day for more than 6 months. Patients were classified as coffee consumers if they regularly drank caffeinated coffee at least once per week over a period of at least 6 months. The same classification was used for caffeinated black tea. Lastly, patients were classified as aspirin users if they took at least two pills per week over a minimum of 6 months.

Duration of smoking, caffeine consumption, and aspirin intake were estimated according to the age the patients started using either substance subtracted from the age at termination. If the patients terminated the consumption after their AAO, the age the patients started was subtracted from their AAO. Periods where the patients stopped regularly consuming were not included in the duration. Smoking dosage was estimated as cigarettes smoked per day within smoking duration time excluding implausible values, so that only values lower than 100 cigarettes per day were included in the analyses. Coffee and black tea dosage was defined as cups per week the patients drank within drinking duration time, excluding all values higher than 100 cups per week from the analysis. Aspirin dosage was defined as pills per week the patients took within aspirin intake duration time. The number of cigarettes for nonsmokers, cups of coffee or black tea for non-drinkers, and pills per week for aspirin non-users was set to zero.

Statistical analysis

For statistical analyses, non-parametric Mann-Whitney *U* test was performed to compare the distribution of AAO between different groups. For correlation analyses, non-parametric Spearman's correlations and linear regression analyses were used to assess correlations and interactions between variables (GraphPad Software Inc., San Diego, CA, USA). Various multi-linear regression models were used to investigate the relationship between environmental factors, age, gender, and potential comorbidities (IBM SPSS Statistics, Stanford, CA, USA) (details of each model are in Supplementary text). Reported *p*-values remain descriptive because they are not corrected for multiple testing and results are exploratory. Patients with missing data on AAO or use of environmental and lifestyle factors were not included in the analyses.

Regression models

Regression model investigating AAO, AAE, environmental factors (binary/dosage/duration)

Age is considered a risk for PD and affects the general dosage and duration of environmental factors. We applied a multiple regression model using AAO as dependent variable and AAE and each environmental factor as covariates. Environmental factors were handled in three different ways: (1) binary (yes-no indication), (2) dosage as a continuous variable, and (3) duration as a continuous variable (IBM SPSS Statistics). Including age at examination as a covariate improves the understanding of how it might influence our models (details in Supplementary text).

Regression model investigating AAO, AAE, gender, environmental factors (binary/dosage/duration) and comorbidities

We estimated a multiple regression model using AAO as dependent variable and further variables as covariates: AAE, gender, and each environmental factor handled in three different ways: (1) binary (yes-no indication), (2) dosage as a continuous variable, and (3) duration as a continuous variable. For the investigation of smoking and aspirin, several potential comorbidities (lung diseases; heart diseases, arthritis, back pain and surgeries with anesthesia) were explored (IBM SPSS Statistics) (details in Supplementary text).

Regression model investigating AAO and combined environmental factors

To evaluate whether the environmental factors show a combined effect, this multiple regression model used the AAO as dependent variable and smoking, coffee drinking, and aspirin intake, all handled binary, as covariates (details in Supplementary text).

Regression model investigating AAO, AAE, gender, combined environmental factors and comorbidities

To evaluate potential confounders and a possible combined effect of all three environmental factors, this multiple regression model adjusted for more covariates, using the AAO as dependent variable and AAE, gender, a selected comorbidity (back pain), and smoking, coffee drinking, and aspirin intake as covariates (details in Supplementary text).

Literature review

We performed a systematic literature review for environmental factors that influence AAO in PD and searched for literature via PubMed that was published before December 14, 2021. We used the free text search terms “Parkinson onset smoking”, resulting in 221 articles,

"Parkinson onset caffeine", resulting in 46 articles, "Parkinson onset coffee", resulting in 52 articles, and "Parkinson onset aspirin", resulting in 13 articles. Detailed descriptions of exclusion criteria are summarized in Supplementary Figure S2.

Results:

Smoking

Patients with iPD, who reported use of tobacco, had a later AAO (*n*=2148; median AAO=63.5 years; *IQR*=56.1-69.1) compared to non-users (*n*=3375; median AAO=60.8 years; *IQR*=53.7-66.7) (*p*<0.0001) (Figure 1A and Table 1). Investigation of possible smoking dosage effects on AAO showed that the number of cigarettes per day was associated with later AAO (*n*=4399, *r*=0.08, *p*<0.0001) (Figure 1B), despite a small correlation strength. Similarly, a longer duration of smoking showed a positive correlation with AAO (*n*=912, *r*=0.07, *p*=0.0328) (Figure 1C), but again with a small correlation strength.

We investigated whether AAE contributed to the correlation between smoking and AAO, as age is considered a risk for PD. When modeled in a linear regression to predict AAO, AAE ($p<1\times10^{-5}$, $\beta>0.9277$, *SE*<0.0168), smoking (binary) (*p*=0.0002, β=0.5354, *SE*=0.1424) and smoking dosage (*p*=0.0016, β=0.0172, *SE*=0.0055) remained in the model as independent predictors, but the smoking duration (*p*=0.5583, β=0.0074, *SE*=0.0127) did not (Supplementary Table S2).

When investigating smoking, we found a correlation between dosage and duration ($p<1\times10^{-5}$, *r*=0.25).

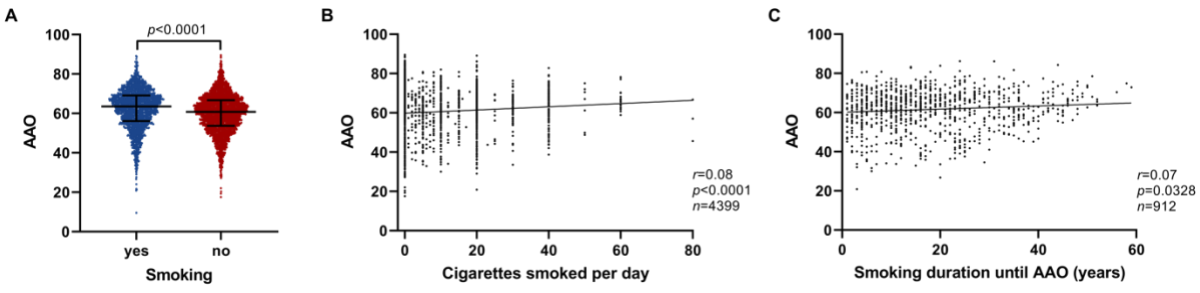


Figure 1: Association of AAO and tobacco use, smoking intensity and smoking duration in iPD. a Scatter plot of AAO of patients with iPD stratified by smoking status. Median values and interquartile ranges (*IQR*) are depicted. **b** Correlation between number of cigarettes smoked per day and AAO of patients with iPD. **c** Correlation between number of years of smoking until AAO and AAO of patients with iPD. *p*-value: exploratory Mann-Whitney *U* test was performed for pairwise comparisons; non-parametric Spearman's correlation and simple linear regression analyses were used to assess interactions between variables; *p*=Spearman's exploratory *p*-value, *r*=Spearman's rank correlation coefficient

Table 1: Association of environmental factors and AAO.

	Yes	**No**	***p*-value**
Tobacco			
n	2148	3375	NA
Median AAO (*IQR*)	63.5 (56.1-69.1)	60.8 (53.7-66.7)	<0.0001
Coffee			
n	3993	1133	NA
Median AAO (*IQR*)	61.9 (54.7-67.6)	59.4 (52.1-65.6)	<0.0001
Black tea			
n	1719	2449	NA
Median AAO (*IQR*)	61.0 (54.1-66.7)	61.3 (53.5-67.2)	0.8228
Aspirin			
n	1003	1989	NA
Median AAO (*IQR*)	64.0 (57.9-69.0)	59.1 (51.8-64.9)	<0.0001

Median AAO stratified by tobacco use, coffee consumption, black tea consumption, and aspirin intake.

To evaluate more potential predictors of AAO, we performed a sensitivity analysis. When modeled in a linear regression to predict AAO (Supplementary text), with covariates smoking binary/dosage or duration, AAE, gender, and lung diseases, smoking (binary) (p=0.0005, β=0.5051, SE=0.1456) and smoking dosage (p=0.0030, β=0.0165, SE=0.0055) showed a positive association with AAO. However, smoking duration (p=0.5741, β=0.0074, SE=0.0131) was not found to be associated with AAO. A positive relationship for AAO with AAE ($p<1\times10^{-5}$, β>0.9254, SE<0.0179) was also observed. We also tested for lung diseases, including chronic obstructive pulmonary disease (COPD) as potential comorbidity, but these did not show any association with AAO (p>0.7642, β>0.0053, SE<0.4432) (Supplementary Table S2).

Caffeine

Patients with iPD who drank coffee regularly had a later AAO (n=3993; median AAO=61.9 years; *IQR*=54.7-67.6) compared to patients with iPD who did not drink coffee at all (n=1133; median AAO=59.4 years; *IQR*=52.1-65.6) (p<0.0001) (Figure 2A and Table 1). Investigation of a possible coffee dosage effect revealed that the number of cups of coffee per week was associated with AAO, although the correlation strength was small (n=4028, r=0.10, p<0.0001) (Figure 2B). Longer coffee drinking duration also showed a positive correlation with AAO (n=2051, r=0.69, p<0.0001) (Figure 2C).

We investigated whether AAE contributed to the correlation between coffee drinking and AAO. When modeled in a linear regression to predict AAO, AAE ($p<1\times10^{-5}$, β>0.8239, SE<0.0122), coffee drinking (binary) ($p<1\times10^{-5}$, β=0.9176, SE=0.1704), coffee drinking dosage ($p=8\times10^{-5}$, β=0.0309, SE=0.0078) and coffee drinking duration ($p<1\times10^{-5}$, β=0.1268, SE=0.0083) all remained in the model as independent predictors (Supplementary Table S2).

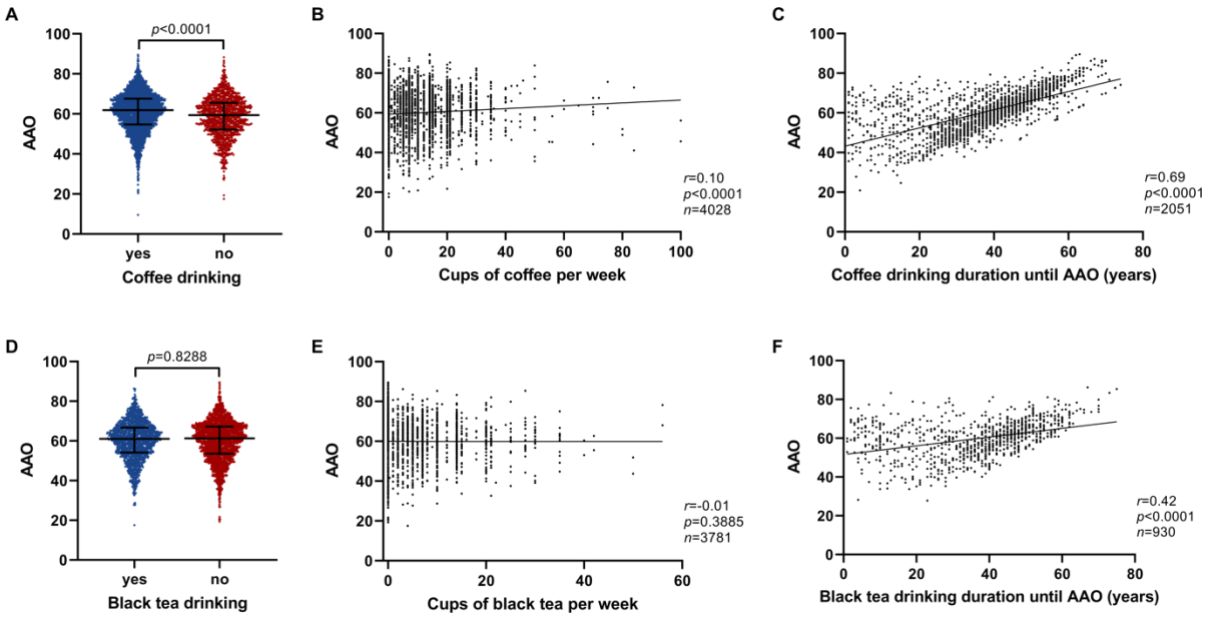


Figure 2: Association of AAO and caffeine consumption, caffeine drinking intensity and caffeine drinking duration in iPD. a Scatter plot of AAO of patients with iPD stratified by coffee consumption. Median values and interquartile ranges (*IQR*) are depicted. **b** Correlation between number of cups of coffee per week and AAO of patients with iPD. **c** Correlation between number of years of coffee drinking until AAO and AAO of patients with iPD. **d** Scatter plot of AAO of patients with iPD stratified by black tea consumption. **e** Correlation between number of cups of black tea per week and AAO of patients with iPD. **f** Correlation between number of years of black tea drinking until AAO and AAO of patients with iPD. *p*-value: exploratory Mann-Whitney *U* test was performed for pairwise comparisons; nonparametric Spearman correlation and simple linear regression analyses were used to assess interactions between variables; *p*=Spearman's exploratory *p*-value, *r*=Spearman's rank correlation coefficient

Again, coffee drinking dosage and duration were correlated ($p<1\times10^{-5}$, $r=0.16$).

We performed a sensitivity analysis to evaluate more potential predictors of AAO. When modeled in a linear regression to predict AAO (Supplementary text), with covariates coffee drinking binary/dosage or duration, AAE and gender, a positive relationship with coffee drinking (binary) ($p<1\times10^{-5}$, $\beta=0.9379$, $SE=0.1750$), coffee drinking dosage ($p=0.0001$, $\beta=0.0321$, $SE=0.0081$) and coffee drinking duration ($p<1\times10^{-5}$, $\beta=0.1276$, $SE=0.0084$) was revealed. In addition, a positive relationship for AAO with AAE ($p<1\times10^{-5}$, $\beta>0.8237$, $SE<0.0125$) was also observed (Supplementary Table S2).

In contrast to the findings for coffee and AAO, black tea drinking was not observed to be associated with AAO (Figure 2D and Table 1). There was also no association between the number of cups of black tea per week and AAO ($n=3781$, $r=-0.01$, $p=0.3885$) (Figure 2E). However, there was a positive correlation of black tea drinking duration with AAO ($n=930$, $r=0.42$, $p<0.0001$) (Figure 2F).

Aspirin

When investigating the effect of anti-inflammatory medication on AAO of patients with iPD, aspirin showed the greatest difference in AAO. Patients with iPD, who reported the use of aspirin, had a 5-year later AAO (*n*=1003; median AAO=64.0 years; *IQR*=57.9-69.0) compared to patients who did not take aspirin (*n*=1989; median AAO=59.1 years; *IQR*=51.8-64.9) (*p*<0.0001) (Figure 3A and Table 1). The difference in AAO for ibuprofen-based non-aspirin medication was small (ibuprofen users: *n*=1087; median AAO=60.6 years; *IQR*=53.2-66.3; ibuprofen non-users: *n*=2008; median AAO=61.1 years; *IQR*=54.2-67.0; *p*=0.0345) or in the case of other anti-inflammatory medication we found no association at all (other anti-inflammatory drug users: *n*=498; median AAO=61.5 years; *IQR*=54.0-66.9; other anti-inflammatory drug non-users: *n*=2393; median AAO=60.7 years; *IQR*=53.7-66.6; *p*=0.2495). As the association of ibuprofen was not as strong as the association with aspirin, we focused on aspirin and AAO for a more in-depth analysis.

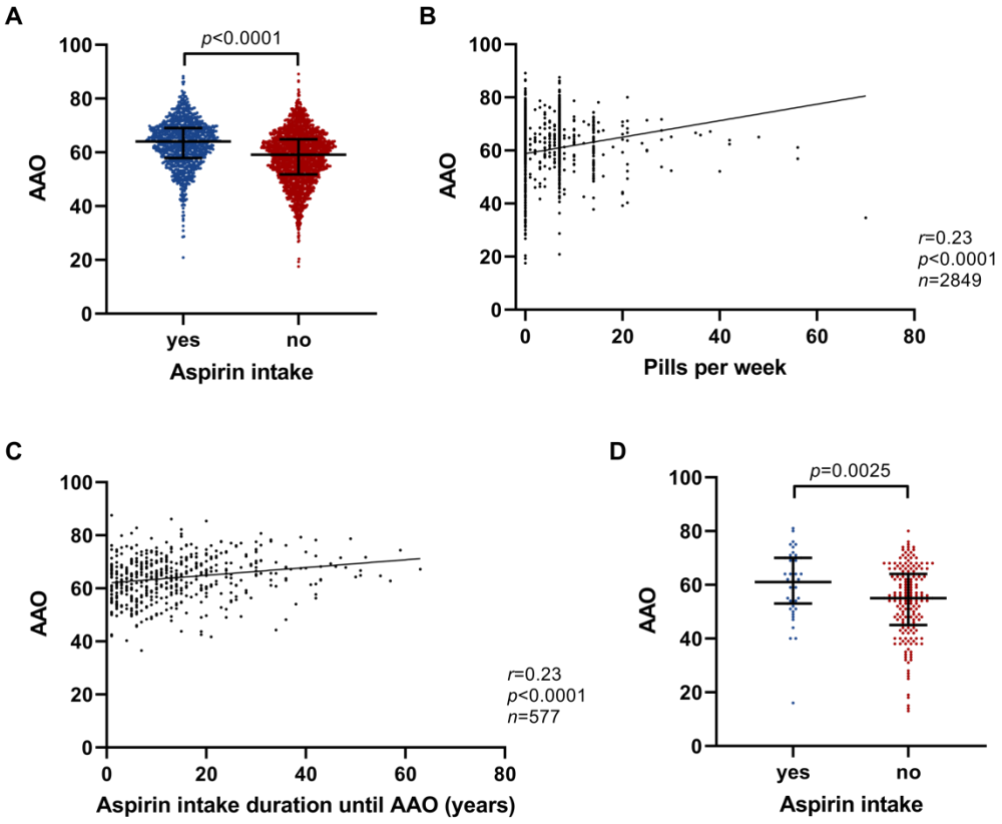


Figure 3: Association of AAO and aspirin intake, aspirin intake intensity and aspirin intake duration in iPD. a Scatter plot of AAO of patients with iPD stratified by aspirin intake. Median values and interquartile ranges (*IQR*) are depicted. **b** Correlation between number of aspirin pills per week and AAO of patients with iPD. **c** Correlation between number of years of aspirin intake until AAO and AAO of patients with iPD. **d** Scatter plot of AAO of patients with iPD from the EPIPARK replication cohort stratified by aspirin intake. *p*-value: exploratory Mann-Whitney *U* test was performed for pairwise comparisons; non-parametric Spearman correlation and simple linear regression analyses were used to assess interactions between variables; *p*=Spearman's exploratory *p*-value, *r*=Spearman's rank correlation coefficient

The number of aspirin pills per week was associated with AAO (n=2849, r=0.23, p<0.0001) (Figure 3B). Likewise, the aspirin intake duration showed an association with AAO (n=577, r=0.23, p<0.0001) (Figure 3C), indicating a later AAO the longer the patients took aspirin before disease onset.

When examining the effect of AAE on aspirin intake and AAO by modeling in a linear regression to predict AAO, AAE (p<1×10^{-5}, β>0.9195, SE<0.0198), aspirin intake (binary) (p=9×10^{-5}, β=0.7654, SE=0.1958) and aspirin intake duration (p=0.0165, β=0.0319, SE=0.0133) remained in the model but the aspirin dosage diminished as independent predictor (p=0.0972, β=0.0315, SE=0.0190) (Supplementary Table S2).

To evaluate more potential predictors of AAO, we performed a sensitivity analysis. When modeled in a linear regression to predict AAO (Supplementary text), with covariates aspirin intake binary/dosage or duration, AAE, gender, and potential comorbidities (heart diseases/arthritis/ back pain/surgeries with anesthesia), aspirin intake (binary) showed a positive relationship with AAO (p<0.0008, β>0.6732, SE<0.2063) as well as aspirin intake duration (p<0.0153, β>0.0338, SE=0.0140). In contrast to this, aspirin intake dosage was not associated with AAO (p>0.1188, β<0.0303, SE<0.0197). However, in all aspirin intake models, a positive relationship for AAO with AAE (p<1×10^{-5}, β>0.9193, SE<0.0213) was observed. In addition, in the models that included aspirin intake (binary) or aspirin intake dosage as covariate, a negative relationship for back pain with AAO (p<0.0264, β<−0.4144, SE<0.1867) was found (Supplementary Table S2).

Replication cohort

Since the aspirin and PD AAO association has not been investigated and published previously, we utilized a separate German iPD cohort to investigate further. In the EPIPARK cohort, patients with iPD who reported the use of at least one aspirin pill per week over a minimal period of one month had a more than 6-year later AAO (n=49; median AAO=61.0 years; IQR=53.0-70.0) compared to patients who did not take aspirin (n=188; median AAO=55.0 years; IQR=45.0-64.0) (p=0.0025) (Figure 3D).

Combined effect of smoking, coffee drinking, and aspirin intake

To investigate whether there is a combined effect of smoking, coffee drinking, and aspirin intake, we used a linear regression model to predict AAO (Supplementary text), showing that all three factors smoking (binary) (p<1×10^{-5}, β=1.8261, SE=0.3767), coffee drinking (binary) (p<1×10^{-5}, β=2.5233, SE=0.4158), as well as aspirin intake (binary) (p<1×10^{-5}, β=4.8768, SE=0.3698), remained in the model as independent predictors (Supplementary Table S2).

To consider more potential predictors of AAO and comorbidities, we performed a sensitivity analysis (Supplementary text). When modeled in a linear regression to predict AAO, with covariates smoking (binary), coffee drinking (binary), aspirin intake (binary), AAE, gender, and back pain, we found a positive relationship with smoking (binary) (p=0.0014, β=0.6400, SE=0.2006), coffee drinking (binary) ($p<1\times10^{-5}$, β=1.1057, SE=0.2222), aspirin intake (binary) (p=0.0003, β=0.7463, SE=0.2041), AAE ($p<1\times10^{-5}$, β=0.9224, SE=0.0109) and a negative relationship with back pain (p=0.0034, β=−0.5435, SE=0.1855) (Supplementary Table S2).

Discussion:

We found an association between the general intake of aspirin, number of pills per week, and aspirin intake duration with later AAO. These results were additionally investigated in a multivariate linear regression model to predict AAO and revealed an association with aspirin intake when examined dichotomous as well as with aspirin intake duration, which was further validated after including more covariates and potential comorbidities. We further replicated our findings concerning aspirin in a separate German iPD cohort (EPIPARK) (Kasten et al., 2013). The effect on PD AAO was not extended to other NSAIDs in the Fox Insight cohort. The difference in AAO for ibuprofen-based non-aspirin medication was only small between users and non-users, and for other anti-inflammatory medication, there was no association at all with AAO. Therefore, we focused our investigations on aspirin. The clinical effect of NSAIDs is still subject to controversial discussion. While some studies indicate a protective effect of NSAIDs or at least an association with PD (Ascherio and Schwarzschild, 2016; Chen et al., 2005), others may see a neuro-protective potential of NSAIDs but not an association with PD at the population level (Auriel et al., 2014; Badshah et al., 2019). No other studies have explored aspirin and AAO in a large iPD cohort so far (Supplementary Figure S2). In addition to the novel findings on aspirin, we replicated previous associations for smoking and caffeine with AAO in PD, summarized in a systematic literature review (Supplementary Figure S2 and Supplementary Table S3) (Benedetti et al., 2000; Cho et al., 2019; Cho et al., 2018; De Reuck et al., 2005; Gallo et al., 2019; Gigante et al., 2018; Gigante et al., 2017; Grandinetti et al., 1994; Greenbaum et al., 2013; Haack et al., 1981; Jimenez-Jimenez et al., 1992; Kandinov et al., 2009; Kuopio et al., 1999; Luth et al., 2020; Maher et al., 2002; Martinez-Rumayor et al., 2009; Mayeux et al., 1994; Neshige et al., 2021; Papapetropoulos et al., 2005; Scott et al., 2005; Tan et al., 2007; Weisskopf et al., 2007; Wijeyekoon et al., 2017; Wilk and Lash, 2007; Yahalom et al., 2020). Thus far, 25 studies analyzed the effect of tobacco or caffeine on PD AAO. These cross-sectional studies have a patient sample size ranging from n=58 to 715.
This effect was further supported by multivariate regression models. When evaluating the independence of smoking and coffee drinking from AAE by pairwise correlations, it showed

that smoking (binary), smoking dosage, coffee drinking (binary), coffee drinking dosage, and coffee drinking duration remained in the model as independent predictors, however, smoking duration did not. These results were robust when including more covariates in the models.

Whether smoking delays AAO is still under debate. In our literature review (Supplementary Figure S2 and Supplementary Table S3) ten studies showed a delay in AAO, while two studies showed an opposite effect, and nine studies did not show a directionality. Gallo *et al*. (Gallo et al., 2019) investigated the risk for PD for smokers and non-smokers in different groups of AAO and showed that there is a prevention of PD onset. However, they argued against a delaying effect of smoking on AAO. An association with a later AAO does not necessarily indicate a causal link. Therefore, we cannot be certain whether the negative association we found in our study is caused by smoking or by other associated factors. Further studies are required to investigate the underlying cause of a later AAO in smokers. In addition, former smokers with PD and current smokers with PD need to be separated to predict a possible long-lasting effect of smoking.

Although the correlation strength for the number of cups of coffee was relatively low, the coffee drinking duration showed a strong correlation, consistent with previous studies (Ascherio and Schwarzschild, 2016; Hu et al., 2007; Paul et al., 2019; Ross et al., 2000), which was also verified in the regression models. Consistent with our study, seven other studies showed that coffee drinkers have a later PD onset. However, two studies showed an opposite effect, and two studies did not report a difference in AAO (Supplementary Table S3). Caffeine is the speculated reason for the protective effect of coffee. Nevertheless, black tea had a more modest association with AAO, likely due to a lower amount of caffeine, which would further explain the strong correlation between a longer black tea drinking duration and a later AAO. The effect of coffee consumption and potential long-term effects need to be investigated in further longitudinal studies.

One strength of our study was the large sample size that provided sufficient power to assess lifestyle factors and PD onset but also allows small magnitude associations to show significance. In addition, online self-report data collections offer many possibilities to promote epidemiological research because of convenience and accessibility for the participants and researchers. A previous study compared self-reported demographic characteristics, symptoms, medical history, and PD medication use of the Fox Insight PD cohort to other in-person observational research study cohorts (Chahine et al., 2020). They found that patterns of responses to patient-reported assessments that were obtained online on the PD cohort of the Fox Insight study were similar to PD cohorts assessed in-person. Patient-reported outcomes are becoming increasingly important to research, therapeutic development, and healthcare delivery, which was already investigated in another previous study on Fox Insight (Dobkin et al., 2020). However, due to the self-report assessments, data may also contain

more subjective perceptions that are difficult to standardize. Additionally, we were limited to the questionnaires and data collected by Fox Insight in this study. This includes the selection of environmental factors as well as the types of questions. Thus, we were unable to assess other potential protective factors and thoroughly investigated smoking, coffee, and aspirin intake. In addition, we were limited to the six-month exposure determination. This period of time might not be sufficient to show a measurable delay in AAO and a higher exposure time might be needed to demonstrate a stronger effect. Nevertheless, these findings may help to acquire a better understanding of this complex disease that can be used to develop specific therapeutic strategies.

This study is a comprehensive assessment of smoking, caffeine, and aspirin intake on the onset of iPD. Besides replicating previous findings in a large self-report American cohort, novel associations of aspirin use with PD AAO were observed. These findings are so far only exploratory; however, they set the stage for future longitudinal assessments on these factors and PD clinical features.

Supplementary Information:

The online version contains supplementary material available at https://doi.org/10.1007/s00415-022-11041-x.

Author contributions:

CG: performed the data analyses and interpretation, performed the regression models, performed the analysis of the EPIPARK cohort, performed the systematic literature review, drafted the manuscript. IK: performed the regression models, critical revision of the manuscript. TL: helped with the analyses, critical revision of the manuscript. BK: helped with the systematic literature review, prepared the data of the EPIPARK cohort for analysis, critical revision of the manuscript. MK, JV, AB: offered comments on the data interpretation, critical revision of the manuscript. AG: critical revision of the manuscript. CK: offered comments on the analyses, critical revision of the manuscript. JT: conception and work design, applied for the Fox Insight DEN access, planned the analyses, data interpretation, drafted the manuscript.

Funding:

Open Access funding enabled and organized by Projekt DEAL. The Fox Insight Study (FI) is funded by The Michael J. Fox Foundation for Parkinson’s Research. We would like to thank the Parkinson’s community for participating in this study to make this research possible. This study was funded by The Michael J. Fox Foundation for Parkinson’s Research (Analysis of

Patient-Reported Outcomes from Fox Insight). JT, MK, and CK are supported by the DFG (FOR 2488).

Data availability:

Data used in the preparation of this article were obtained from the Fox Insight database (https://foxden.michaeljfox.org/insight/explore/fox.jsp) on 18/10/2020. For up-to-date information on the study, visit https://foxden.michaeljfox.org/insight/explore/fox.jsp.

Conflicts of interest:

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval:

Approval was obtained from the Ethics Committee of University of Lübeck.

Consent to participate:

Informed consent was obtained from all individual participants included in the study.

Open Access:

References:

Ascherio, A., and Schwarzschild, M.A. (2016). The epidemiology of Parkinson's disease: risk factors and prevention. Lancet Neurol *15*, 1257-1272.

Ascherio, A., Zhang, S.M., Hernan, M.A., Kawachi, I., Colditz, G.A., Speizer, F.E., and Willett, W.C. (2001). Prospective study of caffeine consumption and risk of Parkinson's disease in men and women. Ann Neurol *50*, 56-63.

Auriel, E., Regev, K., and Korczyn, A.D. (2014). Nonsteroidal anti-inflammatory drugs exposure and the central nervous system. Handb Clin Neurol *119*, 577-584.

Badshah, H., Ikram, M., Ali, W., Ahmad, S., Hahm, J.R., and Kim, M.O. (2019). Caffeine May Abrogate LPS-Induced Oxidative Stress and Neuroinflammation by Regulating Nrf2/TLR4 in Adult Mouse Brains. Biomolecules *9*.

Benedetti, M.D., Bower, J.H., Maraganore, D.M., McDonnell, S.K., Peterson, B.J., Ahlskog, J.E., Schaid, D.J., and Rocca, W.A. (2000). Smoking, alcohol, and coffee consumption preceding Parkinson's disease: a case-control study. Neurology *55*, 1350-1358.

Bloem, B.R., Okun, M.S., and Klein, C. (2021). Parkinson's disease. Lancet *397*, 2284-2303.

Braak, H., Del Tredici, K., Rub, U., de Vos, R.A., Jansen Steur, E.N., and Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging *24*, 197-211.

Chahine, L.M., Chin, I., Caspell-Garcia, C., Standaert, D.G., Brown, E., Smolensky, L., Arnedo, V., Daeschler, D., Riley, L., Korell, M.*, et al.* (2020). Comparison of an Online-Only Parkinson's Disease Research Cohort to Cohorts Assessed In Person. J Parkinsons Dis *10*, 677-691.

Chen, H., Jacobs, E., Schwarzschild, M.A., McCullough, M.L., Calle, E.E., Thun, M.J., and Ascherio, A. (2005). Nonsteroidal antiinflammatory drug use and the risk for Parkinson's disease. Ann Neurol *58*, 963-967.

Cho, B.H., Choi, S.M., and Kim, B.C. (2019). Gender-dependent effect of coffee consumption on tremor severity in de novo Parkinson's disease. BMC Neurol *19*, 194.

Cho, B.H., Choi, S.M., Kim, J.T., and Kim, B.C. (2018). Association of coffee consumption and non-motor symptoms in drug-naive, early-stage Parkinson's disease. Parkinsonism Relat Disord *50*, 42-47.

De Reuck, J., De Weweire, M., Van Maele, G., and Santens, P. (2005). Comparison of age of onset and development of motor complications between smokers and non-smokers in Parkinson's disease. J Neurol Sci *231*, 35-39.

Delamarre, A., and Meissner, W.G. (2017). Epidemiology, environmental risk factors and genetics of Parkinson's disease. Presse Med *46*, 175-181.

Dobkin, R.D., Amondikar, N., Kopil, C., Caspell-Garcia, C., Brown, E., Chahine, L.M., Marras, C., Dahodwala, N., Mantri, S., Standaert, D.G.*, et al.* (2020). Innovative Recruitment Strategies to Increase Diversity of Participation in Parkinson's Disease Research: The Fox Insight Cohort Experience. J Parkinsons Dis *10*, 665-675.

Dorsey, E.R., Sherer, T., Okun, M.S., and Bloem, B.R. (2018). The Emerging Evidence of the Parkinson Pandemic. J Parkinsons Dis *8*, S3-S8.

Gallo, V., Vineis, P., Cancellieri, M., Chiodini, P., Barker, R.A., Brayne, C., Pearce, N., Vermeulen, R., Panico, S., Bueno-de-Mesquita, B.*, et al.* (2019). Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol *48*, 912-925.

Gigante, A.F., Asabella, A.N., Iliceto, G., Martino, T., Ferrari, C., Defazio, G., and Rubini, G. (2018). Chronic coffee consumption and striatal DAT-SPECT findings in Parkinson's disease. Neurol Sci *39*, 551-555.

Gigante, A.F., Martino, T., Iliceto, G., and Defazio, G. (2017). Smoking and age-at-onset of both motor and non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord *45*, 94-96.

Grandinetti, A., Morens, D.M., Reed, D., and MacEachern, D. (1994). Prospective study of cigarette smoking and the risk of developing idiopathic Parkinson's disease. Am J Epidemiol *139*, 1129-1138.

Greenbaum, L., Rigbi, A., Lipshtat, N., Cilia, R., Tesei, S., Asselta, R., Djaldetti, R., Goldwurm, S., and Lerer, B. (2013). Association of nicotine dependence susceptibility gene, CHRNA5, with Parkinson's disease age at onset: gene and smoking status interaction. Parkinsonism Relat Disord *19*, 72-76.

Haack, D.G., Baumann, R.J., McKean, H.E., Jameson, H.D., and Turbek, J.A. (1981). Nicotine exposure and Parkinson disease. Am J Epidemiol *114*, 191-200.

Hu, G., Bidel, S., Jousilahti, P., Antikainen, R., and Tuomilehto, J. (2007). Coffee and tea consumption and the risk of Parkinson's disease. Mov Disord *22*, 2242-2248.

Jimenez-Jimenez, F.J., Mateo, D., and Gimenez-Roldan, S. (1992). Premorbid smoking, alcohol consumption, and coffee drinking habits in Parkinson's disease: a case-control study. Mov Disord *7*, 339-344.

Kandinov, B., Giladi, N., and Korczyn, A.D. (2009). Smoking and tea consumption delay onset of Parkinson's disease. Parkinsonism Relat Disord *15*, 41-46.

Kasten, M., Hagenah, J., Graf, J., Lorwin, A., Vollstedt, E.J., Peters, E., Katalinic, A., Raspe, H., and Klein, C. (2013). Cohort Profile: a population-based cohort to study non-motor symptoms in parkinsonism (EPIPARK). Int J Epidemiol *42*, 128-128k.

Kuopio, A.M., Marttila, R.J., Helenius, H., and Rinne, U.K. (1999). Environmental risk factors in Parkinson's disease. Mov Disord *14*, 928-939.

Luth, T., Konig, I.R., Grunewald, A., Kasten, M., Klein, C., Hentati, F., Farrer, M., and Trinh, J. (2020). Age at Onset of LRRK2 p.Gly2019Ser Is Related to Environmental and Lifestyle Factors. Mov Disord *35*, 1854-1858.

Maher, N.E., Golbe, L.I., Lazzarini, A.M., Mark, M.H., Currie, L.J., Wooten, G.F., Saint-Hilaire, M., Wilk, J.B., Volcjak, J., Maher, J.E., *et al.* (2002). Epidemiologic study of 203 sibling pairs with Parkinson's disease: the GenePD study. Neurology *58*, 79-84.

Mappin-Kasirer, B., Pan, H., Lewington, S., Kizza, J., Gray, R., Clarke, R., and Peto, R. (2020). Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors. Neurology *94*, e2132-e2138.

Martinez-Rumayor, A., Arrieta, O., Sotelo, J., and Garcia, E. (2009). Female gender but not cigarette smoking delays the onset of Parkinson's disease. Clin Neurol Neurosurg *111*, 738-741.

Martino, R., Candundo, H., Lieshout, P.V., Shin, S., Crispo, J.A.G., and Barakat-Haddad, C. (2017). Onset and progression factors in Parkinson's disease: A systematic review. Neurotoxicology *61*, 132-141.

Mayeux, R., Tang, M.X., Marder, K., Cote, L.J., and Stern, Y. (1994). Smoking and Parkinson's disease. Mov Disord *9*, 207-212.

Neshige, S., Ohshita, T., Neshige, R., and Maruyama, H. (2021). Influence of current and previous smoking on current phenotype in Parkinson's disease. J Neurol Sci *427*, 117534.

Papapetropoulos, S., Singer, C., Villar, J.M., Gonzalez, J., and Mash, D.C. (2005). Does cigarette smoking provide clinically significant neuroprotection among patients diagnosed with Parkinson's disease? Mov Disord *20*, 641-642.

Paul, K.C., Chuang, Y.H., Shih, I.F., Keener, A., Bordelon, Y., Bronstein, J.M., and Ritz, B. (2019). The association between lifestyle factors and Parkinson's disease progression and mortality. Mov Disord *34*, 58-66.

Ross, G.W., Abbott, R.D., Petrovitch, H., Morens, D.M., Grandinetti, A., Tung, K.H., Tanner, C.M., Masaki, K.H., Blanchette, P.L., Curb, J.D., *et al.* (2000). Association of coffee and caffeine intake with the risk of Parkinson disease. JAMA *283*, 2674-2679.

San Luciano, M., Tanner, C.M., Meng, C., Marras, C., Goldman, S.M., Lang, A.E., Tolosa, E., Schule, B., Langston, J.W., Brice, A., *et al.* (2020). Nonsteroidal Anti-inflammatory Use and LRRK2 Parkinson's Disease Penetrance. Mov Disord *35*, 1755-1764.

Scott, W.K., Zhang, F., Stajich, J.M., Scott, B.L., Stacy, M.A., and Vance, J.M. (2005). Family-based case-control study of cigarette smoking and Parkinson disease. Neurology *64*, 442-447.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W., *et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Tan, E.K., Chua, E., Fook-Chong, S.M., Teo, Y.Y., Yuen, Y., Tan, L., and Zhao, Y. (2007). Association between caffeine intake and risk of Parkinson's disease among fast and slow metabolizers. Pharmacogenet Genomics *17*, 1001-1005.

Wahner, A.D., Bronstein, J.M., Bordelon, Y.M., and Ritz, B. (2007). Nonsteroidal anti-inflammatory drugs may protect against Parkinson disease. Neurology *69*, 1836-1842.

Weisskopf, M.G., Grodstein, F., and Ascherio, A. (2007). Smoking and cognitive function in Parkinson's disease. Mov Disord *22*, 660-665.

Wijeyekoon, R., Suriyakumara, V., Gamage, R., Fernando, T., Jayasuriya, A., Amarasinghe, D., Gunasekara, H., Sirisena, D., Amaratunga, D., Muthukuda, C., *et al.* (2017). Associations between Lifestyle Factors and Parkinson's Disease in an Urban Sri Lankan Clinic Study. Int Arch Med *10*.

Wilk, J.B., and Lash, T.L. (2007). Risk factor studies of age-at-onset in a sample ascertained for Parkinson disease affected sibling pairs: a cautionary tale. Emerg Themes Epidemiol *4*, 1.

Yahalom, G., Rigbi, A., Israeli-Korn, S., Krohn, L., Rudakou, U., Ruskey, J.A., Benshimol, L., Tsafnat, T., Gan-Or, Z., Hassin-Baer, S., *et al.* (2020). 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.

2.2 Objective 2: The association of coffee drinking, aspirin use, and smoking with clinical severity in patients with PD

The clinical phenotype of patients with PD, including motor signs and non-motor symptoms, can differ due to a variety of factors that are still not fully uncovered. Motor and non-motor signs and symptoms can be assessed with standardized questionnaires, e.g., the Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS), the Non-Motor Symptom Questionnaire (NMSQ), the Non-motor Symptoms Scale (NMSS), the Scales for Outcomes in Parkinson's disease - Autonomic Dysfunction (SCOPA-AUT), the Montreal Cognitive Assessment (MOCA), the Parkinson's Disease Questionnaire (PDQ-39), and many more. These help clinicians to evaluate the patient's physical and mental condition and to trace the disease progression over time. Although most patients experience tremor, bradykinesia, rigidity, imbalance, and cognitive decline, the clinical severity and symptom progression in the individual patients might show different disease courses. Previous studies have shown that, among others, coffee drinking might affect motor and non-motor signs and symptoms, indicating an inverse association with the severity of non-motor symptoms related to mood and cognition (Cho et al., 2018) and tremor severity in male patients with de novo PD (Cho et al., 2019).

In the following study, the relationship between the lifestyle factors coffee drinking, aspirin intake, and smoking and clinical severity was investigated in patients with PD from the Fox Insight cohort. These lifestyle factors have previously shown a protective association with AAO in this study group (Gabbert et al., 2022). Generalized linear regression models were used to assess the relationship between the lifestyle factors and motor/non-motor symptoms, while age, gender, and disease duration were used as further covariates. While coffee drinking showed very limited associations with motor and non-motor symptoms, aspirin intake correlated with motor and non-motor symptoms, indicating a higher probability of experiencing symptoms, e.g. tremor (p=0.0026, β=0.3174), problems with chewing and swallowing (p=0.0358, β=0.1837) and getting up (p=0.0185, β=0.2170), constipation (p=0.0124, β=0.2077), unexplained pains (p=0.0227, β=0.1961), problems remembering ($p<1\times10^{-5}$, β=0.3662), and light-headedness (p=0.0043, β=0.2380). Most of these associations were still robust after including comorbidities such as heart diseases, arthritis, back pain, and surgeries with anesthesia. In addition, smoking was directly associated with clinical severity. Smokers had more problems with motor symptoms, e.g., saliva and drooling (p=0.0106, β=0.1484), chewing and swallowing (p=0.0002, β=0.2243), and freezing (p=0.0212, β=0.1490). Even more pronounced was the association of smoking with non-motor symptoms and symptoms related to mood, indicating that the higher the smoking dosage and duration, the more likely it

was that patients with PD experienced depression ($p<1\times10^{-5}$, $\beta=0.3362$), anxiety ($p=2\times10^{-5}$, $\beta=0.2748$), or hopelessness ($p<1\times10^{-5}$, $\beta=0.3846$).

The study covering this objective is titled “Lifestyle factors and clinical severity of Parkinson’s disease”. It was originally part of the paper of Objective 1 and was later separated as suggested in the review process. In addition, to the previously mentioned data sheets “Environmental Exposure Questionnaires” for all lifestyle factors, “General”, “About You”, and “Your Current Health”, I downloaded the questionnaires “Your Movement Experiences”, “Your Non-movement Experiences”, and “Your Mood” from the Fox Data Exploration Network and harmonized the data accordingly for all >35,000 participants. Using the environmental and lifestyle data questionnaires, I estimated consumers of lifestyle factors and calculated the duration of usage until AAO as described before. After some initial statistical overview performed by Prof. Inke R. König, I performed all generalized linear regression models used in the paper in R, which included all associations between the three lifestyle factors (binary, dosage, duration) with motor- and non-motor symptoms first without and later with potential comorbidities. To expand my knowledge in the use of R and statistical analyses, I participated in a workshop called “Introduction to Statistical Programming with R.” I wrote the first draft of the manuscript, including all tables, and the Supplementary Material, which was supervised and revised by PD Dr. Joanne Trinh. Lastly, in the subsequent peer-review process, I wrote the response letters to the Reviewers and performed all additionally suggested analyses largely independently, as PD Dr. Joanne Trinh was on parental leave by the time the Reviewer’s comments returned from *Scientific Reports*.

Title:

Lifestyle factors and clinical severity of Parkinson’s disease

Authors:

Carolin Gabbert[1], MSc, Prof. Inke R. König[2], Theresa Lüth[1], MSc, Meike Kasten[1,3], MD, Anne Grünewald[1,4], PhD, Christine Klein[1], MD, Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

[2]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany

[3]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany

[4]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[*]Corresponding author

Published in *Scientific Reports*, 2023, DOI: 10.1038/s41598-023-31531-w

Abstract:

Genetic factors, environmental factors, and gene-environment interactions have been found to modify PD risk, age at onset (AAO), and disease progression. The objective of this study was to explore the association of coffee drinking, aspirin intake, and smoking with motor and non-motor symptoms in a cohort of 35,959 American patients with PD from the Fox Insight Study using generalized linear models. Coffee drinkers had fewer problems swallowing, but dosage and duration of coffee intake were not associated with motor or non-motor symptoms. Aspirin intake correlated with more tremor (p=0.0026), problems getting up (p=0.0185), light-headedness (p=0.0043), and problems remembering (p=1×10^{-5}). Smoking was directly associated with symptoms: smokers had more problems with drooling (p=0.0106), swallowing (p=0.0002), and freezing (p<1×10^{-5}). Additionally, smokers had more possibly mood-related symptoms: unexplained pains (p<1×10^{-5}), problems remembering (p=0.0001), and feeling sad (p<1×10^{-5}). Confirmatory and longitudinal studies are warranted to investigate the clinical correlation over time.

Introduction:

An interplay of genetic and environmental factors is known to influence age at onset (AAO) in Parkinson's disease (PD) (Gabbert et al., 2022), the risk for PD, as well as PD progression, and the severity of symptoms (Dunn et al., 2019; Kline et al., 2021; Marras et al., 2019). However, studies that focus on the severity of motor and non-motor symptoms are sparse. Most of them investigate motor and non-motor symptoms by separating participants based on their current coffee drinking or smoking behavior without differentiating between use before or after disease onset (Alves et al., 2004; Cho et al., 2019; Cho et al., 2018; Kandinov et al., 2007; Moccia et al., 2015). Additionally, cohort sizes range between 100 and 300 patients with PD (Alves et al., 2004; Cho et al., 2019; Cho et al., 2018; Kandinov et al., 2007; Moccia et al., 2015), illustrating the importance to investigate the interaction between lifestyle factors and motor symptoms in larger cohorts for better statistical power. There are indications of a protective effect of coffee on motor and non-motor symptoms. An inverse association has been found between the consumption of coffee and the severity of non-motor symptoms related to mood and cognition (Cho et al., 2018). Additionally, it was reported that coffee drinkers had lower tremor scores compared to non-coffee drinkers, also with a dose-dependent relationship between coffee consumption and tremor severity, but this was only significant in male patients with de novo PD (Cho et al., 2019). In contrast to the findings for coffee, no differences in motor or non-motor symptom severity were reported between smokers and non-smokers. There was no significant difference in the duration from motor symptom onset to reaching Hoehn and Yahr stage 3 between smokers and non-smokers (Kandinov et al., 2007).

Additionally, no significant differences were reported in change of disease severity, symptoms of depression, and cognitive impairment between smokers and non-smokers (Alves et al., 2004). In accordance with this, no associations between the Non-Motor Symptoms Questionnaire (NMSQ) total score and smoking status were found. However, following a multinomial logistic regression stepwise model using never smoking as a reference, several non-motor symptoms showed associations with current or former smoking, e.g. lower Unified Parkinson's Disease Rating Scale (UPDRS) part III total scores were associated with smoking (Moccia et al., 2015).

Then again, other studies aim to investigate the immediate effect of coffee or smoking on PD symptoms by specifically administering caffeine or nicotine, which are the suspected agents responsible for the effect of coffee drinking and smoking, respectively. However, studies investigating the specific treatment with caffeine and nicotine after disease onset did not indicate any association with motor symptoms, as motor scores did not significantly differ between treated patients and untreated patients (Postuma et al., 2017; Villafane et al., 2018). In contrast to coffee and smoking, a possible immediate association of aspirin with motor symptoms or motor symptom severity in PD has yet to be investigated.

Tobacco use, coffee consumption, and aspirin intake have already been found to be associated with AAO in the Fox Insight American PD cohort. However, how these specific lifestyle factors affect PD motor and nonmotor symptoms remains unclear, especially when strictly exploring the intake before disease manifestation.

Herein, we performed a cross-sectional exploration of the association of the consumption of coffee, smoking, and the use of aspirin on motor and non-motor symptoms in patients with PD.

Methods:

Demographics and participant examination

Our study sample consists of 35,959 American patients with PD (Supplementary Table S1) from the Fox Insight study (Supplementary Text and Supplementary Figure S1). The Fox Insight study is an ongoing online, longitudinal health study of people with and without PD (Smolensky et al., 2020). The dataset is generated through routine longitudinal assessments, one-time health and disease questionnaires about symptoms, daily activities, and other factors, and, in a subgroup of people with PD, genetic data collection. Fox Insight participants were 18 years of age or older and provided informed consent. In the process of registration, participants were divided into two groups, PD patients and controls, the latter were asked about new diagnoses every three months. PD patients responded to health, non-motor assessments, motor assessments, quality of life, and lifestyle questionnaires. All analyses were performed in accordance with relevant guidelines and regulations. Data collection was performed via an

online data platform. We excluded patients with PD with an AAO lower than 3 years as well as with an age at examination (AAE) lower than 18 years. Most of the patients were of white/Caucasian ethnicity (89.9%) (Supplementary Table S1). PD patients had a mean AAE of 65.7±10.2 *SD* years (range 18.1-119.0 years) and a mean AAO of 60.4±11.0 *SD* years (range 5.1-115.4 years). The mean disease duration until examination at Fox Insight was 5.3±5.6 *SD* years (range 0-64.3 years) and the mean disease duration until their current age was 6.5±5.7 *SD* years (range 0-64.3 years); 40.4% of PD patients were female.

Clinical variables were downloaded from Fox Insight questionnaires: “Your Movement Experiences”, “Your Non-movement Experiences”, “Your Current Health”, and “Your Mood”. Patients report tremor, speech impairment, excess of saliva and drooling, problems chewing and swallowing, problems walking and balance, freezing, and problems getting up for “Your Movement Experiences”. Constipation, unexplained pains, problems remembering, feeling sad, feeling anxious, changed interest in sex, and lightheadedness were reported as “Your Non-movement Experiences”. More specific variables on mood were assessed with “Your Current Health” and “Your Mood” questionnaires. These correspond to participant questionnaires MDS-UPDRS II, NMSQ, and Geriatric Depression Scale (GDS). The PD Risk Factor Questionnaires (PD-RFQ-Us) were used for lifestyle and environmental factors. Each questionnaire and specific variables within are described in detail in the Supplementary Text and Supplementary Table S2.

Lifestyle factors

The intake of coffee, aspirin, and tobacco was estimated from Environmental Exposure Questionnaires and the definitions of coffee drinkers, aspirin users, and smokers were previously described for this cohort (Gabbert et al., 2022). Patients were classified as coffee drinkers if they drank caffeinated coffee at least once per week over a minimum period of six months. Patients were classified as aspirin users if they took at least two pills of aspirin per week for at least six months. Lastly, patients were classified as smokers/tobacco users if they smoked more than 100 cigarettes in their lifetime or if they smoked at least one cigarette per day over a period of at least six months, or if they used smokeless tobacco at least once per day for more than six months.

To assess dosage and long-term effects, duration of caffeine consumption, aspirin intake, and smoking were estimated according to the age the patients started using either substance subtracted from the age at termination or from their AAO if the patients terminated the consumption after their AAO. Periods in between, where the patients stopped regularly consuming, were not included in the duration. Coffee drinking dosage was defined as cups of coffee per week within coffee drinking duration, excluding all values higher than 100 cups per

week from the analysis. Aspirin dosage was defined as aspirin pills per week the patients took within aspirin intake duration time. Smoking dosage was estimated as cigarettes smoked per day during time smoking, excluding implausible values (>100 cigarettes per day). For dosage analyses, the number of cups of coffee for non-drinkers, pills per week for aspirin non-users, and cigarettes for non-smokers was set to zero.

Statistical analysis

Generalized linear regression models were used to estimate the relationship between environmental factors, age, gender, disease duration, and motor/non-motor symptoms (R studio) (details of each model are in the Supplementary Text). Reported *p*-values remain descriptive because they are not corrected for multiple testing and results are exploratory. Patients with missing data on the use of environmental and lifestyle factors or motor and non-motor symptoms were not included in the analyses. Numbers of patients included in each regression model are reported in the tables.

Multiple regression models were estimated to predict the respective symptoms to assess the relationship between environmental factors and motor/non-motor symptoms. To explore potential confounders, we adjusted for covariates by including AAE, gender, disease duration, and potential comorbidities. Environmental factors were separately handled in three different ways: (1) binary (yes-no indication), (2) dosage as a continuous variable, and (3) duration as a continuous variable. Patients indicating no for lifestyle factor had values set to zero. Motor symptoms were dichotomized (yes: scores>1 or no: score=1) because data were not normally distributed (details in Supplementary Text).

Ethics approval

All participants provided informed consent using the Fox Insight website. The Fox Insight study has been approved by the New England Institutional Review board (IRB) (IRB: 120160179; Legacy IRB#: 14-236, Sponsor Protocol Number: 1, Study Title: Fox Insight). Approval was obtained from the Ethics Committee of the University of Lübeck. We confirm that all analyses were performed in accordance with relevant guidelines and regulations.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Results:

Coffee

We first explored the association of coffee drinking before AAO with self-reported motor and nonmotor features (Supplementary Table S2). When coffee drinking was used as a binary yes-no indication in a regression model with covariates AAE, gender, and disease duration (Table 1, Supplementary Table S3) coffee drinkers had fewer problems with chewing and swallowing (*p*=0.0497, *β*=-0.1435, *SE*=0.0731). When cups of coffee per week or coffee drinking duration were used as continuous variables, we found no association with any of the motor symptoms (Table 1).

Table 1: Motor symptoms associated with environmental factors in regression models.

	Coffee			**Aspirin**			**Smoking**		
	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Tremor									
n	4889	3848	1967	2866	2730	547	5269	4201	876
p-value	0.5797	0.0559	0.3860	**0.0026**	**0.0138**	0.5498	0.1366	0.3171	0.5973
β	-0.0473	0.0084	-0.0048	0.3174	0.0287	0.0064	0.1038	0.0038	-0.0035
Speech									
n	4892	3849	1967	2868	2732	547	5272	4203	876
p-value	0.1615	0.7086	0.2585	0.5007	0.0880	0.4461	0.1330	**0.0062**	**0.0454**
β	-0.1021	-0.0013	0.0050	-0.0582	0.0152	-0.0062	0.0893	0.0089	0.0119
Saliva and Drooling									
n	4892	3849	1967	2868	2732	547	5272	4203	876
p-value	0.0823	0.9754	0.9591	0.2339	0.3290	0.5824	**0.0106**	**0.0022**	0.0976
β	-0.1232	0.0001	-0.0002	0.0997	0.0080	0.0044	0.1484	0.0096	0.0093
Chewing and Swallowing									
n	4892	3849	1967	2868	2732	547	5272	4203	876
p-value	**0.0497**	0.6961	0.4785	**0.0358**	**0.0182**	0.9015	**0.0002**	**$<1\times10^{-5}$**	0.0589
β	-0.1435	-0.0014	0.0033	0.1837	0.0201	-0.0010	0.2243	0.0174	0.0107
Walking and Balance									
n	4889	3848	1967	2866	2730	547	5269	4201	876
p-value	0.1156	0.8287	0.6311	0.1056	**0.0106**	0.8477	0.2660	**0.0038**	**4×10^{-5}**
β	-0.1223	-0.0008	0.0022	0.1478	0.0253	-0.0017	0.0698	0.0101	0.0268
Freezing									
n	4889	3848	1967	2866	2730	547	5269	4201	876
p-value	0.2364	0.7292	0.9616	0.2979	0.1481	0.9126	**0.0212**	**0.0052**	**$<1\times10^{-5}$**
β	-0.0935	0.0013	0.0002	0.0996	0.0133	0.0010	0.1490	0.0094	0.0277
Getting up									
n	4889	3848	1967	2866	2730	547	5269	4201	876
p-value	0.4886	0.5482	0.7142	**0.0185**	**0.0182**	0.3039	0.0855	**0.0061**	**$<1\times10^{-5}$**
β	-0.0540	-0.0022	0.0017	0.2170	0.0231	-0.0088	0.1099	0.0098	0.0366

p-value (exploratory): multivariate regression to predict the respective motor symptoms adjusted for covariates by including AAE, gender, and disease duration (time between AAO and current age) in the model. Significant values are in [bold].

Furthermore, coffee did not show an association with the available non-motor symptoms (p>0.05) (Table 2, Supplementary Table S4). When cups of coffee per week were used as a continuous variable, coffee drinking dosage showed a direct correlation with unexplained pains (p=0.0168, β=0.0083, SE=0.0035) (Table 2, Supplementary Table S5). Here, the higher the dosage of coffee, the more unexplained pains were experienced.

Table 2: Non-motor symptoms associated with environmental factors in regression models.

	Coffee			**Aspirin**			**Smoking**		
	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Constipation									
n	4917	3868	1977	2881	2743	548	5297	4221	880
p-value	0.0970	0.9669	0.4438	**0.0124**	**0.0037**	0.6783	0.1994	0.2026	0.6233
β	-0.1164	-0.0001	0.0033	0.2077	0.0251	0.0033	0.0735	0.0039	0.0027
Unexplained Pains									
n	4917	3868	1977	2881	2743	548	5296	4221	880
p-value	0.1623	**0.0168**	0.0883	**0.0227**	**0.0320**	0.8718	**$<1\times10^{-5}$**	**0.0003**	**0.0134**
β	0.1017	0.0083	0.0080	0.1961	0.0178	-0.0013	0.2732	0.0114	0.0140
Problems Remembering									
n	4917	3868	1977	2881	2743	548	5296	4221	880
p-value	0.9859	0.7212	0.5348	**1×10^{-5}**	**0.0005**	0.5864	**0.0001**	**6×10^{-5}**	**0.0017**
β	0.0012	0.0012	-0.0027	0.3662	0.0295	0.0043	0.2176	0.0123	0.0176
Feeling Sad									
n	4914	3865	1976	2880	2742	548	5293	4218	880
p-value	0.2517	0.0627	0.5381	0.0665	**0.0344**	0.4675	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**0.0470**
β	0.0804	0.0063	0.0027	0.1537	0.0177	0.0058	0.3279	0.0152	0.0112
Anxiety									
n	4914	3865	1976	2880	2742	548	5293	4218	880
p-value	0.2199	0.1432	0.5958	0.2999	0.1574	0.8303	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.3910
β	0.0902	0.0052	0.0025	0.0920	0.0118	-0.0018	0.3007	0.0161	0.0049
Changed Interest in Sex									
n	4914	3865	1976	2880	2742	548	5293	4218	880
p-value	0.1032	0.0505	0.6957	**0.0221**	**0.0278**	0.7925	**0.0013**	0.2351	**0.0372**
β	0.1227	0.0069	0.0018	0.2023	0.0184	0.0022	0.1959	0.0038	0.0123
Light-headedness									
n	4913	3864	1976	2880	2742	548	5292	4217	880
p-value	0.2808	0.0561	0.1935	**0.0043**	**0.0060**	0.5288	**0.0005**	**0.0001**	0.1326
β	0.0758	0.0064	-0.0056	0.2380	0.0226	0.0050	0.2000	0.0117	0.0083

p-value (exploratory): multivariate regression to predict the respective non-motor symptoms adjusted for covariates by including AAE, gender, and disease duration (time between AAO and current age) in the model. Significant values are in [bold].

Aspirin

We further explored the association of aspirin intake before AAO with self-reported motor features (Supplementary Table S2). Aspirin intake (binary yes-no indication) showed a direct

association with tremor (p=0.0026, β=0.3174, SE=0.1054), chewing and swallowing problems (p=0.0358, β=0.1837, SE=0.0875), and getting up (p=0.0185, β=0.2170, SE=0.0922) (Table 1, Supplementary Table S3), indicating a higher probability to have tremor or problems with swallowing or getting out of a bed or a chair in the group of aspirin users compared to non-users. We further investigated potential comorbidities by including heart diseases, arthritis, back pain, and surgeries with anesthesia in the regression models. The exploratory association of aspirin intake with most motor symptoms was still robust after including the comorbidities (Supplementary Table S6) with the exception of chewing and swallowing: the association diminished when heart diseases (p=0.1120, β=0.1425, SE=0.0897), arthritis (p=0.0556, β=0.1684, SE=0.0880), back pain (p=0.0591, β=0.1661, SE=0.0880), and surgeries with anesthesia (p=0.0591, β=0.1661, SE=0.0880) were included (Supplementary Table S7). In addition, the association between aspirin intake and getting up diminished when heart diseases were included in the model (p=0.0534, β=0.1819, SE=0.0941) (Supplementary Table S7).

When assessing dosage effects, more aspirin taken per week associated with more problems with tremor (p=0.0138, β=0.0287, SE=0.0117), chewing and swallowing (p=0.0182, β=0.0201, SE=0.0085), walking and balance (p=0.0106, β=0.0253, SE=0.0099), and getting up (p=0.0182, β=0.0231, SE=0.0098) (Table 1, Supplementary Table S3). When potential confounding comorbidities were investigated, the association between the aspirin intake dosage and getting up diminished when back pain was included in the model (p=0.0509, β=0.0196, SE=0.0100) (Supplementary Tables S6, S7).

Aspirin intake duration did not show an association with any of the motor symptoms (Table 1).

In addition, we explored the association between aspirin intake status and non-motor symptoms (Table 2, Supplementary Tables S2, S4). Aspirin intake (binary yes-no indication) exhibited a direct association with constipation (p=0.0124, β=0.2077, SE=0.0831), unexplained pains (p=0.0227, β=0.1961, SE=0.0861), problems remembering (p=1×10^{-5}, β=0.3662, SE=0.0830), changed interest in sex (p=0.0221, β=0.2023, SE=0.0884), and light-headedness (p=0.0043, β=0.2380, SE=0.0833) (Table 2, Supplementary Table S5). Thus, the odds of experiencing constipation, unexplained pains, problems remembering, a changed interest in sex, or feeling light-headed were higher for aspirin users. When more aspirin pills per week were taken, more problems with non-motor symptoms such as constipation (p=0.0037, β=0.0251, SE=0.0086), unexplained pains (p=0.0320, β=0.0178, SE=0.0083), problems remembering (p=0.0005, β=0.0295, SE=0.0085), feeling sad (p=0.0344, β=0.0177, SE=0.0083), changed interest in sex (p=0.0278, β=0.0184, SE=0.0084), and light-headedness (p=0.0060, β=0.0226, SE=0.0082) (Table 2, Supplementary Table S5) were reported. The association between aspirin intake dosage and unexplained pains diminished when arthritis (p=0.0672, β=0.0154, SE=0.0084) and back pain (p=0.0896, β=0.0142, SE=0.0084) were

included (Supplementary Tables S8, S9). In addition, the association between aspirin intake dosage and feeling sad diminished when including heart diseases (p=0.0806, β=0.0147, SE=0.0084), arthritis (p=0.0513, β=0.0164, SE=0.0084), and back pain (p=0.0527, β=0.0163, SE=0.0084) (Supplementary Tables S8, S9). All other motor and non-motor symptoms remained associated with aspirin intake. Aspirin intake duration did not show an association with any of the non-motor symptoms (Table 2).

Smoking

Lastly, we explored the association of smoking before AAO with self-reported motor features (Supplementary Table S2). Smoking directly correlated with excessive saliva and drooling (p=0.0106, β=0.1484, SE=0.0580), chewing and swallowing problems (p=0.0002, β=0.2243, SE=0.0603), and freezing (p=0.0212, β=0.1490, SE=0.0646), when smoking was used as a binary yes-no indication (Table 1, Supplementary Table S3), indicating that smokers had more problems with too much saliva, problems with swallowing, and freezing. In addition, smoking dosage correlated with more problems with speech (p=0.0062, β=0.0089, SE=0.0033), saliva excess and drooling (p=0.0022, β=0.0096, SE=0.0031), chewing and swallowing ($p<1\times10^{-5}$, β=0.0174, SE=0.0031), walking and balance (p=0.0038, β=0.0101, SE=0.0035), freezing (p=0.0052, β=0.0094, SE=0.0034), and getting up (p=0.0061, β=0.0098, SE=0.0036) (Table 1, Supplementary Table S3). Smoking duration also correlated with more problems with speech (p=0.0454, β=0.0119, SE=0.0059), walking and balance ($p=4\times10^{-5}$, β=0.0268, SE=0.0065), freezing ($p<1\times10^{-5}$, β=0.0277, SE=0.0062), and getting up ($p<1\times10^{-5}$, β=0.0366, SE=0.0070) (Table 1, Supplementary Table S3). Thus, more problems with motor symptoms were observed when more cigarettes were smoked per day and when smoking duration was longer. However, when investigating potential comorbidities, the association between smoking duration and speech diminished when heart diseases (p=0.0526, β=0.0115, SE=0.0059) and lung diseases (p=0.0591, β=0.0113, SE=0.0060) were included in the models, although heart diseases (p=0.7074, β=0.0752, SE=0.2002) and lung diseases (p=0.5828, β=0.1105, SE=0.2011) showed no association with speech either (Supplementary Tables S7, S10).

Most strikingly, smoking was directly associated with non-motor symptoms (Table 2, Supplementary Tables S2, S4). Smokers experienced more unexplained pains ($p<1\times10^{-5}$, β=0.2732, SE=0.0595), problems remembering (p=0.0001, β=0.2176, SE=0.0570), feeling sad ($p<1\times10^{-5}$, β=0.3279, SE=0.0579), anxiety ($p<1\times10^{-5}$, β=0.3007, SE=0.0604), changed interest in sex (p=0.0013, β=0.1959, SE=0.0610), and lightheadedness (p=0.0005, β=0.2000, SE=0.0574) (Table 2, Supplementary Table S5). Smoking dosage correlated with more unexplained pains (p=0.0003, β=0.0114, SE=0.0031), problems remembering ($p=6\times10^{-5}$, β=0.0123, SE=0.0030), feeling sad ($p<1\times10^{-5}$, β=0.0152, SE=0.0031), anxiety ($p<1\times10^{-5}$,

β=0.0161, *SE*=0.0031), and light-headedness (p=0.0001, β=0.0117, *SE*=0.0030) (Table 2, Supplementary Table S5). Smoking duration also correlated with more unexplained pains (p=0.0134, β=0.0140, *SE*=0.0057), problems remembering (p=0.0017, β=0.0176, *SE*=0.0056), feeling sad (p=0.0470, β=0.0112, *SE*=0.0056), and changed interest in sex (p=0.0372, β=0.0123, *SE*=0.0059) (Table 2, Supplementary Table S5). Thus, more problems with non-motor symptoms were observed with more cigarettes smoked per day and longer smoking duration. However, the association between smoking duration and feeling sad diminished when heart diseases (p=0.0807, β=0.0099, *SE*=0.0057) and lung diseases (p=0.1103, β=0.0091, *SE*=0.0057) were included in the model. In addition, the association with a changed interest in sex (p=0.0574, β=0.0113, *SE*=0.0059) diminished when heart diseases were included, although heart diseases itself showed no association with a changed interest in sex either (p=0.0795, β=0.3482, *SE*=0.1986) (Supplementary Tables S9, S11).

As the most prominent association for smoking with non-motor symptoms, we additionally explored symptoms specifically related to mood (Supplementary Table S2). More smokers (binary yes-no indication) exhibited depression ($p<1\times10^{-5}$, β=0.3362, *SE*=0.0649), anxiety ($p=2\times10^{-5}$, β=0.2748, *SE*=0.0636), and other factors associated with depression and mood. These include: dropped many activities and interests, life feels empty, getting bored often, being afraid something bad could happen, feeling helpless often, prefer staying at home, feeling to have more memory problems than other people, feeling pretty worthless, and feeling that situation is hopeless ($p<0.0025$, $\beta>0.1955$) (Table 3, Supplementary Tables S12, S13). Likewise, cigarettes smoked per day are associated with more depression ($p<1\times10^{-5}$, β=0.0199, *SE*=0.0033), anxiety ($p<1\times10^{-5}$, β=0.0182, *SE*=0.0032), as well as factors associated with depression and mood ($p<0.0014$, $\beta>0.0105$). Smoking duration also showed the same direction for depression (p=0.0422, β=0.0123, *SE*=0.0061), and factors associated with depression ($p<0.0039$, $\beta>0.0178$) (Table 3, Supplementary Table S13). Thus, all investigated symptoms related to depression and mood were associated with smoking indicating that the higher the smoking dosage and duration, the more likely it was that patients with PD experienced a negative mood. When potential confounding comorbidities were investigated, all symptoms related to mood remained associated with smoking with the exception of depression: the association between smoking duration and depression diminished when heart diseases (p=0.0544, β=0.0117, *SE*=0.0061) and lung diseases (p=0.0584, β=0.0116, *SE*=0.0061) were included, although heart diseases (p=0.1015, β=0.3338, *SE*=0.2038) and lung diseases (p=0.3638, β=0.1839, *SE*=0.2025) showed no association with smoking duration either (Supplementary Tables S14, S15).

Table 3: Symptoms related to mood associated with smoking in regression models.

	Smoking		
	Yes/No	Dosage	Duration
Depression			
n	5223	4151	861
p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**0.0422**
β	0.3362	0.0199	0.0123
Anxiety			
n	5223	4152	861
p-value	**2×10^{-5}**	**$<1\times10^{-5}$**	0.3799
β	0.2748	0.0182	0.0053
Dropped many activities and interests			
n	5205	4152	868
p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**
β	0.3283	0.0165	0.0253
Life feels empty			
n	5201	4151	866
p-value	**0.0004**	**$<1\times10^{-5}$**	**0.0021**
β	0.2843	0.0181	0.0230
Getting bored often			
n	5207	4153	868
p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**
β	0.4263	0.0187	0.0412
Being afraid something bad could happen			
n	5195	4147	864
p-value	**0.0025**	**0.0002**	**0.0039**
β	0.2050	0.0126	0.0184
Feeling helpless often			
n	5193	4144	864
p-value	**2×10^{-5}**	**3×10^{-5}**	**1×10^{-5}**
β	0.3099	0.0151	0.0296
Prefer staying at home			
n	5196	4145	864
p-value	**0.0007**	**$<1\times10^{-5}$**	**0.0017**
β	0.1955	0.0150	0.0179
Feeling to have more memory problems than other people			
n	5206	4156	868
p-value	**0.0019**	**0.0014**	**0.0038**
β	0.2027	0.0105	0.0178
Feeling pretty worthless			
n	5189	4138	862
p-value	**0.0002**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**
β	0.2958	0.0185	0.0370
Feeling that situation is hopeless			
n	5184	4139	862
p-value	**$<1\times10^{-5}$**	**3×10^{-5}**	**0.0001**
β	0.3846	0.0172	0.0304

p-value (exploratory): multivariate regression to predict the respective mood related symptoms adjusted for covariates by including AAE, gender, and disease duration (time between AAO and current age) in the model. Significant values are in [bold].

Discussion:

In our exploratory analysis of the Fox Insight cohort, coffee drinking showed very little association with the severity of motor symptoms that we investigated in generalized linear models while adjusting for AAE, gender, and disease duration. Coffee drinking demonstrated a negative relationship with chewing and swallowing when coffee drinking was used as a binary yes-no indication, indicating fewer problems with chewing and swallowing when drinking coffee. Caffeine has previously been depicted to improve the motor deficits in PD or decelerate PD progression and ameliorate both motor and non-motor early symptoms (Hong et al., 2020; Prediger, 2010). There was a lack of association between non-motor symptoms and general coffee consumption as well as coffee intake duration in our study, however, we found a positive relationship between unexplained pains and the coffee drinking dosage.

Nevertheless, the effect of coffee consumption and potential long-term effects need to be investigated in further confirmatory and longitudinal studies. Long-term effects of coffee drinking may positively impact memory, cognition, and verbal retrieval (Cho et al., 2018; Ritchie et al., 2007; van Gelder et al., 2007).

We found an association between the general intake of aspirin and the number of pills per week with selected motor and non-motor symptoms, indicating more problems with the reported symptoms when using aspirin. In contrast, aspirin intake duration did not show an association with neither motor symptoms nor non-motor symptoms. Nevertheless, it is important to note that aspirin users are a heterogeneous group: the medication can be used for various reasons (e.g. anti-aggregation, cerebrovascular, and cardiovascular problems). Thus, we included heart diseases as an additional comorbidity in our models and the associations with motor and non-motor features for aspirin were still robust. Nevertheless, the potential confounding effect of other comorbidities needs to be considered. The clinical effect of NSAIDs is still subject to controversial discussion. Although anti-inflammatory drugs such as aspirin and ibuprofen are known to reduce the risk for PD or delay PD onset (Ascherio and Schwarzschild, 2016; Chen et al., 2005; Gabbert et al., 2022; Wahner et al., 2007), the clinical impact on motor and non-motor symptom severity and progression remains unclear.

In addition, we found a positive association between the general smoking status, number of cigarettes per day, and smoking duration with selected motor symptoms. Smokers with PD exhibited more problems with speech, too much saliva, chewing and swallowing, walking and balance, freezing, and getting up with a different level of severity depending on the smoking dosage. The underlying cause of more severe motor symptoms in smokers remains unclear. Since smokers in this cohort were on average older than non-smokers but had a shorter disease duration (Supplementary Figures S2-S8), we would rather have expected less severe motor symptoms. In line with our findings, previous studies have reported later motor symptom

onsets in smokers compared to non-smokers (Gigante et al., 2017), however, there was no report on more severe motor symptoms in smokers, but similar baseline motor deficits in smokers and non-smokers (Lee et al., 2017). An excessive amount of saliva and swallowing dysfunctions resulting in drooling are clinically relevant symptoms in PD (Nascimento et al., 2021; Simons, 2017), however, smokers have thicker saliva compared to non-smokers, who tend to have predominantly serous saliva and the amount of saliva was found to decrease with the duration of smoking in the short term (Petrusic et al., 2015). Interestingly, we observe that smokers with PD had increased saliva excess and drooling rather than decreased saliva in the long-term (mean disease duration of 6.5 years SD: 5.7 years). There are also no reports on the effect of smoking on other bradykinetic symptoms like freezing or walking difficulties that would explain the increased problems in smokers. This further highlights the importance to investigate interactions between lifestyle factors and motor symptoms. When extending the investigations to non-motor symptoms, more problems with selected non-motor symptoms were observed when smoking. In addition, smoking correlated with symptoms related to depression and mood. A greater likelihood to experience feelings related to depression was reported when smoking or former smoking compared to never smoking. In general, there is evidence of an established relationship between smoking and mental health, showing that smoking increases the number of days with poor mental health, especially among individuals with more severe illnesses (Plurphanswat et al., 2017; Snell et al., 2021). Still, the causal effect remains unclear. In other words, smoking itself may promote depression and anxiety, or patients with depression are just more likely to smoke and have greater difficulty quitting, which could indicate a reverse causation (Tomita and Manuel, 2020; Weinberger et al., 2017). Reverse causation is a general risk that appears in cross-sectional studies. In addition, the results only show correlations but do not necessarily indicate causalities. To examine whether smoking has an actual impact on motor and non-motor symptoms, longitudinal studies need to be performed in the future to determine a possible long-lasting effect of tobacco use and smoking on PD-related symptoms. The impact of smoking should also be assessed for intake after AAO longitudinally.

One major strength of our study was the large sample size that provided sufficient power to assess lifestyle factors and clinical severity of motor and non-motor symptoms. The variability in age and age at onset did not affect our results after adjustment in our regression models. Although a mean AAO of 60.4 years is within the typical range for patients with PD (Grover et al., 2022), the proportion of patients with early-onset PD (AAO<50 years) is ~17%, slightly higher than expected (Kolicheski et al., 2022). Patients with an earlier PD onset might be more interested in participating in online studies. The same might also apply to individuals with a higher educational level. According to a study in 2020, the Fox Insight cohort has a greater educational attainment as compared to other cohorts (Chahine et al., 2020). When comparing

the educational level in the different subgroups of lifestyle factors, we found smokers to have an overall lower educational level compared to non-smokers (Supplementary Figure S9). In addition, the smoking duration showed a trend of being shorter in the group of smokers with a higher educational level compared to smokers with a lower educational level. In contrast, there was a trend of a longer coffee drinking duration in the group of coffee drinkers with a higher educational level compared to coffee drinkers with a lower educational level (Supplementary Figure S10). There was no difference in the educational level between aspirin users and non-aspirin users (Supplementary Figure S11).

Additionally, as the Fox Insight study collects self-report data online, it offers many possibilities to promote epidemiological research. The convenience and accessibility for the participants and researchers allow easier patient recruitment and higher rates of return. This was previously investigated in a study by comparing the Fox Insight PD cohort's self-reported demographic characteristics, symptoms, medical history, and PD medication use to other in-person observational research study cohorts (Chahine et al., 2020). The authors found that patterns of responses to patient-reported assessments that were obtained online on the PD cohort of the Fox Insight study resembled PD cohorts assessed in person. In addition, patient-reported outcomes are becoming increasingly important to research, therapeutic development, and healthcare delivery, which was already investigated in another previous study on the Fox Insight cohort (Dobkin et al., 2020).

However, we were limited to the questionnaires and data collected by Fox Insight in this study, including the selection of environmental factors and the types of questions. The motor and non-motor symptom scores were subjective evaluations and might differ from assessments by movement disorder specialists as the motor symptom questionnaire enquires about symptom severity over the past week. This way, it might reflect a personal snapshot of the current severity of symptoms but does not take possible off-episodes into account. Motor fluctuations are a major problem in advanced levodopa-treated PD patients, leading to "off" states, in which disability increases (Chou et al., 2018; Dewey, 2004). Adjusting for off-episodes in the regression models might be a crucial point to interpret the scores properly. Unfortunately, we were limited to the data collected by Fox Insight in this study and the numbers of PD patients responding to questionnaires about off-episodes were not sufficient to include in our analyses. In addition, it is important to mention that results were not corrected for multiple testing as analyses were not based on the presence of an 'a priori' hypothesis. Thus, they cannot be interpreted as significant after multiple testing correction and *p*-values remain descriptive.

Nevertheless, these findings may help to acquire a better understanding of this complex disease. This study comprehensively assesses the effect of smoking, coffee drinking, and aspirin intake on clinical symptoms. These findings are so far only exploratory, however, they set the stage for future longitudinal assessments on these factors and PD clinical features.

Data availability:

Data used in the preparation of this article were obtained from the Fox Insight database (https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp) on 18/10/2020. For up-to-date information on the study, visit https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp.

Acknowledgments:

The Fox Insight Study (FI) is funded by The Michael J. Fox Foundation for Parkinson's Research. We would like to thank the Parkinson's community for participating in this study to make this research possible.

Author contributions:

C.G.: formal analysis, investigation, data curation, visualization, writing - original draft. I.K.: methodology, validation, writing - review and editing. T.L.: validation, writing - review and editing. M.K.: writing - review and editing. A.G.: writing - review and editing. C.K.: resources, writing - review and editing. J.T.: conceptualization, methodology, investigation, resources, writing - original draft, supervision, project administration, funding acquisition.

Funding:

Open Access funding enabled and organized by Projekt DEAL. This study was funded by The Michael J. Fox Foundation for Parkinson's Research (Analysis of Patient-Reported Outcomes from Fox Insight). JT, MK, IRK, AG, and CK are supported by the DFG (FOR 2488). In addition, AG is supported by the Luxembourgish Research Fund (FNR) within the framework of the ATTRAC T (Model-IPD, FNR9631103) and INTER programs (ProtectMove II, INTER/DFG/19/14429377).

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/s41598-023-31531-w.

Correspondence and requests for materials should be addressed to J.T.

Reprints and permissions information is available at www.nature.com/reprints.

References:

Alves, G., Kurz, M., Lie, S.A., and Larsen, J.P. (2004). Cigarette smoking in Parkinson's disease: influence on disease progression. Mov Disord *19*, 1087-1092.

Ascherio, A., and Schwarzschild, M.A. (2016). The epidemiology of Parkinson's disease: risk factors and prevention. Lancet Neurol *15*, 1257-1272.

Chahine, L.M., Chin, I., Caspell-Garcia, C., Standaert, D.G., Brown, E., Smolensky, L., Arnedo, V., Daeschler, D., Riley, L., Korell, M.*, et al.* (2020). Comparison of an Online-Only Parkinson's Disease Research Cohort to Cohorts Assessed In Person. J Parkinsons Dis *10*, 677-691.

Chen, H., Jacobs, E., Schwarzschild, M.A., McCullough, M.L., Calle, E.E., Thun, M.J., and Ascherio, A. (2005). Nonsteroidal antiinflammatory drug use and the risk for Parkinson's disease. Ann Neurol *58*, 963-967.

Cho, B.H., Choi, S.M., and Kim, B.C. (2019). Gender-dependent effect of coffee consumption on tremor severity in de novo Parkinson's disease. BMC Neurol *19*, 194.

Cho, B.H., Choi, S.M., Kim, J.T., and Kim, B.C. (2018). Association of coffee consumption and non-motor symptoms in drug-naive, early-stage Parkinson's disease. Parkinsonism Relat Disord *50*, 42-47.

Chou, K.L., Stacy, M., Simuni, T., Miyasaki, J., Oertel, W.H., Sethi, K., Fernandez, H.H., and Stocchi, F. (2018). The spectrum of "off" in Parkinson's disease: What have we learned over 40 years? Parkinsonism Relat Disord *51*, 9-16.

Dewey, R.B., Jr. (2004). Management of motor complications in Parkinson's disease. Neurology *62*, S3-7.

Dobkin, R.D., Amondikar, N., Kopil, C., Caspell-Garcia, C., Brown, E., Chahine, L.M., Marras, C., Dahodwala, N., Mantri, S., Standaert, D.G.*, et al.* (2020). Innovative Recruitment Strategies to Increase Diversity of Participation in Parkinson's Disease Research: The Fox Insight Cohort Experience. J Parkinsons Dis *10*, 665-675.

Dunn, A.R., O'Connell, K.M.S., and Kaczorowski, C.C. (2019). Gene-by-environment interactions in Alzheimer's disease and Parkinson's disease. Neurosci Biobehav Rev *103*, 73-80.

Gabbert, C., Konig, I.R., Luth, T., Kolms, B., Kasten, M., Vollstedt, E.J., Balck, A., Grunewald, A., Klein, C., and Trinh, J. (2022). Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson's disease. J Neurol.

Gigante, A.F., Martino, T., Iliceto, G., and Defazio, G. (2017). Smoking and age-at-onset of both motor and non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord *45*, 94-96.

Grover, S., Kumar Sreelatha, A.A., Pihlstrom, L., Domenighetti, C., Schulte, C., Sugier, P.E., Radivojkov-Blagojevic, M., Lichtner, P., Mohamed, O., Portugal, B.*, et al.* (2022). Genome-wide Association and Meta-analysis of Age at Onset in Parkinson Disease: Evidence From the COURAGE-PD Consortium. Neurology *99*, e698-e710.

Hong, C.T., Chan, L., and Bai, C.H. (2020). The Effect of Caffeine on the Risk and Progression of Parkinson's Disease: A Meta-Analysis. Nutrients *12*.

Kandinov, B., Giladi, N., and Korczyn, A.D. (2007). The effect of cigarette smoking, tea, and coffee consumption on the progression of Parkinson's disease. Parkinsonism Relat Disord *13*, 243-245.

Kline, E.M., Houser, M.C., Herrick, M.K., Seibler, P., Klein, C., West, A., and Tansey, M.G. (2021). Genetic and Environmental Factors in Parkinson's Disease Converge on Immune Function and Inflammation. Mov Disord *36*, 25-36.

Kolicheski, A., Turcano, P., Tamvaka, N., McLean, P.J., Springer, W., Savica, R., and Ross, O.A. (2022). Early-Onset Parkinson's Disease: Creating the Right Environment for a Genetic Disorder. J Parkinsons Dis *12*, 2353-2367.

Lee, Y., Oh, J.S., Chung, S.J., Chung, S.J., Kim, S.J., Nam, C.M., Lee, P.H., Kim, J.S., and Sohn, Y.H. (2017). Does smoking impact dopamine neuronal loss in de novo Parkinson disease? Ann Neurol *82*, 850-854.

Marras, C., Canning, C.G., and Goldman, S.M. (2019). Environment, lifestyle, and Parkinson's disease: Implications for prevention in the next decade. Mov Disord *34*, 801-811.

Moccia, M., Mollenhauer, B., Erro, R., Picillo, M., Palladino, R., and Barone, P. (2015). Non-Motor Correlates of Smoking Habits in de Novo Parkinson's Disease. J Parkinsons Dis *5*, 913-924.

Nascimento, D., Carmona, J., Mestre, T., Ferreira, J.J., and Guimaraes, I. (2021). Drooling rating scales in Parkinson's disease: A systematic review. Parkinsonism Relat Disord *91*, 173-180.

Petrusic, N., Posavac, M., Sabol, I., and Mravak-Stipetic, M. (2015). The Effect of Tobacco Smoking on Salivation. Acta Stomatol Croat *49*, 309-315.

Plurphanswat, N., Kaestner, R., and Rodu, B. (2017). The Effect of Smoking on Mental Health. Am J Health Behav *41*, 471-483.

Postuma, R.B., Anang, J., Pelletier, A., Joseph, L., Moscovich, M., Grimes, D., Furtado, S., Munhoz, R.P., Appel-Cresswell, S., Moro, A.*, et al.* (2017). Caffeine as symptomatic treatment for Parkinson disease (Cafe-PD): A randomized trial. Neurology *89*, 1795-1803.

Prediger, R.D. (2010). Effects of caffeine in Parkinson's disease: from neuroprotection to the management of motor and non-motor symptoms. J Alzheimers Dis *20 Suppl 1*, S205-220.

Ritchie, K., Carriere, I., de Mendonca, A., Portet, F., Dartigues, J.F., Rouaud, O., Barberger-Gateau, P., and Ancelin, M.L. (2007). The neuroprotective effects of caffeine: a prospective population study (the Three City Study). Neurology *69*, 536-545.

Simons, J.A. (2017). Swallowing Dysfunctions in Parkinson's Disease. Int Rev Neurobiol *134*, 1207-1238.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W.*, et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Snell, M., Harless, D., Shin, S., Cunningham, P., and Barnes, A. (2021). A longitudinal assessment of nicotine dependence, mental health, and attempts to quit Smoking: Evidence from waves 1-4 of the Population Assessment of Tobacco and Health (PATH) study. Addict Behav *115*, 106787.

Tomita, A., and Manuel, J.I. (2020). Evidence on the Association Between Cigarette Smoking and Incident Depression From the South African National Income Dynamics Study 2008-2015: Mental Health Implications for a Resource-Limited Setting. Nicotine Tob Res *22*, 118-123.

van Gelder, B.M., Buijsse, B., Tijhuis, M., Kalmijn, S., Giampaoli, S., Nissinen, A., and Kromhout, D. (2007). Coffee consumption is inversely associated with cognitive decline in elderly European men: the FINE Study. Eur J Clin Nutr *61*, 226-232.

Villafane, G., Thiriez, C., Audureau, E., Straczek, C., Kerschen, P., Cormier-Dequaire, F., Van Der Gucht, A., Gurruchaga, J.M., Quere-Carne, M., Evangelista, E.*, et al.* (2018). High-dose transdermal nicotine in Parkinson's disease patients: a randomized, open-label, blinded-endpoint evaluation phase 2 study. Eur J Neurol *25*, 120-127.

Wahner, A.D., Bronstein, J.M., Bordelon, Y.M., and Ritz, B. (2007). Nonsteroidal anti-inflammatory drugs may protect against Parkinson disease. Neurology *69*, 1836-1842.

Weinberger, A.H., Kashan, R.S., Shpigel, D.M., Esan, H., Taha, F., Lee, C.J., Funk, A.P., and Goodwin, R.D. (2017). Depression and cigarette smoking behavior: A critical review of population-based studies. Am J Drug Alcohol Abuse *43*, 416-431.

2.3 Objective 3: Variants in the *GBA1* gene in patients with PD and healthy controls

Variants in the *GBA1* gene are considered the most common genetic risk factor for PD. However, the frequency of *GBA1* variant carriers vastly differs across study groups, which additionally depends on the ethnicity of the patients and controls and the *GBA1* variants that were investigated. The investigation of variants in *GBA1* is further complicated by the nearby pseudogene *GBAP1*, which is located approximately 12 kb downstream of *GBA1* and shares 96% exonic sequence homology with the *GBA1* coding region (Leija-Salazar et al., 2019). The pseudogene confers an increased risk for recombination between homologous regions, resulting in complex genomic structures that are difficult to sequence. Long-read sequencing allows sequencing of the entire *GBA1* gene, including exons and introns, without assessing the pseudogene *GBAP1*.

In this study, Oxford Nanopore long-read sequencing was evaluated as a strategy to determine the frequency of *GBA1* variants in a study group of Norwegian patients with PD and controls. By comparing six analysis pipelines, consisting of two aligners (i.e., NGMLR, Minimap2) and three variant callers (i.e., BCFtools, Clair3, and Pepper-Margin-Deepvariant), the most accurate analysis tools to detect variants in *GBA1* were evaluated. BCFtools performed best with regard to false-positive and false-negative variant calls, independent from the aligner used, resulting in the detection of 13 rare *GBA1* variants in the study group. Pathogenicity scoring predicted two of these variants to be “pathogenic” or “likely pathogenic”, while the other eleven variants were of “uncertain significance”. Of the patients with PD, 4.3% carried one of the most common *GBA1* risk variants, p.N409S or p.L483P, while in comparison, only 1.1% of healthy controls carried one of these variants. To compare these variant frequencies to other studies, a systematic literature review was performed, which included 100 articles and showed *GBA1* frequencies ranging from 0% to 26.3% in patients with PD and 0% to 5.96% in controls in different populations.

The study covering this objective is titled “*GBA1* in Parkinson’s disease: variant detection and pathogenicity scoring matters”. The samples used in the paper were part of a Norwegian cohort from Prof. Jan Aasly, who was the Director of the Department of Neurology at St. Olav's Hospital in Trondheim. I coordinated the shipments and received all available clinical information from patients and controls (>2200 samples). Subsequently, I harmonized the data, sorted the patients, unaffected family members, and healthy controls according to their PD status, known variant carrier status, and family history, and arranged the samples on plates following these criteria. For the DNA sequencing, I coordinated the implementation of the project with Susen Schaake, who performed the sequencing on the Oxford Nanopore GridION in the lab. I developed a data analysis script for the following data analysis, which included filtering for quality, mapping and alignment, variant calling, and annotation. For this, I had to

establish new software tools on the high-performance OMICS cluster of the University, which I coordinated with the IT department. I also had to optimize several software-specific parameters, as those pipelines are highly customizable. To identify rare “pathogenic”/“likely pathogenic”/“uncertain” variants in *GBA1*, I assessed the pathogenicity of all variants using several prediction tools. These variants that were predicted to be “pathogenic”/“likely pathogenic”/“uncertain” were Sanger sequenced by Christoph Much, and I subsequently analyzed the results from the Sanger sequencing. In a sensitivity assessment, I evaluated all variants that were called in the six analysis pipelines and that were validated with Sanger sequencing. To compare the number of patients and controls with and without *GBA1* variants, I performed the statistical analyses. Furthermore, to put the variant frequencies found in this Norwegian study group into a larger context, I conducted a systematic literature review to summarize the frequency of *GBA1* variants across populations. Finally, I wrote the first draft of the manuscript, including the preparation of all tables and figures of the main text and the Supplementary Material, and the response letter in the subsequent peer-review process, supervised and revised by PD Dr. Joanne Trinh.

Title:

GBA1 in Parkinson’s disease: variant detection and pathogenicity scoring matters

Authors:

Carolin Gabbert[1], MSc, Susen Schaake[1], BSc, Theresa Lüth[1], MSc, Christoph Much[1], Christine Klein[1], MD, Jan O. Aasly[2], MD, Matthew J. Farrer, PhD[3], Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
[2]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway
[3]Department of Neurology, University of Florida, Gainesville, FL, USA
[*]Corresponding author

Published in *BMC Genomics*, 2023, DOI: doi.org/10.1186/s12864-023-09417-y

Abstract:

Background

GBA1 variants are the strongest genetic risk factor for Parkinson's disease (PD). However, the pathogenicity of *GBA1* variants concerning PD is still not fully understood. Additionally, the frequency of *GBA1* variants varies widely across populations.

Objectives

To evaluate Oxford Nanopore sequencing as a strategy, to determine the frequency of *GBA1* variants in Norwegian PD patients and controls, and to review the current literature on newly identified variants that add to pathogenicity determination.

Methods

We included 462 Norwegian PD patients and 367 healthy controls. We sequenced the full-length *GBA1* gene on the Oxford Nanopore GridION as an 8.9 kb amplicon. Six analysis pipelines were compared using two aligners (NGMLR, Minimap2) and three variant callers (BCFtools, Clair3, Pepper-Margin-Deepvariant). Confirmation of *GBA1* variants was performed by Sanger sequencing and the pathogenicity of variants was evaluated.

Results

We found 95.8% (115/120) true-positive *GBA1* variant calls, while 4.2% (5/120) variant calls were false-positive, with the NGMLR/Minimap2-BCFtools pipeline performing best. In total, 13 rare *GBA1* variants were detected: two were predicted to be (likely) pathogenic and eleven were of uncertain significance. The odds of carrying one of the two common *GBA1* variants, p.L483P or p.N409S, in PD patients were estimated to be 4.11 times the odds of carrying one of these variants in controls (*OR*=4.11 [1.39, 12.12]).

Conclusions

In conclusion, we have demonstrated that Oxford long-read Nanopore sequencing, along with the NGMLR/Minimap2-BCFtools pipeline is an effective tool to investigate *GBA1* variants. Further studies on the pathogenicity of *GBA1* variants are needed to assess their effect on PD.

Background:

Parkinson's disease (PD) is a multifaceted and highly complex neurodegenerative disorder. Multiple genes have been implicated in PD. Variants in *GBA1* (Glucocerebrosidase A) are considered a common genetic risk factor for PD (Gan-Or et al., 2008; Gan-Or et al., 2018; Greuel et al., 2020; Neumann et al., 2009; Sidransky and Lopez, 2012; Sidransky et al., 2009). Biallelic (homozygous or compound heterozygous) variants in *GBA1* classically cause

Gaucher's disease (GD) and an increased PD risk has been observed in patients with GD and asymptomatic carriers of heterozygous variants (Alcalay et al., 2015; Goker-Alpan et al., 2004; Lwin et al., 2004; Riboldi and Di Fonzo, 2019). Glucocerebrosidase enzymatic activity is reduced in patients with PD who carry a *GBA1* heterozygous variant compared to noncarriers, and it is even lower in *GBA1* homozygotes/compound heterozygotes (Alcalay et al., 2015). Common *GBA1* variants in PD include p.E365K (NM_000157.4, c.1093G>A), p.T408M (NM_000157.4, c.1223 C>T), p.N409S (NM_000157.4, c.1226 A>G), and p.L483P (NM_000157.4, c.1448T>C). However, the classification of the pathogenicity of *GBA1* variants and their effect on PD is still ongoing. For this reason, the reported frequencies of *GBA1* variants across studies are rather inconsistent, with frequencies ranging from 1.8% up to 47% depending on the ethnicity of the samples and the *GBA1* variants investigated (Behl et al., 2021; Gan-Or et al., 2018; Sidransky and Lopez, 2012). In a previous study on the Norwegian population, 311 patients with PD were included and screened for the two common *GBA1* variants (i.e., p.N409S and p.L483P) (Toft et al., 2006). Seven patients (2.3%) that carried a heterozygous *GBA1* variant were found: four of the patients had a p.N409S (1.3%) and three had a p.L483P (1.0%) substitution.

Another challenge that arises when sequencing the *GBA1* gene is the nearby pseudogene *GBAP1*. *GBAP1* shares 96% exonic sequence homology with the *GBA1* coding region with the highest homology between exons 8 and 11. In this region, most pathogenic variants have been reported, usually resulting from recombination events, e.g. gene conversion, fusion, or duplication (Hruska et al., 2008). This complex regional genomic structure complicates PCR and DNA sequencing. To avoid the pseudogene, one method to analyze *GBA1* is by long-read sequencing (Ip et al., 2015). This technology provides full-length *GBA1* sequencing to detect exonic and intronic variants and recombinant alleles in combination with phase information, at high multiplex capacity (Leija-Salazar et al., 2019; Toffoli et al., 2022).

Herein, we have comprehensively characterized *GBA1* in a sample from the Norwegian population by (1) employing and evaluating Oxford Nanopore sequencing as a strategy, (2) determining the frequency of variants within the *GBA1* gene in patients with PD and healthy controls by providing an update on previous reports (Lunde et al., 2018; Toft et al., 2006), and (3) reviewing current literature on newly identified variants that add to pathogenicity determination.

Methods:

Demographics

A sample of 462 Norwegian patients with PD was included in this study (Table 1). All patients were referred by general practitioners and other hospitals and have been clinically examined

and observed longitudinally at the outpatient clinics of three hospitals in Central Norway. One hundred eighty (39%) of the patients were men, and 282 (61%) were women. The mean age at disease onset in the patient group was 60.3 years (*SD*=±9.7 years, range 26 to 88 years). Forty-three out of 462 patients were probands with a family history of PD. Patients with a known genetic cause of PD were not included. In the clinical assessment, patients with PD had an average score of 2.72 (*SD*=±0.86) on the Hoehn and Yahr scale and 357 patients reported a tremor, while 62 had no tremor. In addition, a group of 367 healthy Norwegian individuals (mean age 64.0 years) originating from the same geographic region and without signs of a movement disorder was included to determine the variant frequency in the general population (Table 1). Some of the patients and controls included in this study have been previously included (Toft et al., 2006). However, in this previous report only the p.N409S and the p.L483P have been screened by PCR amplification and subsequent digestion of the PCR product with restriction enzymes and separation of resulting fragments by agarose gel electrophoresis.

Table 1: Demographics of the Norwegian patients with PD and healthy controls.

Full Cohort (*N*=829)	**Patients with PD**	**Healthy Controls**
N	462	367
Male/Female (%)	180/282 (39%/61%)	159/198 (43%/54%)
Mean AAO (*SD*, range)	60.3 (±9.7, 26-88)	NA
Mean Age (*SD*, range	NA	64.0 (±12.2, 30-96)
Family history of PD	43	0

Genetic analysis

Long-read Oxford Nanopore sequencing

We used blood-derived genomic DNA samples from all PD patients and controls. Informed consent was obtained from all participating individuals. We enriched for *GBA1* by amplifying an 8.9 kb sequence, which covered all coding exons, the introns between them, and part of the 3' UTR region (hg38: chr1:155,232,501-155,241,415), described previously (Leija-Salazar et al., 2019; Toffoli et al., 2022) using the LongAmp Taq PCR Kit. Subsequently, 1.3 µg of each patient-derived PCR product was barcoded with the Native 96 Barcoding Kit (EXP-NBD196) and multiplexed. The libraries were generated with the Ligation Sequencing Kit (SQKLSK109) for long-read Nanopore sequencing on R9.4.1 flow cells (FLO-MIN106) on a GridION.

Bioinformatic analyses

Data acquisition and run monitoring was carried out with MinKNOW (version v21.05.25 and later). The integrated Guppy algorithm (version v5.0.16 and later) was used for base-calling with the super-accurate base-calling model, de-multiplexing, and FAST5 and FASTQ file generation. The base-called reads were filtered with Filtlong (v.0.2.0) (https://github.com/rrwick/Filtlong) to only include the best 50% of the reads, based on Phred quality scores (q-score) in the FASTQ files, with a minimum read length of 8 kb. Afterwards, the reads were trimmed with NanoFilt (De Coster et al., 2018) (v2.8.0) and 75 bp were cropped from the front of the reads and 20 bp from the end. Subsequently, the nanopore reads were aligned against the reference sequence (hg38). We used two different aligners: NGMLR (Sedlazeck et al., 2018) (v0.2.7) and Minimap2 (Li, 2021) (v2.22). Then, the alignments were sorted and indexed with SAMtools (Danecek et al., 2021) using v1.9 for the NGMLR alignment, and v1.15 for the Minimap2 alignment. In addition, the coverage for each sample was calculated using SAMtools (Danecek et al., 2021) (v.1.15). The processed BAM files were analyzed with three different variant callers: BCFtools (Danecek et al., 2021) (v1.9), Clair3 (v0.1-r11), and the Pepper-Margin-Deepvariant pipeline (Shafin et al., 2021) (Supplementary Figure S1). Finally, the resulting VCF files containing the SNPs within *GBA1* were annotated using ANNOVAR (Yang and Wang, 2015) (version 2020-06-11).

Sanger sequencing and structural variant detection

Sanger sequencing was performed for all individuals with a rare “pathogenic”/“likely pathogenic”/“uncertain” *GBA1* variant, as previously described (Crossley et al., 2020). Individuals with a rare “pathogenic”/“likely pathogenic”/“uncertain” *GBA1* variant were further examined for possible structural variants (SVs) due to the high exonic sequence homology between *GBA1* and *GBAP1*. We used two sets of primers, as previously described (Toffoli et al., 2022), to detect reciprocal crossovers between the gene and pseudogene resulting in a 20.6 kb deletion or a 20.6 kb duplication using the LongAmp Taq PCR Kit. The resulting PCR products were subsequently run in a 1.5% agarose gel.

Sensitivity assessment

To assess the performance of variant calling, we stratified the detected *GBA1* variants into true-positive, false-positive, and false-negative calls for each data analysis pipeline (NGMLR+BCFtools, NGMLR+Clair3, NGMLR+Pepper-Margin-Deepvariant, Minimap2+BCFtools, Minimap2+Clair3, Minimap2+Pepper-Margin-Deepvariant). True-positive variants were defined as *GBA1* variants detected with Nanopore sequencing that were validated with Sanger sequencing. False-positive variants are those identified with Nanopore

sequencing but not confirmed with Sanger sequencing. False-negative variants were determined as those not called with a specified data analysis pipeline, but later validated with Sanger sequencing. Evaluation of each data analysis pipeline was based on the ratio of false-positive to false-negative calls.

Pathogenicity scoring

GBA1 single nucleotide variants were first filtered based on GnomAD frequency <2%. Pathogenicity classification and scoring were assessed with American College of Medical Genetics and Genomics (ACMG) (Richards et al., 2015) criteria. This included using Varsome (Kopanos et al., 2019), ClinVar (Landrum et al., 2020), SIFT (Ng and Henikoff, 2003), Polyphen2 (Adzhubei et al., 2013), CADD (Rentzsch et al., 2019), and GERP++ (Huber et al., 2020). *GBA1* variants were categorized as “pathogenic”, “likely pathogenic”, or of “uncertain significance” and further validated by Sanger sequencing. The p.E365K variant, which is a known risk factor for PD, was additionally categorized as of “uncertain significance”, despite the “likely benign” score from ACMG. Variants categorized as of “uncertain significance” without information from mutation predictors were not Sanger sequenced.

Statistical analysis

To compare the number of Norwegian PD patients with and without *GBA1* variants to the frequencies of *GBA1* variants in controls, odds ratios were calculated. For this, the number of patients with PD carrying a *GBA1* variant was multiplied by the number of controls without a *GBA1* Sanger sequencing was performed for all individuals with a rare “pathogenic”/“likely pathogenic”/“uncertain” variant and subsequently divided by the number of patients with PD without a *GBA1* variant that was multiplied by the number of controls carrying a *GBA1* variant.

Literature review

We performed a systematic literature review to summarize *GBA1* variants detected in PD and the frequency across different populations (Supplementary Figure S2). We searched for literature via PubMed that was published before August 4, 2022, using the search term “*GBA*” AND “Parkinson” AND “prevalence” OR “*GBA*” AND “Parkinson” AND “frequency”, while setting the species filter to “Human” and the language filter to “English”, resulting in 94 articles. These were screened based on the title, abstract, and full text, excluding all articles not directly screening for variants in the *GBA1* gene in patients with PD. Of these 94 articles, 41 articles were excluded. Reasons for exclusion were reviews or comments without new data (*n*=11), articles that did not perform *GBA1* variant screening or examine *GBA1* variant frequency in their study population (*n*=22), and articles that did not include patients with PD (*n*=10). Multiple

reasons for exclusion were possible. In addition to the articles found via the search term, suitable articles that were referenced in this literature were also included in the overview.

Results:

Long-read sequencing of GBA1

After Nanopore sequencing of the long-range PCR products, we obtained a mean read length of 5.2 kb (*SD*=±2.1 kb) and a mean read quality Phred score of 14.2 (*SD*=±0.4) across all raw sequencing data. After length and quality filtering and read trimming, we obtained a mean read length of 9.0 kb (*SD*=±0.2 kb) and a mean read quality Phred score of 15.7 (*SD*=±0.5). For the filtered samples the mean coverage was 193.1X (*SD*=±187.5X), ranging between 20.6X and 1820.8X across all samples. Nevertheless, the coverage per sample was consistent over all positions (Supplementary Figure S3).

Bioinformatic pipeline comparison

In total, 79 rare *GBA1* variants (gnomAD frequency <2%) were detected in the Nanopore sequencing analysis after filtering. Out of the 79 rare *GBA1* variants, 18 variants were categorized as “pathogenic”, “likely pathogenic”, or of “uncertain significance” (Supplementary Table S1). The remaining 61 rare *GBA1* variants were categorized as “benign” or “likely benign”. The number of *GBA1* variants differed across all six analysis pipelines (Table 2).

BCFtools

With BCFtools, 64 rare annotated *GBA1* variants were detected in 313 samples, resulting in 433 calls in total. The calls were identical with both aligners (i.e., NGMLR and Minimap2). Of these, 15 rare variants, categorized as “pathogenic”, “likely pathogenic”, or of “uncertain significance”, were found in 111 samples (i.e., 120 calls), again independent of the aligner. After Sanger sequencing, 13 variants in 110 samples (i.e., 115 calls) were validated, indicating five false-positive calls but no false-negative calls.

Clair3

For the pipeline using Clair3 as a variant caller after a preceding alignment with NGMLR, 65 rare annotated *GBA1* variants were detected in 308 samples, resulting in 426 calls. Of these calls, 117 calls referred to 14 rare variants, categorized as “pathogenic”, “likely pathogenic”, or of “uncertain significance”, and were detected in 108 samples. After Sanger sequencing, 112 calls in 107 samples were validated, implying that five calls were false-positive, with an additional three calls being false-negative. When Minimap2 was used as an aligner, again 65

rare annotated *GBA1* variants were detected, however, here they were detected in 311 samples (i.e., 429 calls). With this pipeline, 15 rare annotated *GBA1* variants, categorized as "pathogenic", "likely pathogenic", or of "uncertain significance", were found in 110 samples (i.e., 119 calls). Of these, 113 calls in 108 samples could be validated with Sanger sequencing, indicating six false-positive calls and two more false-negative calls.

Pepper-Margin-Deepvariant

Lastly, the Pepper-Margin-Deepvariant pipeline was used for variant calling. After preceding alignment with NGMLR, 453 calls of 72 rare annotated *GBA1* variants were detected in 318 samples, including 123 calls of 17 rare variants, categorized as "pathogenic", "likely pathogenic", or of "uncertain significance", in 114 samples. Only 108 calls in 103 samples were validated with Sanger sequencing, leading to 15 false-positive and additional seven false-negative calls. Similarly, when Minimap2 was used for the alignment, 76 rare annotated *GBA1* variants were detected in 331 samples, resulting in 481 calls. Of these, 18 rare variants, categorized as "pathogenic", "likely pathogenic", or of "uncertain significance", were detected in 120 samples (i.e., 130 calls). Here, 109 calls in 104 samples were Sanger validated, indicating 21 false-positive and six false-negative calls.

Table 2: *GBA1* variants sequenced with Oxford Nanopore and analyzed with six different pipelines using NGMLR and Minimap2 aligners and BCFtools, Clair3, and Pepper-Margin-Deepvariant callers.

Alignment	**NGMLR**			**Minimap2**		
Variant calling	**BCFtools**	**Clair3**	**Pepper-Margin-Deepvariant**	**BCFtools**	**Clair3**	**Pepper-Margin-Deepvariant**
N of calls of the 79 rare *GBA1* variants*	433	426	454	433	429	481
N of calls of the 18 rare "pathogenic"/"likely pathogenic"/"uncertain *GBA1* variants**	120	117	123	120	119	130
N of Sanger validated calls	115	112	108	115	113	109
N of false-positive calls	5	5	15	5	6	21
N of false-negative calls	0	3	7	0	2	6

*Variants were annotated using Annovar; variants with a gnomAD frequency >2% and without information on SNP ID or amino acid change were excluded.
**Variants were classified based on the pathogenicity reported by ACMG, Varsome, Clinvar, SIFT, Polphen2, CADD, and GERP++.

GBA1 variant frequencies in the Norwegian population

In total, 13/18 rare distinct variants within *GBA1* were validated in 462 Norwegians with PD and 367 healthy controls (Figure 1, Supplementary Table S2), whereas 5/18 could not be validated. Two of the 13 rare *GBA1* variants were predicted to be "pathogenic" or "likely

pathogenic" (p.L483P, p.S146X) and eleven *GBA1* variants were of "uncertain significance" (p.G493D, p.N409S, p.T408M, p.A380T, p.R368C, p.E365K, p.D337G, p.S310G, p.R301H, p.R159W, p.R78C). The total carrier frequency of rare *GBA1* variants predicted as "pathogenic", "likely pathogenic", or of "uncertain significance" was 17.1% (79/462) in the PD cases and 8.4% (31/367) in the controls (*OR*=2.24 [1.44, 3.47]) (Table 3). The carrier frequency of known *GBA1* risk variants for PD, including p.L483P, p.N409S, p.T408M, p.E365K, and p.R159W, was 15.2% in the PD cases and 7.9% in the controls (*OR*=2.08 [1.32, 3.29]) (Table 3). With regard to the common and frequently investigated *GBA1* risk variants for PD, the frequency of carrying either a p.L483P or p.N409S variant was 4.3% in patients with PD and 1.1% in healthy controls (*OR*=4.11 [1.39, 12.12]) (Table 3).

In addition, samples with a rare "pathogenic", "likely pathogenic", or "uncertain" *GBA1* variant were further examined for possible structural variants (SVs). However, all samples tested negative for SVs.

Figure 1: Schematic representation of the exonic (gray boxes) and intronic (gray lines) structure of the *GBA1* gene highlighting "pathogenic" (red)/"likely pathogenic" (orange)/"uncertain" (yellow) variants found in our Norwegian series.

Table 3: Number and frequency of patients and controls with *GBA1* variants in our Norwegian series.

Number and frequency of *GBA1* variant carriers	**Rare* "pathogenic"/"likely pathogenic"/"uncertain" *GBA1* variants****			**Known *GBA1* risk variants for PD (p.L483P, p. N409S, p.T408M, p.E365K, p.R159W)**			**Common *GBA1* risk variants for PD (p.L483P, p. N409S)**		
	n	**Percentage**	***OR* (95% *CI*)**	**n**	**Percentage**	***OR* (95% *CI*)**	**n**	**Percentage**	***OR* (95% *CI*)**
PD cases (*n*=462)	79	17.1%	2.24 (1.44, 3.47)	70	15.2%	2.08 (1.32, 3.29)	20	4.3%	4.11 (1.39, 12.12)
Controls (*n*=367)	31	8.4%		29	7.9%		4	1.1%	

*Variants were annotated using Annovar; variants with a gnomAD frequency >2% and without information on SNP ID or amino acid change were excluded.
**Variants were classified based on the pathogenicity reported by ACMG, Varsome, Clinvar, SIFT, Polphen2, CADD, and GERP++.

Literature review

We performed a systematic literature review to summarize *GBA1* variants detected in PD and their frequencies across different populations. In total, 100 articles on *GBA1* variant frequencies across populations were included in the overview (Supplementary Table S3). Most of the studies assessed the p.L483P and p.N409S variants, with p.L483P variant frequencies ranging between 0% and 8.3% for cases with PD and between 0% and 1.4% in the general population. The highest frequency in PD patients was observed in a sample from the Japanese population (Li et al., 2014). For p.N409S, the frequencies ranged from 0% to 26.3% in patients with PD and from 0% to 5.96% in controls, with the highest frequency reported in a sample from the Ashkenazi Jewish population (Aharon-Peretz et al., 2004). In a previous study on the Norwegian population, the frequency of p.L483P was predicted to be 0.5% in patients with PD (2/442) and 1.4% in controls (6/419) (*OR*=0.31 [0.06, 1.56]) (Lunde et al., 2018). For the p.N409S variant, the frequencies reported were 0.2% in patients with PD (1/442) which was comparable to controls (1/419) (*OR*=0.95 [0.06, 15.2]) (Lunde et al., 2018). In addition, the study reported *GBA1* p.E365K, detected in 4.3% of patients with PD (18/442) and in 6.6% of controls (29/419) (*OR*=0.57 [0.13, 1.04]), and p.T408M, found in 1.7% of patients with PD (7/442) and in 3.6% of controls (16/419) (*OR*=0.41 [0.17, 1]).

Discussion:

Variants in the *GBA1* gene are known to affect PD risk, however, the frequency and pathogenicity of these variants are still under debate. The latter is further complicated as the frequencies of *GBA1* variants vary by population. In our study, we used the pathogenicity scoring of ACMG, Varsome, ClinVar, SIFT, Polyphen2, CADD, and GERP++ and categorized the variants found into “pathogenic”, “likely pathogenic”, and variants of “uncertain significance”. Here we report 13 rare variants within *GBA1* in Norwegian PD cases, with two of them predicted to be “pathogenic” or “likely pathogenic”, and eleven *GBA1* variants of “uncertain significance”. In total, we found a “pathogenic” or “likely pathogenic” variant (i.e., p.L483P, p.S146X) in 1.5% of PD cases and in 0% of healthy controls. However, pathogenicity scoring with ACMG does not take the association with PD into account and consequently underestimates the frequency of risk variants in *GBA1*. Therefore, we further investigated known *GBA1* risk variants associated with PD. In our study, 15.2% of patients with PD carried a common *GBA1* risk variant (i.e., p.L483P, p.N409S, p.T408M, p.E365K, and p.R159W), compared to 7.9% of controls.

In a systematic literature review, we evaluated the frequency of variants in the *GBA1* gene in patients with PD and in the general population in 100 studies (Supplementary Table S3). The frequencies of *GBA1* risk variants range between 0% and 26.3% in patients with PD,

highlighting the importance of investigating *GBA1* variants across populations. In a previous study on the Norwegian population, two known risk variants, p.L483P and p.N409S, were investigated among others (Lunde et al., 2018). The frequency of p.L483P was predicted to be 0.5% in patients with PD and 1.4% in controls, while 0.2% of the patients with PD and controls had a p.N409S variant (Lunde et al., 2018). Therefore, the odds of carrying a p.L483P or a p.N409S variant in patients with PD are lower than in healthy controls (*OR*=0.31 [0.06, 1.56], *OR*=0.95 [0.06, 15.2]). In contrast to this, we found slightly higher frequencies in our Norwegian series. The p.L483P variant was found in 1.3% of Norwegian PD cases and in 0% of controls, the p.N409S variant in 2.8% of PD patients and 1.1% of controls, leading to higher odds of carrying these variants in patients with PD compared to controls (*OR*=2.63 [0.85, 8.13]).

These variants are classically found in GD but are also associated with PD risk. However, some variants in *GBA1* show associations with PD but do not cause GD, e.g., the *GBA1* variants p.E365K and p.T408M (Greuel et al., 2020). In our Norwegian series, the p.E365K variant was found in 7.1% of cases with PD and in 2.7% of the general population (*OR*=2.75 [1.33, 5.65]). The p.T408M variant was found in 2.2% of cases with PD and in 3.5% of the general population (*OR*=0.6 [0.26, 1.39]). In a previous study on the Norwegian population, they reported p.E365K in 4.3% of patients with PD and in 6.6% of controls (*OR*=0.57 [0.13, 1.04]), and p.T408M in 1.7% of patients with PD and in 3.6% of controls (*OR*=0.41 [0.17, 1]) (Lunde et al., 2018). Several other studies have also reported higher variant frequencies of p.E365K or p.T408M in their control population compared to their patients with PD (Alcalay et al., 2015; Barkhuizen et al., 2017; Bras et al., 2009; Clark et al., 2007; Crosiers et al., 2016; Graham et al., 2020; Han et al., 2016; Kalinderi et al., 2009; Olszewska et al., 2020), questioning the pathogenicity of these *GBA1* variants and their contribution to causing PD that was previously assessed in two meta-analyses. In the summary of our systematic literature review (Table 4), we have also found on average higher frequencies of these two *GBA1* variants in patients with PD compared to healthy controls. Nevertheless, a better definition of variants associated with PD is strongly needed by classifying variants into categories relevant to the disease.

Although the pathogenicity and the pathomechanism of several *GBA1* variants in patients with PD are still under discussion, gene-targeted therapy might help to treat *GBA1*-PD. So far, there are several strategies for the treatment of *GBA1*-PD with ongoing clinical trials. Some of the therapeutic approaches for *GBA1* include substrate reduction therapy targeting glycosylceramide synthase inhibition, the inhibition of glucocerebrosidase transportation, the development of glucocerebrosidase activators, and gene therapy targeting the replacement of mutated *GBA1* with WT copies of the gene (Senkevich et al., 2022).

Table 4: Summary of frequencies in patients with PD and controls in publications on *GBA1* variants.

Population	Variants	*n* mutation carriers/n total (frequency) in PD	*n* mutation carriers/n total (frequency) in controls	*n* studies included*
White/Caucasian	p.L483P	329/21492 (1.53%)	101/17770 (0.57%)	34
	p.N409S	345/21714 (1.59%)	76/17770 (0.43%)	
	p.T408M	310/15266 (2.03%)	182/14298 (1.27%)	
	p.E365K	668/16408 (4.07%)	239/14849 (1.61%)	
	p.R159W	8/10011 (0.08%)	0/4630 (0%)	
South American	p.L483P	26/1237 (2.10%)	0/1024 (0%)	7
	p.N409S	14/1237 (1.13%)	1/1024 (0.10%)	
	p.T408M	3/635 (0.47%)	0/350 (0%)	
	p.E365K	7/635 (1.10%)	1/350 (0.29%)	
	p.R159W	2/635 (0.31%)	0/350 (0%)	
North African	p.L483P	1/227 (0.44%)	0/177 (0%)	2
	p.N409S	3/589 (0.51%)	3/549 (0.55%)	
	p.T408M	2/227 (0.88%)	0/177 (0%)	
	p.E365K	3/227 (1.32%)	1/177 (0.56%)	
	p.R159W	0/227 (0%)	0/177 (0%)	
Asian	p.L483P	261/10233 (2.55%)	7/5974 (0.12%)	23
	p.N409S	6/3739 (0.16%)	2/3319 (0.06%)	
	p.T408M	0/1963 (0%)	0/1157 (0%)	
	p.E365K	1/2161 (0.05%)	0/1398 (0%)	
	p.R159W	31/4786 (0.65%)	0/3285 (0%)	
Ashkenazi Jewish	p.L483P	8/2689 (0.30%)	6/10175 (0.06%)	9
	p.N409S	428/3117 (13.7%)	382/6795 (5.62%)	
	p.T408M	13/1935 (0.67%)	1/1000 (0.01%)	
	p.E365K	30/1935 (1.55%)	7/1000 (0.07%)	
	p.R159W	0/735 (0%)	0/622 (0%)	

* An overview of all studies included in the summary is provided in Supplementary Table S3.

In addition to variable inclusion criteria, the sequencing method and data analysis, and the chronology when it was performed, has a major influence on the detection rate and *GBA1* variant frequencies reported. Through the years, DNA sequencing technologies evolved tremendously, with new sequencing techniques and especially new prediction tools improving the accuracy of variant detection. Continual refinement in these tools enables more comprehensive identification of variants and highlights the need to re-evaluate known genes as time goes by. Long-read sequencing, in combination with the latest data analysis tools, enabled us to determine the frequency of *GBA1* variants in the Norwegian population with higher precision than before. We evaluated the consensus and accuracy of six different pipelines using two different aligners (NGMLR and Minimap2), as well as three different variant callers (BCFtools, Clair3, and the Pepper-Margin-Deepvariant pipeline). BCFtools performed best with regard to the number of true-positive, false-positive, and false-negative hits, independently from the aligner used. However, one limitation is that we could not fully evaluate pipeline sensitivity. As we only confirmed variants called by our Nanopore analysis pipelines by Sanger sequencing, we underestimate false-negative variants that were not initially called. With Oxford Nanopore technology we assessed the precision of long-read sequencing and

consensus data analysis and detected 115 real *GBA1* variant calls, while five variant calls were false-positives. Thus, >95% of called variants were true-positive. Nanopore long-read sequencing is an accurate tool to detect genetic variations and with further development in flow cells and sequencing kits, the accuracy of variant detection is likely to increase. Another advantage of this technology is the capacity to multiplex samples, which decreases analysis costs of the full 8.9 kb *GBA1* gene to $13 USD per sample, which is lower than other DNA sequencing methods (Yohe and Thyagarajan, 2017). A strength of Oxford Nanopore long-range sequencing is to specifically target the *GBA1* gene without sequencing the pseudogene *GBAP1*, and to detect all disease-causing variants including information on phase (Leija-Salazar et al., 2019; Toffoli et al., 2022).

Conclusions:

In conclusion, we have demonstrated that Oxford Nanopore sequencing is an efficient and scalable tool for investigating *GBA1* variants. We thoroughly evaluated six different data analysis pipelines and found the pipeline consisting of an alignment with either NGMLR or Minimap2 and variant calling with BCFtools to perform best with regard to detecting variants in *GBA1*. With our established and validated workflow, we demonstrated that the frequency of the two common *GBA1* risk variants for PD (i.e., p.L483P and p.N409S) in Norwegian patients with PD is 4.3% and higher than in the general population (1.1%). Furthermore, we reviewed current literature on *GBA1* variant frequencies in PD across populations, thereby adding to pathogenicity determination. Given the importance of this gene, further functional studies on the pathogenicity of *GBA1* are needed to assess their effect on PD.

Supplementary Information:

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09417-y.

Acknowledgments:

This project was supported by the Institute of Neurogenetics at the University of Lübeck. Special thanks to the families and patients who have participated in this study.

Authors' contributions:

C.G. performed the data analysis, data interpretation, and writing of the manuscript. S.S. performed the data acquisition, helped with data interpretation, and reviewed the manuscript. T.L. contributed to the data analysis, data interpretation, and reviewed the manuscript. C.M.

performed the data acquisition, helped with data interpretation, and reviewed the manuscript. C.K. reviewed the manuscript. J.O.A. was responsible for the sample acquisition and reviewed the manuscript. M.J.F. reviewed the manuscript. J.T. was responsible for the conceptualization of the project, data analysis, data interpretation, and writing of the manuscript.

Funding:

This project was supported by the Institute of Neurogenetics at the University of Lübeck. Grant support was received from the DFG RU ProtectMove (FOR2488).
Open Access funding enabled and organized by Projekt DEAL.

Data Availability:

The datasets generated and analyzed during the current study are available in the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) repository under the accession number: PRJNA939498 (SAMN33476944-SAMN33477772).

Declarations:

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. The ethics board of the University of Lübeck, Germany gave ethical approval for this work. Informed consent was obtained from all participating individuals.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References:

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet *Chapter 7*, Unit7 20.

Aharon-Peretz, J., Rosenbaum, H., and Gershoni-Baruch, R. (2004). Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med *351*, 1972-1977.

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.*, et al.* (2015). Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. Brain *138*, 2648-2658.

Barkhuizen, M., Anderson, D.G., van der Westhuizen, F.H., and Grobler, A.F. (2017). A molecular analysis of the GBA gene in Caucasian South Africans with Parkinson's disease. Mol Genet Genomic Med *5*, 147-156.

Behl, T., Kaur, G., Fratila, O., Buhas, C., Judea-Pusta, C.T., Negrut, N., Bustea, C., and Bungau, S. (2021). Cross-talks among GBA mutations, glucocerebrosidase, and alpha-synuclein in GBA-associated Parkinson's disease and their targeted therapeutic approaches: a comprehensive review. Transl Neurodegener *10*, 4.

Bras, J., Paisan-Ruiz, C., Guerreiro, R., Ribeiro, M.H., Morgadinho, A., Januario, C., Sidransky, E., Oliveira, C., and Singleton, A. (2009). Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal. Neurobiol Aging *30*, 1515-1517.

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.*, et al.* (2007). Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. Neurology *69*, 1270-1277.

Crosiers, D., Verstraeten, A., Wauters, E., Engelborghs, S., Peeters, K., Mattheijssens, M., De Deyn, P.P., Theuns, J., Van Broeckhoven, C., and Cras, P. (2016). Mutations in glucocerebrosidase are a major genetic risk factor for Parkinson's disease and increase susceptibility to dementia in a Flanders-Belgian cohort. Neurosci Lett *629*, 160-164.

Crossley, B.M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M.L., Clement, T., and Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. J Vet Diagn Invest *32*, 767-775.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M.*, et al.* (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*.

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. Bioinformatics *34*, 2666-2669.

Gan-Or, Z., Giladi, N., Rozovski, U., Shifrin, C., Rosner, S., Gurevich, T., Bar-Shira, A., and Orr-Urtreger, A. (2008). Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. Neurology *70*, 2277-2283.

Gan-Or, Z., Liong, C., and Alcalay, R.N. (2018). GBA-Associated Parkinson's Disease and Other Synucleinopathies. Curr Neurol Neurosci Rep *18*, 44.

Goker-Alpan, O., Schiffmann, R., LaMarca, M.E., Nussbaum, R.L., McInerney-Leo, A., and Sidransky, E. (2004). Parkinsonism among Gaucher disease carriers. J Med Genet *41*, 937-940.

Graham, O.E.E., Pitcher, T.L., Liau, Y., Miller, A.L., Dalrymple-Alford, J.C., Anderson, T.J., and Kennedy, M.A. (2020). Nanopore sequencing of the glucocerebrosidase (GBA) gene in a New Zealand Parkinson's disease cohort. Parkinsonism Relat Disord *70*, 36-41.

Greuel, A., Trezzi, J.P., Glaab, E., Ruppert, M.C., Maier, F., Jager, C., Hodak, Z., Lohmann, K., Ma, Y., Eidelberg, D.*, et al.* (2020). GBA Variants in Parkinson's Disease: Clinical, Metabolomic, and Multimodal Neuroimaging Phenotypes. Mov Disord *35*, 2201-2210.

Han, F., Grimes, D.A., Li, F., Wang, T., Yu, Z., Song, N., Wu, S., Racacho, L., and Bulman, D.E. (2016). Mutations in the glucocerebrosidase gene are common in patients with Parkinson's disease from Eastern Canada. Int J Neurosci *126*, 415-421.

Hruska, K.S., LaMarca, M.E., Scott, C.R., and Sidransky, E. (2008). Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). Hum Mutat *29*, 567-583.

Huber, C.D., Kim, B.Y., and Lohmueller, K.E. (2020). Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. PLoS Genet *16*, e1008827.

Ip, C.L.C., Loose, M., Tyson, J.R., de Cesare, M., Brown, B.L., Jain, M., Leggett, R.M., Eccles, D.A., Zalunin, V., Urban, J.M.*, et al.* (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Res *4*, 1075.

Kalinderi, K., Bostantjopoulou, S., Paisan-Ruiz, C., Katsarou, Z., Hardy, J., and Fidani, L. (2009). Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Greece. Neurosci Lett *452*, 87-89.

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C.E., Albarca Aguilera, M., Meyer, R., and Massouras, A. (2019). VarSome: the human genomic variant search engine. Bioinformatics *35*, 1978-1980.

Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C.*, et al.* (2020). ClinVar: improvements to accessing data. Nucleic Acids Res *48*, D835-D844.

Leija-Salazar, M., Sedlazeck, F.J., Toffoli, M., Mullin, S., Mokretar, K., Athanasopoulou, M., Donald, A., Sharma, R., Hughes, D., Schapira, A.H.V.*, et al.* (2019). Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. Mol Genet Genomic Med *7*, e564.

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics.

Li, Y., Sekine, T., Funayama, M., Li, L., Yoshino, H., Nishioka, K., Tomiyama, H., and Hattori, N. (2014). Clinicogenetic study of GBA mutations in patients with familial Parkinson's disease. Neurobiol Aging *35*, 935 e933-938.

Lunde, K.A., Chung, J., Dalen, I., Pedersen, K.F., Linder, J., Domellof, M.E., Elgh, E., Macleod, A.D., Tzoulis, C., Larsen, J.P.*, et al.* (2018). Association of glucocerebrosidase polymorphisms and mutations with dementia in incident Parkinson's disease. Alzheimers Dement *14*, 1293-1301.

Lwin, A., Orvisky, E., Goker-Alpan, O., LaMarca, M.E., and Sidransky, E. (2004). Glucocerebrosidase mutations in subjects with parkinsonism. Mol Genet Metab *81*, 70-73.

Neumann, J., Bras, J., Deas, E., O'Sullivan, S.S., Parkkinen, L., Lachmann, R.H., Li, A., Holton, J., Guerreiro, R., Paudel, R.*, et al.* (2009). Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. Brain *132*, 1783-1794.

Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res *31*, 3812-3814.

Olszewska, D.A., McCarthy, A., Soto-Beasley, A.I., Walton, R.L., Magennis, B., McLaughlin, R.L., Hardiman, O., Ross, O.A., and Lynch, T. (2020). Association Between Glucocerebrosidase Mutations and Parkinson's Disease in Ireland. Front Neurol *11*, 527.

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res *47*, D886-D894.

Riboldi, G.M., and Di Fonzo, A.B. (2019). GBA, Gaucher Disease, and Parkinson's Disease: From Genetic to Clinic to New Therapeutic Approaches. Cells *8*.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E.*, et al.* (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 *17*, 405-424.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods *15*, 461-468.

Senkevich, K., Rudakou, U., and Gan-Or, Z. (2022). New therapeutic approaches to Parkinson's disease targeting GBA, LRRK2 and Parkin. Neuropharmacology *202*, 108822.

Shafin, K., Pesout, T., Chang, P.C., Nattestad, M., Kolesnikov, A., Goel, S., Baid, G., Kolmogorov, M., Eizenga, J.M., Miga, K.H.*, et al.* (2021). Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. Nat Methods *18*, 1322-1332.

Sidransky, E., and Lopez, G. (2012). The link between the GBA gene and parkinsonism. Lancet Neurol *11*, 986-998.

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.*, et al.* (2009). Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med *361*, 1651-1661.

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.*, et al.* (2022). Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene. Commun Biol *5*, 670.

Toft, M., Pielsticker, L., Ross, O.A., Aasly, J.O., and Farrer, M.J. (2006). Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population. Neurology *66*, 415-417.

Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc *10*, 1556-1566.

Yohe, S., and Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. Arch Pathol Lab Med *141*, 1544-1557.

2.4 Objective 4: AAO associations of polygenic scores with environmental and lifestyle factors

2.4.1 Objective 4 A: The combined effect of coffee drinking, smoking, aspirin use, and the PGS on AAO in PD

The underlying mechanism that leads to the development of PD is, to this day, still not completely uncovered. About 5% of patients develop PD due to a monogenic cause, while 15% of patients carry a genetic risk variant, such as variants in *GBA1*. However, recent GWAS studies have shown that PD is highly polygenic. The largest meta-GWAS of PD identified 90 independent risk loci that may explain between 16% to 36% of the heritable risk of PD (Nalls et al., 2019). In addition to the genetic risk, a patient's lifestyle may further affect the risk and the age of developing PD. Environmental and lifestyle factors have been repeatedly investigated to assess their association with the risk of PD, AAO, and clinical severity, showing an increased risk for PD for being exposed to or using pesticides, heavy metals, and air pollution (Aune et al., 2023; Elbaz et al., 2004; Fitzmaurice et al., 2014; Gamache et al., 2019; Jafari et al., 2013; Ratner et al., 2014), and a protective effect for PD when using tobacco, coffee, black tea, and aspirin (Casper et al., 2000; Delamarre and Meissner, 2017; Domenighetti et al., 2022b; Fan et al., 2020; Grover et al., 2019; Heilbron et al., 2021; Hong et al., 2020; Larsson and Burgess, 2022; Mappin-Kasirer et al., 2020; Marras et al., 2019; Noyce et al., 2012; Wahner et al., 2007). Nevertheless, interaction studies that assess gene-environment interactions in PD are only sparse.

In the following study, the individual and combined effects of the PGS and the protective lifestyle factors coffee, tobacco, and aspirin on the AAO in patients with PD were evaluated. For this, patients and controls of European descent were included from the AMP-PD and Fox Insight cohorts. Multiple linear regression models were used to assess the association between AAO, PGS, and lifestyle factors. As expected, the PGS, which was calculated based on the nominated 1805 risk SNPs and weights provided by Nalls *et al*., showed a negative correlation with AAO in patients with PD from AMP-PD. In contrast, coffee drinking, tobacco use, and aspirin intake not only showed an individual association with a later AAO but additionally a combined effect, indicating a later AAO the more of these three factors were used. Additive effects were also found for PGS and tobacco, and for PGS and aspirin, however, no interactions were found. Interestingly, in patients with idiopathic PD, aspirin intake showed the greatest association with AAO, while in patients with *GBA1*-PD, only tobacco use was associated with AAO.

The study covering this objective is titled "The combined effect of lifestyle factors and polygenic scores on age at onset in Parkinson's disease". For the Fox Insight cohort, I downloaded the genetic data set, which is provided by 23andMe, from the Fox Data Exploration Network in

addition to the datasets previously described in Objective 1. I established a complete data analysis pipeline on our OMICS cluster for our group that has not been used before and is now frequently used by other students in our group. This pipeline included the conversion of the genetic data from a BGEN format to a PLINK format, performing all necessary steps of quality filtering and filtering for ethnicity, performing a principal component analysis, imputing the data on the Michigan Imputation Server with additional quality filtering, and the actual PGS analysis. Unfortunately, the number of PGS variants included in the Fox Insight genetic dataset was not sufficient to calculate a reliable PGS, even after I included proxy SNPs for the missing PGS SNPs in the analysis. Therefore, only the lifestyle data from Fox Insight were included in the analysis. I am in close contact with representatives from Fox Insight and 23andMe to solve this issue so that the genetic data can be used in future analyses. For the AMP-PD cohort, I filtered the genetic data for quality, performed a principal component analysis, and calculated the PGS. I also filtered the samples for ethnicity and known variant carrier status. For the environmental data, I downloaded and filtered all available data on tobacco and coffee available in AMP-PD. For a more in-depth analysis of the lifestyle factors, I additionally evaluated questionnaires provided by PPMI, a sub-cohort of AMP-PD, and assessed users of coffee, tobacco, and aspirin, similar to the Fox Insight cohort. Leonie Blöbaum, whom I supervised on this project during her Master's thesis, evaluated the PGS SNPs used for the PGS calculation and performed some initial analyses. With regard to the manuscript, I performed all analyses described in the paper except for the standardization of the PGS, which was performed by Sebastian Koch. This further includes all statistical analyses and the preparation of tables and figures used in the main text and Supplementary Material. Lastly, I wrote the first draft of the manuscript, including the Supplementary Material and the response letters in the subsequent peer-review process, supervised and revised by Dr. Joanne Trinh.

Title:

The combined effect of lifestyle factors and polygenic scores on age at onset in Parkinson's disease

Authors:

Carolin Gabbert[1], MSc, Leonie Blöbaum[1], BSc, Theresa Lüth[1], MSc, Prof. Inke R. König[2], Amke Caliebe[3], PhD, Sebastian Koch[3], MSc, Björn-Hergen Laabs[2], PhD, Christine Klein[1], MD, Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
[2]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany
[3]Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel, Germany
[*]Corresponding author

Published as a preprint in medRxiv, 2023, DOI: doi.org/10.1101/2023.08.25.23294466
Under review in *Scientific Reports*

Abstract:

The objective of this study was to investigate the association between a Parkinson's disease (PD)-specific polygenic score (PGS) and protective lifestyle factors on age at onset (AAO) in PD. We included data from 4367 patients with idiopathic PD, 159 patients with *GBA1*-PD, and 3090 healthy controls of European ancestry from AMP-PD, PPMI, and Fox Insight cohorts. The association between PGS and lifestyle factors on AAO was assessed with linear and Cox proportional hazards models. The PGS showed a negative association with AAO (β=-1.07, p=6×10^{-7}) in patients with idiopathic PD. The use of one, two, or three of the protective lifestyle factors showed a reduction in the hazard ratio by 21% (p=0.0001), 44% (p<2×10^{-16}), and 55% (p<2×10^{-16}), compared to no use. An additive effect of aspirin (β=7.62, p=9×10^{-7}) and PGS (β=-1.58, p=0.0149) was found for AAO without an interaction (p=0.9993) in the linear regressions, and similar effects were seen for tobacco. In contrast, no association between aspirin intake and AAO was found in *GBA1*-PD (p>0.05). In our cohort, coffee, tobacco, aspirin, and PGS are independent predictors of PD AAO. Additionally, lifestyle factors seem to have a greater influence on AAO than common genetic risk variants with aspirin presenting the largest effect.

Introduction:

Parkinson's disease (PD) is a complex neurodegenerative disorder. Besides monogenic forms of PD that explain about 5% of PD cases (Jia et al., 2022), GWAS studies have shown that idiopathic PD is highly polygenic (Chang et al., 2017; Nalls et al., 2019). The largest meta-GWAS of PD to date identified 90 independent risk loci across 78 genomic regions that explained between 16% to 36% of the heritable risk of PD (Nalls et al., 2019). That study additionally determined the proportion of SNP-based heritability explained by their PD GWAS and found their 1805 variant polygenic score (PGS) to explain about 26% of PD heritability (Nalls et al., 2019). The calculation of PGSs provides the opportunity to summarize the effect

of the heritable risk to develop the disease on the individual level. Several studies already evaluated the association of PGSs for PD and affection status, age at onset (AAO), or PD-related symptoms (Escott-Price et al., 2015; Han et al., 2020; Huang et al., 2024; Ibanez et al., 2017; Jacobs et al., 2020; Koch et al., 2021; Li et al., 2019; Paul et al., 2018; Pavelka et al., 2022; Reynoso et al., 2023).

In addition to common genetic risk factors, environmental and lifestyle factors have consistently shown an association with PD susceptibility. While some environmental and lifestyle factors, e.g., pesticides, heavy metals, type 2 diabetes mellitus, and traumatic brain injuries, have been reported to increase the risk for PD (Aune et al., 2023; Elbaz et al., 2004; Fitzmaurice et al., 2014; Gamache et al., 2019; Jafari et al., 2013; Ratner et al., 2014), there are also several environmental and lifestyle factors, e.g., smoking, coffee and black tea, NSAIDs, and physical activity, that have been frequently described as protective regarding the risk for PD, AAO and symptom progression (Casper et al., 2000; Delamarre and Meissner, 2017; Domenighetti et al., 2022; Fan et al., 2020; Grover et al., 2019; Heilbron et al., 2021; Hong et al., 2020; Larsson and Burgess, 2022; Mappin-Kasirer et al., 2020; Marras et al., 2019; Noyce et al., 2012; Wahner et al., 2007). We have previously investigated the effect of environmental and lifestyle factors on the AAO in PD in the Fox Insight cohort and found a protective effect for coffee drinking, tobacco use, and aspirin intake, while no or only a marginal difference in AAO was found for black tea drinking, ibuprofen, and other NSAIDs (Gabbert et al., 2022). Interactions between genetic modifiers and lifestyle factors can further affect PD risk. Gene-environment interactions have been shown between the genetic assembly and a patient's lifestyle. Thus far, there are known interactions between *GRIN2A*, *ADORA2A*, and *CYP1A2* and coffee (Chuang et al., 2016; Hamza et al., 2011) as well as between *RXRA*, *SLC17A6*, and *HLA-DRB1* and smoking (Chuang et al., 2017; Lee et al., 2018). In contrast, studies investigating the effect of environmental and lifestyle factors or gene-environment interactions on PD AAO are only sparse. While some studies found a protective effect of coffee and smoking on PD AAO (De Reuck et al., 2005; Gabbert et al., 2022; Gigante et al., 2018; Gigante et al., 2017; Luth et al., 2020; Rosas et al., 2022; Wijeyekoon et al., 2017; Wilk and Lash, 2007; Yahalom et al., 2020), literature on the effect of aspirin on AAO is lacking (Gabbert et al., 2022). It also remains unclear how gene-environment interactions or a genetic predisposition to PD risk together with the presence of certain lifestyle factors influences the AAO in PD.

Herein, we examine AAO associations of the PGS and the combined effect of the established protective lifestyle factors coffee drinking, tobacco use, and aspirin intake in PD. Our rationale in selecting for these three factors lies in 1) the robustness for previous findings on PD risk; 2) our own findings that coffee drinking, tobacco use, and aspirin intake is associated with later AAO, while no association with AAO was found for black tea in this study group; and 3) the

access and availability of this particular data across several datasets. We investigate whether coffee drinking, tobacco use, and aspirin intake are positively associated with the AAO in PD and if these lifestyle factors further have an additive or interactive effect with respect to the PGS on PD AAO.

Materials and methods:

Study demographics

Three datasets containing genetic, environmental, and lifestyle data from the Accelerating Medicine Partnership Parkinson's Disease Knowledge Platform (AMP-PD), the Parkinson's Progression Markers Initiative (PPMI), and the Fox Insight cohort were included in this study. The complete information on the data harmonization of AMP-PD cohorts comprises of eight sub-cohorts in total (BioFIND, HBS, LBD, LCC, PDBP, PPMI, STEADY-PD3 and SURE-PD3). The Fox Investigation for New Discovery of Biomarkers in Parkinson's Disease (BioFIND) is a cross-sectional, multicenter biomarker study designed to discover and verify biomarkers in clinically typical PD (Kang et al., 2016). The Harvard Biomarker Study (HBS) is a large biobank that recruits patients with early-stage PD or mild cognitive impairment to discover new targets for drugs, new genes, and new diagnostics (Mohammadi, 2013). The LBD Study (International Lewy Body Genomics Consortium) performed whole-genome sequencing in large cohorts of Lewy body dementia cases and neurologically healthy controls to study the genetic architecture of this disease and to generate a resource for the scientific community (Chia et al., 2021). The LRRK2 Cohort Consortium (LCC) was created to assemble and investigate groups of people with and without PD who carry variants in the *LRRK2* gene. The National Institute of Neurological Disorders and Stroke (NINDS) Parkinson's Disease Biomarkers Program (PDBP) aims to accelerate the discovery of promising new diagnostic and progression biomarkers for PD. The Parkinson's Progression Markers Initiative (PPMI) is a longitudinal observational study designed to establish biomarker-defined cohorts and identify clinical, imaging, genetic, and biospecimen PD progression markers to accelerate disease-modifying therapeutic trials (Marek et al., 2018). The NINDS funded STEADY-PD III trial (STEADY-PD3) is a Phase 3, parallel group, placebo-controlled study evaluating the efficacy of isradipine 10 mg daily as a disease-modifying agent in early PD for 36 months (Holloway et al., 2018). The Study of URate Elevation in Parkinson's Disease, phase 3 study (SURE-PD3) is a randomized, double-blind, placebo-controlled trial of urate-elevating inosine treatment to slow clinical decline in early PD. The AMP-PD data received from all sub-cohorts was centrally harmonized, curated, quality controlled, and consolidated into one dataset using both automated and manual approaches, which included the aligning of variables between

datasets, decoding of numeric coded variables, clean-up and standardization of medication names, diagnosis, level of education, and the alignment of visit names between cohorts.

The Fox Insight data facilitates discovery, validation, and reproducibility in Parkinson's disease research (Smolensky et al., 2020). The dataset is generated through routine longitudinal assessments (health and medical questionnaires), one-time questionnaires about environmental exposure and healthcare preferences, and genetic data collection. Patient recruitment details for the Fox Insight study have been previously described (Smolensky et al., 2020). Volunteers for the Fox Insight study were recruited through digital channels (e.g. social network ads, search engine marketing, and email newsletters) and on-the-ground recruitment efforts (e.g. research events, clinician referrals). All Fox Insight participants were 18 years of age or older and provided informed consent. Upon registration, participants were divided into patients with PD and healthy controls, whereas the latter were asked about new diagnoses every three months. PD patients responded to health, non-motor assessments, motor assessments, quality of life, and lifestyle questionnaires through twenty questionnaires that are part of routine longitudinal assessments. Detailed questions about lifestyle, personal habits, living and work environments, medication and healthy history are provided in the Environmental Exposure Questionnaires (PD-RFQ-U). The Fox Insight and AMP-PD cohorts are all established data resources from the Michael J. Fox Foundation. Ethical approval was obtained from the Ethics Committee of the University of Lübeck. Patients and healthy controls from PPMI are included in the AMP-PD cohort for genome sequencing and more detailed lifestyle data was documented as part of the PPMI cohort. Information on known genetic mutations were provided by the cohort platforms in the clinical demographics, which included mutation carrier status for LRRK2 p.G2019S and p.R1441G, GBA p.N409S, and SNCA p.A53T in AMP-PD, as well as LRRK2 p.G2019S and GBA p.R535H, p.N409S, and p.E365K mutation carrier status in Fox Insight. As these are some of the most common genetic causes and common genetic risk variants for PD, these were provided by the cohort platforms. Information on other mutations were not included on the platforms. Known mutation carriers were excluded from the group of patients with idiopathic PD and the healthy controls. No other exclusion criteria were applied. In total, 7616 unrelated participants were included in our study: 4367 patients with PD, without a known genetic cause of PD, 159 patients with variants in *GBA1,* which harbor some of the strongest genetic risk variants in PD, and 3090 healthy controls (Table 1). In this study group, the mean AAO of patients with PD without a known genetic cause of PD was 60.5 years (standard deviation, *SD*=±9.7 years, range: 19.3 to 89.1 years) and the mean age at examination (AAE) was 64.7 years (*SD*=±9.0 years, range: 33.0 to 91.5 years). Of the patients with PD, 2480 (56.8%) were men and 1887 (43.2%) were women. Of the 4367 patients with PD, 1986 were from the AMP-PD cohort, of which 386 were from the PPMI subgroup of AMP-PD, and 2381 were from the Fox Insight cohort.

Table 1: Demographics of the study group.

	Patients with PD[a]	**Patients with *GBA1*-PD**	**Healthy Controls**
N_{total}	4367	159	3090
Male/Female (%)	2480/1887 (56.8%/43.2%)	78/81 (49.1%/50.9%)	1495/1595 (48.4%/51.6%)
Coffee drinkers/ non-drinkers (%)	1914/607 (75.9%/24.1%)	118/27 (81.4%/18.6%)	65/11 (85.5%/14.5%)
Tobacco users/ non-users (%)	1468/2563 (36.4%/63.58%)	54/95 (36.2%/63.8%)	310/368 (45.7%/54.3%)
Aspirin users/ non-users (%)	863/1658 (34.2%/65.8%)	28/64 (30.4%/69.6%)	29/47 (38.2%/61.8%)
Mean AAO (*SD*, range)	60.5 (9.7, 19.3-89.1)	61.9 (9.5, 28.5-83.9)	NA
Mean AAE (*SD*, range)	64.7 (9.0, 33.0-91.5)	65.7 (8.9, 32.0-86.1)	69.9 (13.0, 16.0-90.0)
$N_{AMP\text{-}PD}$	1986	NA	3090
Male/Female (%)	1258/728 (63.3%/36.7%)	NA	1495/1595 (48.4%/51.6%)
Coffee drinkers/ non-drinkers (%)	103/37 (73.6%/26.4%)	NA	65/11 (85.5%/14.5%)
Tobacco users/ non-users (%)	663/987 (40.2%/59.8%)	NA	310/368 (45.7%/54.3%)
Aspirin users/non-users (%)	56/84 (40.0%/60.0%)	NA	29/47 (38.2%/61.8%)
Mean AAO (*SD*, range)	60.8 (9.9, 28.0-89.0)	NA	NA
Mean AAE (*SD*, range)	64.5 (9.5, 33.0-90.0)	NA	69.9 (13.0, 16.0-90.0)
N_{PPMI}[b]	386	NA	200
Male/Female (%)	251/135 (65.0%/35.0%)	NA	133/67 (66.5%/33.5%)
Coffee drinkers/ non-drinkers (%)	103/37 (73.6%/26.4%)	NA	65/11 (85.5%/14.5%)
Tobacco users/ non-users (%)	46/94 (32.9%/67.1%)	NA	31/45 (40.8%/59.2%)
Aspirin users/non-users (%)	56/84 (40.0%/60.0%)	NA	29/47 (38.2%/61.8%)
Mean AAO (*SD*, range)	61.5 (9.4, 35.0-85.0)	NA	NA
Mean AAE (*SD*, range)	62.0 (9.5, 35.0-85.5)	NA	61.6 (10.6, 31.0-83.0)
$N_{Fox\ Insight}$	2381	159	NA
Male/Female (%)	1222/1159 (51.3%/48.7%)	78/81 (49.1%/50.9%)	NA
Coffee drinkers/ non-drinkers (%)	1811/570 (76.1%/23.9%)	118/27 (81.4%/18.6%)	NA
Tobacco users/ non-users (%)	805/1576 (33.8%/66.2%)	54/95 (36.2%/63.8%)	NA
Aspirin users/non-users (%)	807/1574 (33.9%/66.1%)	28/64 (30.4%/69.6%)	NA
Mean AAO (*SD*, range)	60.3 (9.5, 19.3-89.1)	61.9 (9.5, 28.5-83.9)	NA
Mean AAE (*SD*, range)	64.9 (8.7, 33.0-91.5)	65.7 (8.9, 32.0-86.1)	NA

[a]Patients with a known genetic cause of PD were excluded.
[b]Samples from the PPMI sub-cohort are already included in AMP-PD and only lifestyle data was added.
Abbreviations: PD, Parkinson's disease; AAO, age at onset; AAE, age at examination; SD, standard deviation.

The group of patients with *GBA1*-PD, who carried one of the *GBA1* variants p.R535H (NM_000157.4, c.1604G>A), p.N409S (NM_000157.4, c.1226A>G), and p.E365K (NM_000157.4, c.1093G>A) consisted of 159 patients with a mean AAO of 61.9 years (*SD*=±9.5 years, range: 28.5 to 83.9 years) and a mean AAE of 65.7 years (*SD*=±8.9 years, range: 32.0 to 86.1 years). Of these, 78 (49.1%) were men and 81 (50.9%) were women.

The group of healthy controls consisted of 3090 participants with a mean AAE of 69.9 years (*SD*=±13.0 years, range: 16.0 to 90.0 years). While 1495 (48.4%) of the controls were men,

1595 (51.6%) were women. All participants included in this study were of white European ancestry as reported in the participant summaries.

Genetic data and polygenic score estimate

AMP-PD genetic data contained whole-genome sequencing (WGS) data from six unified cohorts (BioFIND, HBS, PDBP, PPMI, SURE-PD3, LBD) (Iwaki et al., 2021). All samples of the AMP-PD dataset were processed by the TOPMed Freeze 9 Variant Calling Pipeline for joint genotyping (Iwaki et al., 2021). The genetic dataset from AMP-PD was stored in a binary PLINK format (Purcell et al., 2007). The dataset was filtered using PLINK 1.9 according to standard quality control filtering steps, excluding SNPs with a minor allele frequency <0.01, a missingness per sample >0.02, a missingness per SNP >0.05, and that failed Hardy-Weinberg equilibrium at a threshold of 1×10^{-50}.

For the PGS calculation, a previously proposed composition of 1805 variants associated with PD risk (Koch et al., 2021; Nalls et al., 2019) was used together with the reference alleles and effect sizes. In our study sample, 1725 of the PGS SNPs were included in the AMP-PD data, with additional 13 SNPs that were represented by proxy SNPs. Proxy SNPs were evaluated with *SNiPA* (Arnold et al., 2015) and had to be in a linkage disequilibrium of >0.98 with the SNP of interest. In total, 1738 SNPs were used for the PGS calculation with the PLINK *score* function. The PGS values were subsequently standardized by subtraction of the mean and division by the standard deviation of the PGS among controls (Koch et al., 2021). This standardized PGS was used for all further analyses. Density plots were created with the base-R function *density* and receiver operating characteristic (ROC) curves and corresponding areas under the curve (AUCs) were calculated with the R package *pROC* (R version 4.3.0) (Robin et al., 2011; Wickham, 2016).

In order to perform a principal component analysis (PCA), the unfiltered genetic dataset from AMP-PD was pruned based on the linkage disequilibrium (LD) using an r^2 threshold of 0.3 with the PLINK 1.9 *indep-pairwise* function. Subsequently, the dataset was filtered for a minor allele frequency >0.3 and a genotyping rate >0.99. The PCA was performed using PLINK 1.9 *pca* function.

Lifestyle and environmental data

Available environmental and lifestyle data was harmonized across the unified cohorts in AMP-PD. For this, available clinical assessment data was curated and transformed by aligning variable names from AMP-PD studies to a global mapping file. The harmonization further included simplifying the information on caffeine consumption and use of tobacco, resulting in the indication whether a subject in AMP-PD had ever used caffeinated beverages or tobacco.

Therefore, more detailed environmental and lifestyle data was also obtained from the PPMI sub-cohort separately. PPMI FOUND (Follow up of persons with Neurologic Disease) uses the Parkinson's Disease Risk Factor Questionnaires (PD-RFQ), which collect life-long information on lifestyle and health, including habits, occupation and residence. The Risk Factor Questionnaire was developed by the National Institute of Environmental Health Sciences PD-RFQ Epidemiology Working Group of the Collaborative Centers for Parkinson's Disease Environmental Research and provides a standard assessment tool for general use in epidemiologic studies of PD (https://www.commondataelements.ninds.nih.gov/report-viewer/23723/Risk%20Factor%20Questionnaire%20(RFQ-U)). It has been validated for self-report and interview. In PPMI, the information from the PD-RFQs is captured by telephone or other remote consultation methods. The PPMI FOUND data included detailed information on the consumption of coffee, tobacco, and aspirin. For the Fox Insight cohort, information on the Environmental Exposure Questionnaires for coffee, tobacco, and aspirin, which are also based on the PD-RFQs, were provided through one-time questionnaires that are part of the online clinical assessment (Smolensky et al., 2020). In this study group, patients were classified as coffee consumers if they regularly drank caffeinated coffee at least once per week over a period of at least 6 months. Patients were classified as tobacco users if they have ever used tobacco, or when available, if they smoked more than 100 cigarettes in their lifetime or if they smoked at least one cigarette per day over a minimum period of 6 months or if they used smokeless tobacco at least once per day for more than 6 months. Lastly, patients were classified as aspirin users if they took at least two pills per week over a minimum of 6 months. No distinctions were made between current and former users of these lifestyle factors.

Duration of caffeine consumption, smoking, and aspirin intake were estimated according to the age the patients started using either substance subtracted from the age at termination. If the patients terminated the consumption after their AAO, the age the patients started was subtracted from their AAO. Periods, where the patients stopped regularly consuming, were subtracted from the overall duration.

Coffee drinking dosage was defined as the average number of cups of coffee per week the patients drank within the drinking duration time. Smoking dosage was estimated as cigarettes smoked per day within the smoking duration time. Aspirin dosage was defined as pills per week the patients took within the aspirin intake duration time. The number of cups of coffee for non-drinkers, cigarettes for non-smokers, and pills per week for aspirin non-users was set to zero. In addition, coffee drinking duration for non-drinkers, smoking duration for non-smokers, and aspirin intake duration for aspirin non-users was set to zero.

Statistical analysis

Statistical analyses were performed using R v4.3.0 (R Core Team, 2023). Multiple linear regression models were used to evaluate the association between AAO, PGS, and lifestyle factors in patients with PD. All linear regression models were validated by evaluating diagnostic plots (Residuals vs Fitted, Q-Q Residuals, Scale-Location, and Residuals vs Leverage) and outliers were removed if applicable. In the linear models, AAO was used as the dependent variable and the PGS and/or the lifestyle factors as the independent variables. Estimates (*β*), standard errors (*SE*), and *p*-values were reported. To adjust for potential confounders, sex and the first two principal components (PCs) were included as covariables in the models. Reported *p*-values were not corrected for multiple testing because they did not follow an "a priori" hypothesis, and results were exploratory. Lifestyle factors were handled in three different ways in the regression models: 1) binary (ever-never/yes-no indication), 2) dosage as a continuous variable, and 3) duration as a continuous variable. In a second set of regression models, data from patients with PD and healthy controls was used to estimate Cox proportional hazards models. Here, we modeled AAO from the cumulative number of lifestyle factors used (R package survival), and we used AAO for patients with PD and AAE with censoring for healthy controls. The total number of the three lifestyle factors coffee, tobacco, and aspirin the participants used were included as the numbers zero to three. The sex and the study site were additionally included as covariables since genetic data and, thus, genetic PCs were not available for all participants. Survival plots and forest plots were generated to visualize the Cox proportional hazards model using the *ggsurvplot* (R package *survminer*) and *forest_model* function (R package *forestmodel*). Regression coefficients, hazard ratios (*HR*), 95% confidence intervals (95% *CI*), and *p*-values were reported. To compare the model accuracies of different linear regression models, the adjusted deviance-based R^2 was calculated using the *adjR2* function (R package *glmtoolbox*). To compare the AAO ranges in patients with PD with respect to their PGS, the patients were stratified into quartiles according to their PGS and the difference in AAO between groups was calculated. In addition, to compare the effect sizes of PGS and lifestyle factors on AAO, the PGS was categorized into "low PGS" and "high PGS" according to the median PGS and participants were stratified into the subgroups that either used no protective lifestyle factor or that used all three lifestyle factors.

Results:

Relationship between PGS and AAO

First, to validate the PD-specific PGS in this study group, the PGS values of patients with PD and healthy controls were assessed. In a case-control comparison, the AUC for the ROC

curves of the standardized PGS was 0.67, which was comparable to the AUC obtained in the original study (Nalls et al., 2019).

To analyze the association between the PGS and AAO in patients with PD, a linear regression model including sex and the first two PCs as covariates was used. The PGS showed a negative association with AAO (β=-1.07, *SE*=0.21, $p=6\times10^{-7}$). Thus, if the PGS is increased by one standard deviation (*SD*), the estimated AAO is approximately one year earlier in patients with PD.

We also assessed the AAO ranges in patients with PD by stratifying and comparing the first and last PGS quartiles. Patients with PD in the first PGS quartile had a median AAO of 63 years (range: 31-85 years), while PD patients in the last PGS quartile had a median AAO of 61 years (range: 34-83 years), showing a difference in the median AAO of two years in these two groups.

Relationship between lifestyle factors and AAO

We replicated our previous findings from the Fox Insight cohort (Gabbert et al., 2022) in the AMP-PD/PPMI cohort. In the linear regression model, coffee drinking duration was positively associated with AAO (β=0.19, *SE*=0.04, $p=3\times10^{-5}$). In addition, tobacco use showed a positive association with AAO (β=3.21, *SE*=0.50, $p=2\times10^{-10}$). We also observed positive associations between aspirin use (β=7.35, *SE*=1.50, $p=3\times10^{-6}$), aspirin dosage (β=0.88, *SE*=0.21, $p=7\times10^{-5}$), and aspirin duration (β=0.55, *SE*=0.17, p=0.0013) and the AAO. To investigate the additive effect of the three lifestyle factors, we coded them by the cumulative number of factors the patients consumed. In this linear regression model, the use of three (β=6.34, *SE*=2.86, p=0.0284) protective lifestyle factors showed an association with AAO, while the use of one lifestyle factor (β=0.21, *SE*=2.37, p=0.9284) or two lifestyle factors (β=4.30, *SE*=2.46, p=0.0827) was not associated with AAO. Interestingly, when including all three factors separately in the same model to predict AAO, aspirin was still associated with AAO (β=7.40, *SE*=1.54, $p=4\times10^{-6}$) and the other associations diminished. Although all three protective factors are associated with AAO, aspirin is shown to be a better predictor of AAO when only one lifestyle factor was included in the model (R^2=0.1694; aspirin (yes/no), sex, PC1, and PC2 in the model) compared to coffee and tobacco use (R^2=0.0207, R^2=0.0287; coffee (yes/no) or tobacco (ever/never), sex, PC1, and PC2 in the model).

In a combined analysis of individuals from both AMP-PD/PPMI and Fox Insight using Cox proportional hazards models on the AAO of patients with PD while including the AAE of healthy controls, we first included the lifestyle factors as separate factors. Coffee (*HR*=0.75, 95% *CI*=0.68-0.82, $p=1\times10^{-9}$), tobacco (*HR*=0.78, 95% *CI*=0.72-0.85, $p=8\times10^{-9}$), and aspirin (*HR*=0.66, 95% *CI*=0.60-0.71, $p<2\times10^{-16}$) showed a reduction in the hazard ratio compared to

no use by 25%, 22%, and 34%, respectively (Supplementary Figure S1). In addition, we assessed the dosage and duration for each lifestyle factors in Cox proportional hazards models as a continuous variable, again showing a reduction in the hazard ratio for coffee (dosage: *HR*=0.99, 95% *CI*=0.99-1.00, *p*=0.0006, duration: *HR*=0.99, 95% *CI*=0.98-0.99, $p<2\times10^{-16}$), tobacco (dosage: *HR*=0.99, 95% *CI*=0.99-1.00, $p=8\times10^{-6}$, duration: *HR*=0.99, 95% *CI*=0.98-0.99, $p=9\times10^{-8}$), and aspirin (dosage: *HR*=0.97, 95% *CI*=0.96-0.98, $p=6\times10^{-11}$, duration: *HR*=0.97, 95% *CI*=0.97-0.98, $p=1\times10^{-14}$). To investigate the potential additive effect between all three lifestyle factors, they were coded by the cumulative number of lifestyle factors the participants consumed as above. In the Cox proportional hazards model, the use of one (*HR*=0.79, 95% *CI*=0.70-0.89, *p*=0.0001), two (*HR*=0.56, 95% *CI*=0.49-0.63, $p<2\times10^{-16}$), or three (*HR*=0.45, 95% *CI*=0.39-0.53, $p<2\times10^{-16}$) of the selected lifestyle factors showed a reduction in the hazard ratio compared to the use of none of these lifestyle factors by 21%, 44%, and 55%, respectively, indicating a later AAO when using more lifestyle factors (Table 2, Figure 1, Supplementary Figure S2).

Table 2: Cox proportional hazards model to investigate the additive effects between the use of the lifestyle factors coffee drinking, tobacco use, and aspirin intake (cumulative number (0-3)) on the AAO of PD, while censoring with the AAE of healthy controls.

	Regression coefficient	**Hazard Ratio (95% *CI*)**	***p*-value**	
***n*=2596, *n* events=2521**				
Coffee/Tobacco/Aspirin (1)	-0.2362	0.7896 (0.6992, 0.8918)	**0.0001**	*
Coffee/Tobacco/Aspirin (2)	-0.5855	0.5568 (0.4918, 0.6305)	**$<2\times10^{-16}$**	*
Coffee/Tobacco/Aspirin (3)	-0.7925	0.4527 (0.3851, 0.5321)	**$<2\times10^{-16}$**	*
Sex (Female)	0.1168	1.1239 (1.0366, 1.2186)	**0.0046**	*
Study (Fox Insight)	0.3097	1.3631 (1.1482, 1.6181)	**0.0004**	*

coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Coffee/Tobacco/Aspirin + Sex + Study, data = data).
* *p*-value < 0.05 are highlighted in bold.
Abbreviations: AAO, age at onset; AAE, age at examination; PD, Parkinson's disease; CI, confidence interval.
Note: baseline categories: Coffee/Tobacco/Aspirin = 0, Sex = Male, Study = AMP-PD/PPMI.

We assessed the AAO ranges in the different groups of lifestyle factor exposures. In the subgroup of patients with PD that used no protective lifestyle factor, the median AAO was 57 years (range: 19-78 years), while patients with PD that drank coffee and used tobacco and aspirin had a median AAO of 66 years (range: 38-86 years), indicating a difference in the median AAO of 9 years in these two groups.

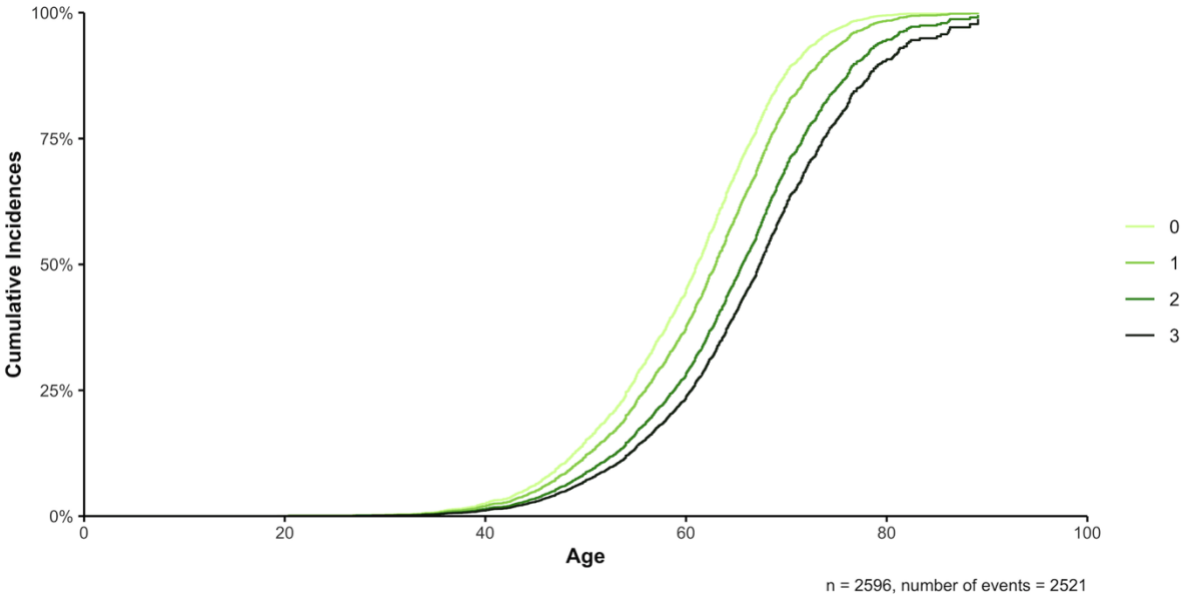


Figure 1: Additive effects of the lifestyle factors coffee drinking, tobacco use, and aspirin intake on the AAO of PD patients, while censoring with the AAE of healthy controls. The different curves describe the cumulative number (0-3) of protective lifestyle factors (coffee drinking, tobacco use, and aspirin intake) the participants used. A Cox proportional hazards model was used to investigate the difference in AAO with respect to the number of protective lifestyle factors used, while censoring with the AAE of healthy controls. The sex and study site were additionally included as covariates (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Coffee/Tobacco/Aspirin + Sex + Study, data = data)).

Additive and interaction effects between PGS and lifestyle factors on AAO

Next, we explored the additive and interactive effects of the PGS and lifestyle factors on AAO in linear regression models. In the additive models including the lifestyle factor coffee drinking, the PGS was not associated with AAO when coffee drinking (binary, dosage, duration) was included as a covariate (Table 3). However, the positive association between the coffee drinking duration and AAO was still robust (β=0.18, *SE*=0.04, $p=5\times10^{-5}$). Further, there were no interactions between the PGS and coffee drinking in the patients with PD.

When we investigated the association between PGS and tobacco use in an additive regression model, both the PGS (β=-1.11, *SE*=0.24, $p=4\times10^{-6}$) and tobacco use (binary ever-never indication) (β=3.16, *SE*=0.50, $p=3\times10^{-10}$) showed associations with AAO (Table 4). However, when dosage of tobacco was included, both the association between tobacco dosage and AAO (β=0.11, *SE*=0.08, *p*=0.2023) as well as between PGS and AAO (β=-1.31, *SE*=0.70, *p*=0.0629) diminished. Similarly, no association between the duration of tobacco use and AAO was found (β=0.17, *SE*=0.10, *p*=0.0910), while the association between the PGS and AAO (β=-1.44, *SE*=0.73 *p*=0.0496) was still robust. In addition, there were no interactions between the PGS and tobacco use with AAO.

Table 3: Linear model on the association of PGS and coffee drinking with AAO in the PD study group.

	Estimate	Standard error	*p*-value	
Coffee drinking (binary) (*n*=139)[1]				
Intercept	56.9608	1.9194	**$<2\times10^{-16}$**	*
PGS	-1.3514	0.6955	0.0541	
Coffee drinking (binary)	1.1929	1.8383	0.5175	
Sex (Male)	3.4321	1.6566	**0.0402**	*
PC1	-8.3114	71.1118	0.9071	
PC2	-28.5586	56.6419	0.6150	
Coffee drinking dosage (*n*=139)[1]				
Intercept	58.1752	1.6649	**$<2\times10^{-16}$**	*
PGS	-1.2006	0.7056	0.0912	
Coffee drinking dosage	-0.0414	0.0816	0.6125	
Sex (Male)	3.7245	1.6396	**0.0247**	*
PC1	-56.0175	67.0364	0.4049	
PC2	-37.4508	57.1142	0.5131	
Coffee drinking duration (*n*=100)[1]				
Intercept	55.3345	1.8344	**$<2\times10^{-16}$**	*
PGS	-1.0068	0.8288	0.2275	
Coffee drinking duration	0.1840	0.0432	**5×10^{-5}**	*
Sex (Male)	-1.1479	1.8432	0.5349	
PC1	101.8483	84.7233	0.2323	
PC2	-79.4870	60.2628	0.1904	
Coffee drinking (binary) (*n*=138)[2]				
Intercept	57.3700	2.0912	**$<2\times10^{-16}$**	*
PGS	-2.3622	1.4141	0.0972	
Coffee drinking (binary)	0.9643	2.2095	0.6632	
Sex (Male)	3.1499	1.6543	0.0591	
PC1	-13.5675	70.8490	0.8484	
PC2	-9.0768	57.0590	0.8739	
PGS:Coffee drinking (binary)	1.0425	1.6328	0.5243	
Coffee drinking dosage (*n*=139)[2]				
Intercept	58.5453	1.7536	**$<2\times10^{-16}$**	*
PGS	-1.7430	1.0617	0.1030	
Coffee drinking dosage	-0.0759	0.0960	0.4305	
Sex (Male)	3.6559	1.6459	**0.0280**	*
PC1	-53.6120	67.2625	0.4269	
PC2	-36.9398	57.2335	0.5198	
PGS:Coffee drinking dosage	0.0595	0.0869	0.4948	
Coffee drinking duration (*n*=100)[2]				
Intercept	55.4035	1.9016	**$<2\times10^{-16}$**	*
PGS	-1.1151	1.1070	0.3164	
Coffee drinking duration	0.1797	0.0522	**0.0009**	*
Sex (Male)	-1.1331	1.8555	0.5429	
PC1	103.6788	86.0536	0.2313	
PC2	-79.7335	60.6014	0.1915	
PGS:Coffee drinking duration	0.0062	0.0419	0.8822	

[1]glm(formula = AAO ~ PGS + Coffee drinking + Sex + PC1 + PC2, family = gaussian, data = data).
[2]glm(formula = AAO ~ PGS * Coffee drinking + Sex + PC1 + PC2, family = gaussian, data = data).
* *p*-value < 0.05 are highlighted in bold.
Abbreviations: PGS, polygenic score; AAO, age at onset; PD, Parkinson's disease; PC, principal component; glm, generalized linear model.

Table 4: Linear model on the association of PGS and tobacco use with AAO in the PD study group.

	Estimate	Standard error	*p*-value	
Tobacco use (binary) (*n*=1650)[1]				
Intercept	60.3997	0.4749	**$<2\times10^{-16}$**	*
PGS	-1.1107	0.2398	**4×10^{-6}**	*
Tobacco use (binary)	3.1643	0.5006	**3×10^{-10}**	*
Sex (Male)	-0.3431	0.5089	0.5002	
PC1	-72.7674	18.6688	**0.0001**	*
PC2	-4.5250	17.0231	0.7904	
Tobacco use dosage (*n*=136)[1]				
Intercept	57.6722	1.5276	**$<2\times10^{-16}$**	*
PGS	-1.3125	0.6996	0.0629	
Tobacco use dosage	0.1061	0.0828	0.2023	
Sex (Male)	3.4475	1.6796	**0.0421**	*
PC1	-46.5124	67.3827	0.4913	
PC2	-35.0421	58.9671	0.5534	
Tobacco use duration (*n*=126)[1]				
Intercept	57.4759	1.5461	**$<2\times10^{-16}$**	*
PGS	-1.4445	0.7282	**0.0496**	*
Tobacco use duration	0.1686	0.0989	0.0910	
Sex (Male)	3.7752	1.7293	**0.0310**	*
PC1	-30.4355	69.4755	0.6621	
PC2	-49.0272	61.5275	0.4271	
Tobacco use (binary) (*n*=1650)[2]				
Intercept	60.3051	0.4916	**$<2\times10^{-16}$**	*
PGS	-0.9605	0.3132	**0.0022**	*
Tobacco use (binary)	3.4028	0.5942	**1×10^{-8}**	*
Sex (Male)	-0.3596	0.5094	0.4804	
PC1	-72.5440	18.6737	**0.0001**	*
PC2	-4.7758	17.0287	0.7792	
PGS:Tobacco use (binary)	-0.3599	0.4829	0.4561	
Tobacco use dosage (*n*=136)[2]				
Intercept	57.7039	1.5443	**$<2\times10^{-16}$**	*
PGS	-1.3800	0.8034	0.0883	
Tobacco use dosage	0.0969	0.0985	0.3272	
Sex (Male)	3.4725	1.6921	**0.0422**	*
PC1	-46.2088	67.6583	0.4959	
PC2	-33.8824	59.5670	0.5705	
PGS:Tobacco use dosage	0.0123	0.0714	0.8630	
Tobacco use duration (*n*=126)[2]				
Intercept	57.4071	1.5621	**$<2\times10^{-16}$**	*
PGS	-1.3076	0.8139	0.1108	
Tobacco use duration	0.1977	0.1251	0.1169	
Sex (Male)	3.7388	1.7381	**0.0335**	*
PC1	-31.6402	69.7955	0.6511	
PC2	-52.7653	62.5193	0.4004	
PGS:Tobacco use duration	-0.0324	0.0848	0.7033	

[1]glm(formula = AAO ~ PGS + Tobacco use + Sex + PC1 + PC2, family = gaussian, data = data).
[2]glm(formula = AAO ~ PGS * Tobacco use + Sex + PC1 + PC2, family = gaussian, data = data).
* *p*-value < 0.05 are highlighted in bold.
Abbreviations: PGS, polygenic score; AAO, age at onset; PD, Parkinson's disease; PC, principal component; glm, generalized linear model.

Table 5: Linear model on the association of PGS and aspirin intake with AAO in the PD study group.

	Estimate	Standard error	*p*-value	
Aspirin intake (binary) (*n* = 140)[1]				
Intercept	55.7965	1.4149	**$<2\times10^{-16}$**	*
PGS	-1.5838	0.6418	**0.0149**	*
Aspirin intake (binary)	7.6159	1.4780	**9×10^{-7}**	*
Sex (Male)	2.4371	1.5181	0.1108	
PC1	-22.9003	61.8372	0.7117	
PC2	-36.6181	52.0627	0.4831	
Aspirin intake dosage (*n* = 134)[1]				
Intercept	56.5418	1.4797	**$<2\times10^{-16}$**	*
PGS	-1.5678	0.6959	**0.0260**	*
Aspirin intake dosage	0.8817	0.2106	**5×10^{-5}**	*
Sex (Male)	2.1982	1.6181	0.1767	
PC1	-34.1742	64.5357	0.5974	
PC2	-18.0128	54.6272	0.7421	
Aspirin intake duration (*n* = 115)[1]				
Intercept	55.2310	1.5322	**$<2\times10^{-16}$**	*
PGS	-1.3545	0.7133	0.0602	
Aspirin intake duration	0.5609	0.1649	**0.0009**	*
Sex (Male)	3.6247	1.7247	**0.0379**	*
PC1	43.9351	72.9233	0.5481	
PC2	26.9794	58.3585	0.6448	
Aspirin intake (binary) (*n* = 140)[2]				
Intercept	55.7961	1.4853	**$<2\times10^{-16}$**	*
PGS	-1.5833	0.8587	0.0674	
Aspirin intake (binary)	7.6168	1.7721	**3×10^{-5}**	*
Sex (Male)	2.4371	1.5249	0.1124	
PC1	-22.8956	62.3001	0.7138	
PC2	-36.6181	52.2582	0.4847	
PGS:Aspirin intake (binary)	-0.0012	1.2985	0.9993	
Aspirin intake dosage (*n* = 134)[2]				
Intercept	56.7069	1.5477	**$<2\times10^{-16}$**	*
PGS	-1.7694	0.8787	**0.0462**	*
Aspirin intake dosage	0.8399	0.2386	**0.0006**	*
Sex (Male)	2.1354	1.6320	0.1931	
PC1	-35.7510	64.8871	0.5826	
PC2	-19.4946	54.9510	0.7234	
PGS: Aspirin intake dosage	0.0738	0.1952	0.7061	
Aspirin intake duration (*n* = 115)[2]				
Intercept	55.3354	1.5472	**$<2\times10^{-16}$**	*
PGS	-1.5324	0.7778	0.0514	
Aspirin intake duration	0.5162	0.1823	**0.0055**	*
Sex (Male)	3.5763	1.7319	**0.0413**	*
PC1	47.1666	73.3544	0.5216	
PC2	25.1460	58.6203	0.6688	
PGS:Aspirin intake duration	0.0809	0.1387	0.5609	

[1]glm(formula = AAO ~ PGS + Aspirin intake + Sex + PC1 + PC2, family = gaussian, data = data).
[2]glm(formula = AAO ~ PGS * Aspirin intake + Sex + PC1 + PC2, family = gaussian, data = data).
* *p*-value < 0.05 are highlighted in bold.
Abbreviations: PGS, polygenic score; AAO, age at onset; PD, Parkinson's disease; PC, principal component; glm, generalized linear model.

We further explored the association between the PGS and aspirin intake on AAO in an additive regression model. When aspirin intake was included as a binary yes-no indication, both the PGS (*β*=-1.58, *SE*=0.64, *p*=0.0149) and aspirin intake (*β*=7.62, *SE*=1.48, $p=9\times10^{-7}$) were associated with AAO in patients with PD (Table 5). Similarly, when the aspirin intake dosage was included in the model, the PGS (*β*=-1.57, *SE*=0.70, *p*=0.0260), as well as aspirin intake dosage (*β*=0.88, *SE*=0.21, $p=5\times10^{-5}$), were associated with AAO. However, when including the aspirin intake duration in the model, aspirin showed an association with AAO (*β*=0.56, *SE*=0.16, *p*=0.0009), while the association between PGS and AAO diminished (*β*=-1.35, *SE*=0.71, *p*=0.0602). There was further no interaction between PGS and aspirin intake in all interaction models.

Impact of PGS and lifestyle factors on AAO

Since the association between PGS and AAO diminished in some models including lifestyle factors as covariables, we investigated the impact PGS and lifestyle factors have on the AAO in PD. In a first approach to compare the effect sizes of PGS and lifestyle factors on AAO, we categorized the PGS into “low PGS” and “high PGS” according to the median PGS and stratified participants into the subgroups that either used no protective lifestyle factor or that used all three lifestyle factors. In the subgroup of participants that used no protective lifestyle factor, a high PGS showed a 3.03 times higher expected hazard of PD as compared to a low PGS (*HR*=3.03, 95% *CI*=1.05-8.78, *p*=0.0409). In contrast, in the subgroup of participants that used all three lifestyle factors, there was no increased hazard ratio (*HR*=1.21, 95% *CI*=0.49-2.99, *p*=0.6863) (Supplementary Figure S3).

We further investigated the model goodness-of-fit of the linear models using the adjusted deviance-based R^2. The model assessing the association between PGS and AAO, while using sex and the first two PCs as covariables, had an adjusted R^2 of 0.0141. In contrast, the linear model evaluating the association between the three lifestyle factors and AAO with the same covariables had an adjusted R^2 of 0.0856, when the lifestyle factors were coded as cumulative quantitative numbers. In the combined linear model, determining the additive association between PGS and the three lifestyle factors with the same covariables as before showed an adjusted R^2 of 0.1039.

Relationship between lifestyle factors and AAO in GBA1-PD

To investigate if the individual and combined effects of the lifestyle factors coffee, tobacco, and aspirin are exclusive to idiopathic PD or if these effects can also be found in patients who carry *GBA1* variants, which are considered some of the strongest genetic risk variants for PD, we examined the relationship between the protective lifestyle factors and AAO in an additional

study group of patients with *GBA1*-PD from Fox Insight. In the linear regression model, coffee drinking duration was positively associated with AAO (*β*=0.24, *SE*=0.05, $p=7\times10^{-7}$) in *GBA1*-PD (Supplementary Table S1). In addition, we observed a positive association between tobacco use and AAO (*β*=3.65, *SE*=1.58, *p*=0.0223) and between aspirin intake duration and AAO (*β*=0.48, *SE*=0.21, *p*=0.0224). When including all three lifestyle factors separately in the same model to predict AAO, only tobacco use was associated with AAO (*β*=6.78, *SE*=2.17, *p*=0.0024), which contrasts with the results found in idiopathic PD. To examine the combined effect of the three lifestyle factors in more detail, we used as influence variable the cumulative number of factors in the Cox proportional hazards model. We observed a protective trend for this variable. With no lifestyle factor as reference, we observed an *HR* of 0.85 for the use of one (95% *CI*=0.40-1.78, *p*=0.6648), *HR* of 0.47 for the use of two (95% *CI*=0.23-0.97, *p*=0.0410), and an *HR* of 0.35 for the use of three lifestyle factors (95% *CI*=0.14-0.86, *p*=0.0216). Of note, the use of one lifestyle factor did not show a significant reduction in the hazard ratio compared to the use of none of these lifestyle factors, which could be a problem of statistical power. However, the use of two or three lifestyle factors showed a significant reduction in the hazard ratio by 53% and 65%, indicating a protective effect on the AAO when using more lifestyle factors.

In an approach to directly compare the relationship of lifestyle factors on AAO between patients with *GBA1*-PD and PD patients without known mutations, we performed the linear regression models including all patients with PD and using the *GBA1* mutation carrier status as another covariate. In the linear regression models, all lifestyle factors showed a positive association with AAO (Supplementary Table S2). Interestingly, an association between *GBA1* mutation carrier status and AAO was further found in the models with coffee drinking dosage (*β*=2.10, *SE*=0.90, *p*=0.0197), tobacco use (binary) (*β*=1.74, *SE*=0.79, *p*=0.0280), and tobacco use duration (*β*=1.84, *SE*=0.88, *p*=0.0357). Therefore, *GBA1* mutation carrier status did not affect the impact of these environmental factors on AAO.

Discussion:

In this study, we have investigated the association between the PGS, calculated based on a previously proposed composition of 1805 variants (Nalls et al., 2019), and the AAO in patients with PD and determined the interaction between the PGS and the lifestyle factors coffee drinking, tobacco use, and aspirin intake on the AAO in PD.

We found that the PGS not only allows discrimination between PD cases and controls(Koch et al., 2021; Nalls et al., 2019) but also showed a negative correlation with AAO (Escott-Price et al., 2015; Ibanez et al., 2017; Koch et al., 2021), indicating that an increase of the PGS by one SD leads to an approximately one-year earlier AAO in patients with PD. This relationship

between PGS and AAO was also robust when adjusting for potentially confounding covariables (i.e., sex and ancestry as represented by the first two principal components). These results demonstrate that the genetic composition, represented by the PGS, adds to understanding the variance in AAO in patients with PD. However, with a range in AAO of 70 years in this study group, more influencing factors and cofounders need to be considered.

The protective effect of environmental and lifestyle factors that decrease the risk of developing PD, influence initial PD-related symptoms and progression, and delay AAO has already been known for years (Checkoway et al., 2002; Chen et al., 2005). However, how these lifestyle factors interact and which combined effect they have on PD AAO remains unresolved. Our group has previously presented a protective effect of coffee, tobacco, and aspirin on the AAO of patients with PD from the Fox Insight study (Gabbert et al., 2022), which we further replicated in the AMP-PD/PPMI study group here. This correlation between lifestyle factors and PD AAO consistently highlights the importance of investigating this interplay further. In a more detailed analysis of the combined effect of the three protective lifestyle factors coffee, tobacco, and aspirin on AAO, we found that the use of either one, two, or three lifestyle factors led to a reduction in the hazard ratio by 21%, 44%, and 55%, respectively, in comparison to no use. As later AAO will tend to be positively associated with lifestyle factor use due to longer observation time, these models were censored with controls to account for this bias. Nevertheless, it is important to take into account that the baseline hazards are already biased as the number of cases and controls are not population representative, which leads to a hazard overestimation for cases and an underestimation for controls. These hazard ratio values are consistent with additive, i.e., independent effects on the logit scale of the lifestyle factors with no synergistic interaction, indicating different underlying mechanisms that lead to the later AAO. Deciphering these mechanisms of action is important to develop suitable therapeutic strategies to delay the AAO of patients with idiopathic PD. In addition, by separating former and current lifestyle factor users possible long-lasting effects could be predicted. Interestingly, aspirin seems to have a larger effect on AAO than coffee or tobacco. The effect of aspirin intake on PD risk is still disputed and the prevalence of NSAIDs use appears to be comparable in the general population and patients with PD (Becker et al., 2011; Noyce et al., 2012; Poly et al., 2019). The prevalence of NSAID intake in the general population varies depending on the definition of NSAID usage, which may include restrictions by categories such as prescription and non-prescription drugs, NSAID doses per pill, and reason for use (e.g., pain (e.g., after a surgery, back pain, headache, menstrual pain), prevention of cardiovascular diseases, or for antipyretic use). In the US general population, the prevalence of aspirin intake, which is the most frequently prescribed NSAID (Al-Azayzih et al., 2020), has been reported to be approximately 50% in adults (Liu et al., 2021; Williams et al., 2015). Therefore, the prevalence of aspirin intake in the general population, including the PD population, is high, although not

as high as for coffee drinking. Given that inflammation is a crucial pathophysiological pathway in PD (Pajares et al., 2020), the anti-inflammatory effect of aspirin might have a protective impact on PD AAO. Although a positive effect of non-steroidal anti-inflammatory drugs on PD risk is contentious (Becker et al., 2011; Noyce et al., 2012; Poly et al., 2019), it is well-known that sustained neuroinflammation leads to the progressive degeneration of dopaminergic neurons (Marogianni et al., 2020). The intake of anti-inflammatory drugs such as aspirin in the prodromal phase, when neuronal degeneration has already started, might, therefore, slow this process, resulting in a later AAO. In addition, studying peripheral immune system alterations and their interactions with aspirin intake and AAO might be beneficial to decipher the underlying mechanisms. The relationship between vascular disorders and PD is controversial, however, it could have played a role in the inter-relationship of aspirin intake, tobacco, and PD. In our cohort, after adjusting for comorbidities such as lung diseases, heart diseases, arthritis, back pain, and surgeries with anesthesia as covariates, our results remain robust (Gabbert et al., 2022). We were limited to the available data on comorbidities in these patients and could not extensively look into the type of vascular disease (i.e. myocardial infarction or stroke). In contrast, the protective effect of aspirin on the AAO of patients with *GBA1*-PD was not as pronounced, indicating that the neuroinflammatory mechanisms leading to neurodegeneration might diverge from patients with idiopathic PD or be masked by the genetic susceptibility in *GBA1*-PD. However, as sample sizes for patients with *GBA1*-PD were small (*n*=159), findings need to be interpreted with caution. Assuming a significance level of 0.05, the power of the coffee assessment is 0.39, of the tobacco assessment, it is 0.6, and of the aspirin assessment, it is 0.34 for the available patients with *GBA1*-PD. To increase the statistical power to 0.8 with a constant effect size, *n*=188 coffee drinkers and *n*=27 non-coffee drinkers, *n*=82 tobacco users and *n*=144 non-tobacco users, and *n*=88 aspirin users and *n*=200 non-aspirin users would be needed. In contrast, for patients with idiopathic PD the power in all three assessments is almost 1. Although the pro-inflammatory signaling does not seem to be related to the PD subtype and there is no evidence of a difference in the immune response between idiopathic PD, monogenic forms of PD (e.g., *LRRK2*-PD), and strong risk factor carriers such as *GBA1*-PD (Thaler et al., 2021), those PD subtypes present with different phenotypes and might need to be treated differently. Thus far, it is not clear how the underlying mechanisms work and if they differ in the different subtypes of PD. However, we have already seen that the effects of environmental and lifestyle factors as well as specific genetic risk factors on AAO vary in different subtypes of PD, especially between monogenic forms of PD and idiopathic PD (Lüth et al., 2023). To follow up on this, future larger-scale studies including patients with monogenic forms of PD or who carry strong risk factors (e.g., *LRRK2*-PD, *GBA1*-PD, or *PRKN/PINK1*-PD), are important to target the effect of anti-inflammatory lifestyle factors on

PD AAO. Especially patients with *PRKN/PINK1*-related PD could be of particular interest as an early AAO is a clinical hallmark of these patients.

In order to investigate the additive and interaction effects of the lifestyle factors together with the PGS on AAO, we applied linear models, showing additive and independent effects of PGS and tobacco use on AAO as well as of PGS and aspirin intake on AAO, with opposite directionality of the PGS and the lifestyle factors. The possibility of an interaction between PGS and lifestyle factors cannot be entirely ruled out. In fact, an interaction between PGS and smoking was reported in two recent studies (Huang et al., 2024; Reynoso et al., 2023). In our study, we found a three times higher expected hazard of PD in the subgroup of participants that used no protective lifestyle factor and had a high PGS as compared to patients with a low PGS. However, the results are thus far only preliminary and need to be interpreted with caution. Nevertheless, they indicate that the PGS is more important for persons that use no protective lifestyle factors. We also found the three lifestyle factors to explain the AAO in patients with PD more accurately than the PGS (Lifestyle factor model: R^2=0.0856, PGS model: R^2=0.0141).

Although there are known gene-environment interactions of coffee and tobacco with PD (Chuang et al., 2017; Chuang et al., 2016; Hamza et al., 2011; Lee et al., 2018), none of the variants included in the calculation of the PGS are located within genes known to show interactions with coffee, smoking, or aspirin. Since this PGS is based on common variants associated with PD risk, it is pathway-independent and different mechanisms can lead to the earlier AAO when having a high PGS or the delayed AAO when using protective lifestyle factors. In contrast, pathway-dependent PGSs, such as the mitochondrial polygenic score, (Billingsley et al., 2019; Lüth et al., 2023) have been shown to interact with lifestyle factors such as pesticides or caffeinated beverages.

Limitations of our study include clinical and genetic data harmonization. The use of data from different cohorts poses the problem of overcoming potential inconsistencies due to differences in the way of data collection. To help overcome this problem, we corrected for the study site in our Cox proportional hazards models including lifestyle data from AMP-PD/PPMI and Fox Insight and also corrected for the first two principal components in all genetic data analyses to account for genetic differences due to ethnic diversity or differences in the type of genetic data collection. Another limitation was that the clinical data provided by the three cohorts sparsely overlapped with genetic data, resulting in small sample sizes in some of the subgroups. Nevertheless, we showed that lifestyle factors have an important effect on PD AAO that is even greater than that of a combined genetic risk. In future studies, this analysis needs to be replicated in a larger study group with diverse ancestral backgrounds. Since the GWAS that was used for the PGS calculation was performed in a European ancestry population, we only included PD patients and controls with European ancestry in our study group. The lack of

ancestry and ethnic diversity in large-scale genetic studies is a well-known problem (Caliebe et al., 2022; Sirugo et al., 2019; Wojcik et al., 2019). To completely unravel the genetic mechanisms that lead to developing PD, future studies must be inclusive of patients from all cultural and genetic backgrounds.

In conclusion, this study is the first to assess the combined effect of the PD-specific PGS together with coffee drinking, tobacco use, and aspirin intake on the AAO of patients with PD and adds to understanding this complex disease. Our results further indicate a potential neuroprotective role of the anti-inflammatory drug aspirin, resulting in a later AAO in PD. Aspirin might play an important protective part in the inflammatory processes that could lead to neurodegeneration in PD. Thus far, these results are only exploratory because they did not follow an “a priori” hypothesis, and the results of the survival analyses only apply to the investigated study group. Therefore, further validation is essential. Nevertheless, our findings underline the importance of investigating both genetic disposition and external influences such as environmental and lifestyle factors to unravel the likelihood of disease manifestation and the variable phenotype presented in patients with PD.

Acknowledgments:

A special thanks to the families and patients who participated in this study.

We would also like to thank Mike Nalls for providing the list of the 1805 common risk variants including the reference alleles and effect sizes that were used in the PGS calculation.

This study was funded by The Michael J. Fox Foundation for Parkinson’s Research (MJFF-019271 and MJFF-021227). This project was further supported by the Institute of Neurogenetics at the University of Lübeck. Grant support was received from the DFG RU ProtectMove (FOR2488).

The AMP PD program is a public-private partnership managed by the Foundation for the National Institutes of Health and funded by the National Institute of Neurological Disorders and Stroke (NINDS) in partnership with the Aligning Science Across Parkinson's (ASAP) initiative; Celgene Corporation, a subsidiary of Bristol-Myers Squibb Company; GlaxoSmithKline plc (GSK); The Michael J. Fox Foundation for Parkinson's Research; Pfizer Inc.; Sanofi US Services Inc.; and Verily Life Sciences. Accelerating Medicines Partnership and AMP are registered service marks of the U.S. Department of Health and Human Services. Clinical data and biosamples used in preparation of this article were obtained from the (i) Michael J. Fox Foundation for Parkinson’s Research (MJFF) and National Institutes of Neurological Disorders and Stroke (NINDS) BioFIND study, (ii) Harvard Biomarkers Study (HBS), (iii) National Institute on Aging (NIA) International Lewy Body Dementia Genetics Consortium Genome Sequencing in Lewy Body Dementia Case-control Cohort (LBD), (iv) MJFF LRRK2 Cohort Consortium

(LCC), (v) NINDS Parkinson's Disease Biomarkers Program (PDBP), (vi) MJFF Parkinson's Progression Markers Initiative (PPMI), and (vii) NINDS Study of Isradipine as a Disease-modifying Agent in Subjects With Early Parkinson Disease, Phase 3 (STEADY-PD3) and (viii) the NINDS Study of Urate Elevation in Parkinson's Disease, Phase 3 (SURE-PD3). BioFIND is sponsored by The Michael J. Fox Foundation for Parkinson's Research (MJFF) with support from the National Institute for Neurological Disorders and Stroke (NINDS). The BioFIND Investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study visit michaeljfox.org/news/biofind. Genome sequence data for the Lewy body dementia case-control cohort were generated at the Intramural Research Program of the U.S. National Institutes of Health. The study was supported in part by the National Institute on Aging (program #: 1ZIAAG000935) and the National Institute of Neurological Disorders and Stroke (program #: 1ZIANS003154). The Harvard Biomarker Study (HBS) is a collaboration of HBS investigators (full list of HBS investigators at https://www.bwhparkinsoncenter.org/biobank/) and funded through philanthropy and NIH and Non-NIH funding sources. The HBS Investigators have not participated in reviewing the data analysis or content of the manuscript. PPMI is sponsored by The Michael J. Fox Foundation for Parkinson's Research and supported by a consortium of scientific partners (full list of all of the PPMI funding partners found at https://www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors). The PPMI investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit www.ppmi-info.org. The Parkinson's Disease Biomarker Program (PDBP) consortium is supported by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health. A full list of PDBP investigators can be found at https://pdbp.ninds.nih.gov/policy. The PDBP investigators have not participated in reviewing the data analysis or content of the manuscript. The Study of Isradipine as a Disease-modifying Agent in Subjects With Early Parkinson Disease, Phase 3 (STEADY-PD3) is funded by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health with support from The Michael J. Fox Foundation and the Parkinson Study Group. For additional study information, visit https://clinicaltrials.gov/ct2/show/study/NCT02168842. The STEADY-PD3 investigators have not participated in reviewing the data analysis or content of the manuscript. The Study of Urate Elevation in Parkinson's Disease, Phase 3 (SURE-PD3) is funded by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health with support from The Michael J. Fox Foundation and the Parkinson Study Group. For additional study information, visit https://clinicaltrials.gov/ct2/show/NCT02642393. The SURE-PD3 investigators have not participated in reviewing the data analysis or content of the manuscript. Funding: PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan,

Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Janssen Neuroscience, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics.
The Fox Insight Study (FI) is funded by The Michael J. Fox Foundation for Parkinson's Research. We would like to thank the Parkinson's community for participating in this study to make this research possible.

Author contributions:

CG performed the data analysis, data interpretation, and writing of the manuscript. LB performed the data analysis, data interpretation, and reviewed the manuscript. TL contributed to the data analysis and interpretation and reviewed the manuscript. IRK helped with the data interpretation and reviewed the manuscript. AC helped with the data interpretation and reviewed the manuscript. SK helped with the data analysis, data interpretation, and reviewed the manuscript. BHL helped with the data interpretation and reviewed the manuscript. CK helped with the data interpretation and reviewed the manuscript. JT was responsible for the conceptualization and design of the project, data interpretation, and writing of the manuscript.

Data availability:

Data used in the preparation of this article were obtained from the Accelerating Medicine Partnership® (AMP®) Parkinson's Disease (AMP PD) Knowledge Platform. For up-to-date information on the study, visit https://www.amp-pd.org. The data that support the findings of this study are available from the Accelerating Medicine Partnership® (AMP®) Parkinson's Disease Knowledge Platform, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request (joanne.trinh@neuro.uni-luebeck.de) and with permission of the Accelerating Medicine Partnership® (AMP®) Parkinson's Disease Knowledge Platform.
Data used in the preparation of this manuscript were obtained from the Fox Insight database (https://foxinsight-info. michaeljfox.org/insight/explore/insight.jsp) on 22/05/2023. For up-to-

date information on the study, visit https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp.

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/access-data-specimens/download-data), RRID:SCR_006431. For up-to-date information on the study, visit www.ppmi-info.org.

Additional information:

Competing interests

All authors declare no financial or non-financial competing interests.

Ethical approval

Approval was obtained from the Ethics Committee of the University of Lübeck. We confirm that all analyses were performed in accordance with relevant guidelines and regulations.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

References:

Al-Azayzih, A., Al-Azzam, S.I., Alzoubi, K.H., Jarab, A.S., Kharaba, Z., Al-Rifai, R.H., and Alnajjar, M.S. (2020). Nonsteroidal Anti-inflammatory Drugs Utilization Patterns and Risk of Adverse Events due to Drug-Drug Interactions among Elderly Patients: A Study from Jordan. Saudi Pharm J *28*, 504-508.

Arnold, M., Raffler, J., Pfeufer, A., Suhre, K., and Kastenmuller, G. (2015). SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics *31*, 1334-1336.

Aune, D., Schlesinger, S., Mahamat-Saleh, Y., Zheng, B., Udeh-Momoh, C.T., and Middleton, L.T. (2023). Diabetes mellitus, prediabetes and the risk of Parkinson's disease: a systematic review and meta-analysis of 15 cohort studies with 29.9 million participants and 86,345 cases. Eur J Epidemiol *38*, 591-604.

Becker, C., Jick, S.S., and Meier, C.R. (2011). NSAID use and risk of Parkinson disease: a population-based case-control study. Eur J Neurol *18*, 1336-1342.

Billingsley, K.J., Barbosa, I.A., Bandres-Ciga, S., Quinn, J.P., Bubb, V.J., Deshpande, C., Botia, J.A., Reynolds, R.H., Zhang, D., Simpson, M.A.*, et al.* (2019). Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. NPJ Parkinsons Dis *5*, 8.

Caliebe, A., Tekola-Ayele, F., Darst, B.F., Wang, X., Song, Y.E., Gui, J., Sebro, R.A., Balding, D.J., Saad, M., Dube, M.P.*, et al.* (2022). Including diverse and admixed populations in genetic epidemiology research. Genet Epidemiol *46*, 347-371.

Casper, D., Yaparpalvi, U., Rempel, N., and Werner, P. (2000). Ibuprofen protects dopaminergic neurons against glutamate toxicity in vitro. Neurosci Lett *289*, 201-204.

Chang, D., Nalls, M.A., Hallgrimsdottir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., International Parkinson's Disease Genomics, C., andMe Research, T., Kerchner, G.A., Ayalon, G.*, et al.* (2017). A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet *49*, 1511-1516.

Checkoway, H., Powers, K., Smith-Weller, T., Franklin, G.M., Longstreth, W.T., Jr., and Swanson, P.D. (2002). Parkinson's disease risks associated with cigarette smoking, alcohol consumption, and caffeine intake. Am J Epidemiol *155*, 732-738.

Chen, H., Jacobs, E., Schwarzschild, M.A., McCullough, M.L., Calle, E.E., Thun, M.J., and Ascherio, A. (2005). Nonsteroidal antiinflammatory drug use and the risk for Parkinson's disease. Ann Neurol *58*, 963-967.

Chia, R., Sabir, M.S., Bandres-Ciga, S., Saez-Atienzar, S., Reynolds, R.H., Gustavsson, E., Walton, R.L., Ahmed, S., Viollet, C., Ding, J.*, et al.* (2021). Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. Nat Genet *53*, 294-303.

Chuang, Y.H., Lee, P.C., Vlaar, T., Mulot, C., Loriot, M.A., Hansen, J., Lill, C.M., Ritz, B., and Elbaz, A. (2017). Pooled analysis of the HLA-DRB1 by smoking interaction in Parkinson disease. Ann Neurol *82*, 655-664.

Chuang, Y.H., Lill, C.M., Lee, P.C., Hansen, J., Lassen, C.F., Bertram, L., Greene, N., Sinsheimer, J.S., and Ritz, B. (2016). Gene-Environment Interaction in Parkinson's Disease: Coffee, ADORA2A, and CYP1A2. Neuroepidemiology *47*, 192-200.

De Reuck, J., De Weweire, M., Van Maele, G., and Santens, P. (2005). Comparison of age of onset and development of motor complications between smokers and non-smokers in Parkinson's disease. J Neurol Sci *231*, 35-39.

Delamarre, A., and Meissner, W.G. (2017). Epidemiology, environmental risk factors and genetics of Parkinson's disease. Presse Med *46*, 175-181.

Domenighetti, C., Sugier, P.E., Sreelatha, A.A.K., Schulte, C., Grover, S., Mohamed, O., Portugal, B., May, P., Bobbili, D.R., Radivojkov-Blagojevic, M.*, et al.* (2022). Mendelian Randomisation Study of Smoking, Alcohol, and Coffee Drinking in Relation to Parkinson's Disease. J Parkinsons Dis *12*, 267-282.

Elbaz, A., Levecque, C., Clavel, J., Vidal, J.S., Richard, F., Amouyel, P., Alperovitch, A., Chartier-Harlin, M.C., and Tzourio, C. (2004). CYP2D6 polymorphism, pesticide exposure, and Parkinson's disease. Ann Neurol *55*, 430-434.

Escott-Price, V., International Parkinson's Disease Genomics, C., Nalls, M.A., Morris, H.R., Lubbe, S., Brice, A., Gasser, T., Heutink, P., Wood, N.W., Hardy, J.*, et al.* (2015). Polygenic risk of Parkinson disease is correlated with disease age at onset. Ann Neurol *77*, 582-591.

Fan, B., Jabeen, R., Bo, B., Guo, C., Han, M., Zhang, H., Cen, J., Ji, X., and Wei, J. (2020). What and How Can Physical Activity Prevention Function on Parkinson's Disease? Oxid Med Cell Longev *2020*, 4293071.

Fitzmaurice, A.G., Rhodes, S.L., Cockburn, M., Ritz, B., and Bronstein, J.M. (2014). Aldehyde dehydrogenase variation enhances effect of pesticides associated with Parkinson disease. Neurology *82*, 419-426.

Gabbert, C., Konig, I.R., Luth, T., Kolms, B., Kasten, M., Vollstedt, E.J., Balck, A., Grunewald, A., Klein, C., and Trinh, J. (2022). Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson's disease. J Neurol.

Gamache, P.L., Haj Salem, I., Roux-Dubois, N., Le Bouthillier, J., Gan-Or, Z., and Dupre, N. (2019). Exposure to Pesticides and Welding Hastens the Age-at-Onset of Parkinson's Disease. Can J Neurol Sci *46*, 711-716.

Gigante, A.F., Asabella, A.N., Iliceto, G., Martino, T., Ferrari, C., Defazio, G., and Rubini, G. (2018). Chronic coffee consumption and striatal DAT-SPECT findings in Parkinson's disease. Neurol Sci *39*, 551-555.

Gigante, A.F., Martino, T., Iliceto, G., and Defazio, G. (2017). Smoking and age-at-onset of both motor and non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord *45*, 94-96.

Grover, S., Lill, C.M., Kasten, M., Klein, C., Del Greco, M.F., and Konig, I.R. (2019). Risky behaviors and Parkinson disease: A mendelian randomization study. Neurology *93*, e1412-e1424.

Hamza, T.H., Chen, H., Hill-Burns, E.M., Rhodes, S.L., Montimurro, J., Kay, D.M., Tenesa, A., Kusel, V.I., Sheehan, P., Eaaswarkhanth, M.*, et al.* (2011). Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. PLoS Genet *7*, e1002237.

Han, Y., Teeple, E., Shankara, S., Sadeghi, M., Zhu, C., Liu, D., FinnGen, Wang, C., Frau, F., Klinger, K.W.*, et al.* (2020). Genome-Wide Polygenic Risk Score Identifies Individuals at Elevated Parkinson's Disease Risk. medRxiv, 2020.2010.2016.20212944.

Heilbron, K., Jensen, M.P., Bandres-Ciga, S., Fontanillas, P., Blauwendraat, C., Nalls, M.A., Singleton, A.B., Smith, G.D., Cannon, P., Noyce, A.J.*, et al.* (2021). Unhealthy Behaviours and Risk of Parkinson's Disease: A Mendelian Randomisation Study. J Parkinsons Dis *11*, 1981-1993.

Holloway, R., Oakes, D., Simuni, T., Hodgeman, K., Greco, B., Henderson, S., Lowell, J., Sharma, S., Tarolli, C., and Biglan, K. (2018). A Phase 3 study of isradipine as a disease modifying agent in patients with early Parkinson disease(STEADY-PD III): Baseline characteristics and study update. Movement Disord *33*, S150-S150.

Hong, C.T., Chan, L., and Bai, C.H. (2020). The Effect of Caffeine on the Risk and Progression of Parkinson's Disease: A Meta-Analysis. Nutrients *12*.

Huang, Y., Chen, Q., Wang, Z., Wang, Y., Lian, A., Zhou, Q., Zhao, G., Xia, K., Tang, B., Li, B.*, et al.* (2024). Risk factors associated with age at onset of Parkinson's disease in the UK Biobank. NPJ Parkinsons Dis *10*, 3.

Ibanez, L., Dube, U., Saef, B., Budde, J., Black, K., Medvedeva, A., Del-Aguila, J.L., Davis, A.A., Perlmutter, J.S., Harari, O.*, et al.* (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.

Iwaki, H., Leonard, H.L., Makarious, M.B., Bookman, M., Landin, B., Vismer, D., Casey, B., Gibbs, J.R., Hernandez, D.G., Blauwendraat, C.*, et al.* (2021). Accelerating Medicines Partnership: Parkinson's Disease. Genetic Resource. Mov Disord *36*, 1795-1804.

Jacobs, B.M., Belete, D., Bestwick, J., Blauwendraat, C., Bandres-Ciga, S., Heilbron, K., Dobson, R., Nalls, M.A., Singleton, A., Hardy, J.*, et al.* (2020). Parkinson's disease determinants, prediction and gene-environment interactions in the UK Biobank. J Neurol Neurosurg Psychiatry *91*, 1046-1054.

Jafari, S., Etminan, M., Aminzadeh, F., and Samii, A. (2013). Head injury and risk of Parkinson disease: a systematic review and meta-analysis. Mov Disord *28*, 1222-1229.

Jia, F., Fellner, A., and Kumar, K.R. (2022). Monogenic Parkinson's Disease: Genotype, Phenotype, Pathophysiology, and Genetic Testing. Genes (Basel) *13*.

Kang, U.J., Goldman, J.G., Alcalay, R.N., Xie, T., Tuite, P., Henchcliffe, C., Hogarth, P., Amara, A.W., Frank, S., Rudolph, A.*, et al.* (2016). The BioFIND study: Characteristics of a clinically typical Parkinson's disease biomarker cohort. Mov Disord *31*, 924-932.

Koch, S., Laabs, B.H., Kasten, M., Vollstedt, E.J., Becktepe, J., Bruggemann, N., Franke, A., Kramer, U.M., Kuhlenbaumer, G., Lieb, W.*, et al.* (2021). Validity and Prognostic Value of a Polygenic Risk Score for Parkinson's Disease. Genes (Basel) *12*.

Larsson, S.C., and Burgess, S. (2022). Appraising the causal role of smoking in multiple diseases: A systematic review and meta-analysis of Mendelian randomization studies. EBioMedicine *82*, 104154.

Lee, P.C., Ahmed, I., Loriot, M.A., Mulot, C., Paul, K.C., Bronstein, J.M., Ritz, B., and Elbaz, A. (2018). Smoking and Parkinson disease: Evidence for gene-by-smoking interactions. Neurology *90*, e583-e592.

Li, W.W., Fan, D.Y., Shen, Y.Y., Zhou, F.Y., Chen, Y., Wang, Y.R., Yang, H., Mei, J., Li, L., Xu, Z.Q., et al. (2019). Association of the Polygenic Risk Score with the Incidence Risk of Parkinson's Disease and Cerebrospinal Fluid alpha-Synuclein in a Chinese Cohort. Neurotox Res 36, 515-522.

Liu, E.Y., Al-Sofiani, M.E., Yeh, H.C., Echouffo-Tcheugui, J.B., Joseph, J.J., and Kalyani, R.R. (2021). Use of Preventive Aspirin Among Older US Adults With and Without Diabetes. JAMA Netw Open *4*, e2112210.

Lüth, T., Gabbert, C., König, I.R., Caliebe, A., Koch, S., Laabs, B.-H., Hentati, F., Sassi, S.B., Amouri, R., Spielmann, M.*, et al.* (2023). Interaction of mitochondrial polygenic score and environmental factors in LRRK2 p.Gly2019Ser parkinsonism. medRxiv, 2023.2001.2002.23284113.

Luth, T., Konig, I.R., Grunewald, A., Kasten, M., Klein, C., Hentati, F., Farrer, M., and Trinh, J. (2020). Age at Onset of LRRK2 p.Gly2019Ser Is Related to Environmental and Lifestyle Factors. Mov Disord *35*, 1854-1858.

Mappin-Kasirer, B., Pan, H., Lewington, S., Kizza, J., Gray, R., Clarke, R., and Peto, R. (2020). Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors. Neurology *94*, e2132-e2138.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C.S., Caspell-Garcia, C., Simuni, T., Jennings, D., Tanner, C.M., Trojanowski, J.Q.*, et al.* (2018). The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. Ann Clin Transl Neurol *5*, 1460-1477.

Marogianni, C., Sokratous, M., Dardiotis, E., Hadjigeorgiou, G.M., Bogdanos, D., and Xiromerisiou, G. (2020). Neurodegeneration and Inflammation-An Interesting Interplay in Parkinson's Disease. Int J Mol Sci *21*.

Marras, C., Canning, C.G., and Goldman, S.M. (2019). Environment, lifestyle, and Parkinson's disease: Implications for prevention in the next decade. Mov Disord *34*, 801-811.

Mohammadi, D. (2013). The Harvard Biomarker Study's big plan. Lancet Neurol *12*, 739-740.

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.*, et al.* (2019). 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.

Noyce, A.J., Bestwick, J.P., Silveira-Moriyama, L., Hawkes, C.H., Giovannoni, G., Lees, A.J., and Schrag, A. (2012). Meta-analysis of early nonmotor features and risk factors for Parkinson disease. Ann Neurol *72*, 893-901.

Pajares, M., A, I.R., Manda, G., Bosca, L., and Cuadrado, A. (2020). Inflammation in Parkinson's Disease: Mechanisms and Therapeutic Implications. Cells *9*.

Paul, K.C., Schulz, J., Bronstein, J.M., Lill, C.M., and Ritz, B.R. (2018). Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. JAMA Neurol *75*, 360-366.

Pavelka, L., Rauschenberger, A., Landoulsi, Z., Pachchek, S., May, P., Glaab, E., Kruger, R., and Consortium, N.-P. (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.

Poly, T.N., Islam, M.M.R., Yang, H.C., and Li, Y.J. (2019). Non-steroidal anti-inflammatory drugs and risk of Parkinson's disease in the elderly population: a meta-analysis. Eur J Clin Pharmacol *75*, 99-108.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J*., et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Ratner, M.H., Farb, D.H., Ozer, J., Feldman, R.G., and Durso, R. (2014). Younger age at onset of sporadic Parkinson's disease among subjects occupationally exposed to metals and pesticides. Interdiscip Toxicol *7*, 123-133.

Reynoso, A., Torricelli, R., Jacobs, B.M., Shi, J., Aslibekyan, S., Norcliffe-Kaufmann, L., Noyce, A.J., and Heilbron, K. (2023). Gene-Environment Interactions for Parkinson's Disease. Ann Neurol.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Muller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 77.

Rosas, I., Moris, G., Coto, E., Blazquez-Estrada, M., Suarez, E., Garcia-Fernandez, C., Martinez, C., Herrera, I.D., Perez-Oliveira, S., Alvarez, V*., et al.* (2022). Smoking is associated with age at disease onset in Parkinson's disease. Parkinsonism Relat Disord *97*, 79-83.

Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. Cell *177*, 26-31.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W*., et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Thaler, A., Omer, N., Giladi, N., Gurevich, T., Bar-Shira, A., Gana-Weisz, M., Goldstein, O., Kestenbaum, M., Shirvan, J.C., Cedarbaum, J.M*., et al.* (2021). Mutations in GBA and LRRK2 Are Not Associated with Increased Inflammatory Markers. J Parkinsons Dis *11*, 1285-1296.

Wahner, A.D., Bronstein, J.M., Bordelon, Y.M., and Ritz, B. (2007). Nonsteroidal anti-inflammatory drugs may protect against Parkinson disease. Neurology *69*, 1836-1842.

Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis. In Use R!, (Cham: Springer International Publishing : Imprint: Springer,), pp. 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color.

Wijeyekoon, R., Suriyakumara, V., Gamage, R., Fernando, T., Jayasuriya, A., Amarasinghe, D., Gunasekara, H., Sirisena, D., Amaratunga, D., Muthukuda, C*., et al.* (2017). Associations between Lifestyle Factors and Parkinson's Disease in an Urban Sri Lankan Clinic Study. Int Arch Med *10*.

Wilk, J.B., and Lash, T.L. (2007). Risk factor studies of age-at-onset in a sample ascertained for Parkinson disease affected sibling pairs: a cautionary tale. Emerg Themes Epidemiol *4*, 1.

Williams, C.D., Chan, A.T., Elman, M.R., Kristensen, A.H., Miser, W.F., Pignone, M.P., Stafford, R.S., and McGregor, J.C. (2015). Aspirin use among adults in the U.S.: results of a national survey. Am J Prev Med *48*, 501-508.

Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L*., et al.* (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514-518.

Yahalom, G., Rigbi, A., Israeli-Korn, S., Krohn, L., Rudakou, U., Ruskey, J.A., Benshimol, L., Tsafnat, T., Gan-Or, Z., Hassin-Baer, S*., et al.* (2020). 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.

2.4.2 *Objective 4 B: The impact of the MGS and tobacco use, caffeine consumption, smoking, and pesticide exposure on AAO in PD*

Impaired mitochondrial function plays a major role in the course of many forms of PD. Several genes have been found to be related to mitochondrial function and to be associated with PD risk. Similar to the PGS, these genes can be estimated in a pathway-dependent mitochondrial polygenic score (MGS), which can be used to assess the risk for PD and PD-related features such as the AAO. Mitochondrial function has previously been shown to be influenced by environmental and lifestyle factors. Especially exposure to pesticides has been shown to cause mitochondrial dysfunction and, therefore, increase the risk for PD. Whether the MGS may also interact with environmental exposure and lifestyle needs to be considered.

In this study, the relationship between the MGS and environmental and lifestyle factors, i.e., smoking, caffeine consumption, and pesticide exposure, with the AAO in patients with PD was investigated. Patients with idiopathic PD (iPD) and patients carrying the LRRK2 p.G2019S variant from the AMP-PD cohort, Fox Insight cohort, and a Tunisian Arab-Berber founder population were included. The MGS was calculated based on variants located in genes described in the secondary mitochondrial function gene list published by Billingsley *et al.* and which were included in all genetic datasets. Multivariate multiple regression analyses were performed to assess the association between MGS and AAO and potential gene-environment interactions. As expected, the MGS was significantly higher in patients with PD compared to the controls. In addition, the MGS was further correlated with a higher AAO in patients with *LRRK2*-PD, especially in those patients of European descent. Interactions between the MGS and environmental and lifestyle factors were further found. Patients with *LRRK2*-PD, who had a higher MGS, had an earlier AAO if they consumed caffeinated soda, while patients with iPD, who had a higher MGS, had an earlier AAO if they were non-smokers.

The study covering this objective is titled “Interaction of Mitochondrial Polygenic Score and Lifestyle Factors in LRRK2 p.Gly2019Ser Parkinsonism”. For the preparation of the Fox Insight genetic dataset, I established a data analysis pipeline as previously described in Objective 2.4.1, as this paper emerged after the PGS analyses. The analysis pipeline comprises the genetic data download, thorough quality control filtering, pruning for linkage disequilibrium and relatedness, and the imputation of variants using the Michigan Imputation Server, which was followed by a second quality control filtering step. For the Tunisian Arab-Berber cohort, we received genetic data from different genotyping arrays (i.e., Affymetrix and Illumina) and in different genome builds (i.e., hg18 and hg19). I harmonized the data, which included a liftover to hg19 of all variants and subsequent quality control filtering. Before I merged the genotyping data from both arrays, I performed an imputation using SHAPEIT and IMPUTE2. Lastly, the merged and imputed dataset was quality control filtered one more time. I took part in all

meetings scheduled to discuss the conceptualization of the manuscript and to evaluate statistical analyses. Theresa Lüth performed the statistical analyses and prepared the tables and figures. The final manuscript was also written by Theresa Lüth and supervised by Dr. Joanne Trinh.

Title:

Interaction of Mitochondrial Polygenic Score and Lifestyle Factors in LRRK2 p.Gly2019Ser Parkinsonism

Authors:

Theresa Lüth[1], MSc, **Carolin Gabbert[1], MSc,** Sebastian Koch[2], MSc, Prof. Inke R. König[3], Make Caliebe[2], PhD, Björn-Hergen Laabs[3], PhD, Faycel Hentati[4], MD, Samia Ben Sassi[4], MD, Rim Amouri[4], MD, Malte Spielmann[5], MD, Christine Klein[1], MD, Anne Grünewald[1,6], PhD, Matthew J. Farrer, PhD[7], Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
[2]Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel, Germany
[3]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany
[4]Neurology Department, National Institute of Neurology, Tunis, Tunisia
[5]Institute of Human Genetics, University of Lübeck, Lübeck, Germany
[6]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[7]Clinical Genomics, University of Florida, Gainsville, FL, USA
[*]Corresponding author

Published in *Movement Disorders*, 2023, DOI: doi.org/10.1002/mds.29563

Abstract:

Background

A mitochondrial polygenic score (MGS) is composed of genes related to mitochondrial function and found to be associated with PD risk.

Objective

To investigate the impact of the MGS and lifestyle/environment on age at onset in LRRK2 p.Gly2019Ser parkinsonism (*LRRK2*-PD) and idiopathic PD (iPD).

Methods

We included *N*=486 patients with *LRRK2*-PD and *N*=9259 with iPD from AMP-PD, Fox Insight, and a Tunisian Arab-Berber founder population. Genotyping data was used to perform the MGS analysis. Additionally, lifestyle/environmental data were obtained from the PD risk factor questionnaire. Linear regression models were used to assess the relationship between MGS, lifestyle/environment, and AAO.

Results

Our derived MGS was significantly higher in PD cases compared to controls (p=1.1×10^{-8}). We observed that higher MGS was significantly associated with earlier AAO in *LRRK2*-PD (p=0.047, β=-1.40) and there was the same trend with a smaller effect size in iPD (p=0.231, β=-0.22). There was a correlation between MGS and AAO in *LRRK2*-PD patients with European descent (p=0.049, r=-0.12), that was visibly less pronounced in Tunisians (p=0.449, r=-0.05). We found that the MGS interacted with caffeinated soda consumption (p=0.003, β=-5.65) in *LRRK2*-PD and with tobacco use (p=0.010, β=1.32) in iPD. Thus, patients with a high MGS had an earlier AAO only if they consumed caffeinated soda or were non-smokers.

Conclusion

The MGS was more strongly associated with earlier AAO in *LRRK2*-PD compared to iPD. Caffeinated soda consumption or tobacco use interacted with MGS to predict AAO. Our study suggests gene-environment interactions as modifiers of AAO in *LRRK2*-PD.

Introduction:

The age at onset (AAO) of Parkinson's disease (PD), as well as the risk of developing the disease, are known to be affected by genetic and environmental factors (Dunn and Kaczorowski, 2019; Gabbert et al., 2022; Kline et al., 2021; Luth et al., 2020; Marras et al., 2019). In terms of genetics, monogenic forms and strong risk factors account for ~10% of PD cases (Klein and Westenberger, 2012). Among these cases, the most common monogenic cause is the Leucine-rich repeat kinase 2 (LRRK2) p.Gly2019Ser mutation. Besides monogenic forms and other genetic variants, PD can be explained by the interplay of complex genetics and lifestyle or environmental factors. One way to assess the cumulative effect of genetic variants on disease risk or AAO is by deriving and using a polygenic score (PGS) (Koch et al., 2021; Nalls et al., 2019). Previously, a mitochondrial polygenic score (MGS) was derived and composed of genes involved in mitophagy, mitochondrial bioenergetics, and

proteostasis pathways (Billingsley et al., 2019). This meant a higher genetic score (a higher cumulative burden) was found to be associated with a higher risk for PD. Biologically, mitochondria are essential key players in PD pathogenesis. In particular, respiratory chain, mitophagy, and mitochondrial biogenesis impairment are associated with PD (Grunewald et al., 2019).

LRRK2 localizes to the cytosol as well as to the mitochondria in the cells. Additionally, fibroblasts derived from patients with *LRRK2*-PD showed reduced NADH dehydrogenase activity and increased mitochondrial mass, mtDNA copy number and nuclear factor erythroid 2-related factor 2 (Nrf2) expression (Delcambre et al., 2020). In macrophages, the LRRK2 p.Gly2019Ser mutation interferes with mitochondrial homeostasis and alters cell death pathways (Weindel et al., 2022). A recent study reported that the seeding of p.Ala53Thr alpha-synuclein oligomerization happens especially at mitochondrial membranes in neurons, which can lead to respiratory chain impairments and a subsequent increase in reactive oxygen species (ROS) (Choi et al., 2022). As alpha-synuclein pathology is an important hallmark in *LRRK2*-PD and iPD, this is another molecular link between PD pathogenesis and mitochondrial impairment.

PD susceptibility has consistently been associated with lifestyle and environmental factors. Several meta-analyses have highlighted the protective association between smoking and PD risk (Noyce et al., 2012). It has been demonstrated that smoking status correlates with later AAO in iPD (Breckenridge et al., 2016; Gabbert et al., 2022; Li et al., 2015). Additionally, caffeine and non-steroidal anti-inflammatory drug (NSAID) consumption was associated with reduced iPD risk and later AAO (Gabbert et al., 2022; Noyce et al., 2012). Analogous to iPD, smoking and caffeine consumption are associated with later onset in *LRRK2*-PD (Luth et al., 2020; Yahalom et al., 2020). In a study including affected and unaffected *LRRK2* mutation carriers, NSAIDs (i.e., aspirin and ibuprofen) users had reduced odds of developing PD (San Luciano et al., 2020). In addition to lifestyle and environment, genetic modifiers of the AAO have been identified as well. There is evidence that variants in the *DNM3* (Trinh et al., 2016) and *CORO1C* (Lai et al., 2021) genes are associated with AAO in *LRRK2*-PD.

The interaction of mitochondrial-related genes, lifestyle, and environment has not been thoroughly investigated. Importantly, mitochondria are at the interface of environmental impacts in the cell. Mitochondrial function can be affected by tobacco use, caffeine consumption or pesticide exposure (Dragicevic et al., 2012; Goncalves et al., 2018; Grunewald et al., 2019; Kanithi et al., 2022). Smoking and vaping have been shown to be associated with mitochondrial gene dysregulation (Tommasi et al., 2021) and there is evidence that caffeine affects mitochondrial bioenergetics (Goncalves et al., 2018) and increases mitochondrial function (Dragicevic et al., 2012). Pesticides like rotenone or paraquat are known to increase mitochondrial dysfunction by inducing redox cycling or binding to complex I, which both result

in the production of reactive oxygen species (ROS) (Nistico et al., 2011). Given the separate relevance of mitochondrial dysfunction and certain lifestyle factors and environmental exposure in PD pathogenesis, we hypothesize that there could be gene-environment interactions modulating their effect on AAO in *LRRK2*-PD and iPD. In our study, we have included patients with iPD and *LRRK2*-PD from AMP-PD, Fox Insight, and a Tunisian Arab-Berber cohort. The AMP-PD and Fox Insight cohorts are publicly available data sets, providing genetic and metadata and are an important resource for PD research. Additionally, as the frequency of LRRK2 p.Gly2019Ser is higher in Ashkenazi Jewish and Tunisian Arab-Berber populations (Trinh et al., 2014b), the Tunisian cohort is specifically relevant to our research question.

Methods:

Study demographics, genetics, and environmental data

Three datasets with genetic, environmental, and lifestyle data were included in this study: AMP-PD, Fox Insight, and a cohort from the Tunisian Arab-Berber population. In total, 9745 patients were included in our study: 486 patients with *LRRK2*-PD (AMP-PD: 127, Fox Insight: 154, Tunisian cohort: 205) and 9259 patients with iPD (AMP-PD: 2077, Fox Insight: 6949, Tunisian cohort: 233). Within the group of patients with iPD, individuals carrying risk variants for PD were excluded (i.e., SNCA p.Ala53Thr, GBA p.Asn370Ser, PRKN p.Arg275Trp) as far as the data was available from the cohorts. For patients with *LRRK2*-PD, the mean AAO was 58.2 years (*SD*=11.1) and the mean age at examination (AAE) was 66.7 years (*SD*=12.4). The mean AAO of patients with iPD was 61.2 years (*SD*=10.2) and the mean AAE was 65.2 years (*SD*=9.6) (Supplementary Table S1).

In order to explore the relatedness of patients from the AMP-PD, Fox Insight cohorts, and Tunisian cohort, we performed an identity-by-descent (IBD) analysis with PLINK and a PI_HAT>0.1875, indicating at least second-degree relatives.

AMP-PD contains whole-genome sequencing (WGS) data from four harmonized cohorts (Iwaki et al., 2021) (Supplementary Text 1). The majority of the patients of the AMP-PD cohort were of European descent (~95%) and the remaining ~5% were of Arab, African American, Hispanic, Asian, Native Hawaiian, or Alaskan descent, with self-reported ethnicity/race. The group of patients with genetic ancestry different from European/white in the AMP-PD cohort is too small and therefore, we have excluded them from our analysis. The Fox Insight dataset is a cohort within the Michael J. Fox Foundation (MJFF) and the genetic data (array-based genotyping) were provided by 23andMe, as previously described (Smolensky et al., 2020). All patients with PD included from the Fox Insight cohort in this study were of self-reported

European ancestry. Lastly, we included a cohort recruited from the Tunisian Arab-Berber population with array-based genotyping data, as previously described (Trinh et al., 2016).
In the Fox Insight and Tunisian cohort, lifestyle and environmental information were assessed with the PD Risk Factor Questionnaire (PD-RFQ) for tobacco use, caffeine consumption, and pesticide exposure (Semple et al., 2004) (Supplementary Text 2, Supplementary Table S2). However, the AMP-PD dataset did not assess environment and lifestyle data with the PD-RFQ. Therefore, available environmental/lifestyle data in the AMP-PD cohort was not used to maintain consistency and we used the more detailed data of the Fox Insight and Tunisian cohorts.

Mitochondrial polygenic score analysis

The genetic datasets from AMP-PD, Fox Insight, and the Tunisian cohort were stored in a binary PLINK format (Purcell et al., 2007). The same quality control filtering steps were applied to all three datasets (minor allele frequency >0.01, missingness per sample <0.02, missingness per SNP <0.05, and Hardy-Weinberg equilibrium $>1\times10^{-50}$) using PLINK v1.9. The Fox Insight dataset was imputed using the Michigan Imputation Server (Das et al., 2016) in combination with the Haplotype Reference Consortium v1.1 reference panel (McCarthy et al., 2016). As the Tunisian dataset is of North African background, we performed the imputation on our in-house computer cluster, using SHAPEIT (O'Connell et al., 2014) and IMPUTE2 (Howie et al., 2009) in combination with the 1000 Genomes Project Phase 3 reference panel (Genomes Project et al., 2015). Genotyping data for AMP-PD was obtained from WGS.
The MGS was calculated using the PLINK score function. Based on the larger secondary mitochondrial function gene list published by Billingsley *et al.* (Billingsley et al., 2019), we calculated the MGS (for a detailed description please see Supplementary Text 3). In order to harmonize the MGS between cohorts, we only used SNPs that were consistently present across all three datasets. Subsequently, we included ~15,000 SNPs for the MGS used in this study (MGS SNPs and corresponding weights can be obtained from: https://github.com/LuethTheresa/MitochondrialPolygenicScoreAndAgeAtOnset). The obtained MGS was standardized to a mean of zero and a standard deviation (SD) of one.

Statistical analysis

Statistical analyses were performed with GraphPad Prism v9.4.0 and R v4.0.3 (R Core Team, 2020; Wickham, 2016). The analysis of the association between MGS and AAO was interpreted for significance in the complete study group, based on the presence of the “a priori” hypothesis on the association between MGS and AAO in *LRRK2*-PD. The association with AAO was tested for significance in *LRRK2*-PD patients and the significance level was set at

p=0.05. All other analyses in this study were exploratory and *p*-values were not corrected for multiple testing.

First, we aimed to investigate the association between AAO and MGS in *LRRK2*-PD and iPD (Supplementary Figure S1). Therefore, we used correlation analyses and multiple regression analyses. In our linear regression models, we used AAO as a dependent variable and the standardized MGS as an independent variable. We included sex and the first five principal components (PC1-5) from the PCA in the regression models to adjust for potential confounders (Supplementary Text 4). Analogous to the MGS, PC1-5 were standardized to a mean of 0 and a standard deviation (*SD*) of 1.

Next, we aimed to investigate interactions between lifestyle, environment (i.e., tobacco use, caffeine consumption, and pesticide exposure), and MGS on AAO. In order to do this, we utilized multiple linear regression models as well. Lifestyle and environmental exposure were set as dichotomous independent variables (yes/no) in our linear regression models. For the investigation of lifestyle and exposure, the Fox Insight and Tunisian cohorts were included, as for AMP-PD the PD-RFQ was not available. All patients from the Fox Insight dataset were of European/white ancestry and all patients from the Tunisian dataset were of Tunisian/Arab ancestry. To visualize potential gene-lifestyle or environment interactions, we performed Kaplan-Meier analyses. To assess the difference in AAO of patients with high or low MGS, a pairwise comparison was performed using the log-rank test. For the stratification, we defined "high MGS" as higher or equal to the median MGS and "low MGS" was defined as lower than the median MGS. Kaplan-Meier analyses were performed for all participants (unstratified for any lifestyle or environmental factor) and stratified by a specific factor (e.g., consumed caffeinated soda yes or no).

Results:

Association between MGS and AAO in PD

We calculated MGS, based on the larger (>1,300 gene names) secondary mitochondrial function gene list published by Billingsley *et al.* (Billingsley et al., 2019), using LDpred2 and our in-house cohorts. We successfully replicated the finding that a higher MGS is associated with a higher risk for PD (*OR*=1.25 per one SD of the MGS, Supplementary Text 3). Additionally, we observed that patients with iPD had a higher MGS compared to healthy controls in the AMP-PD cohort ($p=1.1\times10^{-8}$, Figure 1A).

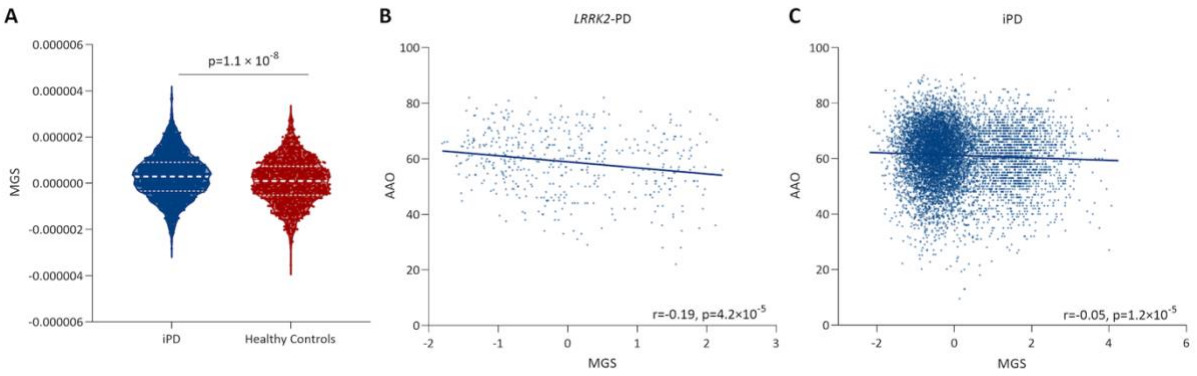


Figure 1: Relationship between Parkinson's affection status, age at onset (AAO) and mitochondrial polygenic score (MGS). (A) The violin plot shows the difference of the MGS between patients with idiopathic PD (iPD) and healthy controls. The dashed lines indicate the median and interquartile range. *p*=Mann-Whitney *U* test exploratory *p*-value. The correlation plots show the association between MGS and AAO in patients with Parkinson's disease carrying the LRRK2 p.Gly2019Ser mutation (*LRRK2*-PD) (B) or patients with iPD (C). *r*=Spearman's rank correlation coefficient, *p*=Spearman's exploratory *p*-value

We analyzed the association between the MGS and the AAO in patients with *LRRK2*-PD. To visualize the relationship between the MGS and AAO we performed a correlation analysis (Figure 1B). The MGS was inversely correlated with the AAO (r=-0.19, p=4.2×10^{-5}, n=477). The higher the MGS, and thereby the higher the cumulative burden of variants associated with mitochondrial dysfunction, the earlier the AAO in *LRRK2*-PD. We then investigated this relationship using multivariable linear regression models and confirmed the significant negative association (β=-1.40, 95% *CI*=-2.77 to -0.02, p=0.047, Table 1). Thus, if the MGS is increased by one SD the AAO is approximately one and a half years earlier in *LRRK2*-PD. As the AAO is earlier in females compared to males (Trinh et al., 2014a) and the AAO and MGS vary between the three cohorts and ethnicities, we included sex and the first five principal components as covariates in the regression models.

Interestingly, when stratifying the data for the two ethnicities/races (i.e., European/white or Tunisian/Arab) to analyze the MGS and AAO relationship, a negative correlation in the same magnitude as before was observed for *LRRK2*-PD patients of European descent (r=-0.12, p=0.049, Supplementary Figure S2A). However, when looking at the patients of Tunisian Arab-Berber descent, the negative correlation is visibly not as pronounced (r=-0.05, p=0.449). Subsequently, we utilized the linear regression model to investigate the association between MGS and AAO in *LRRK2*-PD patients stratified by ethnicities/races as well. The association was in the same direction as in the whole study group in European (β=-0.42, 95% *CI*=-2.64 to 1.81, p=0.715) or Tunisian (β=-1.06, 95% *CI*=-2.67 to 0.55, p=0.197) *LRRK2*-PD patients, but with a smaller effect size and a p-value >0.05. The diminished association could be due to reduced sample sizes in the subgroups.

Table 1: Association between the mitochondrial polygenic score (MGS) and the age at onset (AAO) in patients with *LRRK2*-PD and idiopathic PD.

	***LRRK2*-PD**			**iPD**		
	Estimate	**95% *CI***	***p*-value**	**Estimate**	**95% *CI***	***p*-value**
Complete cohort (*LRRK2*-PD: *N*=473, iPD: *N*=8986)[1]						
MGS	**-1.40**	**-2.77, -0.02**	**0.047***	**-0.22**	**-0.57, 0.14**	**0.231**
Sex Male	1.80	-0.15, 3.75	0.071	0.94	0.51, 1.36	1.5×10^{-5}*
PC1	-0.72	-2.30, 0.85	0.369	-0.44	-0.79, -0.08	0.015*
PC2	-0.35	-1.53, 0.82	0.556	0.21	0.001, 0.42	0.049*
PC3	0.75	-0.46, 1.97	0.225	0.52	0.31, 0.73	1.2×10^{-6}*
PC4	-0.48	-1.47, 0.50	0.400	0.00	-0.21, 0.21	0.975
PC5	0.38	-0.61, 1.37	0.455	0.19	-0.02, 0.40	0.080
Patients with European/white ancestry (*LRRK2*-PD: *N*=269, iPD: *N*=8753)[1]						
MGS	**-0.42**	**-2.64, 1.81**	**0.715**	**-0.38**	**-0.73, -0.02**	**0.038***
Sex Male	0.76	-1.64, 3.15	0.536	0.94	0.51, 1.36	1.4×10^{-5}*
PC1	-0.62	-2.97, 1.73	0.605	-0.02	-0.38, 0.34	0.914
PC2	0.11	-1.42, 1.64	0.890	0.32	0.11, 0.53	0.003*
PC3	-0.12	-1.39, 1.16	0.855	0.00	-0.22, 0.21	0.969
PC4	-1.14	-2.36, 0.08	0.069	-0.04	-0.25, 0.17	0.703
PC5	-0.33	-1.58, 0.92	0.606	0.21	0.002, 0.41	0.052
Patients with Tunisian/Arab ancestry (*LRRK2*-PD: *N*=204, iPD: *N*=233)[1]						
MGS	**-1.06**	**-2.67, 0.55**	**0.197**	**1.18**	**-0.74, 3.10**	**0.231**
Sex Male	3.36	0.07, 6.64	0.046*	-0.89	-4.73, 2.96	0.652
PC1	0.07	-1.57, 1.71	0.934	-0.25	-2.16, 1.65	0.795
PC2	-0.68	-2.33, 0.98	0.424	-0.53	-2.45, 1.39	0.587
PC3	-0.69	-2.33, 0.95	0.411	1.32	-0.62, 3.26	0.183
PC4	0.51	-1.12, 2.14	0.541	1.20	-0.72, 3.11	0.222
PC5	0.81	-0.84, 2.45	0.336	-1.24	-3.17, 0.68	0.206

iPD=idiopathic Parkinson's disease, *LRRK*2-PD= Patients with PD that carry the LRRK2 p.Gly2019Ser variant, MGS=Mitochondrial polygenic score, * *p*-value <0.05
[1]glm(formula = AAO ~ MGS + Sex + PC1 + PC2 + PC3 + PC4 + PC5, family = gaussian)
Baseline categories: Sex=Female

We observed a visibly weaker negative correlation between MGS and AAO in iPD (r=-0.05, $p=1.2\times10^{-5}$, n=9114, Figure 1C) and we could not validate the association using the regression model in the whole study group.

When stratifying the data of the iPD patients for the two ethnicities/races, again the negative correlation between MGS and AAO of patients with European/white ancestry was in the same magnitude as before (r=-0.04, $p=5.2\times10^{-4}$, Supplementary Figure S2B). However, there was a trend for a positive correlation in patients with Tunisian/Arab ancestry (r=0.10, p=0.120). We also used the linear regression model to analyze the association between MGS and AAO in iPD patients stratified by ethnicities/races. The negative association was present in patients with European/white ancestry (β=-0.38, 95% *CI*=-0.73 to -0.02, p=0.038) but not in patients with Tunisian/Arab ancestry (β=1.18, 95% *CI*=-0.74 to 3.10, p=0.231).

Effect of lifestyle factors and environmental exposure on AAO

We focused our analysis on known protective (smoking and caffeine consumption) and risk factors (pesticide exposure) in PD using regression models including sex and study cohort as a covariate.

Tobacco use

We observed no association of smoking with AAO in *LRRK2*-PD (*β*=3.02, 95% *CI*=-1.03 to 7.07, *p*=0.146). In iPD, smoking was associated with later AAO (*β*=1.50, 95% *CI*=0.60 to 2.40, *p*=0.001). Thus, tobacco users had a one and a half years later AAO compared to non-users.

Caffeine consumption

To thoroughly assess the relationship between caffeine and AAO in PD, we analyzed coffee, black tea, green tea, and caffeinated soda consumption. In *LRRK2*-PD, the only caffeinated beverage associated with later AAO was black tea (*β*=5.62, 95% *CI*=1.66 to 9.58, *p*=0.006), meaning that patients that consumed black tea had a five years later AAO compared to *LRRK2*-PD patients that did not consume black tea.

In patients with iPD, coffee consumption was associated with later AAO (*β*=2.20, 95% *CI*=1.10 to 3.29, $p=8.8\times10^{-5}$), meaning that patients that consumed coffee had a two-year later AAO compared to patients that did not. Green tea, however, was not associated with AAO in *LRRK2*-PD or iPD. Caffeinated soda was not associated with AAO in *LRRK2*-PD (*β*=-3.83, 95% *CI*=-7.86 to 0.19, *p*=0.064) but with earlier AAO in iPD (*β*=-2.25, 95% *CI*=-3.33 to -1.17, $p=4.6\times10^{-5}$). Thus, patients with iPD that consumed caffeinated soda had a two years earlier AAO compared to patients that did not.

Pesticide exposure

We did not observe an association between pesticide exposure and AAO in *LRRK2*-PD or iPD.

Interactions between lifestyle, environment, and MGS on AAO

Next, we explored the interaction of lifestyle/environment and MGS on AAO in a linear regression model (Table 2).

Table 2: Interaction between the mitochondrial polygenic score, tobacco use, and caffeine consumption and the age at onset in patients with *LRRK2*-PD and idiopathic PD.

	***LRRK2*-PD**[1]			**iPD**[1]		
	Estimate	**95% *CI***	***p*-value**	**Estimate**	**95% *CI***	***p*-value**
Smoking (*N*=144)				**Smoking (*N*=1759)**		
Lifestyle factor	1.66	-2.85, 6.17	0.472	1.59	0.69, 2.49	0.001*
MGS	0.54	-2.70, 3.79	0.744	-0.78	-1.41, -0.14	0.017*
Sex Male	1.36	-2.87, 5.59	0.530	1.23	0.35, 2.12	0.006*
PC1	-4.59	-9.34, 0.16	0.061	-0.92	-1.50, -0.33	0.002*
PC2	1.76	-0.46, 3.97	0.123	0.01	-0.43, 0.45	0.965
PC3	0.53	-3.16, 4.23	0.778	0.88	0.38, 1.39	0.001*
PC4	0.77	-1.12, 2.66	0.427	0.11	-0.33, 0.55	0.629
PC5	-0.13	-1.87, 1.60	0.879	-0.09	-0.52, 0.35	0.688
MGS:Lifestyle factor	**2.19**	**-1.49, 5.88**	**0.246**	**1.32**	**0.32, 2.32**	**0.010***
Coffee consumption (*N*=136)				**Coffee consumption (*N*=1676)**		
Lifestyle factor	4.77	-0.57, 10.11	0.083	2.20	1.10, 3.29	9.1×10^{-5}*
MGS	4.09	0.36, 7.83	0.034*	-0.13	-1.17, 0.91	0.808
Sex Male	3.88	0.07, 7.70	0.048*	0.70	-0.22, 1.61	0.135
PC1	-3.54	-8.36, 1.29	0.153	-1.02	-1.62, -0.41	0.001*
PC2	1.22	-1.10, 3.54	0.304	-0.11	-0.57, 0.34	0.626
PC3	0.03	-3.81, 3.87	0.988	0.76	0.24, 1.28	0.004*
PC4	1.39	-0.59, 3.36	0.171	0.07	-0.39, 0.52	0.776
PC5	0.18	-1.63, 1.99	0.845	-0.08	-0.52, 0.37	0.736
MGS:Lifestyle factor	**-4.06**	**-8.26, 0.13**	**0.060**	**-0.01**	**-1.19, 1.16**	**0.981**
Caffeinated soda consumption (*N*=133)				**Caffeinated soda consumption (*N*=1406)**		
Lifestyle factor	0.27	-4.48, 5.01	0.913	-2.24	-3.33, -1.16	5.0×10^{-5}*
MGS	3.69	0.54, 6.84	0.023*	-0.31	-1.24, 0.61	0.508
Sex Male	3.50	-0.20, 7.20	0.066	0.48	-0.51, 1.48	0.342
PC1	-3.63	-8.30, 1.04	0.131	-1.07	-1.72, -0.42	0.001*
PC2	0.80	-1.50, 3.10	0.498	-0.07	-0.57, 0.42	0.769
PC3	0.52	-3.23, 4.27	0.786	0.75	0.19, 1.32	0.009*
PC4	0.95	-0.94, 2.83	0.329	0.24	-0.25, 0.74	0.339
PC5	0.27	-1.49, 2.03	0.761	-0.27	-0.75, 0.22	0.283
MGS:Lifestyle factor	**-5.65**	**-9.37, -1.94**	**0.003***	**0.08**	**-1.04, 1.21**	**0.883**

iPD=idiopathic Parkinson's disease, *LRRK2*-PD= Patients with PD that carry the LRRK2 p.Gly2019Ser variant, MGS=Mitochondrial polygenic score, * *p*-value <0.05
[1]glm(formula = AAO ~ MGS * Lifestyle factor + Sex + PC1 + PC2 + PC3 + PC4 + PC5, family = gaussian)
Baseline categories: Sex=Female

MGS and tobacco use

We detected no interaction between MGS and tobacco use in *LRRK2*-PD. However, there was an interaction between MGS and tobacco use in patients with iPD (*β*=1.32, 95% *CI*=0.32 to 2.32, *p*=0.010, Table 2). In other words, our results suggest that the association between the MGS and AAO is dependent upon tobacco use. To further investigate this interaction we analyzed the association between MGS and AAO on tobacco users and non-users separately.

There was no association in tobacco users. On the other hand, we saw a negative association in tobacco non-users (β=-0.79, 95% *CI*=-1.41 to -0.17, *p*=0.013, Supplementary Table S3). Thus, an increase of one SD in the MGS is associated with approximately one year earlier AAO in patients with iPD that did not use tobacco.

To visualize this interaction, we performed Kaplan-Meier analyses, which showed an earlier AAO in iPD patients with a high MGS who had not used tobacco (Figure 2B). The median AAO of patients that did not use tobacco and had a high MGS was 61.7 years, compared to iPD patients with a low MGS that did not use tobacco at 62.3 years (*p*=0.115).

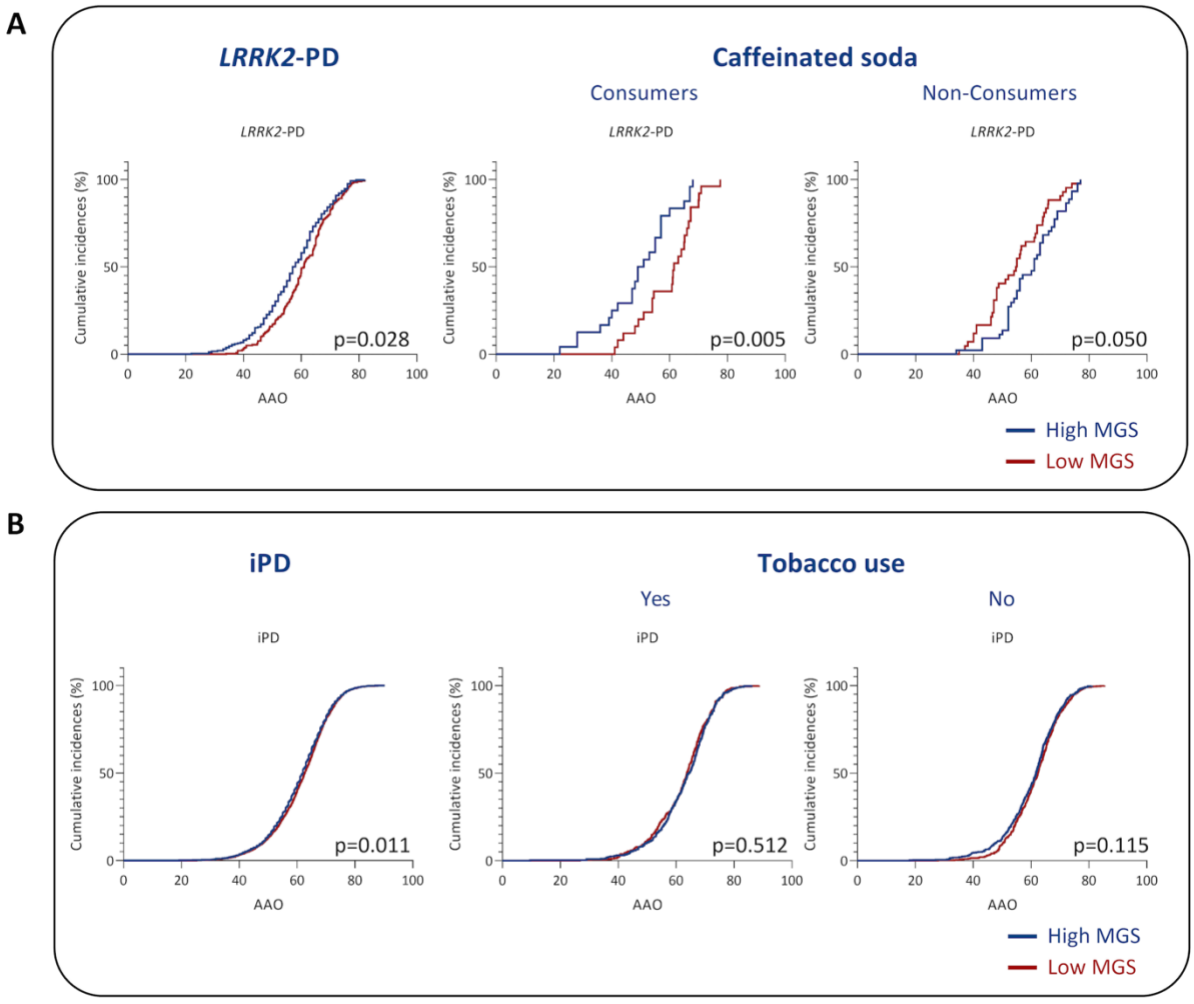


Figure 2: Relationship between age at onset (AAO), mitochondrial polygenic score (MGS) and caffeine consumption or tobacco use. (A) Kaplan-Meier plots show the difference in AAO of patients with *LRRK2*-PD and high MGS or low MGS. The patients were plotted unstratified and stratified by caffeinated soda consumption. (B) Kaplan-Meier plots show the difference in AAO of patients with iPD and high MGS or low MGS. The patients were plotted unstratified and stratified by tobacco use. *p*=Log-rank test exploratory *p*-value

Thus, tobacco non-users with a high MGS had a ~six months earlier median AAO than iPD patients with a low MGS. Interestingly, the AAO was later in iPD patients with a high MGS that did use tobacco (high MGS: median AAO=64.4, low MGS: median AAO=63.7). In comparison, the median AAO was ~seven months earlier in iPD patients with a high MGS, unstratified for any lifestyle factors.

MGS and caffeine consumption

Next, we investigated the interaction between caffeine consumption, including coffee, black tea, green tea, and caffeinated soda.

There was a trend for an interaction between MGS and coffee consumption with the AAO of *LRRK2*-PD patients (*β*=-4.06, 95% *CI*=-8.26 to 0.13, *p*=0.060). In addition, we detected a more pronounced interaction between MGS and caffeinated soda consumption (*β*=-5.65, 95% *CI*=-9.37 to -1.94, *p*=0.003) in *LRRK2*-PD. To further investigate this interaction we analyzed the association between MGS and AAO in caffeinated soda drinkers and non-drinkers separately. There was a trend for a negative association in patients with *LRRK2*-PD that consumed caffeinated soda (*β*=-2.30, 95% *CI*=-7.99 to 3.39, *p*=0.432, Supplementary Table S4). On the other hand, we saw a positive association between the MGS and AAO in patients that did not consume caffeinated soda (*β*=3.86, 95% *CI*=0.36 to 7.35, *p*=0.034). To also visualize this potential interaction, we performed Kaplan-Meier analyses, which showed an earlier AAO in patients with a high MGS who consumed caffeinated soda (Figure 2A) in *LRRK2*-PD. The median AAO of patients that consumed caffeinated soda and had a high MGS was 50.0 years, compared to *LRRK2*-PD patients with a low MGS that consumed caffeinated soda at 61.5 years (*p*=0.005). Thus, caffeinated soda consumers with a high MGS had a ~eleven years earlier median AAO than *LRRK2*-PD patients with a low MGS. In comparison, the median AAO was only ~four years earlier in *LRRK2*-PD patients, unstratified for any lifestyle factor. Analogously to the MGS and tobacco use interaction in iPD, the AAO was later in *LRRK2*-PD patients with a high MGS that did not consume caffeinated soda (high MGS: median AAO=61.0, low MGS: median AAO=54.85).

There was no interaction between the MGS, coffee or caffeinated soda consumption in iPD. Furthermore, there was also no interaction between MGS, black tea, and green tea in neither *LRRK2*-PD or iPD (Supplementary Table S5).

MGS and pesticide exposure

Lastly, we investigated the interaction between MGS and pesticide exposure in a work and a non-work setting. However, we did not detect an interaction between MGS and pesticide exposure in *LRRK2*-PD or iPD (Supplementary Table S5).

Discussion:

Gene-environment interactions are relevant as onset modifiers of *LRRK2*-PD and iPD. The main strength of this study is the size of the study cohort consisting of three large cohorts. In addition, we utilized the thorough overlap of genetic, lifestyle, and environmental data of two cohorts to comprehensively investigate the relationship between MGS and AAO in PD. We see a robust relationship between the MGS and AAO in *LRRK2*-PD even after adjusting for potentially confounding covariates (i.e., sex, cohort, or ethnicity represented by principal components 1-5). To our knowledge, we demonstrate a novel association between MGS and earlier AAO in *LRRK2*-PD. Furthermore, the diverse ethnic background of the patients in this study shows population-specific effects of the MGS. Though we see an overall association between the MGS and AAO, when separating the cohorts, the association was found to be more pronounced in the European cohorts and visibly weaker in the Tunisian/Arab cohort in the correlation analysis. It is well known that population- or ethnic-specific background is a key factor in polygenic scores and it is important for future studies to be inclusive of patients from diverse backgrounds (Caliebe et al., 2022; Duncan et al., 2019; Martin et al., 2019). To illustrate the importance of the study population in genetic scores, we performed a principal component analysis (PCA) using common SNPs. Additionally, we included the publicly available 1000 Genomes Project dataset as a validation for the clustering of the populations. In the PCA, the AMP-PD cohort clustered together with the Fox Insight cohort and the European samples of the 1000 Genomes Project, as both consist of patients of mainly European/white ancestry (Supplementary Figure S3, Supplementary Text 4). In the study that constructed the MGS that we used, the dataset consisted of participants of European ancestry (Billingsley et al., 2019). However, the frequency of LRRK2 p.Gly2019Ser is higher in Ashkenazi Jewish and Tunisian Arab-Berber populations (Trinh et al., 2014b). This highlights the importance of deriving an MGS from these two founder populations, as it would be pertinent to further understanding the MGS effect. Combined international efforts will be required to generate, evaluate, and estimate an MGS in diverse populations. The lack of diverse cohorts in large-scale genetic studies is a well-known problem (Sirugo et al., 2019; Wojcik et al., 2019), but more diversity is essential to overcome such limitations of polygenic scores.

Limitations of our study include potential bias that comes from different data reported in the three cohorts. In terms of genetics, genotyping data were either obtained from arrays (Fox Insight and Tunisian cohort) or WGS (AMP-PD) that could contribute to batch effects. The AAO was earlier in the Tunisian cohort compared to the AMP-PD and Fox Insight cohorts. We observed that the percentage of early-onset PD patients (EOPD, AAO<50 years) is higher than expected (Kolicheski et al., 2022) within the Tunisian cohort, with ~28% in the *LRRK2*-PD group and ~32% in the iPD group. This difference could be a result of the different genetic and

cultural backgrounds of the patients. However, we included the first five principal components as covariates in our analysis to adjust and counteract potential biases. Additionally, we compared the association between MGS and AAO in EOPD and LOPD. The association between MGS and AAO observed for patients with *LRRK2*-PD, was in the same direction in patients with EOPD (β=-1.13, 95% *CI*=-2.61 to 0.35, p=0.136, Supplementary Table S6) and LOPD (β=-0.86, 95% *CI*=-1.99 to 0.27, p=0.138). Nevertheless, potential biases and confounding factors may impact the results of studies on age at onset, such as cultural differences, medical health availability, and family history of PD.

Relatedness remains problematic for the Tunisian cohort. After investigating IBD, the Fox Insight cohort had no related patients but there were three *LRRK2*-PD patients who were related within the AMP-PD cohort and eight *LRRK2*-PD patients in the Tunisian cohort. One in three marriages are consanguineous within the Tunisian population (Trinh et al., 2016) and patients included in this study were partly recruited from families. Still, the minority of patients in our study (n=11) are closely related and account for a minor fraction of the sample size.

The main environmental/lifestyle questionnaire used in our study is the validated PD-RFQ. However, the PD-RFQ was only available from the Fox Insight and Tunisian cohort. To harmonize the data as much as possible, AMP-PD was not included in our environment/lifestyle analyses. The PD-RFQ, though validated, also has its own caveats. For example, pesticide exposure in a non-work setting includes any exposure to chemicals utilized to kill insects, other pests, plants, weeds, mold, or mildew used in the house, garden, or on pets, which leads to an inflation of individual exposure. Diverse cultural preferences also exist that may not be captured by the lifestyle questionnaires: one example is the main source of caffeine intake (i.e., coffee, tea, or soda), which varies significantly in different countries (Reyes and Cornelis, 2018). To overcome this caveat, we stratified our data for ethnicity/race and study cohort and performed interaction analyses only on iPD patients of the Fox Insight cohort that were all of European/white ancestry. Still, there was a trend for an interaction between MGS and tobacco use in predicting AAO (β=0.86, 95% *CI*=-0.02 to 1.75, p=0.056, Supplementary Table S7).

A previous positive association between an estimated MGS and AAO was detected in addition to an association with PD risk (Billingsley et al., 2019; Dehestani et al., 2022). However, herein we report a negative association with onset. It is important to note that we observe a negative association with AAO in patients with *LRRK2*-PD (r=-0.19, p=4.2×10^{-5}) and there was the same trend in iPD (r=-0.05, p=1.2×10^{-5}) with a small effect size. There may be three underlying reasons: 1) this study is based on a newly built MGS with the same mitochondrial-related gene list but alternative resulting variants and weights, 2) different cohorts for the idiopathic PD analyses were utilized (i.e., Harvard Biomarker Study, McGill Parkinson's, Oslo Parkinson's Disease Study, Parkinson's Disease Biomarker's Program, Parkinson's Progression Markers

Initiative, Spanish Parkinson's (IPDGC) part2, and German GWAS), 3) lastly, the effect size was larger in *LRRK2*-PD, a more homogeneous cohort compared to iPD. The causes of the disease in iPD patients can be much more diverse and this heterogeneity may overshadow the subtle effect of the MGS, which may only be valid for certain subtypes of iPD. Another potential explanation is that mitochondrial biological implications are strongly related to disease onset in *LRRK2*-PD but not in iPD. Mitochondrial abnormalities are involved in the pathogenesis of *LRRK2*-PD, such as reduced NADH dehydrogenase activity, increased mitochondrial mass, mtDNA copy number, and nuclear factor erythroid 2-related factor 2 (Nrf2) expression (Delcambre et al., 2020). Thus, an additional mitochondrial burden, reflected in a higher MGS, could lead to an earlier AAO in patients with *LRRK2*-PD.

Mitochondrial function can be affected by tobacco use, caffeine consumption, or pesticide exposure (Dragicevic et al., 2012; Goncalves et al., 2018; Grunewald et al., 2019; Kanithi et al., 2022). We, among others, have reported that caffeinated soda intake was associated with earlier AAO (Luth et al., 2020) or increased PD risk (Tanner et al., 2017). Hence, caffeinated soda appears to be different from other caffeinated beverages and potentially caffeine-independent mechanisms are driving these effects. For patients with *LRRK2*-PD, there was an interaction between MGS and caffeinated soda consumption. The median AAO was ~11 years earlier in patients with a high MGS that consumed caffeinated soda. As the median AAO of *LRRK2*-PD patients unstratified for any lifestyle factor was only ~4 years earlier, our data support a gene-lifestyle interaction between caffeine intake and MGS. Caffeine consumption is reported as a protective factor in PD, except for caffeinated soda, as described above. However, in rats, there is evidence that treatment with caffeine induces mitochondrial dysfunction in the neonatal brain (Kasala et al., 2020).

In addition to caffeine, tobacco use interacted with the MGS in patients with iPD exclusively. In contrast to caffeinated soda, tobacco is a protective factor in PD (Gabbert et al., 2022; Luth et al., 2020; Noyce et al., 2012). This could explain why we only observe an earlier AAO in iPD patients with a high MGS that did not use tobacco, as the absence of this protective factor may enhance the vulnerability for a higher MGS.

Our results underline the importance of including lifestyle and environment when investigating genetic associations with AAO or disease risk. Gene-lifestyle or gene-environment interactions could significantly influence the association with these traits. A recent study demonstrated that GWAS analyses could be affected by gene-environment correlations across geographic regions. The genetic correlations with socioeconomic status-related traits were significantly reduced when controlling for geographic regions (Abdellaoui et al., 2022). Likewise, our study shows the differences between Tunisian Arab-Berbers and European/white ancestry though a more refined investigation is warranted.

In conclusion, there was an association between the MGS and earlier AAO in patients with *LRRK2*-PD and iPD, but with a visibly smaller effect size in the latter one. Furthermore, we detected gene-lifestyle interactions in *LRRK2*-PD and iPD. Thus, lifestyle and environmental factors may interact with the MGS and affect its impact on the AAO in PD (Figure 3). Our results highlight the importance of functional studies investigating the underlying molecular mechanisms leading to the interaction between MGS, caffeine consumption, and tobacco use.

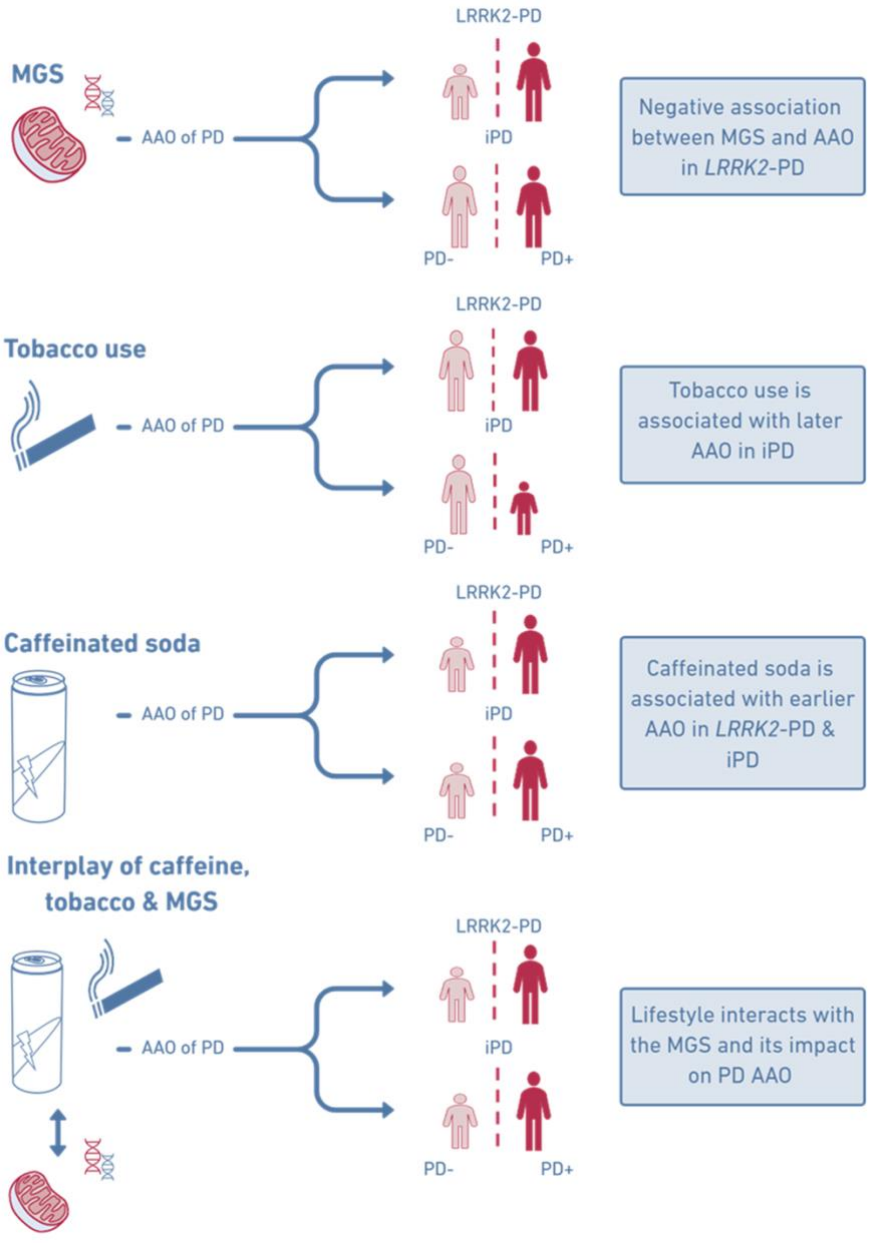


Figure 3: Summary of results. Caffeine consumption can be associated with later or earlier AAO in PD, depending on the beverage. Furthermore, there is evidence for gene and lifestyle interactions, as in caffeine consumers and patients that did not use tobacco, the effect of the MGS on AAO in PD is more pronounced.

Acknowledgments:

A special thanks to the families and patients who participated in this study.

This project was supported by the DFG RU ProtectMove (DFG FOR2488), the Michael J. Fox Foundation (MJFF-021227 & MJFF-019271), and the Else Kröner-Fresenius-Stiftung.

The Fox Insight Study (FI) is funded by The Michael J. Fox Foundation for Parkinson's Research. We would like to thank the Parkinson's community and 23andMe research participants and employees for making this research possible. Data used in the preparation of this manuscript were obtained from the Fox Insight database (https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp) on 18/10/2020. For up-to-date information on the study, visit https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp.

Data used in the preparation of this article were obtained from the Accelerating Medicine Partnership® (AMP®) Parkinson's Disease (AMP PD) Knowledge Platform. For up-to-date information on the study, visit https://www.amp-pd.org.

The AMP® PD program is a public-private partnership managed by the Foundation for the National Institutes of Health and funded by the National Institute of Neurological Disorders and Stroke (NINDS) in partnership with the Aligning Science Across Parkinson's (ASAP) initiative; Celgene Corporation, a subsidiary of Bristol-Myers Squibb Company; GlaxoSmithKline plc (GSK); The Michael J. Fox Foundation for Parkinson's Research ; Pfizer Inc.; Sanofi US Services Inc.; and Verily Life Sciences. ACCELERATING MEDICINES PARTNERSHIP and AMP are registered service marks of the U.S. Department of Health and Human Services.

Clinical data and biosamples used in preparation of this article were obtained from the (i) Michael J. Fox Foundation for Parkinson's Research (MJFF) and National Institutes of Neurological Disorders and Stroke (NINDS) BioFIND study, (ii) Harvard Biomarkers Study (HBS), (iii) National Institute on Aging (NIA) International Lewy Body Dementia Genetics Consortium Genome Sequencing in Lewy Body Dementia Case-control Cohort (LBD), (iv) MJFF LRRK2 Cohort Consortium (LCC), (v) NINDS Parkinson's Disease Biomarkers Program (PDBP), (vi) MJFF Parkinson's Progression Markers Initiative (PPMI), and (vii) NINDS Study of Isradipine as a Disease-modifying Agent in Subjects With Early Parkinson Disease, Phase 3 (STEADY-PD3) and (viii) the NINDS Study of Urate Elevation in Parkinson's Disease, Phase 3 (SURE-PD3).

BioFIND is sponsored by The Michael J. Fox Foundation for Parkinson's Research (MJFF) with support from the National Institute for Neurological Disorders and Stroke (NINDS). The BioFIND Investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit michaeljfox.org/news/biofind.

Genome sequence data for the Lewy body dementia case-control cohort were generated at the Intramural Research Program of the U.S. National Institutes of Health. The study was

supported in part by the National Institute on Aging (program #: 1ZIAAG000935) and the National Institute of Neurological Disorders and Stroke (program #: 1ZIANS003154).
The Harvard Biomarker Study (HBS) is a collaboration of HBS investigators [full list of HBS investigators found at https://www.bwhparkinsoncenter.org/biobank/ and funded through philanthropy and NIH and Non-NIH funding sources. The HBS Investigators have not participated in reviewing the data analysis or content of the manuscript.
PPMI is sponsored by The Michael J. Fox Foundation for Parkinson's Research and supported by a consortium of scientific partners: [list the full names of all of the PPMI funding partners found at https://www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors]. The PPMI investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit www.ppmi-info.org.
The Parkinson's Disease Biomarker Program (PDBP) consortium is supported by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health. A full list of PDBP investigators can be found at https://pdbp.ninds.nih.gov/policy. The PDBP investigators have not participated in reviewing the data analysis or content of the manuscript.
The Study of Isradipine as a Disease-modifying Agent in Subjects With Early Parkinson Disease, Phase 3 (STEADY-PD3) is funded by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health with support from The Michael J. Fox Foundation and the Parkinson Study Group. For additional study information, visit https://clinicaltrials.gov/ct2/show/study/NCT02168842. The STEADY-PD3 investigators have not participated in reviewing the data analysis or content of the manuscript.
The Study of Urate Elevation in Parkinson's Disease, Phase 3 (SURE-PD3) is funded by the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health with support from The Michael J. Fox Foundation and the Parkinson Study Group. For additional study information, visit https://clinicaltrials.gov/ct2/show/NCT02642393. The SURE-PD3 investigators have not participated in reviewing the data analysis or content of the manuscript.

Conflicts of Interest:

CK serves as a medical advisor to Centogene and Retromer Therapeutics and received speaking honoraria from Desitin. The remaining authors declare no conflict of interest.

Funding:

This project was supported by the DFG RU ProtectMove (DFG FOR2488), the Michael J. Fox Foundation (MJFF-021227 & MJFF-019271), and the Else Kröner-Fresenius-Stiftung.

Author contributions:

Conception and design of the study: J.T., A.G., I.K. and C.K; Acquisition and analysis of data: T.L., C.G., I.K., A.M., S.K., B. H. L., F.H., S.B. S., R. A., M.S., M. J. F. and J. T.; Drafting a significant portion of the manuscript or figures: T.L. and J.T.

References:

Abdellaoui, A., Dolan, C.V., Verweij, K.J.H., and Nivard, M.G. (2022). Gene-environment correlations across geographic regions affect genome-wide association studies. Nat Genet *54*, 1345-1354.

Billingsley, K.J., Barbosa, I.A., Bandres-Ciga, S., Quinn, J.P., Bubb, V.J., Deshpande, C., Botia, J.A., Reynolds, R.H., Zhang, D., Simpson, M.A*., et al.* (2019). Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. NPJ Parkinsons Dis *5*, 8.

Breckenridge, C.B., Berry, C., Chang, E.T., Sielken, R.L., Jr., and Mandel, J.S. (2016). Association between Parkinson's Disease and Cigarette Smoking, Rural Living, Well-Water Consumption, Farming and Pesticide Use: Systematic Review and Meta-Analysis. PLoS One *11*, e0151841.

Caliebe, A., Tekola-Ayele, F., Darst, B.F., Wang, X., Song, Y.E., Gui, J., Sebro, R.A., Balding, D.J., Saad, M., Dube, M.P*., et al.* (2022). Including diverse and admixed populations in genetic epidemiology research. Genet Epidemiol *46*, 347-371.

Choi, M.L., Chappard, A., Singh, B.P., Maclachlan, C., Rodrigues, M., Fedotova, E.I., Berezhnov, A.V., De, S., Peddie, C.J., Athauda, D*., et al.* (2022). Pathological structural conversion of alpha-synuclein at the mitochondria induces neuronal toxicity. Nat Neurosci *25*, 1134-1148.

Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M*., et al.* (2016). Next-generation genotype imputation service and methods. Nat Genet *48*, 1284-1287.

Dehestani, M., Liu, H., Sreelatha, A.A.K., Schulte, C., Bansal, V., and Gasser, T. (2022). Mitochondrial and autophagy-lysosomal pathway polygenic risk scores predict Parkinson's disease. Mol Cell Neurosci *121*, 103751.

Delcambre, S., Ghelfi, J., Ouzren, N., Grandmougin, L., Delbrouck, C., Seibler, P., Wasner, K., Aasly, J.O., Klein, C., Trinh, J*., et al.* (2020). Mitochondrial Mechanisms of LRRK2 G2019S Penetrance. Front Neurol *11*, 881.

Dragicevic, N., Delic, V., Cao, C., Copes, N., Lin, X., Mamcarz, M., Wang, L., Arendash, G.W., and Bradshaw, P.C. (2012). Caffeine increases mitochondrial function and blocks melatonin signaling to mitochondria in Alzheimer's mice and cells. Neuropharmacology *63*, 1368-1379.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., and Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun *10*, 3328.

Dunn, A.R., and Kaczorowski, C.C. (2019). Regulation of intrinsic excitability: Roles for learning and memory, aging and Alzheimer's disease, and genetic diversity. Neurobiol Learn Mem *164*, 107069.

Gabbert, C., Konig, I.R., Luth, T., Kolms, B., Kasten, M., Vollstedt, E.J., Balck, A., Grunewald, A., Klein, C., and Trinh, J. (2022). Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson's disease. J Neurol.

Genomes Project, C., 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*., et al.* (2015). A global reference for human genetic variation. Nature *526*, 68-74.

Goncalves, D.F., de Carvalho, N.R., Leite, M.B., Courtes, A.A., Hartmann, D.D., Stefanello, S.T., da Silva, I.K., Franco, J.L., Soares, F.A.A., and Dalla Corte, C.L. (2018). Caffeine and acetaminophen association: Effects on mitochondrial bioenergetics. Life Sci *193*, 234-241.

Grunewald, A., Kumar, K.R., and Sue, C.M. (2019). New insights into the complex role of mitochondria in Parkinson's disease. Prog Neurobiol *177*, 73-93.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet *5*, e1000529.

Iwaki, H., Leonard, H.L., Makarious, M.B., Bookman, M., Landin, B., Vismer, D., Casey, B., Gibbs, J.R., Hernandez, D.G., Blauwendraat, C*., et al.* (2021). Accelerating Medicines Partnership: Parkinson's Disease. Genetic Resource. Mov Disord *36*, 1795-1804.

Kanithi, M., Junapudi, S., Shah, S.I., Matta Reddy, A., Ullah, G., and Chidipi, B. (2022). Alterations of Mitochondrial Network by Cigarette Smoking and E-Cigarette Vaping. Cells *11*.

Kasala, S., Briyal, S., Prazad, P., Ranjan, A.K., Stefanov, G., Donovan, R., and Gulati, A. (2020). Exposure to Morphine and Caffeine Induces Apoptosis and Mitochondrial Dysfunction in a Neonatal Rat Brain. Front Pediatr *8*, 593.

Klein, C., and Westenberger, A. (2012). Genetics of Parkinson's disease. Cold Spring Harb Perspect Med *2*, a008888.

Kline, E.M., Houser, M.C., Herrick, M.K., Seibler, P., Klein, C., West, A., and Tansey, M.G. (2021). Genetic and Environmental Factors in Parkinson's Disease Converge on Immune Function and Inflammation. Mov Disord *36*, 25-36.

Koch, S., Laabs, B.H., Kasten, M., Vollstedt, E.J., Becktepe, J., Bruggemann, N., Franke, A., Kramer, U.M., Kuhlenbaumer, G., Lieb, W.*, et al.* (2021). Validity and Prognostic Value of a Polygenic Risk Score for Parkinson's Disease. Genes (Basel) *12*.

Kolicheski, A., Turcano, P., Tamvaka, N., McLean, P.J., Springer, W., Savica, R., and Ross, O.A. (2022). Early-Onset Parkinson's Disease: Creating the Right Environment for a Genetic Disorder. J Parkinsons Dis *12*, 2353-2367.

Lai, D., Alipanahi, B., Fontanillas, P., Schwantes-An, T.H., Aasly, J., Alcalay, R.N., Beecham, G.W., Berg, D., Bressman, S., Brice, A.*, et al.* (2021). Genomewide Association Studies of LRRK2 Modifiers of Parkinson's Disease. Ann Neurol *90*, 76-88.

Li, X., Li, W., Liu, G., Shen, X., and Tang, Y. (2015). Association between cigarette smoking and Parkinson's disease: A meta-analysis. Arch Gerontol Geriatr *61*, 510-516.

Luth, T., Konig, I.R., Grunewald, A., Kasten, M., Klein, C., Hentati, F., Farrer, M., and Trinh, J. (2020). Age at Onset of LRRK2 p.Gly2019Ser Is Related to Environmental and Lifestyle Factors. Mov Disord *35*, 1854-1858.

Marras, C., Canning, C.G., and Goldman, S.M. (2019). Environment, lifestyle, and Parkinson's disease: Implications for prevention in the next decade. Mov Disord *34*, 801-811.

Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet *51*, 584-591.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K.*, et al.* (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet *48*, 1279-1283.

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.*, et al.* (2019). 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.

Nistico, R., Mehdawy, B., Piccirilli, S., and Mercuri, N. (2011). Paraquat- and rotenone-induced models of Parkinson's disease. Int J Immunopathol Pharmacol *24*, 313-322.

Noyce, A.J., Bestwick, J.P., Silveira-Moriyama, L., Hawkes, C.H., Giovannoni, G., Lees, A.J., and Schrag, A. (2012). Meta-analysis of early nonmotor features and risk factors for Parkinson disease. Ann Neurol *72*, 893-901.

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I.*, et al.* (2014). A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet *10*, e1004234.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

R Core Team (2020). A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria).

Reyes, C.M., and Cornelis, M.C. (2018). Caffeine in the Diet: Country-Level Consumption and Guidelines. Nutrients *10*.

San Luciano, M., Tanner, C.M., Meng, C., Marras, C., Goldman, S.M., Lang, A.E., Tolosa, E., Schule, B., Langston, J.W., Brice, A.*, et al.* (2020). Nonsteroidal Anti-inflammatory Use and LRRK2 Parkinson's Disease Penetrance. Mov Disord *35*, 1755-1764.

Semple, S.E., Dick, F., Cherrie, J.W., and Geoparkinson Study, G. (2004). Exposure assessment for a population-based case-control study combining a job-exposure matrix with interview data. Scand J Work Environ Health *30*, 241-248.

Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. Cell *177*, 26-31.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W.*, et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Tanner, C., Marras, C., Meng, C., Marder, K., Bressman, S., Saunders-Pullman, R., Alcalay, R., Tolosa, E., Brice, A., Goldman, S.*, et al.* (2017). Caffeinated Drinks, LRRK2 Genotype and PD. Mov Disord [abstract].

Tommasi, S., Pabustan, N., Li, M., Chen, Y., Siegmund, K.D., and Besaratinia, A. (2021). A novel role for vaping in mitochondrial gene dysregulation and inflammation fundamental to disease development. Sci Rep *11*, 22773.

Trinh, J., Amouri, R., Duda, J.E., Morley, J.F., Read, M., Donald, A., Vilarino-Guell, C., Thompson, C., Szu Tu, C., Gustavsson, E.K.*, et al.* (2014a). Comparative study of Parkinson's disease and leucine-rich repeat kinase 2 p.G2019S parkinsonism. Neurobiol Aging *35*, 1125-1131.

Trinh, J., Guella, I., and Farrer, M.J. (2014b). Disease penetrance of late-onset parkinsonism: a meta-analysis. JAMA Neurol *71*, 1535-1539.

Trinh, J., Gustavsson, E.K., Vilarino-Guell, C., Bortnick, S., Latourelle, J., McKenzie, M.B., Tu, C.S., Nosova, E., Khinda, J., Milnerwood, A.*, et al.* (2016). DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. Lancet Neurol *15*, 1248-1256.

Weindel, C.G., Martinez, E.L., Zhao, X., Mabry, C.J., Bell, S.L., Vail, K.J., Coleman, A.K., VanPortfliet, J.J., Zhao, B., Wagner, A.R.*, et al.* (2022). Mitochondrial ROS promotes susceptibility to infection via gasdermin D-mediated necroptosis. Cell *185*, 3214-3231 e3223.

Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis. In Use R!, (Cham: Springer International Publishing : Imprint: Springer,), pp. 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color.

Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L.*, et al.* (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514-518.

Yahalom, G., Rigbi, A., Israeli-Korn, S., Krohn, L., Rudakou, U., Ruskey, J.A., Benshimol, L., Tsafnat, T., Gan-Or, Z., Hassin-Baer, S.*, et al.* (2020). 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.

3. CONCLUSION

The relationship and substantial influence of genetic, environmental, and lifestyle factors on the AAO and clinical severity of different forms of PD (i.e., idiopathic PD, *GBA1*-PD, *LRRK2*-PD) was investigated. First, lifestyle factors were identified as modifiers of AAO in patients with idiopathic PD. In addition to the known protective lifestyle factors smoking and coffee drinking, aspirin intake was associated with a later AAO in PD, which was further replicated in an independent study cohort. Second, these three protective lifestyle factors were further found to be associated with clinical severity in patients with PD. Aspirin users and smokers had a higher probability of experiencing motor- and non-motor symptoms. In smokers, the association with more severe problems was even more pronounced for non-motor symptoms related to mood, such as anxiety and depression. Third, different Nanopore long-read sequencing analysis workflows were evaluated to analyze variants in *GBA1*. By using two aligners and three variant callers, the most accurate analysis pipeline was determined (i.e., NGMLR/Minimap2+BCFtools), leading to the detection of 13 rare *GBA1* variants that were categorized as "pathogenic", "likely pathogenic", or "uncertain significance". Ten of these rare variants were more frequently found in patients with PD than healthy controls in this study group. In a systematic literature review of 100 articles assessing *GBA1* variant frequencies in PD, variant frequencies between 0% and 26.3% were found in patients in PD, depending on the *GBA1* variant, the population, and the sequencing method. Lastly, investigations of 1) the pathway-independent polygenic score and 2) the pathway-dependent mitochondrial polygenic score in combination with environmental and lifestyle factors in idiopathic PD, *GBA1*-PD, and *LRRK2*-PD were performed. While PGS, MGS, and caffeinated soda were associated with an earlier AAO, coffee drinking, tobacco use, and aspirin intake showed individual and combined associations with a later AAO. Additional additive effects between tobacco use and PGS, as well as aspirin intake and PGS on AAO, were demonstrated without an interaction. In contrast, there was an interaction between tobacco use and MGS in iPD and between caffeinated soda and MGS in *LRRK2*-PD.

Thus, the patient’s genetic disposition and external influences such as environmental and lifestyle factors can, separately and in combination, affect the course of disease in patients with PD (Figure 5).

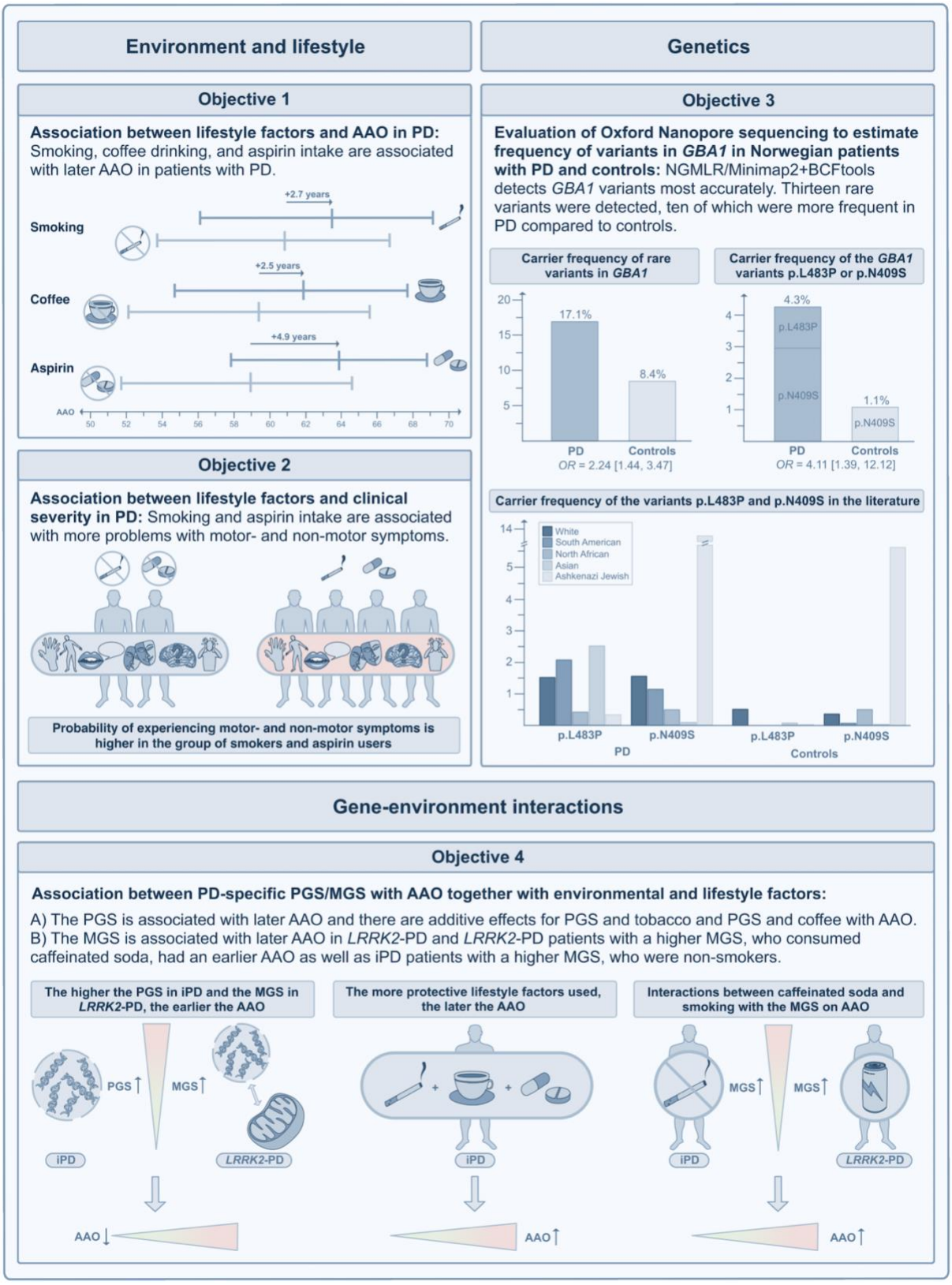


Figure 5: Overview of the main findings. The main findings of the objectives 1-4 are displayed in the four panels.

3.1 Lifestyle factors modifying AAO in PD

PD is a multifactorial disorder that can present with different phenotypes depending on the patient's lifestyle and exposure to the environment. Although the relationship with the risk of developing PD is increasingly well characterized, associations with the age at disease onset still lack in-depth analyses. In this objective, the aim was to analyze environmental and lifestyle factors as AAO modifiers in patients with PD (Gabbert et al., 2022).

Patients with PD who regularly smoked, drank coffee, or took aspirin before their PD onset had a later AAO compared to patients who did not use or consume these substances, with a 3-, 2-, and 5-year later median AAO, respectively. There were additional correlations with a later AAO, the more and the longer the patients used to smoke, drink coffee, or take aspirin. In this study group, these associations could not be extended to other caffeinated beverages, e.g., black tea, and other anti-inflammatory drugs, e.g., ibuprofen. While the protective effect of smoking and coffee on PD AAO has been previously demonstrated in other studies that were summarized in a systematic literature review, the protective effect of aspirin on PD AAO, which was found in the initial study group and an independent replication cohort, has not been previously described. The clinical impact of NSAIDs on patients with PD is still uncertain (Ascherio and Schwarzschild, 2016; Auriel et al., 2014; Badshah et al., 2019; Chen et al., 2005). Inflammation is suspected to play a role in the pathophysiological pathway of PD. However, whether it affects PD pathogenesis or accelerates progression has yet to be clarified (Marogianni et al., 2020). It is known that the production of inflammatory factors can contribute to neuronal death (Tansey et al., 2022), and prolonged neuroinflammation might result from ongoing cell death, which might further impact PD progression (Pajares et al., 2020). Therefore, the use of NSAIDs might protect neuronal integrity, however, an association between NSAIDs and the risk of PD has not been found thus far (Poly et al., 2019). Whether there is a causal relation between the intake of aspirin and the later AAO due to the anti-inflammatory effect needs to be determined. In a sensitivity analysis, potential comorbidities that could be the reason for the regular aspirin intake were included in the analyses but did not explain the association with AAO. Interestingly, the protective effect that was found for aspirin could not be extended to other NSAIDs, suggesting either different mechanisms of action or different confounders. It is further necessary to note that all lifestyle factors assessments were self-reported by the patients and, therefore, subjective perceptions, which could influence at least the dosage and the duration of usage.

Nevertheless, a consistent trend was seen for all three lifestyle factors (i.e., coffee, smoking, and aspirin), which have not only correlated with later AAO when assessed as binary variables but additionally when the dosage and duration of usage before AAO were considered. Although the underlying mechanisms of action are not entirely unraveled, higher and

consistent consumption of these factors might have a positive effect in delaying the disease development and need to be assessed in future investigations. It further needs to be investigated with caution whether the consumption of coffee, smoking, and aspirin intake may be used for PD prevention and therapeutic strategies (e.g., in prodromal PD). Despite the protective association between coffee, smoking, and aspirin with PD susceptibility and AAO, these lifestyle factors also have side effects. Although moderate consumption of caffeinated coffee can be part of a healthy lifestyle, a high caffeine intake may also induce anxiety, insomnia, and nervousness (Lara, 2010; van Dam et al., 2020). More severe side effects have further been found for smoking and NSAID use. Tobacco use constitutes several health hazards, including an increased risk of cardiovascular, respiratory, and gastrointestinal disorders, as well as an increased risk for cancer, a decreased immune response, and an overall reduction in quality of life (Le Foll et al., 2022; Mishra et al., 2015). In addition to the protective effects of NSAIDs, which can reduce pain and inflammation and protect against cancer and heart attacks, NSAIDs have been implicated in gastrointestinal, cardiovascular, hepatic, renal, cerebral, and pulmonary complications and bleedings leading to severe organ damage (Al-Azayzih et al., 2020; Bindu et al., 2020; Harirforoosh et al., 2013). Therefore, treatment with these lifestyle factors does not seem justifiable at this research stage (Ascherio and Schwarzschild, 2016).

These results add new insight to our understanding of PD and aspirin should be considered as a potential AAO modifier in iPD in addition to the known protective lifestyle factors smoking and coffee. Additional analyses in populations with a more diverse ethnic and genetic background are needed to assess these findings in other PD genotypes and patients of non-European ancestry, as it has been shown that there can be ethnic and cultural differences in lifestyle factor use and PD phenotype (Abbas et al., 2018; Ben-Joseph et al., 2020; Reyes and Cornelis, 2018; Ritz et al., 2007; Sauerbier et al., 2017; Trinh et al., 2014; Wright Willis et al., 2010).

3.2 Lifestyle factors modifying clinical severity in PD

After determining the association between coffee drinking, aspirin intake, and smoking with later AAO in patients with PD, the relationship between these lifestyle factors and clinical severity, i.e., motor- and non-motor signs and symptoms, was investigated (Gabbert et al., 2023b).

Clinical severity and symptom progression in patients with PD can vary vastly and the period from first subtle symptoms until PD diagnosis might span several decades (Bloem et al., 2021). Even after the onset of PD, motor and non-motor signs and symptoms might progress at

different rates, and it is suggested that environmental and lifestyle factors can influence the occurrence of symptoms and their progression.

In the group of patients with PD who regularly smoked or took aspirin, the probability of experiencing motor or non-motor symptoms was higher compared to the group of patients who did not smoke or take aspirin. Coffee drinking only showed a negative association with the probability of experiencing problems with chewing and swallowing and a positive association with unexplained pains the more cups of coffee were drunk. Therefore, although coffee drinking was correlated with a later age at symptom onset, it did not show a clear relationship with the occurrence of different symptoms. Patients with PD drinking coffee or caffeine, in general, have previously been presented with a lower rate of PD progression (Cho et al., 2019; Hong et al., 2020; Paul et al., 2019). In contrast to the findings for coffee, in the group of patients using aspirin or smoking, more patients had problems with motor and non-motor symptoms compared to the group of patients who did not take aspirin or smoke. Interestingly, while the associations between smoking and clinical severity were relatively comparable for the binary assessment, dosage, and duration of smoking, no association between aspirin intake duration and motor- and non-motor symptoms was found. This finding could indicate that the disease progression is not affected by the time the patients took aspirin but rather by the amount. However, while aspirin intake was defined as intake before PD onset, a change in medication or additional confounders might also play a role in the symptom progression. A longitudinal assessment of the motor and non-motor symptoms with a simultaneous evaluation of the consumption of lifestyle factors and other changes in the patient's life might be crucial to comprehensively assess the relationship between lifestyle factors and clinical severity in PD. It is further necessary to consider that both lifestyle factors and PD symptoms were self-reported in this study. Although the outcome of the assessment of motor- and non-motor symptoms has previously been shown to be comparable to in-person observational research study cohorts (Chahine et al., 2020), the consumption of lifestyle factors was answered to the best of the patient's memory and might differ from the precise consumption. The dosage and duration might be affected by recall biases. The dosage can change over time and the usage period might not always be as consistent, which complicates the evaluation of the possible impact of lifestyle factors. Nevertheless, investigating lifestyle factors and their impact on clinical severity in PD is essential to acquire a better understanding of the complex disease mechanisms.

In conclusion, these results show that aspirin intake and smoking might modify clinical severity of PD and that a patient's lifestyle before the development of PD symptoms can affect not only the AAO but also the course of the disease.

3.3 *GBA1* risk variants in PD

Genetic risk variants can increase the probability of developing PD and variants in *GBA1* are considered the most common genetic risk factors for PD. In this objective, the aim was to evaluate Oxford Nanopore long-read sequencing as a strategy, to determine the frequency of *GBA1* variants in a study group of Norwegian patients with PD and controls, and to compare these variant frequencies to other studies and populations in a systematic literature review (Gabbert et al., 2023c).

Long-read Nanopore sequencing has been proven to detect variants in *GBA1* accurately. However, the choice of software tools plays an essential role in the analysis. While Clair3 (Zheng et al., 2022) and Pepper-Margin-Deepvariant (Shafin et al., 2021) had more false-positive calls and even showed false-negative calls, BCFtools (Danecek et al., 2021) presented with the most accurate variant detection, independent from the aligner (i.e., NGMLR (Sedlazeck et al., 2018) or Minimap2 (Li, 2021). However, the pipeline sensitivity could not be fully evaluated in this study as Sanger sequencing was only performed to confirm variants called with Nanopore sequencing, and the number of false-negative variants might be underestimated. Nevertheless, Oxford Nanopore long-read sequencing accuracy has improved in recent years, further impacting future studies as variant detection becomes even more accurate. The technology further has the advantage that it is both cost and time-efficient. Only a small amount of DNA is needed for the analysis, and due to the capacity of multiplexing samples, it is possible to sequence large cohorts in a short amount of time at comparably low costs. With the additional advantage of sequencing the entire *GBA1* gene without the nearby pseudogene, long-read sequencing technology is most suitable for assessing variants in *GBA1*. With this analysis pipeline, 13 rare distinct variants within *GBA1* were validated, which were found in 17.1% of Norwegian patients with PD and 8.4% of healthy controls. The odds of carrying one of the two more commonly investigated *GBA1* variants, p.L483P or p. N409S, in cases with PD were estimated to be more than four times the odds of carrying one of these variants in controls, which is higher than for most other populations, as investigated in a systematic literature review including 100 publications. However, other variants, e.g., p.T408M, which has been described as a PD risk variant and is not found in the context of Gaucher's disease (Greuel et al., 2020), were found in a lower frequency in PD cases compared to healthy controls. Until today, the pathogenicity of variants in *GBA1* in the context of PD has not been completely evaluated. Therefore, a thorough assessment of all variants in *GBA1* is needed to correctly classify variants relevant to PD. In addition, how variants in *GBA1* further influence the PD phenotype, including the AAO, signs and symptoms, and disease progression, needs to be determined. In this study, patients with *GBA1*-PD had an average 3-year earlier AAO compared to patients without a *GBA1* variant. In recent studies, *GBA1* variant

carriers were found to show altered metabolomic and neuroimaging findings, suggesting a more severe PD pathology compared to other PD cases (Greuel et al., 2020; Onal et al., 2024; Szlepak et al., 2024). This highlights the importance of investigating genetic risk variants in addition to established monogenic forms of PD. A better understanding of the influence of genetic risk variants on PD and how these differ across populations will help improve treatment.

In conclusion, these results underline the importance of an accurate data analysis pipeline and further show that variants in *GBA1* are more frequently found in Norwegian patients with PD than healthy controls.

3.4 Combined effect of lifestyle and genetic risk factors in PD

Lastly, to obtain a more comprehensive understanding of the combined effects of environmental and lifestyle factors and genetics, the cumulative burden of SNPs associated with PD risk (PGS) or with mitochondrial function (MGS) was calculated and investigated together with environmental and lifestyle factors to assess the AAO in patients with PD (Gabbert et al., 2023a; Lüth et al., 2023).

Besides monogenic forms of PD, common risk variants with an individually small effect size have shown a considerable effect on PD risk when taken together. In the first study of this objective, it was demonstrated that the PGS inversely correlated with AAO in patients with iPD, which is in line with other studies that have investigated PGS and AAO (Huang et al., 2024; Pavelka et al., 2022). The lifestyle factors coffee, tobacco, and aspirin, which have previously individually presented with an association with a later AAO (Gabbert et al., 2022), have also shown a combined effect, indicating that the more of these three factors were used by the patients, the later the AAO. In addition, tobacco use and PGS, as well as aspirin intake and PGS, have shown additive effects on AAO with opposite directionality for the lifestyle factors and PGS. However, no interaction was found in this group. In contrast, an interaction between PGS and smoking (Reynoso et al., 2023), as well as between PGS and smoking on AAO (Huang et al., 2024), has been previously presented in other studies with larger sample sizes. Interestingly, in patients with *GBA1*-PD, no additive effect for coffee, tobacco, and aspirin was found. Only tobacco use demonstrated a robust association with AAO when all three factors were included in the same regression model. Although these findings need to be replicated before reaching a clear conclusion, they might point to different inflammatory mechanisms reflected by the varied effects of aspirin on iPD and *GBA1*-PD AAO.

In addition to the general PD-specific PGS, pathway-dependent PGS, which are based on genetic variants assigned to genes involved in disease pathways, can be used to assess PD traits. A large-scale PGS analysis, focusing on annotated gene sets representative of curated

pathways, identified multiple biological pathways associated with PD risk through common genetic variation and further nominated several signal transduction mechanisms affecting protein misfolding and aggregation, adaptive and innate immune response, vesicular-mediated transport, and lipid metabolism on the risk for PD (Bandres-Ciga et al., 2020). Lysosomal PGS have been shown to predict PD status and have further been associated with a faster progression of cognitive decline in PD (Dehestani et al., 2022; Tunold et al., 2024). Another important mechanism in the development of PD is impaired mitochondrial function, which can be assessed using a mitochondria-specific PGS (MGS) to capture the cumulative burden of SNPs associated with mitochondrial function. Previous investigations have shown that MGS can predict PD status and correlate with an earlier AAO in patients with PD (Arena et al., 2024; Billingsley et al., 2019; Dehestani et al., 2022). Similarly, in the second study of this objective, the MGS was not only significantly higher in patients with PD compared to controls but also correlated with AAO in patients with *LRRK2*-PD and a trend in the same direction in patients with iPD. In contrast to the findings for PGS, the MGS interacted with lifestyle factors. The impact of the MGS on AAO depends on certain lifestyle factors, which is due to the fact that mitochondria are the interface of the environmental impact on a cell. In patients with iPD, an interaction between MGS and smoking on AAO was found, while in patients with *LRRK2*-PD, an interaction between MGS and caffeinated soda was found. Thus, non-smoking iPD patients who had a higher MGS had a 6-month earlier AAO compared to patients with a lower MGS. *LRRK2*-PD patients, who had a higher MGS and used to consume caffeinated soda, had an 11-year earlier AAO compared to patients with a lower MGS. In comparison, the median AAO was only four years earlier in *LRRK2*-PD patients, unstratified for any lifestyle factor. Caffeinated soda itself has previously been associated with an earlier AAO in *LRRK2*-PD patients (Luth et al., 2020). In contrast to the protective findings for coffee, the results on caffeinated soda suggest a caffeine-independent effect on the AAO. Caffeinated soda has been shown to interact with genes related to mitochondrial function and, therefore, presents with a damaging effect on mitochondria, which could influence PD progression.

These findings highlight the immense effect lifestyle factors can have on the AAO in PD and further demonstrate how important gene-environment and gene-lifestyle interactions might be. Gene-environment interactions potentially have a major impact on PD phenotypes. However, replication of findings is only limited, as interactions might vary in different study groups and across populations with diverse genetic and ethnic backgrounds. This is especially true since there is a general lack of diversity in genetic PD studies (Schumacher-Schuh et al., 2022). In addition, gene-environment interactions are most frequently investigated in the context of PD risk. Nevertheless, gene-environment interactions that affect AAO or clinical severity in PD need to be investigated to shed light on mechanisms influencing disease progression.

In conclusion, these results demonstrate that the cumulative genetic burden and the patient's lifestyle have an impact on the AAO in PD, especially for genetic variants associated with mitochondrial function.

4. OUTLOOK

The future perspectives of these studies focus on expanding the findings to a more detailed assessment of lifestyle factors, a longitudinal assessment of motor and non-motor signs and symptoms, comparing findings in different PD genotypes and phenotypes, and expanding the analyses to more ethnically diverse populations.

For an in-depth investigation of the relationship between the lifestyle factors smoking, coffee drinking, and aspirin intake with AAO, current and former users at the time of disease onset will be separated for the analyses. This allows for assessing possible long-lasting effects of these lifestyle factors on AAO. Since many patients change their habits after receiving the PD diagnosis, it is also essential to investigate whether the clinical severity of symptoms differs depending on whether patients have continued to smoke, drink coffee, or take aspirin or have stopped the consumption. For a more precise assessment, a longitudinal analysis of motor and non-motor signs and symptoms will be performed and evaluated using mixed linear effect models. Thus, the change in symptom severity can be investigated over time, which results in more detailed outcomes than a cross-sectional assessment. In addition, other environmental and lifestyle factors, e.g., alcohol consumption, pesticide exposure, and physical activity, will be included in the analyses.

To analyze gene-environment interactions in patients with *GBA1*-PD, a large genetic data cohort with additional data available on environmental and lifestyle factors will be investigated to compare the results on PGS and lifestyle factors found in iPD to patients with *GBA1*-PD. We further plan to extend these analyses to monogenic forms of PD, i.e., *LRRK2*-PD, as the MGS findings have shown that the gene-environment and gene-lifestyle interactions can differ in iPD and *LRRK2*-PD.

Another critical aspect of the future perspectives is to perform the above analyses in populations with diverse ethnic backgrounds. As most of our previous results are based on patients with White/European ancestry and a Western lifestyle, we will compare the findings across populations to evaluate potential population-specific effects of genetic variations or environmental and lifestyle factors on PD risk, AAO, and symptom severity. These extensive investigations will facilitate a better comprehension of PD pathogenesis.

5. REFERENCES

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

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*., et al.* (2015). A global reference for human genetic variation. Nature *526*, 68-74.

Abbas, M.M., Xu, Z., and Tan, L.C.S. (2018). Epidemiology of Parkinson's Disease-East Versus West. Mov Disord Clin Pract *5*, 14-28.

Ahmed, I., Lee, P.C., Lill, C.M., Searles Nielsen, S., Artaud, F., Gallagher, L.G., Loriot, M.A., Mulot, C., Nacfer, M., Liu, T*., et al.* (2014). Lack of replication of the GRIN2A-by-coffee interaction in Parkinson disease. PLoS Genet *10*, e1004788.

Al-Azayzih, A., Al-Azzam, S.I., Alzoubi, K.H., Jarab, A.S., Kharaba, Z., Al-Rifai, R.H., and Alnajjar, M.S. (2020). Nonsteroidal Anti-inflammatory Drugs Utilization Patterns and Risk of Adverse Events due to Drug-Drug Interactions among Elderly Patients: A Study from Jordan. Saudi Pharm J *28*, 504-508.

Albin, R., and Grotewold, N. (2023). What is the Parkinson Pandemic? Mov Disord *38*, 2141-2144.

Aligning Science Across Parkinson's. Uncovering the roots of Parkinson's disease, together. https://parkinsonsroadmap.org/#. Accessed April 16, 2024.

Alves, G., Larsen, J.P., Emre, M., Wentzel-Larsen, T., and Aarsland, D. (2006). Changes in motor subtype and risk for incident dementia in Parkinson's disease. Mov Disord *21*, 1123-1130.

Ardui, S., Ameur, A., Vermeesch, J.R., and Hestand, M.S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res *46*, 2159-2168.

Arena, G., Landoulsi, Z., Grossmann, D., Payne, T., Vitali, A., Delcambre, S., Baron, A., Antony, P., Boussaad, I., Bobbili, D.R*., et al.* (2024). Polygenic Risk Scores Validated in Patient-Derived Cells Stratify for Mitochondrial Subtypes of Parkinson's Disease. Ann Neurol.

Armstrong, M.J., and Okun, M.S. (2020). Diagnosis and Treatment of Parkinson Disease: A Review. JAMA *323*, 548-560.

Ascherio, A., and Schwarzschild, M.A. (2016). The epidemiology of Parkinson's disease: risk factors and prevention. Lancet Neurol *15*, 1257-1272.

Aune, D., Schlesinger, S., Mahamat-Saleh, Y., Zheng, B., Udeh-Momoh, C.T., and Middleton, L.T. (2023). Diabetes mellitus, prediabetes and the risk of Parkinson's disease: a systematic review and meta-analysis of 15 cohort studies with 29.9 million participants and 86,345 cases. Eur J Epidemiol *38*, 591-604.

Auriel, E., Regev, K., and Korczyn, A.D. (2014). Nonsteroidal anti-inflammatory drugs exposure and the central nervous system. Handb Clin Neurol *119*, 577-584.

Badshah, H., Ikram, M., Ali, W., Ahmad, S., Hahm, J.R., and Kim, M.O. (2019). Caffeine May Abrogate LPS-Induced Oxidative Stress and Neuroinflammation by Regulating Nrf2/TLR4 in Adult Mouse Brains. Biomolecules *9*.

Balestrino, R., and Schapira, A.H.V. (2020). Parkinson disease. Eur J Neurol *27*, 27-42.

Ball, N., Teo, W.P., Chandra, S., and Chapman, J. (2019). Parkinson's Disease and the Environment. Front Neurol *10*, 218.

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*., et al.* (2020). Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. Acta Neuropathol *140*, 341-358.

Barbeau, A., Dallaire, L., Buu, N.T., Poirier, J., and Rucinska, E. (1985). Comparative behavioral, biochemical and pigmentary effects of MPTP, MPP+ and paraquat in Rana pipiens. Life Sci *37*, 1529-1538.

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*., et al.* (2021). Gene-corrected p.A30P SNCA patient-derived isogenic neurons rescue neuronal branching and function. Sci Rep *11*, 21946.

Barkhuizen, M., Anderson, D.G., and Grobler, A.F. (2016). Advances in GBA-associated Parkinson's disease--Pathology, presentation and therapies. Neurochem Int *93*, 6-25.

Barreto, G.E., Iarkov, A., and Moran, V.E. (2014). Beneficial effects of nicotine, cotinine and its metabolites as potential agents for Parkinson's disease. Front Aging Neurosci *6*, 340.

Bata-Garcia, J.L., Tun-Coba, L., Alvarez-Cervera, F.J., Villanueva-Toledo, J.R., Heredia-Lopez, F.J., and Gongora-Alfaro, J.L. (2010). Improvement of postural adjustment steps in hemiparkinsonian rats chronically treated with caffeine is mediated by concurrent blockade of A1 and A2A adenosine receptors. Neuroscience *166*, 590-603.

Ben-Joseph, A., Marshall, C.R., Lees, A.J., and Noyce, A.J. (2020). Ethnic Variation in the Manifestation of Parkinson's Disease: A Narrative Review. J Parkinsons Dis *10*, 31-45.

Benedetti, M.D., Bower, J.H., Maraganore, D.M., McDonnell, S.K., Peterson, B.J., Ahlskog, J.E., Schaid, D.J., and Rocca, W.A. (2000). Smoking, alcohol, and coffee consumption preceding Parkinson's disease: a case-control study. Neurology *55*, 1350-1358.

Benowitz, N.L. (2009). Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol *49*, 57-71.

Berg, D., Postuma, R.B., Adler, C.H., Bloem, B.R., Chan, P., Dubois, B., Gasser, T., Goetz, C.G., Halliday, G., Joseph, L.*, et al.* (2015). MDS research criteria for prodromal Parkinson's disease. Mov Disord *30*, 1600-1611.

Betarbet, R., Sherer, T.B., and Greenamyre, J.T. (2002). Animal models of Parkinson's disease. Bioessays *24*, 308-318.

Bettiol, S.S., Rose, T.C., Hughes, C.J., and Smith, L.A. (2015). Alcohol Consumption and Parkinson's Disease Risk: A Review of Recent Findings. J Parkinsons Dis *5*, 425-442.

Billingsley, K.J., Bandres-Ciga, S., Saez-Atienzar, S., and Singleton, A.B. (2018). Genetic risk factors in Parkinson's disease. Cell Tissue Res *373*, 9-20.

Billingsley, K.J., Barbosa, I.A., Bandres-Ciga, S., Quinn, J.P., Bubb, V.J., Deshpande, C., Botia, J.A., Reynolds, R.H., Zhang, D., Simpson, M.A.*, et al.* (2019). Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. NPJ Parkinsons Dis *5*, 8.

Bindu, S., Mazumder, S., and Bandyopadhyay, U. (2020). Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. Biochem Pharmacol *180*, 114147.

Blauwendraat, C., Bras, J.M., Nalls, M.A., Lewis, P.A., Hernandez, D.G., Singleton, A.B., and International Parkinson's Disease Genomics, C. (2018). Coding variation in GBA explains the majority of the SYT11-GBA Parkinson's disease GWAS locus. Mov Disord *33*, 1821-1823.

Blauwendraat, C., Reed, X., Krohn, L., Heilbron, K., Bandres-Ciga, S., Tan, M., Gibbs, J.R., Hernandez, D.G., Kumaran, R., Langston, R.*, et al.* (2020). Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. Brain *143*, 234-248.

Bloem, B.R., Okun, M.S., and Klein, C. (2021). Parkinson's disease. Lancet *397*, 2284-2303.

Borghammer, P. (2023). The brain-first vs. body-first model of Parkinson's disease with comparison to alternative models. J Neural Transm (Vienna) *130*, 737-753.

Bottigliengo, D., Foco, L., Seibler, P., Klein, C., Konig, I.R., and Del Greco, M.F. (2022). A Mendelian randomization study investigating the causal role of inflammation on Parkinson's disease. Brain *145*, 3444-3453.

Braak, H., Del Tredici, K., Rub, U., de Vos, R.A., Jansen Steur, E.N., and Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging *24*, 197-211.

Brett, B.L., Gardner, R.C., Godbout, J., Dams-O'Connor, K., and Keene, C.D. (2022). Traumatic Brain Injury and Risk of Neurodegenerative Disorder. Biol Psychiatry *91*, 498-507.

Brothers, H.M., Marchalant, Y., and Wenk, G.L. (2010). Caffeine attenuates lipopolysaccharide-induced neuroinflammation. Neurosci Lett *480*, 97-100.

Burke, R.E., Dauer, W.T., and Vonsattel, J.P. (2008). A critical evaluation of the Braak staging scheme for Parkinson's disease. Ann Neurol *64*, 485-491.

Calabresi, P., Mechelli, A., Natale, G., Volpicelli-Daley, L., Di Lazzaro, G., and Ghiglieri, V. (2023). Alpha-synuclein in Parkinson's disease and other synucleinopathies: from overt neurodegeneration back to early synaptic dysfunction. Cell Death Dis *14*, 176.

Casper, D., Yaparpalvi, U., Rempel, N., and Werner, P. (2000). Ibuprofen protects dopaminergic neurons against glutamate toxicity in vitro. Neurosci Lett *289*, 201-204.

Chahine, L.M., Chin, I., Caspell-Garcia, C., Standaert, D.G., Brown, E., Smolensky, L., Arnedo, V., Daeschler, D., Riley, L., Korell, M.*, et al.* (2020). Comparison of an Online-Only Parkinson's Disease Research Cohort to Cohorts Assessed In Person. J Parkinsons Dis *10*, 677-691.

Chang, D., Nalls, M.A., Hallgrimsdottir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., International Parkinson's Disease Genomics, C., andMe Research, T., Kerchner, G.A., Ayalon, G.*, et al.* (2017). A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet *49*, 1511-1516.

Chen, H., Jacobs, E., Schwarzschild, M.A., McCullough, M.L., Calle, E.E., Thun, M.J., and Ascherio, A. (2005). Nonsteroidal antiinflammatory drug use and the risk for Parkinson's disease. Ann Neurol *58*, 963-967.

Cho, B.H., Choi, S.M., and Kim, B.C. (2019). Gender-dependent effect of coffee consumption on tremor severity in de novo Parkinson's disease. BMC Neurol *19*, 194.

Cho, B.H., Choi, S.M., Kim, J.T., and Kim, B.C. (2018). Association of coffee consumption and non-motor symptoms in drug-naive, early-stage Parkinson's disease. Parkinsonism Relat Disord *50*, 42-47.

Chuang, Y.H., Lee, P.C., Vlaar, T., Mulot, C., Loriot, M.A., Hansen, J., Lill, C.M., Ritz, B., and Elbaz, A. (2017). Pooled analysis of the HLA-DRB1 by smoking interaction in Parkinson disease. Ann Neurol *82*, 655-664.

Chuang, Y.H., Lill, C.M., Lee, P.C., Hansen, J., Lassen, C.F., Bertram, L., Greene, N., Sinsheimer, J.S., and Ritz, B. (2016). Gene-Environment Interaction in Parkinson's Disease: Coffee, ADORA2A, and CYP1A2. Neuroepidemiology *47*, 192-200.

Cilia, R., Tunesi, S., Marotta, G., Cereda, E., Siri, C., Tesei, S., Zecchinelli, A.L., Canesi, M., Mariani, C.B., Meucci, N.*, et al.* (2016). Survival and dementia in GBA-associated Parkinson's disease: The mutation matters. Ann Neurol *80*, 662-673.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M.*, et al.* (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*.

De Carolis, L., Galli, S., Bianchini, E., Rinaldi, D., Raju, M., Calio, B., Alborghetti, M., and Pontieri, F.E. (2023). Age at Onset Influences Progression of Motor and Non-Motor Symptoms during the Early Stage of Parkinson's Disease: A Monocentric Retrospective Study. Brain Sci *13*.

De Reuck, J., De Weweire, M., Van Maele, G., and Santens, P. (2005). Comparison of age of onset and development of motor complications between smokers and non-smokers in Parkinson's disease. J Neurol Sci *231*, 35-39.

Dehestani, M., Liu, H., Sreelatha, A.A.K., Schulte, C., Bansal, V., and Gasser, T. (2022). Mitochondrial and autophagy-lysosomal pathway polygenic risk scores predict Parkinson's disease. Mol Cell Neurosci *121*, 103751.

Delamarre, A., and Meissner, W.G. (2017). Epidemiology, environmental risk factors and genetics of Parkinson's disease. Presse Med *46*, 175-181.

Di Fonzo, A., Wu-Chou, Y.H., Lu, C.S., van Doeselaar, M., Simons, E.J., Rohe, C.F., Chang, H.C., Chen, R.S., Weng, Y.H., Vanacore, N.*, et al.* (2006). A common missense variant in the LRRK2 gene, Gly2385Arg, associated with Parkinson's disease risk in Taiwan. Neurogenetics *7*, 133-138.

Domenighetti, C., Douillard, V., Sugier, P.E., Sreelatha, A.A.K., Schulte, C., Grover, S., May, P., Bobbili, D.R., Radivojkov-Blagojevic, M., Lichtner, P.*, et al.* (2022a). The Interaction between HLA-DRB1 and Smoking in Parkinson's Disease Revisited. Mov Disord *37*, 1929-1937.

Domenighetti, C., Sugier, P.E., Sreelatha, A.A.K., Schulte, C., Grover, S., Mohamed, O., Portugal, B., May, P., Bobbili, D.R., Radivojkov-Blagojevic, M.*, et al.* (2022b). Mendelian Randomisation Study of Smoking, Alcohol, and Coffee Drinking in Relation to Parkinson's Disease. J Parkinsons Dis *12*, 267-282.

Dorsey, E.R., Sherer, T., Okun, M.S., and Bloem, B.R. (2018). The Emerging Evidence of the Parkinson Pandemic. J Parkinsons Dis *8*, S3-S8.

Dudbridge, F. (2016). Polygenic Epidemiology. Genet Epidemiol *40*, 268-272.

Elbaz, A., Levecque, C., Clavel, J., Vidal, J.S., Richard, F., Amouyel, P., Alperovitch, A., Chartier-Harlin, M.C., and Tzourio, C. (2004). CYP2D6 polymorphism, pesticide exposure, and Parkinson's disease. Ann Neurol *55*, 430-434.

Eriksson, A.K., Lofving, S., Callaghan, R.C., and Allebeck, P. (2013). Alcohol use disorders and risk of Parkinson's disease: findings from a Swedish national cohort study 1972-2008. BMC Neurol *13*, 190.

Fan, B., Jabeen, R., Bo, B., Guo, C., Han, M., Zhang, H., Cen, J., Ji, X., and Wei, J. (2020). What and How Can Physical Activity Prevention Function on Parkinson's Disease? Oxid Med Cell Longev *2020*, 4293071.

Farrer, M.J., Stone, J.T., Lin, C.H., Dachsel, J.C., Hulihan, M.M., Haugarvoll, K., Ross, O.A., and Wu, R.M. (2007). Lrrk2 G2385R is an ancestral risk factor for Parkinson's disease in Asia. Parkinsonism Relat Disord *13*, 89-92.

Fitzmaurice, A.G., Rhodes, S.L., Cockburn, M., Ritz, B., and Bronstein, J.M. (2014). Aldehyde dehydrogenase variation enhances effect of pesticides associated with Parkinson disease. Neurology *82*, 419-426.

Fox, S.H., Katzenschlager, R., Lim, S.Y., Barton, B., de Bie, R.M.A., Seppi, K., Coelho, M., Sampaio, C., and Movement Disorder Society Evidence-Based Medicine, C. (2018). International Parkinson and movement disorder

society evidence-based medicine review: Update on treatments for the motor symptoms of Parkinson's disease. Mov Disord *33*, 1248-1266.

Gabbert, C., Blobaum, L., Luth, T., Konig, I.R., Caliebe, A., Koch, S., Bjorn-Hergen, L., Klein, C., and Trinh, J. (2023a). The combined effect of lifestyle factors and polygenic scores on age at onset in Parkinson's disease. medRxiv.

Gabbert, C., Konig, I.R., Luth, T., Kasten, M., Grunewald, A., Klein, C., and Trinh, J. (2023b). Lifestyle factors and clinical severity of Parkinson's disease. Sci Rep *13*, 9537.

Gabbert, C., Konig, I.R., Luth, T., Kolms, B., Kasten, M., Vollstedt, E.J., Balck, A., Grunewald, A., Klein, C., and Trinh, J. (2022). Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson's disease. J Neurol.

Gabbert, C., Schaake, S., Luth, T., Much, C., Klein, C., Aasly, J.O., Farrer, M.J., and Trinh, J. (2023c). GBA1 in Parkinson's disease: variant detection and pathogenicity scoring matters. BMC Genomics *24*, 322.

Gagne, J.J., and Power, M.C. (2010). Anti-inflammatory drugs and risk of Parkinson disease: a meta-analysis. Neurology *74*, 995-1002.

Gamache, P.L., Haj Salem, I., Roux-Dubois, N., Le Bouthillier, J., Gan-Or, Z., and Dupre, N. (2019). Exposure to Pesticides and Welding Hastens the Age-at-Onset of Parkinson's Disease. Can J Neurol Sci *46*, 711-716.

Gan-Or, Z., Amshalom, I., Kilarski, L.L., Bar-Shira, A., Gana-Weisz, M., Mirelman, A., Marder, K., Bressman, S., Giladi, N., and Orr-Urtreger, A. (2015). Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology *84*, 880-887.

Gao, X., Chen, H., Schwarzschild, M.A., and Ascherio, A. (2011). Use of ibuprofen and risk of Parkinson disease. Neurology *76*, 863-869.

Gigante, A.F., Asabella, A.N., Iliceto, G., Martino, T., Ferrari, C., Defazio, G., and Rubini, G. (2018). Chronic coffee consumption and striatal DAT-SPECT findings in Parkinson's disease. Neurol Sci *39*, 551-555.

Gigante, A.F., Martino, T., Iliceto, G., and Defazio, G. (2017). Smoking and age-at-onset of both motor and non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord *45*, 94-96.

Global Parkinson's Genetics Program. About GP2. https://gp2.org/about-gp2/. Accessed April 16, 2024.

Grant, H., Lantos, P.L., and Parkinson, C. (1980). Cerebral damage in paraquat poisoning. Histopathology *4*, 185-195.

Greenland, J.C., Williams-Gray, C.H., and Barker, R.A. (2019). The clinical heterogeneity of Parkinson's disease and its therapeutic implications. Eur J Neurosci *49*, 328-338.

Greuel, A., Trezzi, J.P., Glaab, E., Ruppert, M.C., Maier, F., Jager, C., Hodak, Z., Lohmann, K., Ma, Y., Eidelberg, D.*, et al.* (2020). GBA Variants in Parkinson's Disease: Clinical, Metabolomic, and Multimodal Neuroimaging Phenotypes. Mov Disord *35*, 2201-2210.

Grover, S., Kumar Sreelatha, A.A., Pihlstrom, L., Domenighetti, C., Schulte, C., Sugier, P.E., Radivojkov-Blagojevic, M., Lichtner, P., Mohamed, O., Portugal, B.*, et al.* (2022). Genome-wide Association and Meta-analysis of Age at Onset in Parkinson Disease: Evidence From the COURAGE-PD Consortium. Neurology *99*, e698-e710.

Grover, S., Lill, C.M., Kasten, M., Klein, C., Del Greco, M.F., and Konig, I.R. (2019). Risky behaviors and Parkinson disease: A mendelian randomization study. Neurology *93*, e1412-e1424.

Grunewald, A., Kumar, K.R., and Sue, C.M. (2019). New insights into the complex role of mitochondria in Parkinson's disease. Prog Neurobiol *177*, 73-93.

Gustavsson, E.K., Follett, J., Trinh, J., Barodia, S.K., Real, R., Liu, Z., Grant-Peters, M., Fox, J.D., Appel-Cresswell, S., Stoessl, A.J.*, et al.* (2024). RAB32 Ser71Arg in autosomal dominant Parkinson's disease: linkage, association, and functional analyses. Lancet Neurol *23*, 603-614.

Hammond, E.C. (1966). Smoking in relation to the death rates of one million men and women. Natl Cancer Inst Monogr *19*, 127-204.

Hamza, T.H., Chen, H., Hill-Burns, E.M., Rhodes, S.L., Montimurro, J., Kay, D.M., Tenesa, A., Kusel, V.I., Sheehan, P., Eaaswarkhanth, M.*, et al.* (2011). Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. PLoS Genet *7*, e1002237.

Hamza, T.H., Hill-Burns, E.M., Scott, W.K., Vance, J.M., Factor, S.A., Zabetian, C.P., and Payami, H. (2014). Glutamate receptor gene GRIN2A, coffee, and Parkinson disease. PLoS Genet *10*, e1004774.

Harirforoosh, S., Asghar, W., and Jamali, F. (2013). Adverse effects of nonsteroidal antiinflammatory drugs: an update of gastrointestinal, cardiovascular and renal complications. J Pharm Pharm Sci *16*, 821-847.

Harper, J.W., Ordureau, A., and Heo, J.M. (2018). Building and decoding ubiquitin chains for mitophagy. Nat Rev Mol Cell Biol *19*, 93-108.

Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics *107*, 1-8.

Heilbron, K., Jensen, M.P., Bandres-Ciga, S., Fontanillas, P., Blauwendraat, C., Nalls, M.A., Singleton, A.B., Smith, G.D., Cannon, P., Noyce, A.J.*, et al.* (2021). Unhealthy Behaviours and Risk of Parkinson's Disease: A Mendelian Randomisation Study. J Parkinsons Dis *11*, 1981-1993.

Hill-Burns, E.M., Singh, N., Ganguly, P., Hamza, T.H., Montimurro, J., Kay, D.M., Yearout, D., Sheehan, P., Frodey, K., McLear, J.A.*, et al.* (2013). A genetic basis for the variable effect of smoking/nicotine on Parkinson's disease. Pharmacogenomics J *13*, 530-537.

Ho, P.W., Leung, C.T., Liu, H., Pang, S.Y., Lam, C.S., Xian, J., Li, L., Kung, M.H., Ramsden, D.B., and Ho, S.L. (2020). Age-dependent accumulation of oligomeric SNCA/alpha-synuclein from impaired degradation in mutant LRRK2 knockin mouse model of Parkinson disease: role for therapeutic activation of chaperone-mediated autophagy (CMA). Autophagy *16*, 347-370.

Hollenbach, J.A., Norman, P.J., Creary, L.E., Damotte, V., Montero-Martin, G., Caillier, S., Anderson, K.M., Misra, M.K., Nemat-Gorgani, N., Osoegawa, K.*, et al.* (2019). A specific amino acid motif of HLA-DRB1 mediates risk and interacts with smoking history in Parkinson's disease. Proc Natl Acad Sci U S A *116*, 7419-7424.

Hong, C.T., Chan, L., and Bai, C.H. (2020). The Effect of Caffeine on the Risk and Progression of Parkinson's Disease: A Meta-Analysis. Nutrients *12*.

Hruska, K.S., LaMarca, M.E., Scott, C.R., and Sidransky, E. (2008). Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). Hum Mutat *29*, 567-583.

Hu, G., Bidel, S., Jousilahti, P., Antikainen, R., and Tuomilehto, J. (2007). Coffee and tea consumption and the risk of Parkinson's disease. Mov Disord *22*, 2242-2248.

Hu, T., Chitnis, N., Monos, D., and Dinh, A. (2021). Next-generation sequencing technologies: An overview. Hum Immunol *82*, 801-811.

Huang, S.N., Ruan, H.Z., Chen, M.Y., Zhou, G., and Qian, Z.M. (2018). Aspirin increases ferroportin 1 expression by inhibiting hepcidin via the JAK/STAT3 pathway in interleukin 6-treated PC-12 cells. Neurosci Lett *662*, 1-5.

Huang, Y., Chen, Q., Wang, Z., Wang, Y., Lian, A., Zhou, Q., Zhao, G., Xia, K., Tang, B., Li, B.*, et al.* (2024). Risk factors associated with age at onset of Parkinson's disease in the UK Biobank. NPJ Parkinsons Dis *10*, 3.

Ibarra-Gutierrez, M.T., Serrano-Garcia, N., and Orozco-Ibarra, M. (2023). Rotenone-Induced Model of Parkinson's Disease: Beyond Mitochondrial Complex I Inhibition. Mol Neurobiol *60*, 1929-1948.

International HapMap, C. (2003). The International HapMap Project. Nature *426*, 789-796.

International Schizophrenia, C., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748-752.

Iwaki, H., Blauwendraat, C., Makarious, M.B., Bandres-Ciga, S., Leonard, H.L., Gibbs, J.R., Hernandez, D.G., Scholz, S.W., Faghri, F., International Parkinson's Disease Genomics, C.*, et al.* (2020). Penetrance of Parkinson's Disease in LRRK2 p.G2019S Carriers Is Modified by a Polygenic Risk Score. Mov Disord *35*, 774-780.

Jackson, L.M., Woodruff, B.K., Tremblay, C., Shill, H.A., Beach, T.G., Serrano, G.E., and Adler, C.H. (2024). Parkinson's Disease Associated with G2019S LRRK2 Mutations without Lewy Body Pathology. Mov Disord Clin Pract.

Jacobs, B.M., Belete, D., Bestwick, J., Blauwendraat, C., Bandres-Ciga, S., Heilbron, K., Dobson, R., Nalls, M.A., Singleton, A., Hardy, J.*, et al.* (2020). Parkinson's disease determinants, prediction and gene-environment interactions in the UK Biobank. J Neurol Neurosurg Psychiatry *91*, 1046-1054.

Jafari, S., Etminan, M., Aminzadeh, F., and Samii, A. (2013). Head injury and risk of Parkinson disease: a systematic review and meta-analysis. Mov Disord *28*, 1222-1229.

Jankovic, J., McDermott, M., Carter, J., Gauthier, S., Goetz, C., Golbe, L., Huber, S., Koller, W., Olanow, C., Shoulson, I.*, et al.* (1990). Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. Neurology *40*, 1529-1534.

Jia, F., Fellner, A., and Kumar, K.R. (2022). Monogenic Parkinson's Disease: Genotype, Phenotype, Pathophysiology, and Genetic Testing. Genes (Basel) *13*.

Jimenez-Jimenez, F.J., Alonso-Navarro, H., Garcia-Martin, E., and Agundez, J.A.G. (2019). Alcohol consumption and risk for Parkinson's disease: a systematic review and meta-analysis. J Neurol *266*, 1821-1834.

Johansen, K.K., Torp, S.H., Farrer, M.J., Gustavsson, E.K., and Aasly, J.O. (2018). A Case of Parkinson's Disease with No Lewy Body Pathology due to a Homozygous Exon Deletion in Parkin. Case Rep Neurol Med *2018*, 6838965.

Kahn, H.A. (1966). The Dorn study of smoking and mortality among U.S. veterans: report on eight and one-half years of observation. Natl Cancer Inst Monogr *19*, 1-125.

Kalda, A., Yu, L., Oztas, E., and Chen, J.F. (2006). Novel neuroprotection by caffeine and adenosine A(2A) receptor antagonists in animal models of Parkinson's disease. J Neurol Sci *248*, 9-15.

Kandinov, B., Giladi, N., and Korczyn, A.D. (2009). Smoking and tea consumption delay onset of Parkinson's disease. Parkinsonism Relat Disord *15*, 41-46.

Kessler, II, and Diamond, E.L. (1971). Epidemiologic studies of Parkinson's disease. I. Smoking and Parkinson's disease: a survey and explanatory hypothesis. Am J Epidemiol *94*, 16-25.

Kim, J.M., Lee, J.Y., Kim, H.J., Kim, J.S., Shin, E.S., Cho, J.H., Park, S.S., and Jeon, B.S. (2010). The LRRK2 G2385R variant is a risk factor for sporadic Parkinson's disease in the Korean population. Parkinsonism Relat Disord *16*, 85-88.

Koch, S., Laabs, B.H., Kasten, M., Vollstedt, E.J., Becktepe, J., Bruggemann, N., Franke, A., Kramer, U.M., Kuhlenbaumer, G., Lieb, W.*, et al.* (2021). Validity and Prognostic Value of a Polygenic Risk Score for Parkinson's Disease. Genes (Basel) *12*.

Laabs, B.H., Klein, C., Pozojevic, J., Domingo, A., Bruggemann, N., Grutz, K., Rosales, R.L., Jamora, R.D., Saranza, G., Diesta, C.C.E.*, et al.* (2021). Identifying genetic modifiers of age-associated penetrance in X-linked dystonia-parkinsonism. Nat Commun *12*, 3216.

Lange, L.M., Gonzalez-Latapi, P., Rajalingam, R., Tijssen, M.A.J., Ebrahimi-Fakhari, D., Gabbert, C., Ganos, C., Ghosh, R., Kumar, K.R., Lang, A.E.*, et al.* (2022). Nomenclature of Genetic Movement Disorders: Recommendations of the International Parkinson and Movement Disorder Society Task Force - An Update. Mov Disord *37*, 905-935.

Langston, J.W. (2017). The MPTP Story. J Parkinsons Dis *7*, S11-S19.

Langston, J.W., Ballard, P., Tetrud, J.W., and Irwin, I. (1983). Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. Science *219*, 979-980.

Langston, J.W., Irwin, I., Langston, E.B., and Forno, L.S. (1984). 1-Methyl-4-phenylpyridinium ion (MPP+): identification of a metabolite of MPTP, a toxin selective to the substantia nigra. Neurosci Lett *48*, 87-92.

Lara, D.R. (2010). Caffeine, mental health, and psychiatric disorders. J Alzheimers Dis *20 Suppl 1*, S239-248.

Larsson, S.C., and Burgess, S. (2022). Appraising the causal role of smoking in multiple diseases: A systematic review and meta-analysis of Mendelian randomization studies. EBioMedicine *82*, 104154.

Le Foll, B., Piper, M.E., Fowler, C.D., Tonstad, S., Bierut, L., Lu, L., Jha, P., and Hall, W.D. (2022). Tobacco and nicotine use. Nat Rev Dis Primers *8*, 19.

Lee, P.C., Ahmed, I., Loriot, M.A., Mulot, C., Paul, K.C., Bronstein, J.M., Ritz, B., and Elbaz, A. (2018). Smoking and Parkinson disease: Evidence for gene-by-smoking interactions. Neurology *90*, e583-e592.

Leija-Salazar, M., Sedlazeck, F.J., Toffoli, M., Mullin, S., Mokretar, K., Athanasopoulou, M., Donald, A., Sharma, R., Hughes, D., Schapira, A.H.V.*, et al.* (2019). Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. Mol Genet Genomic Med *7*, e564.

Lesage, S., Anheim, M., Condroyer, C., Pollak, P., Durif, F., Dupuits, C., Viallet, F., Lohmann, E., Corvol, J.C., Honore, A.*, et al.* (2011). Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. Hum Mol Genet *20*, 202-210.

Lesage, S., Durr, A., Tazir, M., Lohmann, E., Leutenegger, A.L., Janin, S., Pollak, P., Brice, A., and French Parkinson's Disease Genetics Study, G. (2006). LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. N Engl J Med *354*, 422-423.

Lewis, S.J., Foltynie, T., Blackwell, A.D., Robbins, T.W., Owen, A.M., and Barker, R.A. (2005). Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. J Neurol Neurosurg Psychiatry *76*, 343-348.

Li, C., Ou, R., Chen, Y., Gu, X., Wei, Q., Cao, B., Zhang, L., Hou, Y., Liu, K., Chen, X.*, et al.* (2021). Genetic Modifiers of Age at Onset for Parkinson's Disease in Asians: A Genome-Wide Association Study. Mov Disord *36*, 2077-2084.

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics.

Li, J., Yang, D., Li, Z., Zhao, M., Wang, D., Sun, Z., Wen, P., Dai, Y., Gou, F., Ji, Y.*, et al.* (2023). PINK1/Parkin-mediated mitophagy in neurodegenerative diseases. Ageing Res Rev *84*, 101817.

Lill, C.M., Roehr, J.T., McQueen, M.B., Kavvoura, F.K., Bagade, S., Schjeide, B.M., Schjeide, L.M., Meissner, E., Zauft, U., Allen, N.C.*, et al.* (2012). Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS Genet *8*, e1002548.

Lim, S.Y., and Klein, C. (2024). Parkinson's Disease is Predominantly a Genetic Disease. J Parkinsons Dis *14*, 467-482.

Lin, C.H., Wu, R.M., Tai, C.H., Chen, M.L., and Hu, F.C. (2011). Lrrk2 S1647T and BDNF V66M interact with environmental factors to increase risk of Parkinson's disease. Parkinsonism Relat Disord *17*, 84-88.

Ling, X., Wang, C., Li, L., Pan, L., Huang, C., Zhang, C., Huang, Y., Qiu, Y., Lin, F., and Huang, Y. (2023). Third-generation sequencing for genetic disease. Clin Chim Acta *551*, 117624.

Liu, H., Dehestani, M., Blauwendraat, C., Makarious, M.B., Leonard, H., Kim, J.J., Schulte, C., Noyce, A., Jacobs, B.M., Foote, I.*, et al.* (2022). Polygenic Resilience Modulates the Penetrance of Parkinson Disease Genetic Risk Factors. Ann Neurol *92*, 270-278.

Liu, Y., Bai, H., Gen, S., Zhang, H., Wang, S., Hua, L., Yang, X., Zhang, S., Li, J., and Wang, Y. (2020). Interaction between SNCA gene polymorphisms and T2DM with Parkinson's disease. Acta Neurol Scand *142*, 443-448.

Liu, Y., Zeng, X., Hui, Y., Zhu, C., Wu, J., Taylor, D.H., Ji, J., Fan, W., Huang, Z., and Hu, J. (2015). Activation of alpha7 nicotinic acetylcholine receptors protects astrocytes against oxidative stress-induced apoptosis: implications for Parkinson's disease. Neuropharmacology *91*, 87-96.

Lüth, T., Gabbert, C., König, I.R., Caliebe, A., Koch, S., Laabs, B.-H., Hentati, F., Sassi, S.B., Amouri, R., Spielmann, M.*, et al.* (2023). Interaction of mitochondrial polygenic score and environmental factors in LRRK2 p.Gly2019Ser parkinsonism. medRxiv, 2023.2001.2002.23284113.

Luth, T., Konig, I.R., Grunewald, A., Kasten, M., Klein, C., Hentati, F., Farrer, M., and Trinh, J. (2020). Age at Onset of LRRK2 p.Gly2019Ser Is Related to Environmental and Lifestyle Factors. Mov Disord *35*, 1854-1858.

Manthripragada, A.D., Schernhammer, E.S., Qiu, J., Friis, S., Wermuth, L., Olsen, J.H., and Ritz, B. (2011). Non-steroidal anti-inflammatory drug use and the risk of Parkinson's disease. Neuroepidemiology *36*, 155-161.

Mappin-Kasirer, B., Pan, H., Lewington, S., Kizza, J., Gray, R., Clarke, R., and Peto, R. (2020). Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors. Neurology *94*, e2132-e2138.

Marees, A.T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., and Derks, E.M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res *27*, e1608.

Markey, S.P., Johannessen, J.N., Chiueh, C.C., Burns, R.S., and Herkenham, M.A. (1984). Intraneuronal generation of a pyridinium metabolite may cause drug-induced parkinsonism. Nature *311*, 464-467.

Marogianni, C., Sokratous, M., Dardiotis, E., Hadjigeorgiou, G.M., Bogdanos, D., and Xiromerisiou, G. (2020). Neurodegeneration and Inflammation-An Interesting Interplay in Parkinson's Disease. Int J Mol Sci *21*.

Marras, C., Canning, C.G., and Goldman, S.M. (2019). Environment, lifestyle, and Parkinson's disease: Implications for prevention in the next decade. Mov Disord *34*, 801-811.

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.*, et al.* (2016). Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force. Mov Disord *31*, 436-457.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K.*, et al.* (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet *48*, 1279-1283.

McCulloch, C.C., Kay, D.M., Factor, S.A., Samii, A., Nutt, J.G., Higgins, D.S., Griffith, A., Roberts, J.W., Leis, B.C., Montimurro, J.S.*, et al.* (2008). Exploring gene-environment interactions in Parkinson's disease. Hum Genet *123*, 257-265.

Mishra, A., Chaturvedi, P., Datta, S., Sinukumar, S., Joshi, P., and Garg, A. (2015). Harmful effects of nicotine. Indian J Med Paediatr Oncol *36*, 24-31.

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.*, et al.* (2019). 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.

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.*, et al.* (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet *46*, 989-993.

Nehlig, A., Daval, J.L., and Debry, G. (1992). Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects. Brain Res Brain Res Rev *17*, 139-170.

Nicholatos, J.W., Francisco, A.B., Bender, C.A., Yeh, T., Lugay, F.J., Salazar, J.E., Glorioso, C., and Libert, S. (2018). Nicotine promotes neuron survival and partially protects from Parkinson's disease by suppressing SIRT6. Acta Neuropathol Commun *6*, 120.

Noyce, A.J., Bestwick, J.P., Silveira-Moriyama, L., Hawkes, C.H., Giovannoni, G., Lees, A.J., and Schrag, A. (2012). Meta-analysis of early nonmotor features and risk factors for Parkinson disease. Ann Neurol *72*, 893-901.

Nyren, P., and Lundin, A. (1985). Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal Biochem *151*, 504-509.

Onal, G., Yalcin-Cakmakli, G., Ozcelik, C.E., Boussaad, I., Seker, U.O.S., Fernandes, H.J.R., Demir, H., Kruger, R., Elibol, B., Dokmeci, S.*, et al.* (2024). Variant-specific effects of GBA1 mutations on dopaminergic neuron proteostasis. J Neurochem.

Outeiro, T.F., Alcalay, R.N., Antonini, A., Attems, J., Bonifati, V., Cardoso, F., Chesselet, M.F., Hardy, J., Madeo, G., McKeith, I.*, et al.* (2023). Defining the Riddle in Order to Solve It: There Is More Than One "Parkinson's Disease". Mov Disord *38*, 1127-1142.

Ozelius, L.J., Senthil, G., Saunders-Pullman, R., Ohmann, E., Deligtisch, A., Tagliati, M., Hunt, A.L., Klein, C., Henick, B., Hailpern, S.M.*, et al.* (2006). LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. N Engl J Med *354*, 424-425.

Pajares, M., A, I.R., Manda, G., Bosca, L., and Cuadrado, A. (2020). Inflammation in Parkinson's Disease: Mechanisms and Therapeutic Implications. Cells *9*.

Pan, H., Liu, Z., Ma, J., Li, Y., Zhao, Y., Zhou, X., Xiang, Y., Wang, Y., Zhou, X., He, R.*, et al.* (2023). Genome-wide association study using whole-genome sequencing identifies risk loci for Parkinson's disease in Chinese population. NPJ Parkinsons Dis *9*, 22.

Pan, L., Meng, L., He, M., and Zhang, Z. (2021). Tau in the Pathophysiology of Parkinson's Disease. J Mol Neurosci *71*, 2179-2191.

Parain, K., Hapdey, C., Rousselet, E., Marchand, V., Dumery, B., and Hirsch, E.C. (2003). Cigarette smoke and nicotine protect dopaminergic neurons against the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine Parkinsonian toxin. Brain Res *984*, 224-232.

Park, H.J., Lee, P.H., Ahn, Y.W., Choi, Y.J., Lee, G., Lee, D.Y., Chung, E.S., and Jin, B.K. (2007). Neuroprotective effect of nicotine on dopaminergic neurons by anti-inflammatory action. Eur J Neurosci *26*, 79-89.

Parkinson, J. (2002). An essay on the shaking palsy. 1817. J Neuropsychiatry Clin Neurosci *14*, 223-236; discussion 222.

Parkinson Progression Marker, I. (2011). The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol *95*, 629-635.

Paul, K.C., Chuang, Y.H., Shih, I.F., Keener, A., Bordelon, Y., Bronstein, J.M., and Ritz, B. (2019). The association between lifestyle factors and Parkinson's disease progression and mortality. Mov Disord *34*, 58-66.

Pavelka, L., Rauschenberger, A., Landoulsi, Z., Pachchek, S., May, P., Glaab, E., Kruger, R., and Consortium, N.-P. (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.

Peng, B., Yang, Q., R, B.J., Liu, Y., Akbar, M., Song, B.J., Zhou, S., and Wang, X. (2020). Role of Alcohol Drinking in Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis. Int J Mol Sci *21*.

Poewe, W., Seppi, K., Tanner, C.M., Halliday, G.M., Brundin, P., Volkmann, J., Schrag, A.E., and Lang, A.E. (2017). Parkinson disease. Nat Rev Dis Primers *3*, 17013.

Poly, T.N., Islam, M.M.R., Yang, H.C., and Li, Y.J. (2019). Non-steroidal anti-inflammatory drugs and risk of Parkinson's disease in the elderly population: a meta-analysis. Eur J Clin Pharmacol *75*, 99-108.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R.*, et al.* (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045-2047.

Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E.*, et al.* (2015). MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord *30*, 1591-1601.

Ratner, M.H., Farb, D.H., Ozer, J., Feldman, R.G., and Durso, R. (2014). Younger age at onset of sporadic Parkinson's disease among subjects occupationally exposed to metals and pesticides. Interdiscip Toxicol *7*, 123-133.

Rees, K., Stowe, R., Patel, S., Ives, N., Breen, K., Clarke, C.E., and Ben-Shlomo, Y. (2011). Non-steroidal anti-inflammatory drugs as disease-modifying agents for Parkinson's disease: evidence from observational studies. Cochrane Database Syst Rev, CD008454.

Ren, Y., Liu, W., Jiang, H., Jiang, Q., and Feng, J. (2005). Selective vulnerability of dopaminergic neurons to microtubule depolymerization. J Biol Chem *280*, 34105-34112.

Reyes, C.M., and Cornelis, M.C. (2018). Caffeine in the Diet: Country-Level Consumption and Guidelines. Nutrients *10*.

Reynoso, A., Torricelli, R., Jacobs, B.M., Shi, J., Aslibekyan, S., Norcliffe-Kaufmann, L., Noyce, A.J., and Heilbron, K. (2023). Gene-Environment Interactions for Parkinson's Disease. Ann Neurol.

Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics *13*, 278-289.

Rhodes, S.L., Fitzmaurice, A.G., Cockburn, M., Bronstein, J.M., Sinsheimer, J.S., and Ritz, B. (2013). Pesticides that inhibit the ubiquitin-proteasome system: effect measure modification by genetic variation in SKP1 in Parkinson׳s disease. Environ Res *126*, 1-8.

Ritz, B., Ascherio, A., Checkoway, H., Marder, K.S., Nelson, L.M., Rocca, W.A., Ross, G.W., Strickland, D., Van Den Eeden, S.K., and Gorell, J. (2007). Pooled analysis of tobacco use and risk of Parkinson disease. Arch Neurol *64*, 990-997.

Rizig, M., Bandres-Ciga, S., Makarious, M.B., Ojo, O.O., Crea, P.W., Abiodun, O.V., Levine, K.S., Abubakar, S.A., Achoru, C.O., Vitale, D*., et al.* (2023). Identification of genetic risk loci and causal insights associated with Parkinson's disease in African and African admixed populations: a genome-wide association study. Lancet Neurol *22*, 1015-1025.

Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. Science *281*, 363, 365.

Rosler, T.W., Salama, M., Shalash, A.S., Khedr, E.M., El-Tantawy, A., Fawi, G., El-Motayam, A., El-Seidy, E., El-Sherif, M., El-Gamal, M*., et al.* (2018). K-variant BCHE and pesticide exposure: Gene-environment interactions in a case-control study of Parkinson's disease in Egypt. Sci Rep *8*, 16525.

Ruan, S., Xie, J., Wang, L., Guo, L., Li, Y., Fan, W., Ji, R., Gong, Z., Xu, Y., Mao, J*., et al.* (2023). Nicotine alleviates MPTP-induced nigrostriatal damage through modulation of JNK and ERK signaling pathways in the mice model of Parkinson's disease. Front Pharmacol *14*, 1088957.

Samii, A., Etminan, M., Wiens, M.O., and Jafari, S. (2009). NSAID use and the risk of Parkinson's disease: systematic review and meta-analysis of observational studies. Drugs Aging *26*, 769-779.

San Luciano, M., Tanner, C.M., Meng, C., Marras, C., Goldman, S.M., Lang, A.E., Tolosa, E., Schule, B., Langston, J.W., Brice, A*., et al.* (2020). Nonsteroidal Anti-inflammatory Use and LRRK2 Parkinson's Disease Penetrance. Mov Disord *35*, 1755-1764.

Sanders, L.H., Paul, K.C., Howlett, E.H., Lawal, H., Boppana, S., Bronstein, J.M., Ritz, B., and Greenamyre, J.T. (2017). Editor's Highlight: Base Excision Repair Variants and Pesticide Exposure Increase Parkinson's Disease Risk. Toxicol Sci *158*, 188-198.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74*, 5463-5467.

Sauerbier, A., Lenka, A., Aris, A., and Pal, P.K. (2017). Nonmotor Symptoms in Parkinson's Disease: Gender and Ethnic Differences. Int Rev Neurobiol *133*, 417-446.

Schumacher-Schuh, A.F., Bieger, A., Okunoye, O., Mok, K.Y., Lim, S.Y., Bardien, S., Ahmad-Annuar, A., Santos-Lobato, B.L., Strelow, M.Z., Salama, M*., et al.* (2022). Underrepresented Populations in Parkinson's Genetics Research: Current Landscape and Future Directions. Mov Disord *37*, 1593-1604.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods *15*, 461-468.

See, W.Z.C., Naidu, R., and Tang, K.S. (2024). Paraquat and Parkinson's Disease: The Molecular Crosstalk of Upstream Signal Transduction Pathways Leading to Apoptosis. Curr Neuropharmacol *22*, 140-151.

Serratos, I.N., Hernandez-Perez, E., Campos, C., Aschner, M., and Santamaria, A. (2022). An Update on the Critical Role of alpha-Synuclein in Parkinson's Disease and Other Synucleinopathies: from Tissue to Cellular and Molecular Levels. Mol Neurobiol *59*, 620-642.

Shafin, K., Pesout, T., Chang, P.C., Nattestad, M., Kolesnikov, A., Goel, S., Baid, G., Kolmogorov, M., Eizenga, J.M., Miga, K.H*., et al.* (2021). Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. Nat Methods *18*, 1322-1332.

Shin, W.H., and Chung, K.C. (2020). Death-associated Protein Kinase 1 Phosphorylates alpha-Synuclein at Ser129 and Exacerbates Rotenone-induced Toxic Aggregation of alpha-Synuclein in Dopaminergic SH-SY5Y Cells. Exp Neurobiol *29*, 207-218.

Sia, M.W., Foo, J.N., Saffari, S.E., Wong, A.S., Khor, C.C., Yuan, J.M., Tan, E.K., Koh, W.P., and Tan, L.C. (2021). Polygenic Risk Scores in a Prospective Parkinson's Disease Cohort. Mov Disord *36*, 2936-2940.

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.*, et al.* (2009). Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med *361*, 1651-1661.

Simon, D.K., Wu, C., Tilley, B.C., Lohmann, K., Klein, C., Payami, H., Wills, A.M., Aminoff, M.J., Bainbridge, J., Dewey, R.*, et al.* (2017). Caffeine, creatine, GRIN2A and Parkinson's disease progression. J Neurol Sci *375*, 355-359.

Singh, K., Singh, S., Singhal, N.K., Sharma, A., Parmar, D., and Singh, M.P. (2010). Nicotine- and caffeine-mediated changes in gene expression patterns of MPTP-lesioned mouse striatum: Implications in neuroprotection mechanism. Chem Biol Interact *185*, 81-93.

Skrahina, V., Gaber, H., Vollstedt, E.J., Forster, T.M., Usnich, T., Curado, F., Bruggemann, N., Paul, J., Bogdanovic, X., Zulbahar, S.*, et al.* (2021). The Rostock International Parkinson's Disease (ROPAD) Study: Protocol and Initial Findings. Mov Disord *36*, 1005-1010.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W.*, et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Szlepak, T., Kossev, A.P., Csaban, D., Illes, A., Udvari, S., Balicza, P., Borsos, B., Takats, A., Klivenyi, P., and Molnar, M.J. (2024). GBA-associated Parkinson's disease in Hungary: clinical features and genetic insights. Neurol Sci *45*, 2671-2679.

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat Rev Genet *20*, 467-484.

Tan, E.K., Zhao, Y., Tan, L., Lim, H.Q., Lee, J., Yuen, Y., Pavanni, R., Wong, M.C., Fook-Chong, S., and Liu, J.J. (2007). Analysis of LRRK2 Gly2385Arg genetic variant in non-Chinese Asians. Mov Disord *22*, 1816-1818.

Tanner, C.M., Kamel, F., Ross, G.W., Hoppin, J.A., Goldman, S.M., Korell, M., Marras, C., Bhudhikanok, G.S., Kasten, M., Chade, A.R.*, et al.* (2011). Rotenone, paraquat, and Parkinson's disease. Environ Health Perspect *119*, 866-872.

Tansey, M.G., Wallings, R.L., Houser, M.C., Herrick, M.K., Keating, C.E., and Joers, V. (2022). Inflammation and immune dysfunction in Parkinson disease. Nat Rev Immunol *22*, 657-673.

The Accelerating Medicines Partnership in Parkinson’s Disease (AMP-PD). About AMP-PD Overview. https://amp-pd.org/about. Accessed April 16, 2024.

The Accelerating Medicines Partnership in Parkinson’s Disease (AMP-PD). Unified AMP-PD Cohorts. https://amp-pd.org/unified-cohorts. Accessed April 16, 2024.

The Michael J. Fox Foundation. Fox Insight. https://www.michaeljfox.org/fox-insight. Accessed April 16, 2024.

The Michael J. Fox Foundation. Our promise. https://www.michaeljfox.org/our-promise. Accessed April 16, 2024.

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.*, et al.* (2022). Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene. Commun Biol *5*, 670.

Trinh, J., Guella, I., and Farrer, M.J. (2014). Disease penetrance of late-onset parkinsonism: a meta-analysis. JAMA Neurol *71*, 1535-1539.

Tunold, J.A., Tan, M.M.X., Toft, M., Ross, O., van de Berg, W.D.J., and Pihlstrom, L. (2024). Lysosomal Polygenic Burden Drives Cognitive Decline in Parkinson's Disease with Low Alzheimer Risk. Mov Disord *39*, 596-601.

U. K. Parkinson's Disease Consortium, Wellcome Trust Case Control Consortium, Spencer, C.C., Plagnol, V., Strange, A., Gardner, M., Paisan-Ruiz, C., Band, G., Barker, R.A., Bellenguez, C.*, et al.* (2011). Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. Hum Mol Genet *20*, 345-353.

van Dam, R.M., Hu, F.B., and Willett, W.C. (2020). Coffee, Caffeine, and Health. N Engl J Med *383*, 369-378.

Vellingiri, B., Suriyanarayanan, A., Abraham, K.S., Venkatesan, D., Iyer, M., Raj, N., and Gopalakrishnan, A.V. (2022). Influence of heavy metals in Parkinson's disease: an overview. J Neurol *269*, 5798-5811.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vieira, S.R.L., and Schapira, A.H.V. (2021). Glucocerebrosidase mutations: A paradigm for neurodegeneration pathways. Free Radic Biol Med *175*, 42-55.

Voelkerding, K.V., Dames, S.A., and Durtschi, J.D. (2009). Next-generation sequencing: from basic research to diagnostics. Clin Chem *55*, 641-658.

Wahner, A.D., Bronstein, J.M., Bordelon, Y.M., and Ritz, B. (2007). Nonsteroidal anti-inflammatory drugs may protect against Parkinson disease. Neurology *69*, 1836-1842.

Wang, C., Cai, Y., Zheng, Z., Tang, B.S., Xu, Y., Wang, T., Ma, J., Chen, S.D., Langston, J.W., Tanner, C.M.*, et al.* (2012). Penetrance of LRRK2 G2385R and R1628P is modified by common PD-associated genetic variants. Parkinsonism Relat Disord *18*, 958-963.

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., and Au, K.F. (2021). Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol *39*, 1348-1365.

Williams-Gray, C.H., Foltynie, T., Brayne, C.E., Robbins, T.W., and Barker, R.A. (2007). Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. Brain *130*, 1787-1798.

Williams-Gray, C.H., Mason, S.L., Evans, J.R., Foltynie, T., Brayne, C., Robbins, T.W., and Barker, R.A. (2013). The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. J Neurol Neurosurg Psychiatry *84*, 1258-1264.

Wright Willis, A., Evanoff, B.A., Lian, M., Criswell, S.R., and Racette, B.A. (2010). Geographic and ethnic variation in Parkinson disease: a population-based study of US Medicare beneficiaries. Neuroepidemiology *34*, 143-151.

Xie, Y.X., Bezard, E., and Zhao, B.L. (2005). Investigating the receptor-independent neuroprotective mechanisms of nicotine in mitochondria. J Biol Chem *280*, 32405-32412.

Yahalom, G., Rigbi, A., Israeli-Korn, S., Krohn, L., Rudakou, U., Ruskey, J.A., Benshimol, L., Tsafnat, T., Gan-Or, Z., Hassin-Baer, S.*, et al.* (2020). 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.

Yang, J., Lv, D.J., Li, L.X., Wang, Y.L., Qi, D., Chen, J., Mao, C.J., Wang, F., Liu, Y., Hu, L.F.*, et al.* (2019). Nicotine improved the olfactory impairment in MPTP-induced mouse model of Parkinson's disease. Neurotoxicology *73*, 175-182.

Yoon, E., Ahmed, S., Li, R., Bandres-Ciga, S., Blauwendraat, C., Dustin, I., Scholz, S., Hallett, M., and Ehrlich, D. (2023). Association of polygenic risk score with response to deep brain stimulation in Parkinson's disease. BMC Neurol *23*, 143.

Zhang, D., Jiang, H., and Xie, J. (2014). Alcohol intake and risk of Parkinson's disease: a meta-analysis of observational studies. Mov Disord *29*, 819-822.

Zhang, X., Gao, F., Wang, D., Li, C., Fu, Y., He, W., and Zhang, J. (2018). Tau Pathology in Parkinson's Disease. Front Neurol *9*, 809.

Zheng, Z., Li, S., Su, J., Leung, A.W., Lam, T.W., and Luo, R. (2022). Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nat Comput Sci *2*, 797-803.

Zhou, F.M., Wilson, C.J., and Dani, J.A. (2002). Cholinergic interneuron characteristics and nicotinic properties in the striatum. J Neurobiol *53*, 590-605.

6. APPENDIX

Overview of related publications included in each thesis objective.

Objective	Publication	Link to publication	Link to Supplementary material
1:	Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson's disease	https://link.springer.com/article/10.1007/s00415-022-11041-x	https://static-content.springer.com/esm/art%3A10.1007%2Fs00415-022-11041-x/MediaObjects/415_2022_11041_MOESM1_ESM.pdf
2:	Lifestyle factors and clinical severity of Parkinson's disease	https://www.nature.com/articles/s41598-023-31531-w	https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-023-31531-w/MediaObjects/41598_2023_31531_MOESM1_ESM.pdf
3:	GBA1 in Parkinson's disease: variant detection and pathogenicity scoring matters	https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-023-09417-y	https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09417-y/MediaObjects/12864_2023_9417_MOESM1_ESM.pdf
4 A:	The combined effect of lifestyle factors and polygenic scores on age at onset in Parkinson's disease	https://www.medrxiv.org/content/10.1101/2023.08.25.23294466v1	
4 B:	Interaction of Mitochondrial Polygenic Score and Lifestyle Factors in LRRK2 p.Gly2019Ser Parkinsonism	https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mds.29563	https://movementdisorders.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fmds.28238&file=md s28238-sup-0001-Supinfo.docx

SUPPLEMENTARY MATERIAL FOR

Coffee, smoking and aspirin are associated with age at onset in idiopathic Parkinson's disease

Carolin Gabbert[1], MSc, Prof. Inke R. König[2], Theresa Lüth[1], MSc, Beke Kolms[1], BSc, Meike Kasten[1,3], MD, Eva-Juliane Vollstedt[1], MD, Alexander Balck[1], MD, Fox Insight Study, Anne Grünewald[1,4], PhD, Christine Klein[1], MD, Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Lübeck, Germany
[2]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany
[3]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany
[4]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

*** Correspondence:**
Joanne Trinh
University of Lübeck
Ratzeburger Allee 160
23538 Lübeck, Germany
Email: joanne.trinh@neuro.uni-luebeck.de
Tel.: +49-451-31018202

Supplementary Table S1: Demographics of the Fox Insight participants

Full Cohort (n=35,963)	Patients with PD
Male (%)	18,349 (51.0%)
Female (%)	14,528 (40.4%)
Ethnicity:	
White/Caucasian (%)	32,332 (89.9%)
Black/African American (%)	369 (1.0%)
American Indian/Alaska Native (%)	393 (1.1%)
Asian (%)	691 (1.9%)
Native Hawaiian/Other Pacific Islander (%)	47 (0.1%)
Hispanic/Latino/Spanish Origin (%)	1,692 (4.7%)
Mean AAO (SD)	60.4 (11.0)
Median AAO (IQR)	61.3 (53.6-68.1)
Mean AAE (SD)	65.7 (10.2)
Median AAE (IQR)	66.7 (59.6-72.6)
Mean Current Age (SD)	66.9 (10.2)
Median Current Age (IQR)	68.0 (60.8-73.7)
Mean Disease Duration until Examination (SD)	5.3 (5.6)
Median Disease Duration until Examination (IQR)	3.5 (1.2-7.6)
Mean Disease Duration until Current Age (SD)	6.5 (5.7)
Median Disease Duration until Current Age (IQR)	5.0 (2.5-8.9)

Supplementary text:

Fox Insight study:

The Fox Insight study is an ongoing online, longitudinal health study of people with and without PD with targeted enrollment set to at least 125,000 individuals (Smolensky et al., 2020). The data is a rich data set facilitating discovery, validation, and reproducibility in PD research. The dataset is generated through routine longitudinal assessments (health and medical questionnaires evaluated at regular cycles); one-time health and disease questionnaires about symptoms, daily activities, and other factors; and, in a subgroup of people with PD, genetic data collection. Qualified researchers can explore, analyze, and download patient-reported outcomes (PROs) data and PD-related genetic variants at https://foxden.michaeljfox.org. The full Fox Insight genetic data set, including approximately 650,000 single nucleotide polymorphisms (SNPs) per participant, can be requested separately with institutional review. Fox Insight participants were 18 years of age or older and provided informed consent. In the process of registration, participants were divided into two groups, PD patients and controls, the latter were asked about new diagnoses every three months. PD patients responded to health, non-motor assessments, motor assessments, quality of life, and lifestyle questionnaires. The PD-RFQ-U on “Smoking and Tobacco” questionnaire was used to evaluate smoking, the PD-RFQ-U on “Caffeine” to evaluate coffee drinking and black tea

drinking, and the PD-RFQ-U on “Anti-inflammatory Medication History” for anti-inflammatory drug intake. The survey on “Your Health History” was used to include possible comorbidities in our models. All of these data were self-reported by the patients. For each environmental or lifestyle factor the corresponding datasets were downloaded from the FoxDEN website (https://foxden.michaeljfox.org/insight/explore/fox.jsp) (log:18/10/2020).

Statistical analysis:

For a first statistical analysis, non-parametric Mann-Whitney U test was performed to compare the distribution of AAO between different groups. For correlation analyses, non-parametric Spearman correlations and linear regression analyses were used to assess correlations and interactions between variables (GraphPad Software Inc., San Diego, CA, USA). For a more in-depth analysis, we performed multilinear regression models to investigate the relationship between environmental factors, age, disease duration, motor/non-motor symptoms and potential comorbidities (IBM SPSS Statistics).

Regression model investigating AAO, AAE, environmental factors (binary/dosage/duration):

→ glm(formula = AAO ~ AAE + EnvFactorBinary, family = gaussian, data = data)

→ glm(formula = AAO ~ AAE + EnvFactorDosage, family = gaussian, data = data)

→ glm(formula = AAO ~ AAE + EnvFactorDuration, family = gaussian, data = data)

Regression model investigating AAO, AAE, gender, environmental factors (binary/dosage/duration) and comorbidities:

→ glm(formula = AAO ~ AAE + Gender + EnvFactorBinary (+Comorbidity), family = gaussian, data = data)

→ glm(formula = AAO ~ AAE + Gender + EnvFactorDosage (+Comorbidity), family = gaussian, data = data)

→ glm(formula = AAO ~ AAE + Gender + EnvFactorDuration (+Comorbidity), family = gaussian, data = data)

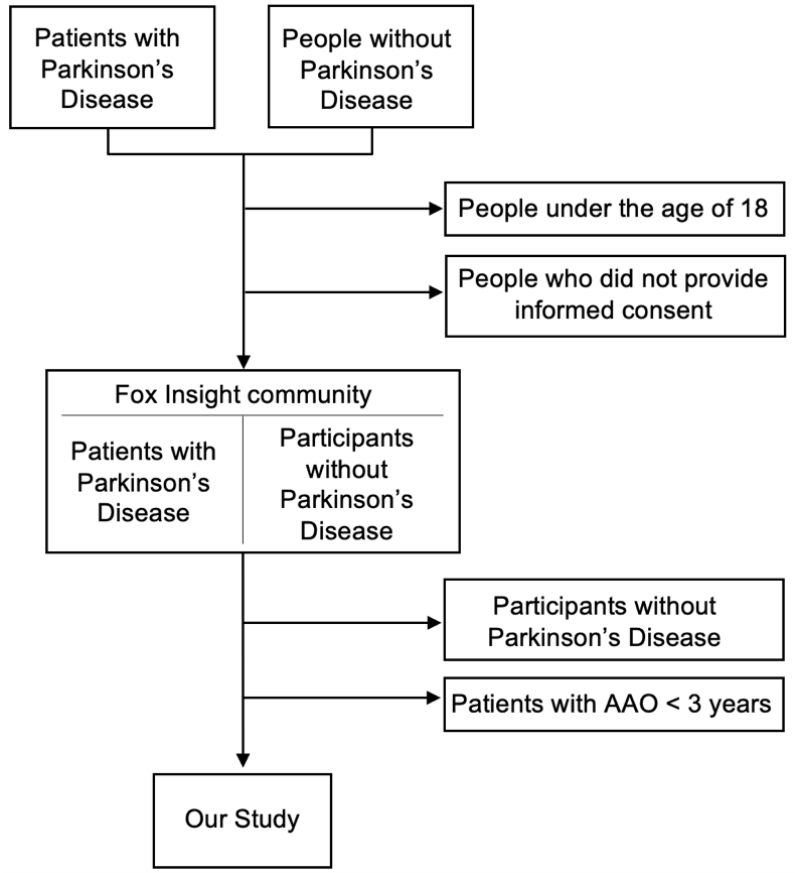
Regression model investigating AAO and combined environmental factors:

→ glm(formula = AAO ~ SmokingBinary + CoffeeBinary + AspirinBinary, family = gaussian, data = data)

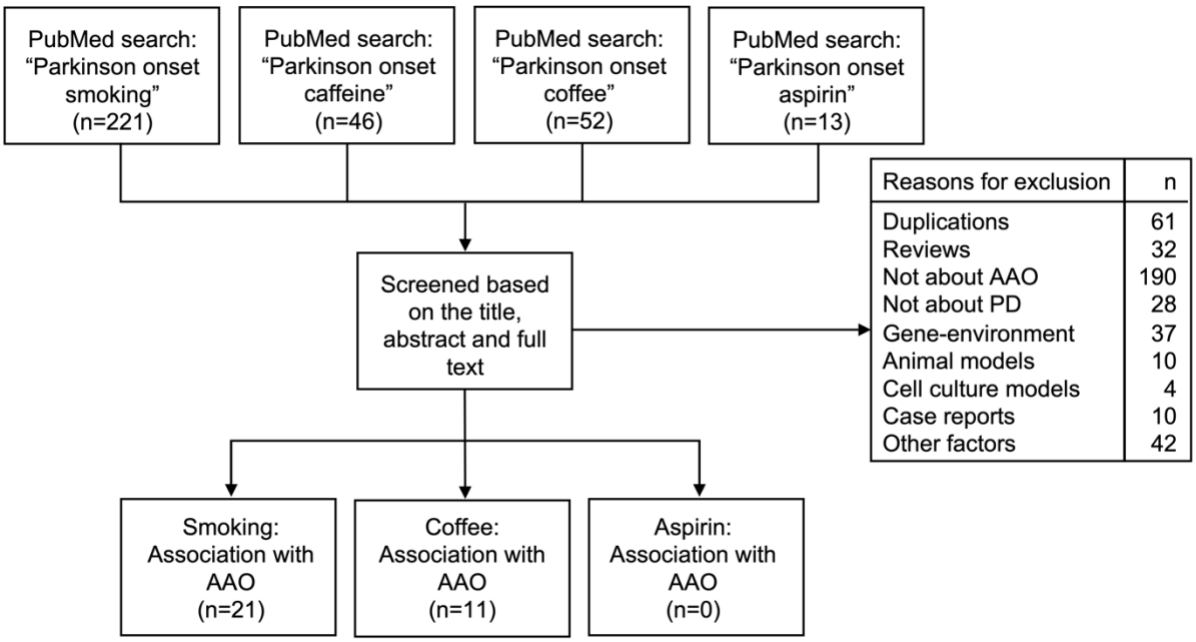
Regression model investigating AAO, AAE, gender, combined environmental factors and comorbidities:

→ glm(formula = AAO ~ AAE + Gender + SmokingBinary + CoffeeBinary + AspirinBinary + BackPain, family = gaussian, data = data)

Exclusion and inclusion criteria:



Supplementary Figure S1: Work flow of the inclusion and exclusion criteria of participant recruitment from the Fox Insight Study and for this study.



Supplementary Figure S2: Work flow of the literature search in PubMed. We searched for literature via PubMed that was published before December 14, 2021. We used the free text search terms "Parkinson onset smoking", resulting in 221 articles, "Parkinson onset caffeine", resulting in 46 articles, "Parkinson onset coffee", resulting in 52 articles and "Parkinson onset aspirin", resulting in 13 articles. These were screened based on the title, abstract and full text, excluding all articles not directly investigating smoking, coffee drinking or aspirin intake and their influence on AAO in PD. Reasons for exclusion and the number of excluded articles are shown (multiple reasons for exclusion were possible). In the end, 21 articles were obtained that described a relationship between AAO and smoking, and 11 articles describing a relationship between AAO and coffee, with an overlap of 7 articles. There was no study that investigated the association between AAO and aspirin. Reasons for exclusion: Duplications: duplicates between the different search terms; Reviews: review articles without new investigations; Not about AAO: articles not examining AAO or articles not investigating the association between AAO and smoking, caffeine/coffee or aspirin; Not about PD: articles about other diseases or about symptoms or parts of PD; Gene-environment: studies investigating gene-environment interactions and the influence of environmetal factors on specific genotypes; Animal models: articles studying animals, not humans; Cell culture models: studies only on cells; Case reports: descriptions of single cases; Other factors: studies investigating other environmental factors than smoking, coffee/caffeine or aspirin in PD or only adjusting for smoking, coffee/caffeine or aspirin in regression models to predict other factors.

Supplementary Table S2: Generalized linear models. Regression models for AAO and environmental factors smoking, coffee drinking and aspirin intake in the Fox Insight cohort

Dependent variable: AAO

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9343	0.0075	$<1\times10^{-5}$
Smoking (binary)	0.5354	0.1424	0.0002

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9277	0.0083	$<1\times10^{-5}$
Smoking Dosage	0.0172	0.0055	0.0016

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9543	0.0168	$<1\times10^{-5}$
Smoking Duration	0.0074	0.0127	0.5583

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9298	0.0077	$<1\times10^{-5}$
Coffee drinking (binary)	0.9176	0.1704	$<1\times10^{-5}$

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9348	0.0086	$<1\times10^{-5}$
Coffee drinking Dosage	0.0309	0.0078	8×10^{-5}

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.8239	0.0122	$<1\times10^{-5}$
Coffee drinking Duration	0.1268	0.0083	$<1\times10^{-5}$

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9272	0.0103	$<1\times10^{-5}$
Aspirin intake (binary)	0.7654	0.1958	9×10^{-5}

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9318	0.0103	$<1\times10^{-5}$
Aspirin intake Dosage	0.0315	0.0190	0.0972

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9195	0.0198	$<1\times10^{-5}$
Aspirin intake Duration	0.0319	0.0133	0.0165

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9334	0.0078	$<1\times10^{-5}$
Gender	0.0234	0.1419	0.8688
Lung Disease	0.0608	0.2028	0.7642
Smoking (binary)	0.5051	0.1456	0.0005

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9254	0.0086	$<1\times10^{-5}$
Gender	0.0292	0.1564	0.8518
Lung Disease	0.0332	0.2239	0.8822
Smoking Dosage	0.0165	0.0055	0.0030

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9627	0.0179	$<1\times10^{-5}$
Gender	0.2611	0.3259	0.4230
Lung Disease	0.0053	0.4432	0.9905
Smoking Duration	0.0074	0.0131	0.5741

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9279	0.0080	$<1\times10^{-5}$
Gender	-0.0178	0.1450	0.9026
Coffee drinking (binary)	0.9379	0.1750	$<1\times10^{-5}$

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9340	0.0089	$<1\times10^{-5}$
Gender	0.1385	0.1635	0.3969
Coffee drinking Dosage	0.0321	0.0081	0.0001

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.8237	0.0125	$<1\times10^{-5}$
Gender	0.2130	0.1900	0.2623
Coffee drinking Duration	0.1276	0.0084	$<1\times10^{-5}$

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9250	0.0108	$<1\times10^{-5}$
Gender	0.0306	0.1860	0.8695
Aspirin intake (binary)	0.6979	0.2063	0.0007
Heart Disease	0.0228	0.2747	0.9340

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9259	0.0109	$<1\times10^{-5}$
Gender	0.0100	0.1879	0.9575
Aspirin intake (binary)	0.6791	0.2016	0.0008
Arthritis	0.0650	0.1911	0.7337

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9282	0.0107	$<1\times10^{-5}$
Gender	0.0252	0.1849	0.8917
Aspirin intake (binary)	0.7082	0.2012	0.0004
Back Pain	-0.4855	0.1830	0.0080

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9264	0.0107	$<1\times10^{-5}$
Gender	0.0108	0.1861	0.9538
Aspirin intake (binary)	0.6732	0.2018	0.0008
Surgeries with Anesthesia	0.1781	0.3219	0.5801

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9274	0.0109	$<1\times10^{-5}$
Gender	0.0504	0.1888	0.7893
Aspirin intake Dosage	0.0263	0.0197	0.1811
Heart Disease	0.2443	0.2827	0.3875

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9292	0.0109	$<1\times10^{-5}$
Gender	0.0082	0.1906	0.9657
Aspirin intake Dosage	0.0267	0.0194	0.1693
Arthritis	0.1215	0.1950	0.5331

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9319	0.0107	$<1\times10^{-5}$
Gender	0.0309	0.1875	0.8693
Aspirin intake Dosage	0.0303	0.0194	0.1188
Back Pain	-0.4144	0.1867	0.0264

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9300	0.0108	$<1\times10^{-5}$
Gender	0.0149	0.1887	0.9372
Aspirin intake Dosage	0.0267	0.0194	0.1703
Surgeries with Anesthesia	0.2440	0.3229	0.4499

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9193	0.0213	$<1\times10^{-5}$
Gender	0.2667	0.3301	0.4191
Aspirin intake Duration	0.0353	0.0140	0.0114
Heart Disease	0.2950	0.3396	0.3850

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9204	0.0212	$<1\times10^{-5}$
Gender	0.1529	0.3344	0.6474
Aspirin intake Duration	0.0338	0.0140	0.0153
Arthritis	0.3836	0.3166	0.2257

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9221	0.0210	$<1\times10^{-5}$
Gender	0.2600	0.3293	0.4297
Aspirin intake Duration	0.0359	0.0140	0.0103
Back Pain	-0.3227	0.3078	0.2945

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9231	0.0210	$<1\times10^{-5}$
Gender	0.2654	0.3294	0.4204
Aspirin intake Duration	0.0362	0.0140	0.0097
Surgeries with Anesthesia	-0.7760	0.7107	0.2748

Covariates	Regression coefficient β	Standard error	*p*-value
Smoking (binary)	1.8261	0.3767	$<1\times10^{-5}$
Coffee drinking (binary)	2.5233	0.4158	$<1\times10^{-5}$
Aspirin intake (binary)	4.8768	0.3698	$<1\times10^{-5}$

Covariates	Regression coefficient β	Standard error	*p*-value
AAE	0.9224	0.0109	$<1\times10^{-5}$
Gender	0.1592	0.1881	0.3975
Smoking (binary)	0.6400	0.2006	0.0014
Coffee drinking (binary)	1.1057	0.2222	$<1\times10^{-5}$
Aspirin intake (binary)	0.7463	0.2041	0.0003
Back Pain	-0.5435	0.1855	0.0034

Supplementary Table S3: Publications on AAO and smoking/coffee of the literature search in PubMed

Authors, Year (PMID)	*n* (PD/control)	*n* (smokers with PD/non-smokers with PD)	*n* (coffee drinkers with PD/coffee non-drinkers with PD)	Effect on AAO
Benedetti et al., 2000 (11087780)	196/196	NA	NA	**Smoking:** - Ever-smokers: median AAO 70 years - Never-smokers: median AAO 71.5 years → Similar AAO **Coffee:** - Coffee-drinkers: median AAO 72 years - Non-drinkers median AAO 64 years → Later AAO
Kandinov et al., 2009 (18434232)	278/0	111/167	180/98	**Smoking:** - Smokers: 1-9 pack-years: mean AAO 57.1 years, ≥10 pack-years: mean AAO 61.2 years - Non-smokers: mean AAO 57.2 years (p=0.7; p=0.04) - A higher amount of cigarettes smoked per day, showed a later AAO → Later AAO (when ≥10 pack-years) **Coffee:** - Coffee-drinkers: <2 daily cups: mean AAO 58 years, 2-3 daily cups: mean AAO 57.6 years, >3 daily cups: mean AAO 55 years - Coffee non-drinkers: mean AAO 59.5 years → Earlier AAO (with dosage effect)
Luth et al., 2020 (32875616)	342/57 (142 with LRRK2 G2019S PD) and 57 mutation carriers)	112 /199 (41 smokers with LRRK2 G2019S PD and 85 non-smokers with LRRK2 G2019S PD)	182 /130 (62 coffee drinkers with LRRK2 G2019S PD and 63 coffee non-drinkers with LRRK2 G2019S PD)	**Smoking:** - LRRK2 PD: Smokers: median AAO 60 years Non-smokers: median AAO 52 years (p=0.0215) - iPD: Smokers: median AAO 55 years, Non-smokers: median 53.5 years (p=0.7906) - Number of cigarettes per day correlated with AAO (p=0.0296) as well as smoking duration (p<0.0001) → Later AAO (in LRRK2 PD) **Coffee:** - LRRK2 PD: Coffee-drinkers: median AAO 55 years Non-drinkers: median AAO 52 years (p=0.5439) - iPD: Coffee-drinkers: median AAO 55 years Non-drinkers: median AAO 52 years (p=0.3279) → Trend to a later AAO, but no significant difference

Maher et al., 2002 (11781409)	396/0	81/337	401/17	**Smoking:** - Among siblings who smoked, pack-years of smoking was related to later age at onset (p=0.0001) → Later AAO **Coffee:** - The mean age at onset did not differ according to exposure to coffee (p=0.79) → Similar AAO
Wijeyekoon et al., 2017 (29057010)	144/102	NA	NA	**Smoking:** - Ever-smoking in males associated with delayed AAO (p=0.048) → Later AAO (in males) **Coffee:** - Regular coffee drinking associated with later AAO (p<0.001) → Later AAO
Wilk and Lash, 2007 (17408493)	NA	NA	NA	**Smoking:** - Current cigarette smoking is a predictor of PD age-at-onset and age at-enrollment in both the full and age-restricted samples. → Later AAO **Coffee:** - Coffee drinking is a predictor of PD age-at-onset and age at-enrollment in both the full and age-restricted samples. → Later AAO
Yahalom et al., 2020 (32310186)	225/0 (65 with LRRK2 G2019S PD and 60 with GBA N370S PD)	98 /127 (26 with LRRK2 G2019S PD and 39 mutation carriers and 27 with GBA N370S PD and 33 mutation carriers)	199 /25 (56 with LRRK2 G2019S PD and 9 mutation carriers and 53 with GBA N370S PD and 6 mutation carriers)	**Smoking:** - Smoking associated with AAO (p=0.032) → Later AAO **Coffee:** - Consumption level of coffee (p=0.001) significantly associated with PD AAO → Later AAO the higher the amount of coffee
De Reuck et al., 2005 (15792818)	512/0	184/328	NA	**Smoking:** - Ever-smokers: mean AAO 65.9 years - Never-smokers: mean AAO 62.4 years (p=0.001) → Later AAO
Gallo et al., 2019 (30462234)	715/213,818	312/404	NA	**Smoking:** - The risk does not vary over the follow-up period, and this argues against a delaying effect of smoking on PD onset
Gigante et al., 2017 (28988683)	262/0	111/151	NA	**Smoking:** - Ever-smokers: mean AAO 61.7 years - Never-smokers: mean AAO 59.3 years (p=0.03) → Later AAO

Grandinetti et al., 1994 (8209872)	58/0	25/33	NA	**Smoking:** - Ever-smokers: mean AAO 69.2 years - Never-smokers: mean AAO 70.2 years (*p*=0.49) → Similar AAO
Greenbaum et al., 2013 (22884254)	677/0	239/438	NA	**Smoking:** - Ever-smokers: mean AAO 54.4 years - Never-smokers: mean AAO 55.8 years (*p*=NS) → Similar AAO
Haack et al., 1981 (7304554)	237/474	87/150	223/14 (coffee or tea)	**Smoking:** - Smokers: mean AAO 52.7 years (men: 54.4 years, women: 42.5) - Non-smokers: mean AAO 57.8 years (men: 58.9 years, women: 55.4) → Earlier AAO
Jimenez-Jimenez et al., 1992 (1484528)	128/256	NA	NA	**Smoking:** - No significant correlation between premorbid levels of exposure to cigarette smoking and AAO in the PD group.
Kuopio et al., 1999 (10584666)	123/246	NA	NA	**Smoking:** - Ever-smokers: AAO 65.0 years - Never-smokers: AAO 61.8 years (*p*=0.051) → Later AAO - Male ever-smokers: AAO 64.8 years - Male never-smokers: AAO 58.3 years (*p*=0.009) → Later AAO
Martinez-Rumayor et al., 2009 (19695769)	247/0	91/156	NA	**Smoking:** - Smokers: mean AAO 55 years - Non-smokers: mean AAO 59 years (*p*=0.08) → Later AAO
Mayeux et al., 1994 (8196685)	150/180	69/81	NA	**Smoking:** - Smokers: AAO 64.2 years - Non-smokers: AAO 64.5 years (p=0.88) → Similar AAO
Neshige et al., 2021 (34130061)	110/110	67/43	NA	**Smoking:** - Ever-smokers: AAO 63 years - Non-smokers: AAO 65 years (*p*=0.28) → Similar AAO
Papapetropoulos et al., 2005 (15747366)	113/0	58/55 (14 smokers and 44 ex-smokers)	NA	**Smoking:** - Smokers: AAO 58.6 years - Ex-smokers: AAO 64.9 years - Non-smokers: AAO 65.5 years (*p*=0.006 and *p*=0.011) → Earlier AAO
Scott et al., 2005 (15699372)	143/168	NA	NA	**Smoking:** - Ever-smokers: mean AAO 55.7 years - Non-smokers: mean AAO 55.4 years - 140 sibships → Similar AAO

Weisskopf et al., 2007 (17266085)	137/466	128/153 (16 current smokers and 112 past smokers)	NA	**Smoking:** - Current smokers: AAO 65.3 years - Past smokers: AAO 67.1 years - Never-smokers: AAO 66.7 years → Similar AAO
Cho et al., 2018 (29449185)	196/0	NA	136/60	**Coffee:** - Coffee drinkers: mean AAO 63.4 years - Non-coffee drinkers: mean AAO 67.3 years (p=0.008) → Earlier AAO
Cho et al., 2019 (31412802)	284/0	NA	204/80 Males: 120/27 Females: 84/53	**Coffee:** - Coffee drinkers: AAO 62.8 years - Non-coffee drinkers: AAO 67.0 years (p=0.001) → Earlier AAO - Male coffee drinkers: AAO 62.6 years - Male non-coffee drinkers: AAO 69.8 years (p=0.001) → Earlier AAO - Female coffee drinkers: AAO 63.2 years - Female non-coffee drinkers: AAO 65.5 years (p=NS) → Similar AAO
Gigante et al., 2018 (29362953)	83/0	12/71	71/12	**Coffee:** - Number of coffee drinking years associated with a significant increase in AAO (p<0.001) → Later AAO (dosage effect)
Tan et al., 2007 (18075470)	418/468	81/337	401/17	**Coffee:** - Significant association between caffeine intake and the onset of PD ($p=2.01\times10^{-5}$) - Dosage effect, showing a later AAO the higher the caffeine consumption → Later AAO (with dosage effect)

References:

Benedetti, M.D., Bower, J.H., Maraganore, D.M., McDonnell, S.K., Peterson, B.J., Ahlskog, J.E., Schaid, D.J., and Rocca, W.A. (2000). Smoking, alcohol, and coffee consumption preceding Parkinson's disease: a case-control study. Neurology *55*, 1350-1358.

Cho, B.H., Choi, S.M., and Kim, B.C. (2019). Gender-dependent effect of coffee consumption on tremor severity in de novo Parkinson's disease. BMC Neurol *19*, 194.

Cho, B.H., Choi, S.M., Kim, J.T., and Kim, B.C. (2018). Association of coffee consumption and non-motor symptoms in drug-naive, early-stage Parkinson's disease. Parkinsonism Relat Disord *50*, 42-47.

De Reuck, J., De Weweire, M., Van Maele, G., and Santens, P. (2005). Comparison of age of onset and development of motor complications between smokers and non-smokers in Parkinson's disease. J Neurol Sci *231*, 35-39.

Gallo, V., Vineis, P., Cancellieri, M., Chiodini, P., Barker, R.A., Brayne, C., Pearce, N., Vermeulen, R., Panico, S., Bueno-de-Mesquita, B.*, et al.* (2019). Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol *48*, 912-925.

Gigante, A.F., Asabella, A.N., Iliceto, G., Martino, T., Ferrari, C., Defazio, G., and Rubini, G. (2018). Chronic coffee consumption and striatal DAT-SPECT findings in Parkinson's disease. Neurol Sci *39*, 551-555.

Gigante, A.F., Martino, T., Iliceto, G., and Defazio, G. (2017). Smoking and age-at-onset of both motor and non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord *45*, 94-96.

Grandinetti, A., Morens, D.M., Reed, D., and MacEachern, D. (1994). Prospective study of cigarette smoking and the risk of developing idiopathic Parkinson's disease. Am J Epidemiol *139*, 1129-1138.

Greenbaum, L., Rigbi, A., Lipshtat, N., Cilia, R., Tesei, S., Asselta, R., Djaldetti, R., Goldwurm, S., and Lerer, B. (2013). Association of nicotine dependence susceptibility gene, CHRNA5, with Parkinson's disease age at onset: gene and smoking status interaction. Parkinsonism Relat Disord *19*, 72-76.

Haack, D.G., Baumann, R.J., McKean, H.E., Jameson, H.D., and Turbek, J.A. (1981). Nicotine exposure and Parkinson disease. Am J Epidemiol *114*, 191-200.

Jimenez-Jimenez, F.J., Mateo, D., and Gimenez-Roldan, S. (1992). Premorbid smoking, alcohol consumption, and coffee drinking habits in Parkinson's disease: a case-control study. Mov Disord *7*, 339-344.

Kandinov, B., Giladi, N., and Korczyn, A.D. (2009). Smoking and tea consumption delay onset of Parkinson's disease. Parkinsonism Relat Disord *15*, 41-46.

Kuopio, A.M., Marttila, R.J., Helenius, H., and Rinne, U.K. (1999). Environmental risk factors in Parkinson's disease. Mov Disord *14*, 928-939.

Luth, T., Konig, I.R., Grunewald, A., Kasten, M., Klein, C., Hentati, F., Farrer, M., and Trinh, J. (2020). Age at Onset of LRRK2 p.Gly2019Ser Is Related to Environmental and Lifestyle Factors. Mov Disord *35*, 1854-1858.

Maher, N.E., Golbe, L.I., Lazzarini, A.M., Mark, M.H., Currie, L.J., Wooten, G.F., Saint-Hilaire, M., Wilk, J.B., Volcjak, J., Maher, J.E.*, et al.* (2002). Epidemiologic study of 203 sibling pairs with Parkinson's disease: the GenePD study. Neurology *58*, 79-84.

Martinez-Rumayor, A., Arrieta, O., Sotelo, J., and Garcia, E. (2009). Female gender but not cigarette smoking delays the onset of Parkinson's disease. Clin Neurol Neurosurg *111*, 738-741.

Mayeux, R., Tang, M.X., Marder, K., Cote, L.J., and Stern, Y. (1994). Smoking and Parkinson's disease. Mov Disord *9*, 207-212.

Neshige, S., Ohshita, T., Neshige, R., and Maruyama, H. (2021). Influence of current and previous smoking on current phenotype in Parkinson's disease. J Neurol Sci *427*, 117534.

Papapetropoulos, S., Singer, C., Villar, J.M., Gonzalez, J., and Mash, D.C. (2005). Does cigarette smoking provide clinically significant neuroprotection among patients diagnosed with Parkinson's disease? Mov Disord *20*, 641-642.

Scott, W.K., Zhang, F., Stajich, J.M., Scott, B.L., Stacy, M.A., and Vance, J.M. (2005). Family-based case-control study of cigarette smoking and Parkinson disease. Neurology *64*, 442-447.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W.*, et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Tan, E.K., Chua, E., Fook-Chong, S.M., Teo, Y.Y., Yuen, Y., Tan, L., and Zhao, Y. (2007). Association between caffeine intake and risk of Parkinson's disease among fast and slow metabolizers. Pharmacogenet Genomics *17*, 1001-1005.

Weisskopf, M.G., Grodstein, F., and Ascherio, A. (2007). Smoking and cognitive function in Parkinson's disease. Mov Disord *22*, 660-665.

Wijeyekoon, R., Suriyakumara, V., Gamage, R., Fernando, T., Jayasuriya, A., Amarasinghe, D., Gunasekara, H., Sirisena, D., Amaratunga, D., Muthukuda, C.*, et al.* (2017). Associations between Lifestyle Factors and Parkinson's Disease in an Urban Sri Lankan Clinic Study. Int Arch Med *10*.

Wilk, J.B., and Lash, T.L. (2007). Risk factor studies of age-at-onset in a sample ascertained for Parkinson disease affected sibling pairs: a cautionary tale. Emerg Themes Epidemiol *4*, 1.

Yahalom, G., Rigbi, A., Israeli-Korn, S., Krohn, L., Rudakou, U., Ruskey, J.A., Benshimol, L., Tsafnat, T., Gan-Or, Z., Hassin-Baer, S.*, et al.* (2020). 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.

SUPPLEMENTARY MATERIAL FOR

Lifestyle factors and clinical severity of Parkinson's disease

Carolin Gabbert[1], MSc, Prof. Inke R. König[2], Theresa Lüth[1], MSc, Meike Kasten[1,3], MD, Fox Insight Study[4], Anne Grünewald[1,5], PhD, Christine Klein[1], MD, Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Lübeck, Germany
[2]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany
[3]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany
[4]The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA
[5]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

*** Correspondence:**
Joanne Trinh
University of Lübeck
Ratzeburger Allee 160
23538 Lübeck, Germany
Email: joanne.trinh@neuro.uni-luebeck.de
Tel.: +49-451-31018202

Supplementary Table S1: Demographics of the Fox Insight participants.

Full study group (n=35,959)	Patients with PD
Male (%)	18,349 (51.0%)
Female (%)	14,528 (40.4%)
Ethnicity:	
White/Caucasian (%)	32,332 (89.9%)
Black/African American (%)	369 (1.0%)
American Indian/Alaska Native (%)	393 (1.1%)
Asian (%)	691 (1.9%)
Native Hawaiian/Other Pacific Islander (%)	47 (0.1%)
Hispanic/Latino/Spanish Origin (%)	1,692 (4.7%)
Mean AAO (SD)	60.4 (11.0)
Median AAO (IQR)	61.3 (53.6-68.1)
Mean AAE (SD)	65.7 (10.2)
Median AAE (IQR)	66.7 (59.6-72.6)
Mean Current Age (SD)	66.9 (10.2)
Median Current Age (IQR)	68.0 (60.8-73.7)
Mean Disease Duration until Examination (SD)	5.3 (5.6)
Median Disease Duration until Examination (IQR)	3.5 (1.2-7.6)
Mean Disease Duration until Current Age (SD)	6.5 (5.7)
Median Disease Duration until Current Age (IQR)	5.0 (2.5-8.9)

Supplementary Text:

Fox Insight study:

The Fox Insight study is an ongoing online, longitudinal health study of people with and without PD with targeted enrollment set to at least 125,000 individuals (Smolensky et al., 2020). The data is a rich data set facilitating discovery, validation, and reproducibility in PD research. The dataset is generated through routine longitudinal assessments (health and medical questionnaires evaluated at regular cycles); one-time health and disease questionnaires about symptoms, daily activities, and other factors; and, in a subgroup of people with PD, genetic data collection. Qualified researchers can explore, analyze, and download patient-reported outcomes (PROs) data and PD-related genetic variants at https://foxden.michaeljfox.org. The full Fox Insight genetic data set, including approximately 650,000 single nucleotide polymorphisms (SNPs) per participant, can be requested separately with institutional review.

Fox Insight participants were 18 years of age or older and provided informed consent. In the process of registration, participants were divided into two groups, PD patients and controls, the latter were asked about new diagnoses every three months. PD patients responded to health, non-motor assessments, motor assessments, quality of life, and lifestyle questionnaires. These questionnaires are based on the Movement Disorders Society – Unified

Parkinson's disease Rating Scale (MDS-UPDRS) Part II, the Non-Motor Symptoms Questionnaire (NMSQ), and the Geriatric Depression Scale (GDS).
The PD-RFQ-U on "Smoking and Tobacco" questionnaire was used to evaluate smoking, the PD-RFQ-U on "Caffeine" to evaluate coffee drinking, and the PD-RFQ-U on "Anti-inflammatory Medication History" for anti-inflammatory drug intake. The surveys on "Your Movement Experiences" (MDS-UPDRS Part II; The scores range from 1 to 5, with higher scores indicating more severe symptoms) and "Your Non-Movement Experiences" (NMSQ; Scores of 0 and 1) were used to assess the association between smoking, coffee drinking, and aspirin intake with motor and non-motor symptoms. Finally, the surveys on "Your Current Health" and "Your Mood" (GDS; Scores of 0 and 1) were used to examine the association between smoking and mood, anxiety, and depression. All of these data were self-reported by the patients.
For each environmental or lifestyle factor, the corresponding datasets were downloaded from the FoxDEN website (https://foxden.michaeljfox.org/insight/explore/fox.jsp) (log:18/10/2020).

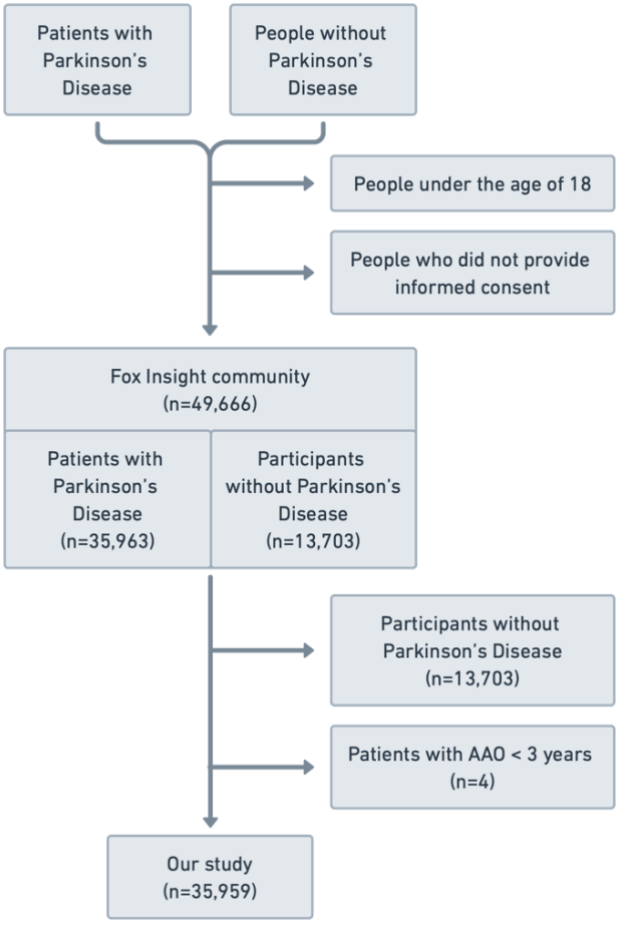
Statistical analysis:
We performed multilinear regression models to investigate the relationship between environmental factors, age, disease duration, and motor/non-motor symptoms (R studio).

Regression model investigating environmental factors and motor/non-motor symptoms adjusted for AAE, gender, disease duration, and potential comorbidities (heart diseases and lung diseases for smoking, and heart diseases, arthritis, back pain, and surgeries with anesthesia for aspirin):

→ glm(formula = MotorSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + EnvFactorYes, family = binomial, data = data)
→ glm(formula = MotorSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + EnvFactorDosage, family = binomial, data = data)
→ glm(formula = MotorSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + EnvFactorDuration, family = binomial, data = data)
→ glm(formula = NonMotorSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + EnvFactorYes, family = binomial, data = data)
→ glm(formula = NonMotorSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + EnvFactorDosage, family = binomial, data = data)
→ glm(formula = NonMotorSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + EnvFactorDuration, family = binomial, data = data)
→ glm(formula = MoodSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + SmokingYes, family = binomial, data = data)

→ glm(formula = MoodSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + SmokingDosage, family = binomial, data = data)

→ glm(formula = MoodSymptomYes ~ AAE + Gender + DiseaseDuration (+ Comorbidity) + SmokingDuration, family = binomial, data = data)



Supplementary Figure S1: Workflow of the inclusion and exclusion criteria of participant recruitment from the Fox Insight study and for this study.

Supplementary Table S2: Clinical variables.

Your Movement Experiences	Your Non-Movement Experiences	Your Current Health	Your Mood
Tremor: Over the past week, have you usually had shaking or tremor?	Have you experienced constipation (less than three bowel movements a week) or having to strain to pass a stool in the last month?	Do you currently have depression?	Have you dropped many of your activities and interests?
Speech: Over the past week, have you had problems with your speech?	Have you experienced unexplained pains (not due to known conditions such as arthritis) in the last month?	Have you had anxiety?	Do you feel that your life is empty?
Saliva and Drooling: Over the past week, have you usually had too much saliva during when you are awake or when you sleep?	Have you experienced problems remembering things that have happened recently or forgetting to do things in the last month?	/	Do you often get bored?
Chewing and Swallowing: Over the past week, have you usually had problems swallowing pills or eating meals? Do you need your pills cut or crushed or your meals to be made soft, chopped or blended to avoid choking?	Have you experienced feeling sad, 'low' or 'blue' in the last month?	/	Are you afraid that something bad is going to happen to you?
Walking and Balance: Over the past week, have you usually had problems with balance and walking?	Have you experienced feeling anxious, frightened or panicky in the last month?	/	Do you often feel helpless?
Freezing: Over the past week, on your usual day when walking, do you suddenly stop or freeze as if your feet are stuck to the floor?	Have you experienced feeling less interested in sex or more interested in sex in the last month?	/	Do you prefer to stay at home, rather than going out and doing new things?
Getting out of bed, a care, or a deep chair: Over the past week, have you usually had trouble getting out of a bed, a car seat, or a deep chair?	Have you experienced feeling light-headed, dizzy or weak standing from sitting or lying in the last month?	/	Do you feel you have more problems with memory than most people?
/	/	/	Do you feel pretty worthless the way you are now?
/	/	/	Do you feel that your situation is hopeless?

Motor and non-motor symptoms that were used from the surveys “Your Movement Experiences”, “Your Non-Movement Experiences”, “Your Current Health”, and “Your Mood”.

Supplementary Table S3: Generalized linear models on motor symptoms. Regression models for motor symptoms associated with coffee drinking, aspirin intake, and smoking in the Fox Insight cohort.

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	0.0036	0.0034	0.2984
Gender	-0.0193	0.0612	0.7521
Disease Duration	0.0735	0.0063	$<1\times10^{-5}$
Coffee drinking (binary)	-0.1435	0.0731	0.0497

Dependent variable: Tremor

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0257	0.0057	$<1\times10^{-5}$
Gender	-0.2811	0.0947	0.0030
Disease Duration	-0.0307	0.0087	0.0004
Aspirin intake (binary)	0.3174	0.1054	0.0026

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0044	0.0047	0.3435
Gender	-0.0027	0.0806	0.9733
Disease Duration	0.0722	0.0084	$<1\times10^{-5}$
Aspirin intake (binary)	0.1837	0.0875	0.0358

Dependent variable: Getting Up

Covariates	Estimate	Standard error	*p*-value
AAE	0.0215	0.0048	$<1\times10^{-5}$
Gender	0.0432	0.0832	0.6034
Disease Duration	0.1040	0.0109	$<1\times10^{-5}$
Aspirin intake (binary)	0.2170	0.0922	0.0185

Dependent variable: Tremor

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0249	0.0057	2×10^{-5}
Gender	-0.3021	0.0966	0.0018
Disease Duration	-0.0329	0.0089	0.0002
Aspirin intake Dosage	0.0287	0.0117	0.0138

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0042	0.0048	0.3729
Gender	-0.0326	0.0824	0.6927
Disease Duration	0.0736	0.0087	$<1\times10^{-5}$
Aspirin intake Dosage	0.0201	0.0085	0.0182

Dependent variable: Walking and Balance

Covariates	Estimate	Standard error	*p*-value
AAE	0.0024	0.0048	0.6178
Gender	-0.0070	0.0846	0.9339
Disease Duration	0.1141	0.0114	$<1\times10^{-5}$
Aspirin intake Dosage	0.0253	0.0099	0.0106

Dependent variable: Getting Up

Covariates	Estimate	Standard error	*p*-value
AAE	0.0218	0.0048	$<1\times10^{-5}$
Gender	0.0565	0.0845	0.5039
Disease Duration	0.1038	0.0111	$<1\times10^{-5}$
Aspirin intake Dosage	0.0231	0.0098	0.0182

Dependent variable: Saliva and Drooling

Covariates	Estimate	Standard error	*p*-value
AAE	0.0114	0.0031	0.0003
Gender	-0.5069	0.0563	$<1\times10^{-5}$
Disease Duration	0.0518	0.0059	$<1\times10^{-5}$
Smoking (binary)	0.1484	0.0580	0.0106

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	0.0004	0.0033	0.9042
Gender	-0.0550	0.0588	0.3495
Disease Duration	0.0706	0.0059	$<1\times10^{-5}$
Smoking (binary)	0.2243	0.0603	0.0002

Dependent variable: Freezing

Covariates	Estimate	Standard error	p-value
AAE	0.0006	0.0035	0.8747
Gender	-0.0676	0.0630	0.2829
Disease Duration	0.1107	0.0064	$<1\times10^{-5}$
Smoking (binary)	0.1490	0.0646	0.0212

Dependent variable: Speech

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0017	0.0036	0.6296
Gender	-0.5920	0.0649	$<1\times10^{-5}$
Disease Duration	0.0931	0.0076	$<1\times10^{-5}$
Smoking Dosage	0.0089	0.0033	0.0062

Dependent variable: Saliva and Drooling

Covariates	Estimate	Standard error	*p*-value
AAE	0.0092	0.0035	0.0085
Gender	-0.5331	0.0632	$<1\times10^{-5}$
Disease Duration	0.0490	0.0066	$<1\times10^{-5}$
Smoking Dosage	0.0096	0.0031	0.0022

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	0.0007	0.0037	0.8560
Gender	0.0369	0.0663	0.5778
Disease Duration	0.0721	0.0067	$<1\times10^{-5}$
Smoking Dosage	0.0174	0.0031	$<1\times10^{-5}$

Dependent variable: Walking and Balance

Covariates	Estimate	Standard error	*p*-value
AAE	0.0070	0.0037	0.0608
Gender	0.0507	0.0683	0.4577
Disease Duration	0.1015	0.0087	$<1\times10^{-5}$
Smoking Dosage	0.0101	0.0035	0.0038

Dependent variable: Freezing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0012	0.0040	0.7601
Gender	-0.0413	0.0714	0.5632
Disease Duration	0.1137	0.0073	$<1\times10^{-5}$
Smoking Dosage	0.0094	0.0034	0.0052

Dependent variable: Getting Up

Covariates	Estimate	Standard error	*p*-value
AAE	0.0240	0.0038	$<1\times10^{-5}$
Gender	0.0747	0.0690	0.2790
Disease Duration	0.0898	0.0086	$<1\times10^{-5}$
Smoking Dosage	0.0098	0.0036	0.0061

Dependent variable: Speech

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0068	0.0080	0.3913
Gender	-0.7464	0.1449	$<1\times10^{-5}$
Disease Duration	0.0847	0.0186	$<1\times10^{-5}$
Smoking Duration	0.0119	0.0059	0.0454

Dependent variable: Walking and Balance

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0018	0.0083	0.8289
Gender	0.1699	0.1534	0.2682
Disease Duration	0.0981	0.0209	$<1\times10^{-5}$
Smoking Duration	0.0268	0.0065	4×10^{-5}

Dependent variable: Freezing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0150	0.0089	0.0918
Gender	-0.1832	0.1603	0.2531
Disease Duration	0.0981	0.0168	$<1\times10^{-5}$
Smoking Duration	0.0277	0.0062	$<1\times10^{-5}$

Dependent variable: Getting Up

Covariates	Estimate	Standard error	*p*-value
AAE	0.0185	0.0084	0.0268
Gender	0.0647	0.1569	0.6801
Disease Duration	0.0851	0.0208	4×10^{-5}
Smoking Duration	0.0366	0.0070	$<1\times10^{-5}$

Supplementary Table S4: Non-motor symptoms associated with environmental factors. Percentage of patients stratified by coffee consumption, aspirin intake, and smoking status and non-motor symptoms.

		Coffee			**Aspirin**			**Smoking**		
		yes	no	*p*-value	yes	no	*p*-value	yes	no	*p*-value
Constipation	yes	53.5% (*n*=2127)	57.0% (*n*=643)	0.0970	55.2% (*n*=553)	51.6% (*n*=1027)	**0.0124**	55.1% (*n*=1174)	53.6% (*n*=1801)	0.1994
	no	46.5% (*n*=1847)	43.0% (*n*=486)		44.8% (*n*=449)	48.4% (*n*=962)		44.9% (*n*=957)	46.4% (*n*=1559)	
Unexplained Pains	yes	37.2% (*n*=1480)	37.5% (*n*=423)	0.1623	37.8% (*n*=379)	38.5% (*n*=765)	**0.0227**	40.0% (*n*=852)	35.7% (*n*=1199)	**$<1\times10^{-5}$**
	no	62.8% (*n*=2494)	62.5% (*n*=706)		62.2% (*n*=623)	61.5% (*n*=1224)		60.0% (*n*=1278)	64.3% (*n*=2161)	
Problems Remembering	yes	47.0% (*n*=1869)	46.8% (*n*=528)	0.9859	50.1% (*n*=502)	41.9% (*n*=834)	**1×10^{-5}**	50.6% (*n*=1078)	44.9% (*n*=1510)	**0.0001**
	no	53.0% (*n*=2105)	53.2% (*n*=601)		49.9% (*n*=500)	58.1% (*n*=1155)		49.4% (*n*=1052)	55.1% (*n*=1850)	
Feeling Sad	yes	50.2% (*n*=1993)	50.3% (*n*=567)	0.2517	47.9% (*n*=480)	50.1% (*n*=995)	0.0665	53.9% (*n*=1148)	48.0% (*n*=1611)	**$<1\times10^{-5}$**
	no	49.8% (*n*=1980)	49.7% (*n*=560)		52.1% (*n*=522)	49.9% (*n*=993)		46.1% (*n*=982)	52.0% (*n*=1746)	
Anxiety	yes	36.2% (*n*=1438)	35.9% (*n*=405)	0.2199	32.9% (*n*=330)	37.2% (*n*=740)	0.2999	38.7% (*n*=825)	34.2% (*n*=1149)	**$<1\times10^{-5}$**
	no	63.8% (*n*=2535)	64.1% (*n*=722)		67.1% (*n*=672)	62.8% (*n*=1248)		61.3% (*n*=1305)	65.8% (*n*=2208)	
Changed Interest in Sex	yes	34.6% (*n*=1376)	32.3% (*n*=364)	0.1032	34.9% (*n*=350)	32.6% (*n*=649)	**0.0221**	35.7% (*n*=761)	32.4% (*n*=1089)	**0.0013**
	no	65.4% (*n*=2597)	67.7% (*n*=763)		65.1% (*n*=652)	67.4% (*n*=1339)		64.3% (*n*=1369)	67.6% (*n*=2268)	
Light-headedness	yes	43.4% (*n*=1723)	41.7% (*n*=470)	0.2808	45.1% (*n*=452)	40.7% (*n*=810)	**0.0043**	45.7% (*n*=974)	41.4% (*n*=1390)	**0.0005**
	no	56.6% (*n*=2250)	58.3% (*n*=656)		54.9% (*n*=550)	59.3% (*n*=1178)		54.3% (*n*=1156)	58.6% (*n*=1966)	

P-value (exploratory): Multivariate regression to predict the respective non-motor symptoms adjusted for covariates by including AAE, gender, and disease duration (time between AAO and current age).

Supplementary Table S5: Generalized linear models on non-motor symptoms. Regression models for non-motor symptoms associated with coffee drinking, aspirin intake, and smoking in the Fox Insight cohort.

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0305	0.0038	$<1\times10^{-5}$
Gender	0.4508	0.0687	$<1\times10^{-5}$
Disease Duration	0.0118	0.0066	0.0749
Coffee drinking Dosage	0.0083	0.0035	0.0168

Dependent variable: Constipation

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0003	0.0044	0.9542
Gender	0.0580	0.0759	0.4447
Disease Duration	0.0315	0.0079	7×10^{-5}
Aspirin intake (binary)	0.2077	0.0831	0.0124

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0302	0.0046	$<1\times10^{-5}$
Gender	0.3524	0.0786	$<1\times10^{-5}$
Disease Duration	0.0181	0.0077	0.0188
Aspirin intake (binary)	0.1961	0.0861	0.0227

Dependent variable: Problems Remembering

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0065	0.0044	0.1399
Gender	-0.1202	0.0763	0.1150
Disease Duration	0.0129	0.0076	0.0894
Aspirin intake (binary)	0.3662	0.0830	1×10^{-5}

Dependent variable: Changed Interest in Sex

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0338	0.0047	$<1\times10^{-5}$
Gender	-0.5160	0.0817	$<1\times10^{-5}$
Disease Duration	0.0187	0.0079	0.0179
Aspirin intake (binary)	0.2023	0.0884	0.0221

Dependent variable: Light-headedness

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0117	0.0044	0.0085
Gender	0.0299	0.0766	0.6960
Disease Duration	0.0013	0.0076	0.8615
Aspirin intake (binary)	0.2380	0.0833	0.0043

Dependent variable: Constipation

Covariates	Estimate	Standard error	*p*-value
AAE	0.0021	0.0044	0.6421
Gender	0.0674	0.0775	0.3846
Disease Duration	0.0316	0.0082	0.0001
Aspirin intake Dosage	0.0251	0.0086	0.0037

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0298	0.0046	$<1\times10^{-5}$
Gender	0.3633	0.0804	$<1\times10^{-5}$
Disease Duration	0.0191	0.0080	0.0163
Aspirin intake Dosage	0.0178	0.0083	0.0320

Dependent variable: Problems Remembering

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0054	0.0045	0.2262
Gender	-0.1140	0.0779	0.1432
Disease Duration	0.0133	0.0078	0.0887
Aspirin intake Dosage	0.0295	0.0085	0.0005

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0354	0.0046	$<1\times10^{-5}$
Gender	0.3975	0.0783	$<1\times10^{-5}$
Disease Duration	0.0168	0.0079	0.0331
Aspirin intake Dosage	0.0177	0.0083	0.0344

Dependent variable: Changed Interest in Sex

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0321	0.0048	$<1\times10^{-5}$
Gender	-0.5400	0.0836	$<1\times10^{-5}$
Disease Duration	0.0171	0.0082	0.0358
Aspirin intake Dosage	0.0184	0.0084	0.0278

Dependent variable: Light-headedness

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0108	0.0045	0.0157
Gender	0.0416	0.0782	0.5949
Disease Duration	0.0035	0.0078	0.6538
Aspirin intake Dosage	0.0226	0.0082	0.0060

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0320	0.0032	$<1\times10^{-5}$
Gender	0.3859	0.0579	$<1\times10^{-5}$
Disease Duration	0.0149	0.0055	0.0069
Smoking (binary)	0.2732	0.0595	$<1\times10^{-5}$

Dependent variable: Problems Remembering

Covariates	Estimate	Standard error	*p*-value
AAE	0.0018	0.0031	0.5498
Gender	-0.2018	0.0555	0.0003
Disease Duration	0.0155	0.0054	0.0039
Smoking (binary)	0.2176	0.0570	0.0001

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0299	0.0032	$<1\times10^{-5}$
Gender	0.3498	0.0560	$<1\times10^{-5}$
Disease Duration	0.0139	0.0054	0.0100
Smoking (binary)	0.3279	0.0579	$<1\times10^{-5}$

Dependent variable: Anxiety

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0395	0.0033	$<1\times10^{-5}$
Gender	0.3907	0.0587	$<1\times10^{-5}$
Disease Duration	-0.0005	0.0057	0.9244
Smoking (binary)	0.3007	0.0604	$<1\times10^{-5}$

Dependent variable: Changed Interest in Sex

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0335	0.0033	$<1\times10^{-5}$
Gender	-0.5699	0.0600	$<1\times10^{-5}$
Disease Duration	0.0175	0.0056	0.0019
Smoking (binary)	0.1959	0.0610	0.0013

Dependent variable: Light-headedness

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0124	0.0031	6×10^{-5}
Gender	0.0706	0.0558	0.2062
Disease Duration	0.0004	0.0054	0.9412
Smoking (binary)	0.2000	0.0574	0.0005

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0320	0.0036	$<1\times10^{-5}$
Gender	0.4477	0.0653	$<1\times10^{-5}$
Disease Duration	0.0136	0.0063	0.0306
Smoking Dosage	0.0114	0.0031	0.0003

Dependent variable: Problems Remembering

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0002	0.0034	0.9481
Gender	-0.0999	0.0624	0.1091
Disease Duration	0.0189	0.0061	0.0020
Smoking Dosage	0.0123	0.0030	6×10^{-5}

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0291	0.0035	$<1\times10^{-5}$
Gender	0.3590	0.0629	$<1\times10^{-5}$
Disease Duration	0.0134	0.0061	0.0282
Smoking Dosage	0.0152	0.0031	$<1\times10^{-5}$

Dependent variable: Anxiety

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0377	0.0037	$<1\times10^{-5}$
Gender	0.3896	0.0662	$<1\times10^{-5}$
Disease Duration	-0.0008	0.0065	0.9022
Smoking Dosage	0.0161	0.0031	$<1\times10^{-5}$

Dependent variable: Light-headedness

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0115	0.0035	0.0009
Gender	0.1096	0.0629	0.0812
Disease Duration	-0.0002	0.0061	0.9740
Smoking Dosage	0.0117	0.0030	0.0001

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	p-value
AAE	-0.0482	0.0081	$<1\times10^{-5}$
Gender	0.2455	0.1420	0.0838
Disease Duration	0.0167	0.0144	0.2476
Smoking Duration	0.0140	0.0057	0.0134

Dependent variable: Problems Remembering

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0048	0.0076	0.5312
Gender	-0.4815	0.1392	0.0005
Disease Duration	-0.0150	0.0143	0.2951
Smoking Duration	0.0176	0.0056	0.0017

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0460	0.0082	$<1\times10^{-5}$
Gender	0.2244	0.1407	0.1107
Disease Duration	0.0024	0.0143	0.8691
Smoking Duration	0.0112	0.0056	0.0470

Dependent variable: Changed Interest in Sex

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0451	0.0082	$<1\times10^{-5}$
Gender	-0.7188	0.1524	$<1\times10^{-5}$
Disease Duration	0.0249	0.0147	0.0901
Smoking Duration	0.0123	0.0059	0.0372

Supplementary Table S6: Motor symptoms associated with aspirin intake in regression models including potential comorbidities.

		Heart Diseases			**Arthritis**			**Back Pain**			**Surgeries with Anesthesia**		
		Aspirin											
		Yes/No	Dosage	Duration	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Tremor	*n*	2854	2718	546	2849	2713	545	2850	2714	546	2852	2716	546
	p-value	**0.0040**	**0.0212**	0.5592	**0.0031**	**0.0163**	0.5430	**0.0045**	**0.0210**	0.5687	**0.0029**	**0.0149**	0.5788
	β	0.3098	0.0270	0.0063	0.3123	0.0281	0.0066	0.3007	0.0268	0.0061	0.3149	0.0284	0.0060
Speech	*n*	2856	2720	546	2851	2715	545	2852	2716	546	2854	2718	546
	p-value	0.4252	0.1185	0.3991	0.4522	0.1143	0.3428	0.4883	0.1146	0.3807	0.4293	0.1144	0.2840
	β	-0.0704	0.0141	-0.0069	-0.0653	0.0141	-0.0078	-0.0601	0.0141	-0.0072	-0.0687	0.0141	-0.0088
Saliva and Drooling	*n*	2856	2720	546	2851	2715	545	2852	2716	546	2854	2718	546
	p-value	0.3411	0.4267	0.5394	0.2549	0.3234	0.5455	0.2677	0.3498	0.5707	0.2371	0.2968	0.5351
	β	0.0816	0.0066	0.0049	0.0959	0.0082	0.0049	0.0933	0.0077	0.0046	0.0996	0.0087	0.0050
Chewing and Swallowing	*n*	2856	2720	546	2851	2715	545	2852	2716	546	2854	2718	546
	p-value	0.1120	**0.0374**	0.9540	0.0556	**0.0287**	0.8214	0.0591	**0.0331**	0.8354	0.0591	**0.0256**	0.8129
	β	0.1425	0.0180	-0.0005	0.1684	0.0188	-0.0019	0.1661	0.0182	-0.0018	0.1661	0.0191	-0.0020
Walking and Balance	*n*	2854	2718	546	2849	2713	545	2850	2714	546	2852	2716	546
	p-value	0.3443	**0.0330**	0.8998	0.1785	**0.0211**	0.7816	0.1729	**0.0280**	0.7627	0.1074	**0.0086**	0.7662
	β	0.0883	0.0212	-0.0011	0.1246	0.0233	-0.0024	0.1264	0.0223	-0.0027	0.1477	0.0263	-0.0026
Freezing	*n*	2854	2718	546	2849	2713	545	2850	2714	546	2852	2716	546
	p-value	0.3930	0.1566	0.8882	0.4965	0.2760	0.9973	0.3731	0.2004	0.9827	0.3905	0.1797	0.9949
	β	0.0836	0.0132	0.0013	0.0657	0.0101	-3×10^{-5}	0.0857	0.0118	0.0002	0.0826	0.0124	-6×10^{-5}
Getting up	*n*	2854	2718	546	2849	2713	545	2850	2714	546	2852	2716	546
	p-value	0.0534	**0.0308**	0.3426	**0.0391**	**0.0367**	0.3113	**0.0375**	0.0509	0.2534	**0.0252**	**0.0188**	0.2057
	β	0.1819	0.0214	-0.0081	0.1924	0.0209	-0.0087	0.1950	0.0196	-0.0099	0.2074	0.0233	-0.0110

P-value (exploratory): Multivariate regression to predict the respective motor symptoms adjusted for covariates by including AAE, gender, disease duration (time between AAO and current age), and comorbidities in the model.

Supplementary Table S7: Generalized linear models on motor symptoms including potential comorbidities. Regression models for motor symptoms with a change in outcome for aspirin intake and smoking in the Fox Insight cohort while adjusting for potential comorbidities.

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0055	0.0048	0.2485
Gender	0.0084	0.0811	0.9178
Disease Duration	0.0729	0.0085	$<1\times10^{-5}$
Heart Diseases (binary)	0.2666	0.1173	0.0231
Aspirin intake (binary)	0.1425	0.0897	0.1120

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0061	0.0048	0.2068
Gender	-0.0500	0.0823	0.5437
Disease Duration	0.0724	0.0085	$<1\times10^{-5}$
Arthritis (binary)	0.2401	0.0833	0.0040
Aspirin intake (binary)	0.1684	0.0880	0.0556

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0049	0.0047	0.3007
Gender	-0.0042	0.0811	0.9582
Disease Duration	0.0708	0.0085	$<1\times10^{-5}$
Back Pain (binary)	0.3360	0.0805	3×10^{-5}
Aspirin intake (binary)	0.1661	0.0880	0.0591

Dependent variable: Chewing and Swallowing

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0055	0.0048	0.2487
Gender	-0.0278	0.0814	0.7328
Disease Duration	0.0725	0.0084	$<1\times10^{-5}$
Surgeries with Anesthesia (binary)	0.4380	0.1503	0.0036
Aspirin intake (binary)	0.1661	0.0880	0.0591

Dependent variable: Getting Up

Covariates	Estimate	Standard error	*p*-value
AAE	0.0202	0.0048	3×10^{-5}
Gender	0.0546	0.0835	0.5289
Disease Duration	0.1039	0.0109	$<1\times10^{-5}$
Heart Diseases (binary)	0.2775	0.1300	0.0328
Aspirin intake (binary)	0.1819	0.0941	0.0534

Dependent variable: Getting Up

Covariates	Estimate	Standard error	*p*-value
AAE	0.0202	0.0049	4×10^{-5}
Gender	0.0397	0.0858	0.6433
Disease Duration	0.1006	0.0112	$<1\times10^{-5}$
Back Pain (binary)	0.7012	0.0858	$<1\times10^{-5}$
Aspirin intake Dosage	0.0196	0.0100	0.0509

Dependent variable: Speech

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0079	0.0082	0.3311
Gender	-0.7369	0.1458	$<1\times10^{-5}$
Disease Duration	0.0841	0.0186	$<1\times10^{-5}$
Heart Diseases (binary)	0.0752	0.2002	0.7074
Smoking Duration	0.0115	0.0059	0.0526

Dependent variable: Speech

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0074	0.0080	0.3550
Gender	-0.7417	0.1451	$<1\times10^{-5}$
Disease Duration	0.0839	0.0186	$<1\times10^{-5}$
Lung Diseases (binary)	0.1105	0.2011	0.5828
Smoking Duration	0.0113	0.0060	0.0591

Supplementary Table S8: Non-motor symptoms associated with aspirin intake in regression models including potential comorbidities.

		Heart Diseases			**Arthritis**			**Back Pain**			**Surgeries with Anesthesia**		
		Aspirin											
		Yes/No	Dosage	Duration	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Constipation	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	**0.0264**	**0.0055**	0.6126	**0.0136**	**0.0040**	0.6505	**0.0200**	**0.0083**	0.7533	**0.0144**	**0.0036**	0.6749
	β	0.1888	0.0244	0.0041	0.2061	0.0252	0.0036	0.1954	0.0231	0.0026	0.2044	0.0254	0.0034
Unexplained Pains	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	**0.0220**	**0.0338**	0.7971	**0.0383**	0.0672	0.7426	**0.0414**	0.0896	0.6581	**0.0275**	**0.0419**	0.7705
	β	0.2020	0.0179	-0.0022	0.1799	0.0154	-0.0028	0.1788	0.0142	-0.0038	0.1909	0.0170	-0.0024
Problems Remembering	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	**0.0003**	**0.0033**	0.6288	**2×10^{-5}**	**0.0012**	0.7542	**2×10^{-5}**	**0.0014**	0.7882	**3×10^{-5}**	**0.0011**	0.7528
	β	0.3080	0.0249	0.0038	0.3545	0.0276	0.0025	0.3554	0.0271	0.0021	0.3481	0.0278	0.0025
Feeling Sad	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	0.2404	0.0806	0.3990	0.1073	0.0513	0.4817	0.0946	0.0527	0.4985	0.0807	**0.0350**	0.4154
	β	0.1009	0.0147	0.0068	0.1360	0.0164	0.0057	0.1414	0.0163	0.0055	0.1472	0.0177	0.0066
Anxiety	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	0.5224	0.2396	0.9046	0.3557	0.1719	0.8528	0.3786	0.2036	0.8032	0.2894	0.1349	0.8282
	β	0.0582	0.0100	-0.0010	0.0824	0.0115	-0.0016	0.0787	0.0107	-0.0021	0.0946	0.0126	-0.0019
Changed Interest in Sex	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	**0.0379**	**0.0309**	0.7100	**0.0178**	**0.0185**	0.6927	**0.0266**	**0.0310**	0.7401	**0.0210**	**0.0223**	0.6549
	β	0.1882	0.0184	0.0032	0.2105	0.0199	0.0034	0.1968	0.0182	0.0028	0.2051	0.0193	0.0038
Light-headedness	*n*	2865	2727	547	2860	2722	546	2861	2723	547	2863	2725	547
	p-value	**0.0396**	**0.0296**	0.5597	**0.0083**	**0.0130**	0.6834	**0.0092**	**0.0161**	0.6425	**0.0084**	**0.0108**	0.6331
	β	0.1761	0.0180	0.0046	0.2218	0.0206	0.0033	0.2193	0.0198	0.0037	0.2211	0.0210	0.0038

P-value (exploratory): Multivariate regression to predict the respective non-motor symptoms adjusted for covariates by including AAE, gender, disease duration (time between AAO and current age), and comorbidities in the model.

Supplementary Table S9: Generalized linear models on non-motor symptoms including potential comorbidities. Regression models for non-motor symptoms with a change in outcome for aspirin intake and smoking in the Fox Insight cohort while adjusting for potential comorbidities.

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0343	0.0048	$<1\times10^{-5}$
Gender	0.3041	0.0822	0.0002
Disease Duration	0.0195	0.0081	0.0156
Arthritis (binary)	0.3817	0.0842	$<1\times10^{-5}$
Aspirin intake Dosage	0.0154	0.0084	0.0672

Dependent variable: Unexplained Pains

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0332	0.0048	$<1\times10^{-5}$
Gender	0.3706	0.0820	$<1\times10^{-5}$
Disease Duration	0.0159	0.0082	0.0516
Back Pain (binary)	0.7359	0.0825	$<1\times10^{-5}$
Aspirin intake Dosage	0.0142	0.0084	0.0896

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0385	0.0047	$<1\times10^{-5}$
Gender	0.4128	0.0790	$<1\times10^{-5}$
Disease Duration	0.0170	0.0079	0.0327
Heart Diseases (binary)	0.3383	0.1187	0.0044
Aspirin intake Dosage	0.0147	0.0084	0.0806

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0399	0.0047	$<1\times10^{-5}$
Gender	0.3368	0.0799	3×10^{-5}
Disease Duration	0.0161	0.0080	0.0430
Arthritis (binary)	0.3183	0.0821	0.0001
Aspirin intake Dosage	0.0164	0.0084	0.0513

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0382	0.0047	$<1\times10^{-5}$
Gender	0.3930	0.0790	$<1\times10^{-5}$
Disease Duration	0.0133	0.0080	0.0955
Back Pain (binary)	0.4246	0.0790	$<1\times10^{-5}$
Aspirin intake Dosage	0.0163	0.0084	0.0527

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0497	0.0085	$<1\times10^{-5}$
Gender	0.2623	0.1422	0.0650
Disease Duration	0.0015	0.0143	0.9170
Heart Diseases (binary)	0.4377	0.1920	0.0226
Smoking Duration	0.0099	0.0057	0.0807

Dependent variable: Feeling Sad

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0464	0.0083	$<1\times10^{-5}$
Gender	0.2334	0.1412	0.0985
Disease Duration	0.0016	0.0144	0.9112
Lung Diseases (binary)	0.4308	0.1954	0.0275
Smoking Duration	0.0091	0.0057	0.1103

Dependent variable: Changed Interest in Sex

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0487	0.0085	$<1\times10^{-5}$
Gender	-0.6888	0.1536	$<1\times10^{-5}$
Disease Duration	0.0239	0.0148	0.1064
Heart Diseases (binary)	0.3482	0.1986	0.0795
Smoking Duration	0.0113	0.0059	0.0574

Supplementary Table S10: Motor symptoms associated with smoking in regression models including potential comorbidities.

		Heart Diseases			**Lung Diseases**		
		Smoking					
		Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Tremor	*n*	5249	4183	872	5245	4180	872
	p-value	0.1666	0.3558	0.4804	0.1477	0.3326	0.4715
	β	0.0967	0.0035	-0.0047	0.1012	0.0037	-0.0049
Speech	*n*	5252	4185	872	5248	4182	872
	p-value	0.1640	**0.0076**	0.0526	0.1672	**0.0078**	0.0591
	β	0.0829	0.0087	0.0115	0.0823	0.0087	0.0113
Saliva and Drooling	*n*	5252	4185	872	5248	4182	872
	p-value	**0.0142**	**0.0028**	0.0964	**0.0151**	**0.0031**	0.1609
	β	0.1427	0.0094	0.0094	0.1415	0.0093	0.0080
Chewing and Swallowing	*n*	5252	4185	872	5248	4182	872
	p-value	**0.0004**	**$<1\times10^{-5}$**	0.0765	**0.0004**	**$<1\times10^{-5}$**	0.1281
	β	0.2140	0.0169	0.0101	0.2162	0.0171	0.0088
Walking and Balance	*n*	5249	4183	872	5245	4180	872
	p-value	0.3366	**0.0075**	**9×10^{-5}**	0.3210	**0.0064**	**0.0003**
	β	0.0604	0.0094	0.0268	0.0625	0.0096	0.0241
Freezing	*n*	5249	4183	872	5245	4180	872
	p-value	**0.0313**	**0.0065**	**2×10^{-5}**	**0.0274**	**0.0087**	**9×10^{-5}**
	β	0.1395	0.0092	0.0270	0.1432	0.0088	0.0247
Getting up	*n*	5249	4183	872	5249	4180	872
	p-value	0.1130	**0.0090**	**$<1\times10^{-5}$**	0.1085	**0.0094**	**$<1\times10^{-5}$**
	β	0.1015	0.0093	0.0356	0.1029	0.0093	0.0349

P-value (exploratory): Multivariate regression to predict the respective motor symptoms adjusted for covariates by including AAE, gender, disease duration (time between AAO and current age), and comorbidities in the model.

Supplementary Table S11: Non-motor symptoms associated with smoking in regression models including potential comorbidities.

		Heart Diseases			Lung Diseases		
		Smoking					
		Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Constipation	*n*	5272	4200	876	5268	4197	876
	p-value	0.2324	0.2212	0.6463	0.2215	0.2081	0.6950
	β	0.0686	0.0037	0.0025	0.0702	0.0038	0.0022
Unexplained Pains	*n*	5271	4200	876	5267	4197	876
	p-value	**$<1\times10^{-5}$**	**0.0005**	**0.0205**	**$<1\times10^{-5}$**	**0.0005**	**0.0224**
	β	0.2695	0.0110	0.0132	0.2710	0.0109	0.0131
Problems Remembering	*n*	5271	4200	876	5267	4197	876
	p-value	**0.0002**	**0.0002**	**0.0018**	**0.0001**	**9×10^{-5}**	**0.0047**
	β	0.2165	0.0117	0.0176	0.2217	0.0120	0.0161
Feeling Sad	*n*	5269	4198	876	5265	4195	876
	p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.0807	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.1103
	β	0.3200	0.0145	0.0099	0.3243	0.0149	0.0091
Anxiety	*n*	5269	4198	876	5265	4195	876
	p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.4953	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.5394
	β	0.2969	0.0157	0.0039	0.2986	0.0158	0.0035
Changed Interest in Sex	*n*	5269	4198	876	5265	4195	876
	p-value	**0.0013**	0.2643	0.0574	**0.0014**	0.2353	**0.0287**
	β	0.1973	0.0036	0.0113	0.1951	0.0038	0.0131
Light-headedness	*n*	5268	4197	876	5264	4194	876
	p-value	**0.0007**	**0.0003**	0.1949	**0.0005**	**0.0002**	0.1439
	β	0.1953	0.0111	0.0072	0.2011	0.0115	0.0082

P-value (exploratory): Multivariate regression to predict the respective non-motor symptoms adjusted for covariates by including AAE, gender, disease duration (time between AAO and current age), and comorbidities in the model.

Supplementary Table S12: Symptoms related to mood associated with smoking status. Percentage of patients stratified by smoking status and symptoms related to mood.

		Smoking		
		yes	no	*p*-value
Depression	yes	29.8% (*n*=627)	24.2% (*n*=799)	**<1×10^{-5}**
	no	70.2% (*n*=1475)	75.8% (*n*=2509)	
Anxiety	yes	31.5% (*n*=661)	27.2% (*n*=900)	**2×10^{-5}**
	no	68.5% (*n*=1440)	72.8% (*n*=2409)	
Dropped many activities and interests	yes	40.1% (*n*=837)	33.0% (*n*=1087)	**<1×10^{-5}**
	no	59.9% (*n*=1252)	67.0% (*n*=2204)	
Life feels empty	yes	16.0% (*n*=333)	13.2% (*n*=433)	**0.0004**
	no	84.0% (*n*=1751)	86.8% (*n*=2858)	
Getting bored often	yes	29.4% (*n*=615)	22.3% (*n*=734)	**<1×10^{-5}**
	no	70.6% (*n*=1475)	77.7% (*n*=2558)	
Being afraid something bad could happen	yes	25.0% (*n*=521)	22.5% (*n*=739)	**0.0025**
	no	75.0% (*n*=1559)	77.5% (*n*=2551)	
Feeling helpless often	yes	21.8% (*n*=453)	17.8% (*n*=584)	**2×10^{-5}**
	no	78.2% (*n*=1628)	82.2% (*n*=2703)	
Prefer staying at home	yes	54.6% (*n*=1136)	50.0% (*n*=1644)	**0.0007**
	no	45.4% (*n*=946)	50.0% (*n*=1645)	
Feeling to have more memory problems than other people	yes	28.4% (*n*=592)	25.0% (*n*=823)	**0.0019**
	no	71.6% (*n*=1494)	75.0% (*n*=2472)	
Feeling pretty worthless	yes	17.0% (*n*=354)	14.1% (*n*=462)	**0.0002**
	no	83.0% (*n*=1727)	85.9% (*n*=2821)	
Feeling that situation is hopeless	yes	14.8% (*n*=308)	11.3% (*n*=372)	**<1×10^{-5}**
	no	85.2% (*n*=1767)	88.7% (*n*=2912)	

P-value (exploratory): Multivariate regression to predict the respective mood symptoms adjusted for covariates by including AAE, gender, and disease duration (time between AAO and current age).

Supplementary Table S13: Generalized linear models on mood. Regression models for symptoms related to mood and associated with smoking in the Fox Insight cohort.

Dependent variable: Depression

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0176	0.0035	$<1\times10^{-5}$
Gender	0.3748	0.0638	$<1\times10^{-5}$
Disease Duration	0.0065	0.0060	0.2815
Smoking (binary)	0.3362	0.0649	$<1\times10^{-5}$

Dependent variable: Anxiety

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0284	0.0034	$<1\times10^{-5}$
Gender	0.3113	0.0621	$<1\times10^{-5}$
Disease Duration	0.0064	0.0059	0.2788
Smoking (binary)	0.2748	0.0636	2×10^{-5}

Dependent variable: Dropped many activities and interests

Covariates	Estimate	Standard error	*p*-value
AAE	0.0034	0.0033	0.2995
Gender	-0.0815	0.0585	0.1636
Disease Duration	0.0360	0.0057	$<1\times10^{-5}$
Smoking (binary)	0.3283	0.0598	$<1\times10^{-5}$

Dependent variable: Feeling that life feels empty

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0190	0.0043	1×10^{-5}
Gender	-0.0355	0.0798	0.6564
Disease Duration	0.0161	0.0074	0.0291
Smoking (binary)	0.2843	0.0809	0.0004

Dependent variable: Getting bored often

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0287	0.0036	$<1\times10^{-5}$
Gender	-0.2916	0.0654	$<1\times10^{-5}$
Disease Duration	0.0334	0.0060	$<1\times10^{-5}$
Smoking (binary)	0.4263	0.0661	$<1\times10^{-5}$

Dependent variable: Being afraid something bad could happen

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0284	0.0036	$<1\times10^{-5}$
Gender	0.2330	0.0661	0.0004
Disease Duration	-0.0188	0.0070	0.0069
Smoking (binary)	0.2050	0.0677	0.0025

Dependent variable: Feeling helpless often

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0109	0.0039	0.0050
Gender	0.0886	0.0708	0.2108
Disease Duration	0.0329	0.0064	$<1\times10^{-5}$
Smoking (binary)	0.3099	0.0720	2×10^{-5}

Dependent variable: Prefer staying at home

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0105	0.0031	0.0007
Gender	-0.1151	0.0559	0.0394
Disease Duration	0.0051	0.0055	0.3462
Smoking (binary)	0.1955	0.0576	0.0007

Dependent variable: Feeling to have more memory problems than other people

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0235	0.0035	$<1\times10^{-5}$
Gender	-0.4046	0.0643	$<1\times10^{-5}$
Disease Duration	0.0102	0.0061	0.0938
Smoking (binary)	0.2027	0.0652	0.0019

Dependent variable: Feeling pretty worthless

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0192	0.0042	$<1\times10^{-5}$
Gender	0.0245	0.0779	0.7536
Disease Duration	0.0279	0.0070	7×10^{-5}
Smoking (binary)	0.2958	0.0791	0.0002

Dependent variable: Feeling that situation is hopeless

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0184	0.0046	6×10^{-5}
Gender	0.1564	0.0838	0.0620
Disease Duration	0.0397	0.0072	$<1\times10^{-5}$
Smoking (binary)	0.3846	0.0848	$<1\times10^{-5}$

Dependent variable: Depression

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0153	0.0040	0.0001
Gender	0.4080	0.0727	$<1\times10^{-5}$
Disease Duration	0.0038	0.0070	0.5885
Smoking Dosage	0.0199	0.0033	$<1\times10^{-5}$

Dependent variable: Anxiety

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0269	0.0039	$<1\times10^{-5}$
Gender	0.3355	0.0703	$<1\times10^{-5}$
Disease Duration	0.0054	0.0068	0.4266
Smoking Dosage	0.0182	0.0032	$<1\times10^{-5}$

Dependent variable: Dropped many activities and interests

Covariates	Estimate	Standard error	*p*-value
AAE	0.0005	0.0037	0.8925
Gender	-0.0813	0.0664	0.2206
Disease Duration	0.0395	0.0065	$<1\times10^{-5}$
Smoking Dosage	0.0165	0.0031	$<1\times10^{-5}$

Dependent variable: Feeling that life feels empty

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0190	0.0049	0.0001
Gender	0.0257	0.0906	0.7764
Disease Duration	0.0144	0.0085	0.0916
Smoking Dosage	0.0181	0.0038	$<1\times10^{-5}$

Dependent variable: Getting bored often

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0313	0.0041	$<1\times10^{-5}$
Gender	-0.2616	0.0744	0.0004
Disease Duration	0.0330	0.0069	$<1\times10^{-5}$
Smoking Dosage	0.0187	0.0033	$<1\times10^{-5}$

Dependent variable: Being afraid something bad could happen

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0267	0.0040	$<1\times10^{-5}$
Gender	0.3142	0.0744	2×10^{-5}
Disease Duration	-0.0203	0.0079	0.0101
Smoking Dosage	0.0126	0.0034	0.0002

Dependent variable: Feeling helpless often

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0116	0.0045	0.0096
Gender	0.1538	0.0812	0.0583
Disease Duration	0.0357	0.0074	$<1\times10^{-5}$
Smoking Dosage	0.0151	0.0036	3×10^{-5}

Dependent variable: Prefer staying at home

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0122	0.0035	0.0004
Gender	-0.0469	0.0628	0.4552
Disease Duration	0.0066	0.0062	0.2882
Smoking Dosage	0.0150	0.0032	$<1\times10^{-5}$

Dependent variable: Feeling to have more memory problems than other people

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0250	0.0039	$<1\times10^{-5}$
Gender	-0.3760	0.0725	$<1\times10^{-5}$
Disease Duration	0.0122	0.0069	0.0779
Smoking Dosage	0.0105	0.0033	0.0014

Dependent variable: Feeling pretty worthless

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0210	0.0049	2×10^{-5}
Gender	0.0356	0.0890	0.6894
Disease Duration	0.0337	0.0080	3×10^{-5}
Smoking Dosage	0.0185	0.0038	$<1\times10^{-5}$

Dependent variable: Feeling that situation is hopeless

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0203	0.0053	0.0001
Gender	0.1846	0.0966	0.0562
Disease Duration	0.0429	0.0082	$<1\times10^{-5}$
Smoking Dosage	0.0172	0.0041	3×10^{-5}

Dependent variable: Depression

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0262	0.0083	0.0017
Gender	0.1939	0.1528	0.2045
Disease Duration	0.0056	0.0156	0.7187
Smoking Duration	0.0123	0.0061	0.0422

Dependent variable: Dropped many activities and interests

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0109	0.0079	0.1678
Gender	-0.1527	0.1435	0.2875
Disease Duration	0.0345	0.0146	0.0186
Smoking Duration	0.0253	0.0057	$<1\times10^{-5}$

Dependent variable: Feeling that life feels empty

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0262	0.0104	0.0118
Gender	-0.1131	0.1953	0.5626
Disease Duration	0.0215	0.0184	0.2429
Smoking Duration	0.0230	0.0075	0.0021

Dependent variable: Getting bored often

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0371	0.0087	2×10^{-5}
Gender	-0.5031	0.1619	0.0019
Disease Duration	0.0324	0.0153	0.0343
Smoking Duration	0.0412	0.0063	$<1\times10^{-5}$

Dependent variable: Being afraid something bad could happen

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0395	0.0087	$<1\times10^{-5}$
Gender	0.0247	0.1619	0.8786
Disease Duration	-0.0443	0.0195	0.0231
Smoking Duration	0.0184	0.0064	0.0039

Dependent variable: Feeling helpless often

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0278	0.0095	0.0036
Gender	-0.3514	0.1790	0.0496
Disease Duration	0.0366	0.0177	0.0390
Smoking Duration	0.0296	0.0068	1×10^{-5}

Dependent variable: Prefer staying at home

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0246	0.0079	0.0019
Gender	-0.2025	0.1406	0.1499
Disease Duration	0.0054	0.0144	0.7049
Smoking	0.0179	0.0057	0.0017

Dependent variable: Feeling to have more memory problems than other people

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0259	0.0085	0.0022
Gender	-0.2679	0.1582	0.0903
Disease Duration	-0.0203	0.0173	0.2403
Smoking Duration	0.0178	0.0062	0.0038

Dependent variable: Feeling pretty worthless

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0351	0.0102	0.0006
Gender	-0.2937	0.1918	0.1256
Disease Duration	0.0342	0.0175	0.0506
Smoking Duration	0.0370	0.0072	$<1\times10^{-5}$

Dependent variable: Feeling hopeless

Covariates	Estimate	Standard error	*p*-value
AAE	-0.0367	0.0108	0.0007
Gender	0.1368	0.2033	0.5011
Disease Duration	0.0241	0.0195	0.2158
Smoking Duration	0.0304	0.0078	0.0001

Supplementary Table S14: Generalized linear models on symptoms related to mood including potential comorbidities. Regression models for symptoms related to mood with a change in outcome for smoking in the Fox Insight cohort while adjusting for potential comorbidities.

		Heart Diseases			**Lung Diseases**		
		Smoking					
		Yes/No	Dosage	Duration	Yes/No	Dosage	Duration
Depression	*n*	5213	4143	859	5213	4143	859
	p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.0544	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	0.0584
	β	0.3354	0.0197	0.0117	0.3405	0.0197	0.0116
Anxiety	*n*	5213	4144	859	5213	4144	859
	p-value	**2×10^{-5}**	**$<1\times10^{-5}$**	0.4005	**2×10^{-5}**	**$<1\times10^{-5}$**	0.3173
	β	0.2725	0.0179	0.0051	0.2713	0.0178	0.0061
Dropped many activities and interests	*n*	5189	4138	866	5188	4138	866
	p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**1×10^{-5}**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**1×10^{-5}**
	β	0.3234	0.0158	0.0249	0.3278	0.0161	0.0251
Life feels empty	*n*	5185	4137	864	5184	4137	864
	p-value	**0.0006**	**1×10^{-5}**	**0.0034**	**0.0005**	**$<1\times10^{-5}$**	**0.0018**
	β	0.2773	0.0170	0.0220	0.2840	0.0180	0.0237
Getting bored often	*n*	5191	4137	866	5190	4139	866
	p-value	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**2×10^{-5}**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**
	β	0.4187	0.0179	0.0403	0.4244	0.0187	0.0422
Being afraid something bad could happen	*n*	5179	4133	862	5179	4133	862
	p-value	**0.0037**	**0.0007**	**0.0060**	**0.0026**	**0.0004**	**0.0076**
	β	0.1972	0.0117	0.0176	0.2046	0.0122	0.0172
Feeling helpless often	*n*	5177	4130	862	5177	4130	862
	p-value	**2×10^{-5}**	**7×10^{-5}**	**3×10^{-5}**	**2×10^{-5}**	**4×10^{-5}**	**2×10^{-5}**
	β	0.3053	0.0143	0.0284	0.3105	0.0147	0.0293
Prefer staying at home	*n*	5180	4131	862	5180	4131	862
	p-value	**0.0008**	**$<1\times10^{-5}$**	**0.0015**	**0.0006**	**$<1\times10^{-5}$**	**0.0020**
	β	0.1938	0.0148	0.0183	0.1970	0.0149	0.0178
Feeling to have more memory problems than other people	*n*	5190	4142	866	5190	4142	866
	p-value	**0.0018**	**0.0027**	**0.0046**	**0.0016**	**0.0020**	**0.0097**
	β	0.2040	0.0100	0.0175	0.2069	0.0103	0.0161
Feeling pretty worthless	*n*	5173	4124	860	5173	4124	860
	p-value	**0.0003**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**	**0.0002**	**$<1\times10^{-5}$**	**$<1\times10^{-5}$**
	β	0.2860	0.0172	0.0358	0.2965	0.0184	0.0375
Feeling that situation is hopeless	*n*	5168	4125	860	5168	4125	860
	p-value	**1×10^{-5}**	**0.0001**	**0.0001**	**$<1\times10^{-5}$**	**4×10^{-5}**	**0.0001**
	β	0.3749	0.0161	0.0298	0.3847	0.0171	0.0301

P-value (exploratory): Multivariate regression to predict the respective symptoms related to mood adjusted for covariates by including AAE, gender, disease duration (time between AAO and current age), and comorbidities in the model.

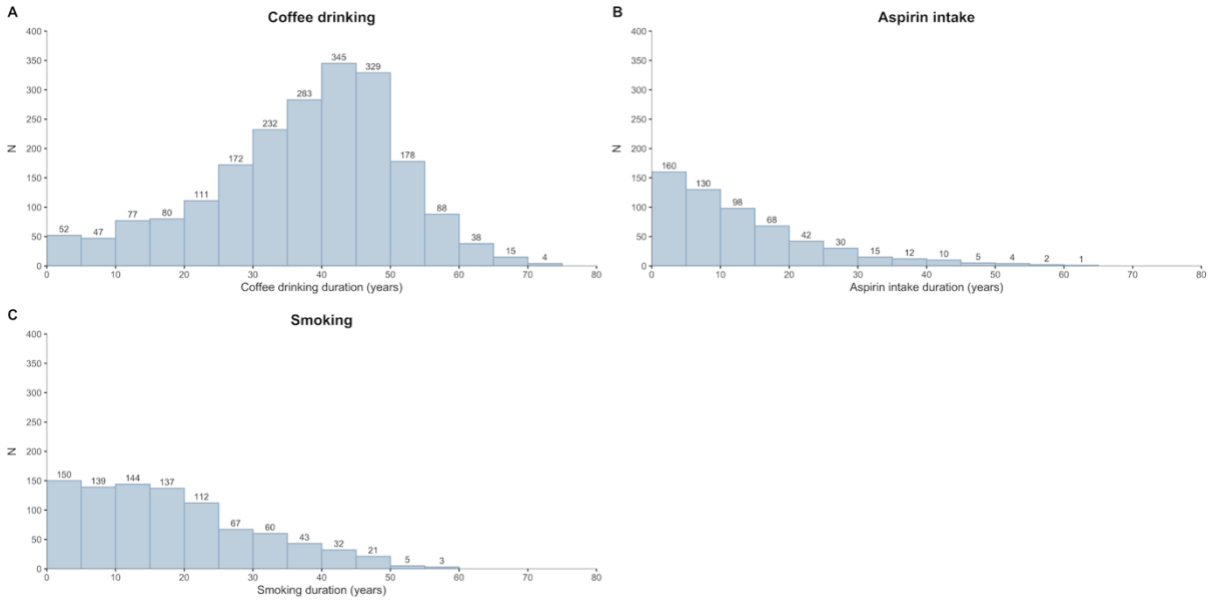
Supplementary Table S15: Generalized linear models on mood. Regression models for symptoms related to mood and significant for smoking in the Fox Insight cohort.

Dependent variable: Depression

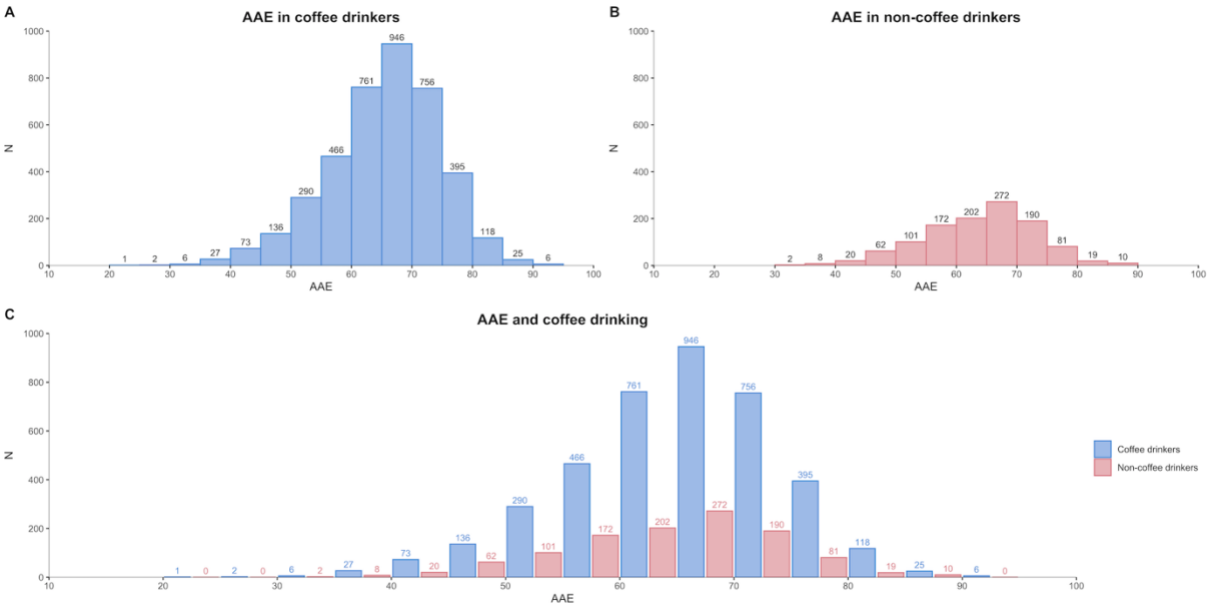
Covariates	Estimate	Standard error	*p*-value
AAE	-0.0291	0.0085	0.0007
Gender	0.2161	0.1542	0.1610
Disease Duration	0.0063	0.0155	0.6852
Heart Diseases (binary)	0.3338	0.2038	0.1015
Smoking Duration	0.0117	0.0061	0.0544

Dependent variable: Depression

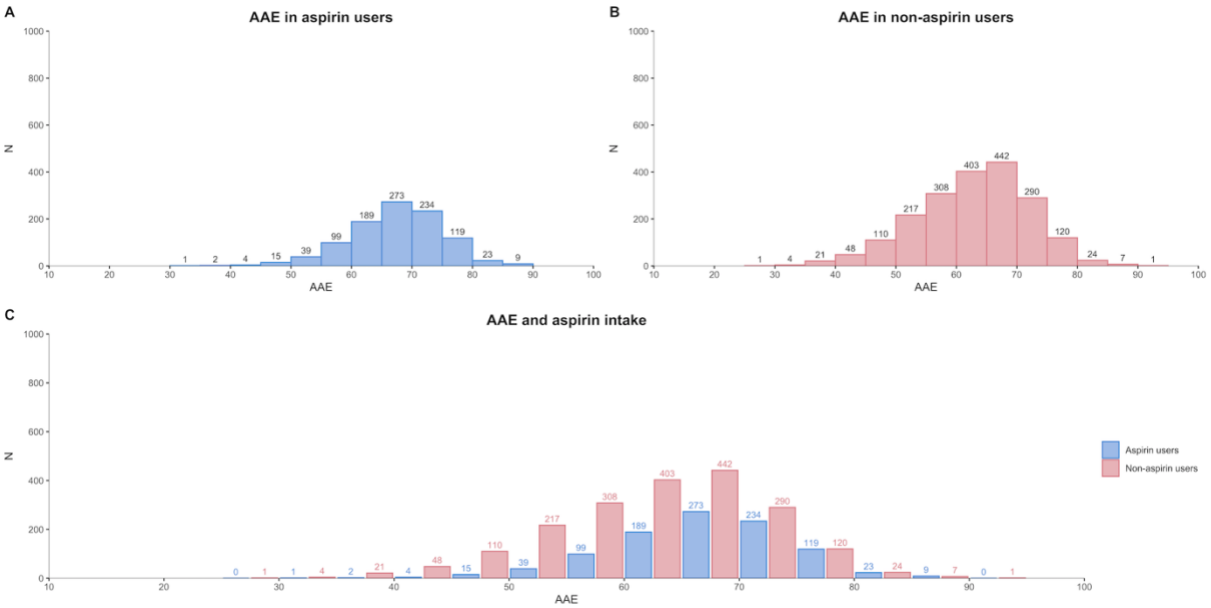
Covariates	Estimate	Standard error	*p*-value
AAE	-0.0264	0.0083	0.0016
Gender	0.1906	0.1529	0.2127
Disease Duration	0.0063	0.0155	0.6851
Lung Diseases (binary)	0.1839	0.2025	0.3638
Smoking Duration	0.0116	0.0061	0.0584



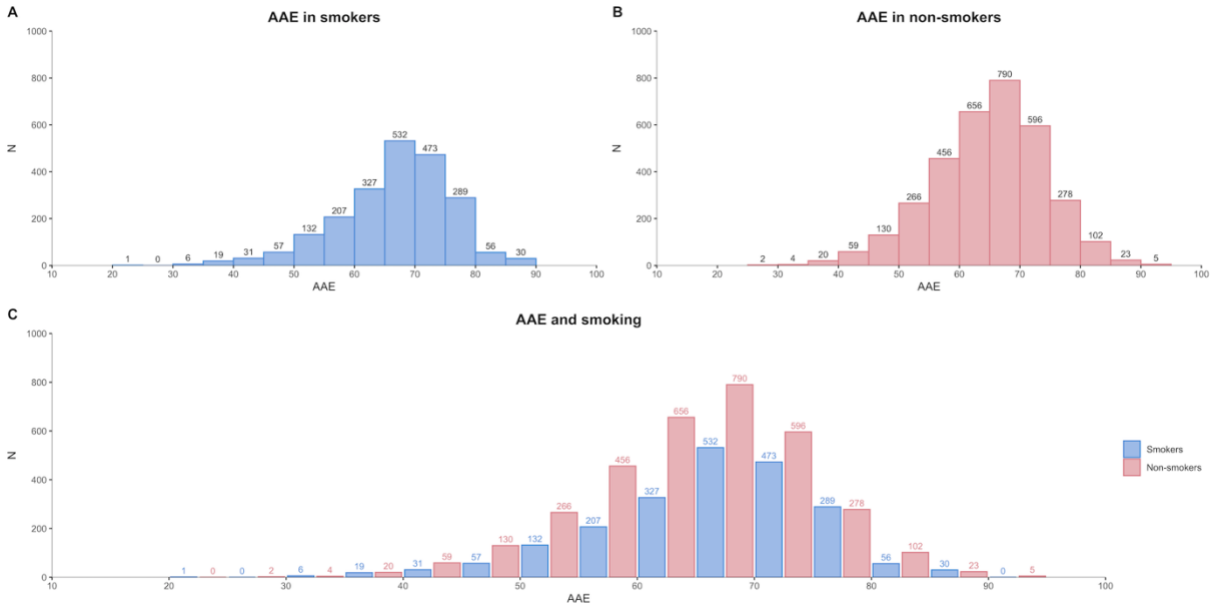
Supplementary Figure S2: Distribution of coffee drinking duration, aspirin intake duration, and smoking duration in the Fox Insight cohort. (**A**) Histogram of the coffee drinking duration in years. (**B**) Histogram of the aspirin intake duration in years. (**C**) Histogram of the smoking duration in years.



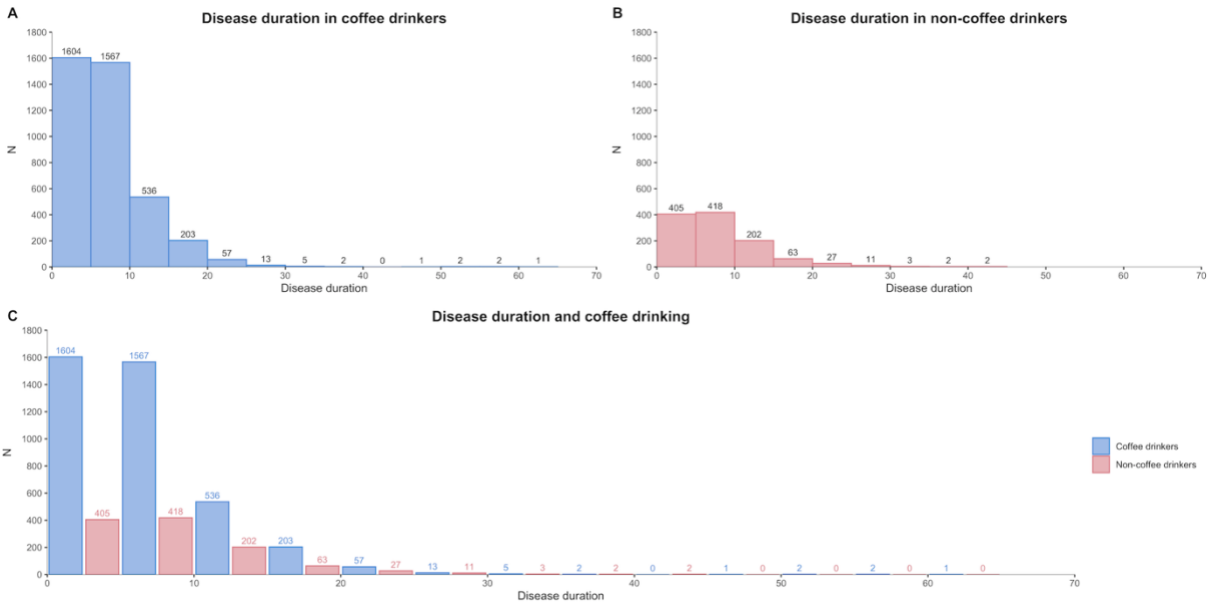
Supplementary Figure S3: Distribution of AAE in the subgroups of coffee drinkers and non-coffee drinkers. (**A**) Histogram of the AAE in the subgroup of coffee drinkers. (**B**) Histogram of the AAE in the subgroup of non-coffee drinkers (**C**) Histogram of the AAE in the subgroups of coffee drinkers and non-coffee drinkers.



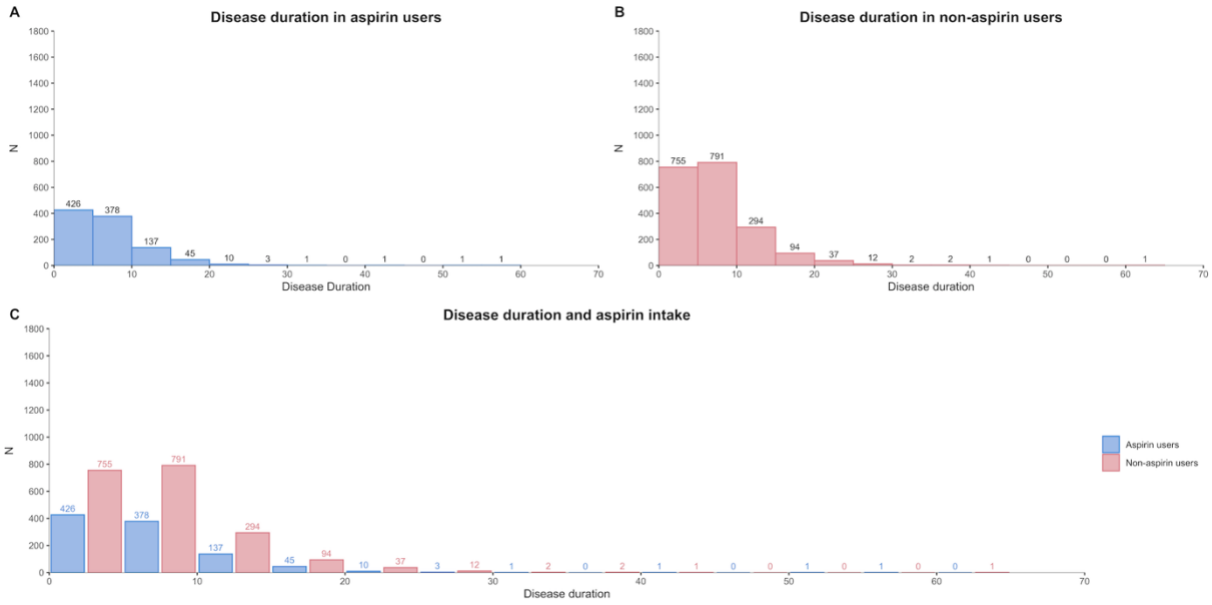
Supplementary Figure S4: Distribution of AAE in the subgroups of aspirin users and non-aspirin users. (**A**) Histogram of the AAE in the subgroup of aspirin users. (**B**) Histogram of the AAE in the subgroup of non-aspirin users (**C**) Histogram of the AAE in the subgroups of aspirin users and non-aspirin users.



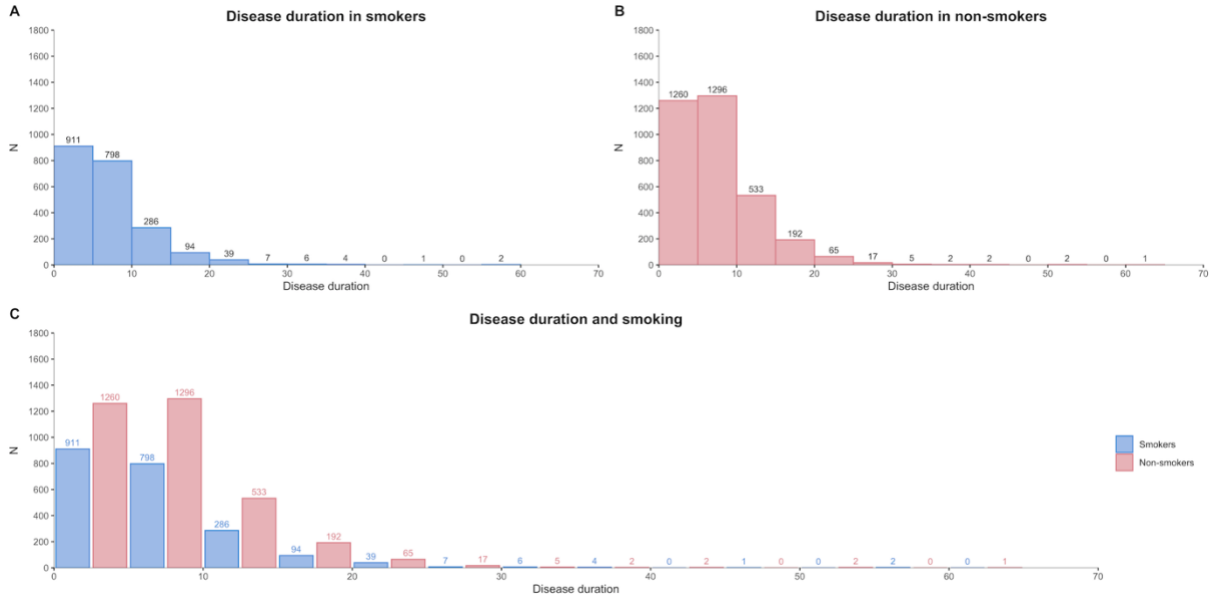
Supplementary Figure S5: Distribution of AAE in the subgroups of smokers and non-smokers. (**A**) Histogram of the AAE in the subgroup of smokers. (**B**) Histogram of the AAE in the subgroup of non-smokers (**C**) Histogram of the AAE in the subgroups of smokers and non-smokers.



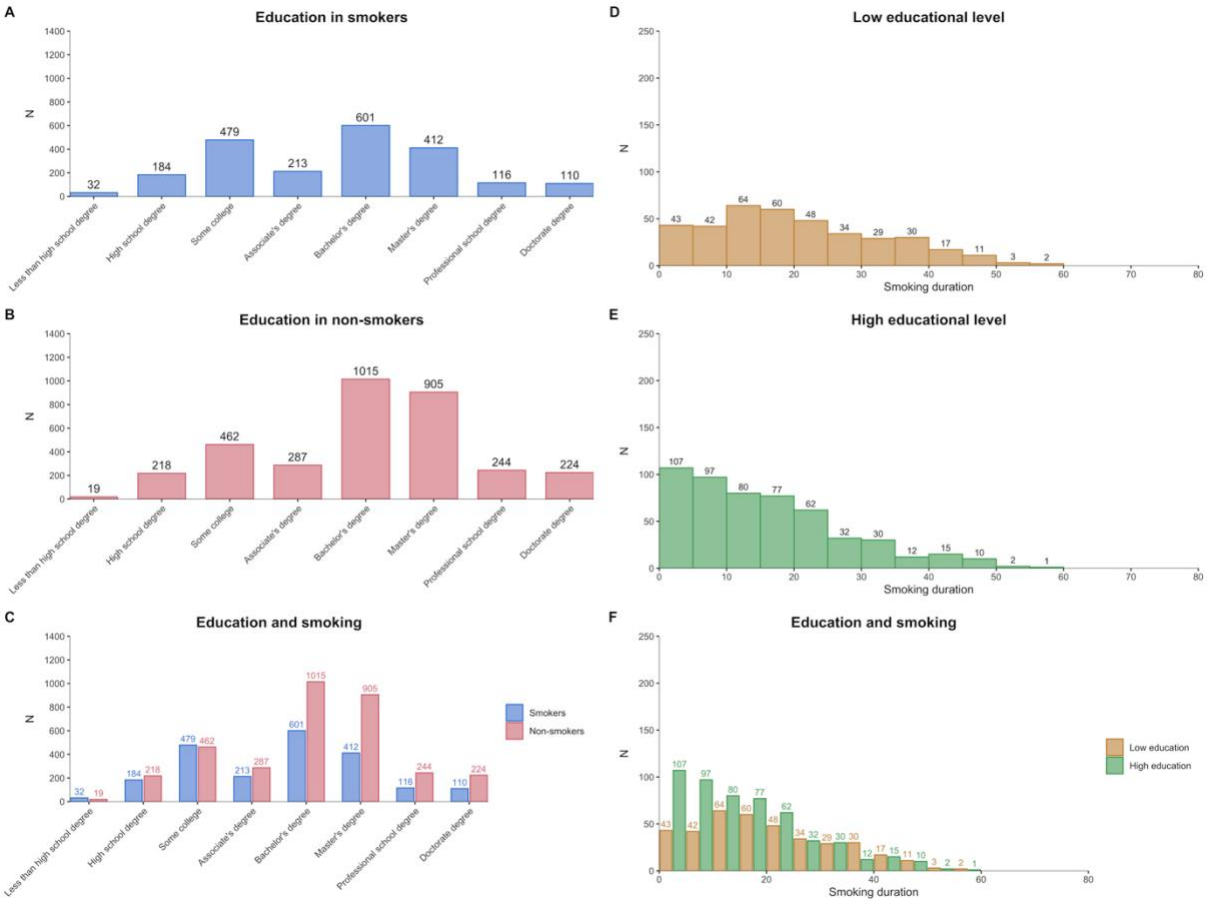
Supplementary Figure S6: Distribution of disease duration in the subgroups of coffee drinkers and non-coffee drinkers. (**A**) Histogram of the disease duration in the subgroup of coffee drinkers. (**B**) Histogram of disease duration in the subgroup of non-coffee drinkers (**C**) Histogram of disease duration in the subgroups of coffee drinkers and non-coffee drinkers.



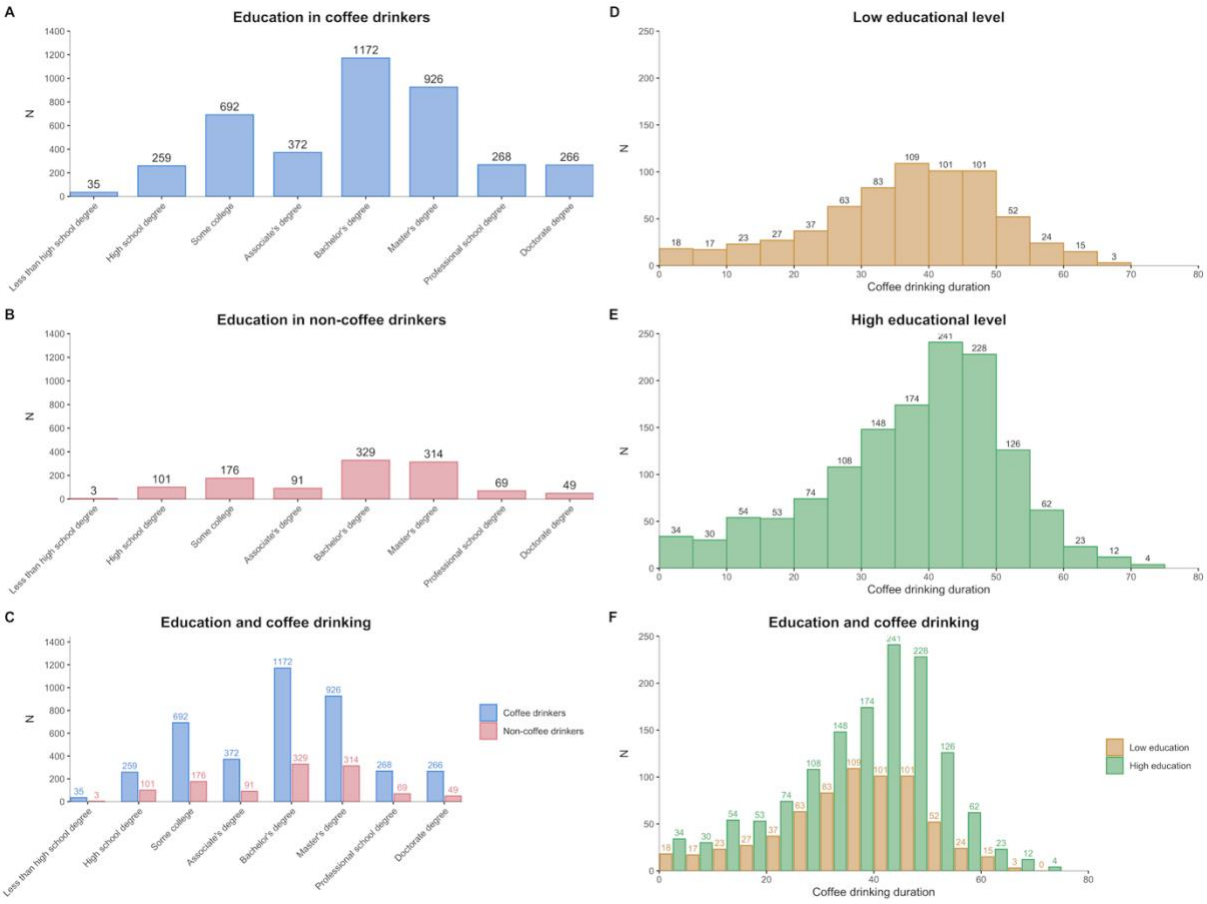
Supplementary Figure S7: Distribution of disease duration in the subgroups of aspirin users and non-aspirin users. (**A**) Histogram of the disease duration in the subgroup of aspirin users. (**B**) Histogram of the disease duration in the subgroup of non-aspirin users (**C**) Histogram of the disease duration in the subgroups of aspirin users and non-aspirin users.



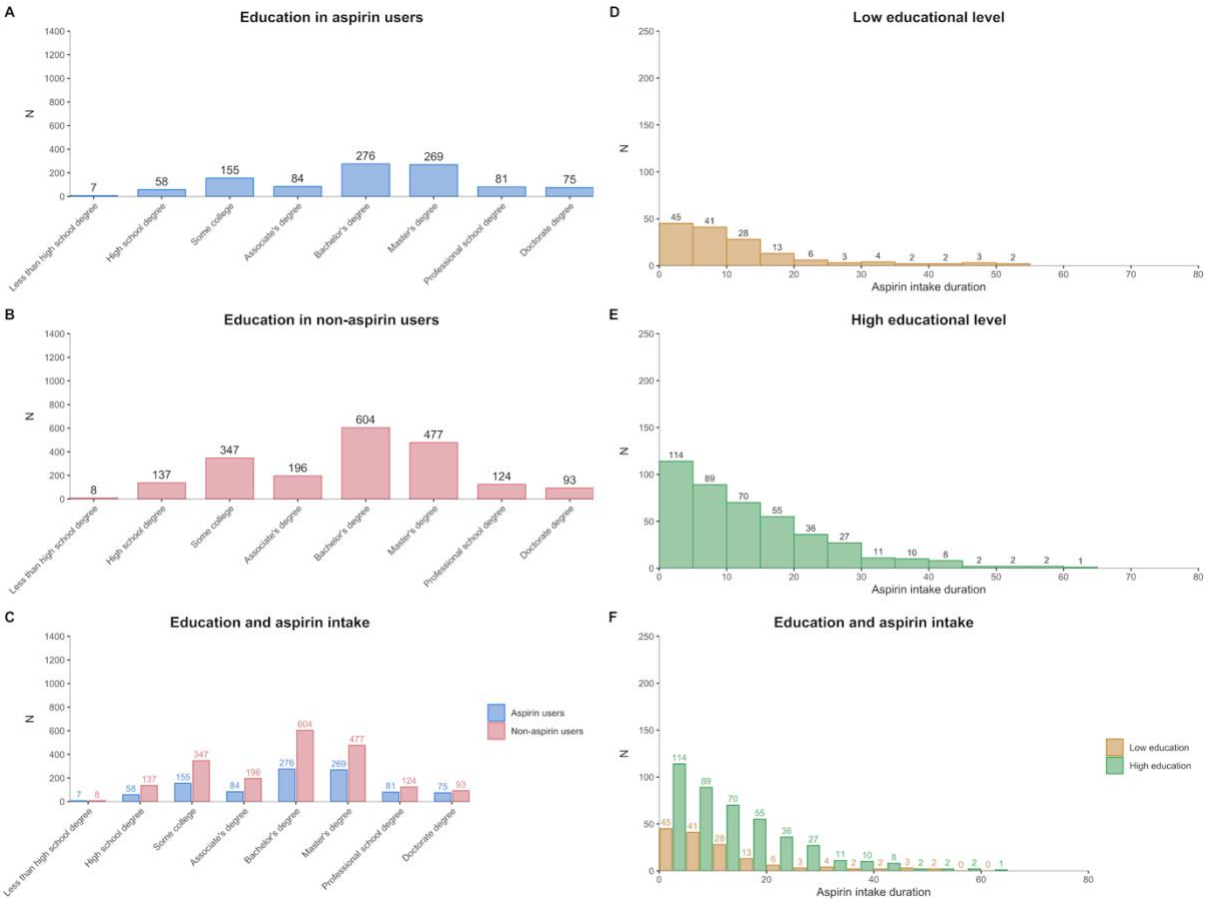
Supplementary Figure S8: Distribution of disease duration in the subgroups of smokers and non-smokers. (**A**) Histogram of the disease duration in the subgroup of smokers. (**B**) Histogram of the disease duration in the subgroup of non-smokers (**C**) Histogram of the disease duration in the subgroups of smokers and non-smokers.



Supplementary Figure S9: Education and smoking. (**A**) Distribution of the educational level in the subgroup of smokers. (**B**) Distribution of the educational level in the subgroup of non-smokers (**C**) Distribution of the educational level in the subgroups of smokers and non-smokers. (**D**) Histogram of the smoking duration in the subgroup of smokers with a low educational level (i.e., less than a high school degree, high school degree, some college, and Associate's degree). (**E**) Histogram of the smoking duration in the subgroup of smokers with a high educational level (i.e., Bachelor's degree, Master's degree, professional school degree, and doctorate degree) (**F**) Histogram of the smoking duration in the subgroups of smokers with a low educational level and with a high educational level.



Supplementary Figure S10: Education and coffee drinking. (**A**) Distribution of the educational level in the subgroup of coffee drinkers. (**B**) Distribution of the educational level in the subgroup of non-coffee drinkers (**C**) Distribution of the educational level in the subgroups of coffee drinkers and non-coffee drinkers. (**D**) Histogram of the coffee drinking duration in the subgroup of coffee drinkers with a low educational level (i.e., less than a high school degree, high school degree, some college, and Associate's degree). (**E**) Histogram of the coffee drinking duration in the subgroup of coffee drinkers with a high educational level (i.e., Bachelor's degree, Master's degree, professional school degree, and doctorate degree) (**F**) Histogram of the coffee drinking duration in the subgroups of coffee drinkers with a low educational level and with a high educational level.



Supplementary Figure S11: Education and aspirin intake. (**A**) Distribution of the educational level in the subgroup of aspirin users. (**B**) Distribution of the educational level in the subgroup of non-aspirin users (**C**) Distribution of the educational level in the subgroups of aspirin users and non-aspirin users. (**D**) Histogram of the aspirin intake duration in the subgroup of aspirin users with a low educational level (i.e., less a than high school degree, high school degree, some college, and Associate's degree). (**E**) Histogram of the aspirin intake duration in the subgroup of aspirin users with a high educational level (i.e., Bachelor's degree, Master's degree, professional school degree, and doctorate degree) (**F**) Histogram of the aspirin intake duration in the subgroups of aspirin users with a low educational level and with a high educational level.

References:

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W.*, et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

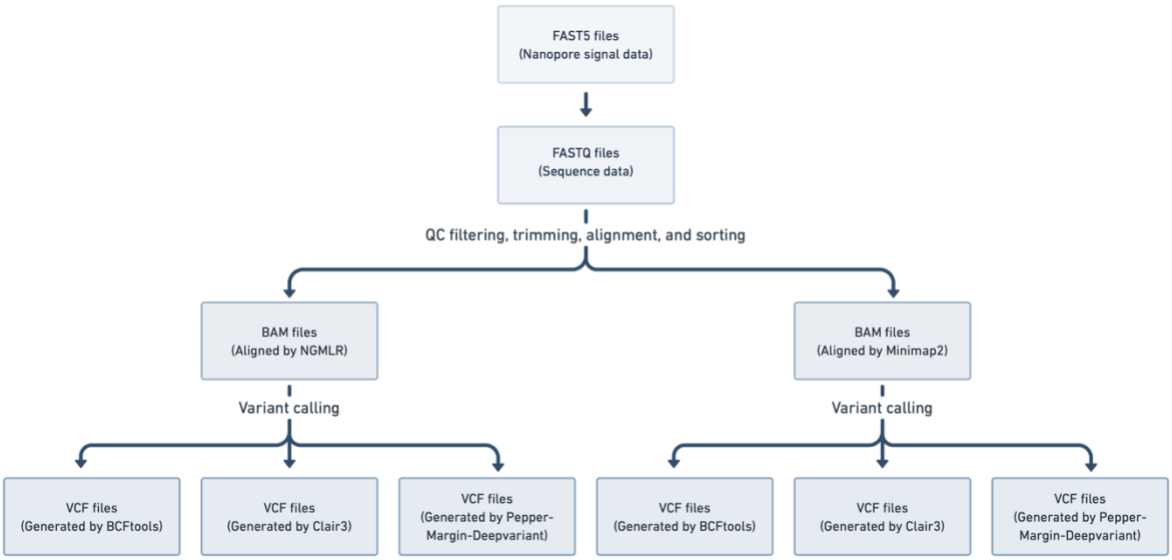
SUPPLEMENTARY MATERIAL FOR

GBA1 in Parkinson's disease: variant detection and pathogenicity scoring matters

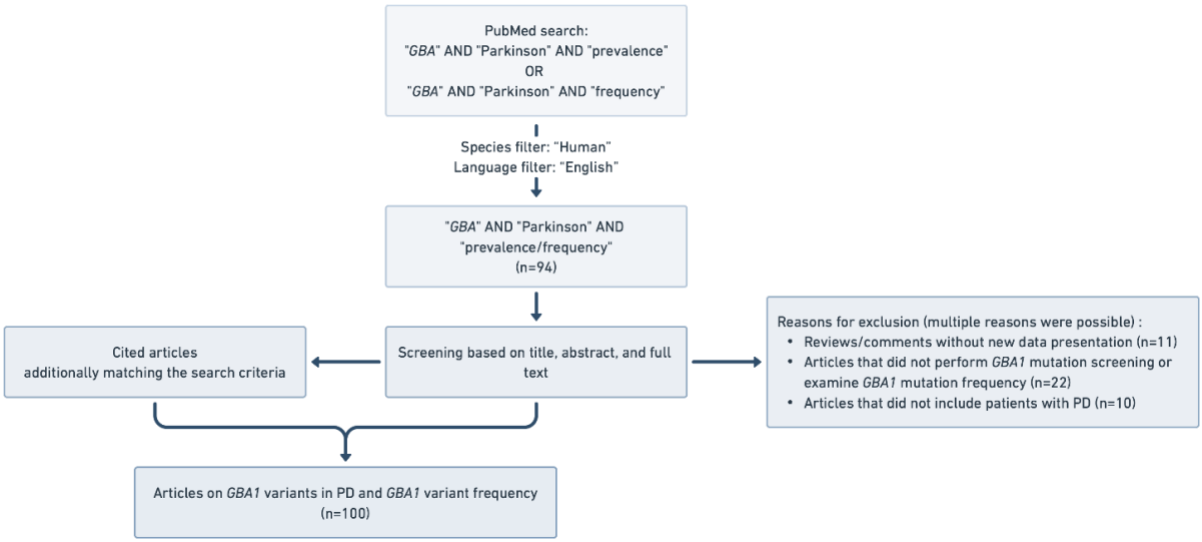
Carolin Gabbert[1], MSc, Susen Schaake[1], BSc, Theresa Lüth[1], MSc, Christoph Much[1], Christine Klein[1], MD, Jan O. Aasly[2], MD, Matthew J. Farrer[3], PhD, Joanne Trinh[1*], PhD

[1]Institute of Neurogenetics, University of Lübeck, Lübeck, Germany
[2]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway
[3]Department of Neurology, University of Florida, Gainesville, Florida

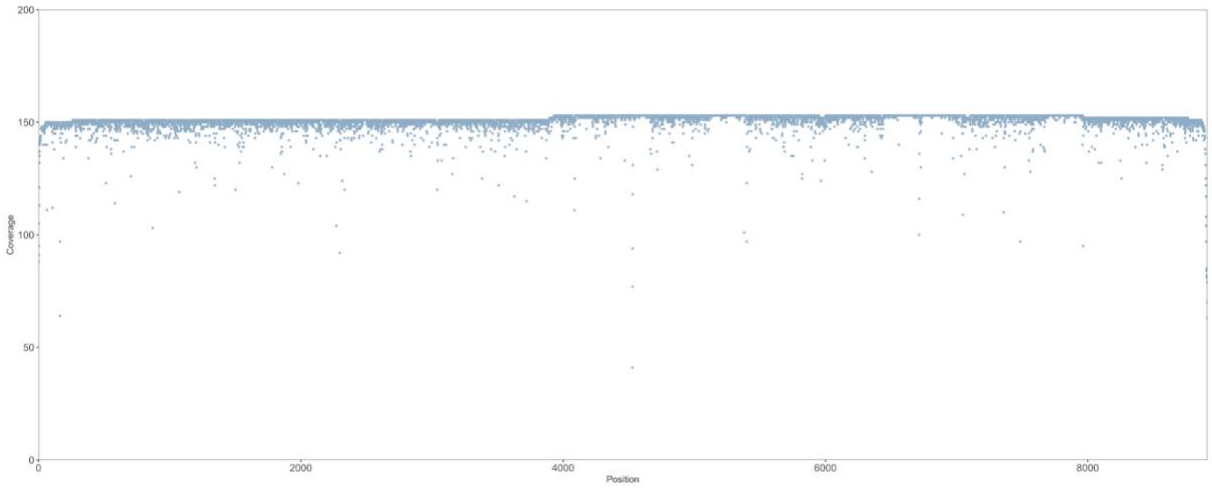
*** Correspondence:**
Joanne Trinh
University of Lübeck
Ratzeburger Allee 160
23538 Lübeck, Germany
Email: joanne.trinh@neuro.uni-luebeck.de
Tel.: +49-451-31018202



Supplementary Figure S1: Laboratory workflow data analysis pipeline of how the data was processed and which aligners (i.e. NGMLR and Minimap2) and variant callers (i.e. BCFtools, Clair3, Pepper-Margin-Deepvariant).



Supplementary Figure S2: Workflow of the literature search in PubMed. We searched for literature via PubMed that was published before August 4, 2022, using the search term "*GBA*" AND "Parkinson" AND "prevalence" OR "*GBA*" AND "Parkinson" AND "frequency", while setting the species filter to “Human” and the language filter to “English”, resulting in 94 articles. These were screened based on the title, abstract and full text, excluding all articles not directly screening for variants in the *GBA1* gene in patients with PD. Reasons for exclusion were reviews, or comments without new data (n=11), articles that did not perform *GBA1* variant screening or examine *GBA1* variant frequency in their study population (n=22), and articles that were not about PD or did not include patients with PD (n=10) (multiple reasons for exclusion were possible). In addition to the articles found via the search term, suitable articles that were referenced in this literature were also included in the overview. In the end, 100 articles on *GBA1* variant frequencies across populations were included in the overview.



Supplementary Figure S3: Exemplary coverage plot. The coverage is shown over the full *GBA1* amplicon (8.9 kb)

Supplementary Table S1: Overview over all rare *GBA1* variants detected with Nanopore sequencing

Position chr1 (hg38)	REF	ALT	Region	SNP ID	Variant information *GBA* (NM_000157.4)	AA change	ACMG classification	ClinVar	SIFT	Poly-phen2 HDIV	Poly-phen2 HVAR	CADD raw	CADD phred	GERP++	gnomAD (All)
155232724	C	T	intergenic	rs578095239	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0002
155232838	C	T	intergenic	rs561692357	.	.	Uncertain significance 1 points = 1 P - 0 B	.	.	.	.	.	.	.	.
155232871	T	C	intergenic	rs529155642	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0006
155232982	G	A	intergenic	rs530421143	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.001
155233134	C	T	intergenic	rs557608543	.	.	Uncertain significance 0 points = 1 P - 1 B	.	.	.	.	.	.	.	0.0005
155233268	C	G	intergenic	rs367634752	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0016
155233286	C	T	intergenic	rs1042594015	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	.
155233287	G	A	intergenic	rs368998334	.	.	Likely benign -6 points = 0 P - 6 B	.	.	.	.	.	.	.	0.0032
155233514	A	G	downstream	rs2361530	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	.
155233517	C	T	downstream	rs2361531	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	.
155233521	T	A	downstream	rs2361532	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	.
155233531	G	A	downstream	rs2361533	.	.	Uncertain significance 0 points = 1 P - 1 B	.	.	.	.	.	.	.	.
155233541	G	T	downstream	rs4024047	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.00003238
155233543	G	A	downstream	rs4024048	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.00003241
155233549	C	T	downstream	rs4024049	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.00003258
155233612	A	G	downstream	rs2142046	.	.	Benign -10 points = 0 P - 10 B	.	.	.	.	.	.	.	0.0024
155233639	G	A	downstream	rs2142045	.	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0005
155234414	G	A	downstream	rs201209118	.	.	Benign -10 points = 0 P - 10 B	.	.	.	.	.	.	.	0
155234893	A	G	UTR3	rs368275143	c.*102T>C	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0113
155234903	C	T	UTR3	rs708606	c.*92G>A	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0022
155235104	G	A	intronic	rs374690110	c.1506-4C>T	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0002
155235206	G	A	exonic	rs371779859	c.1494C>T	p.Val498=	Likely benign -2 points = 1 P - 3 B	.	.	.	.	.	.	.	0.00009689
155235222	C	T	exonic	.	c.1478G>A	p.Gly493Asp	Uncertain significance 1 points = 1 P - 0 B	.	T	D	P	4.834	24.8	3.16	.

155235252	A	G	exonic	rs421016	c.1448T>C	p.Leu483Pro	Pathogenic 11 points = 11 P – 0 B	Pathogenic; risk factor	D	P	P	4.842	24.8	3.16	0.0007
155235302	A	G	exonic	.	c.1398T>C	p.Ile466=	Likely benign -2 points = 1 P - 3 B	.	.	.	.	.	.	.	.
155235344	G	A	intronic	rs569191841	c.1389-33C>T	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.00003234
155235379	A	G	intronic	rs2974924	c.1389-68T>C	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0105
155235587	C	T	intronic	rs12752133	c.1388+94G>A	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	0.0127
155235791	G	A	exonic	rs201499639	c.1278C>T	p.Pro426=	Likely benign -2 points = 1 P - 3 B	Likely benign	.	.	.	.	.	.	0.0000323
155235843	T	C	exonic	rs76763715	c.1226A>G	p.Asn409Ser	Uncertain significance 5 points = 5 P – 0 B	Pathogenic/Likely pathogenic; risk factor	D	P	B	3.202	22.7	3.53	0.0017
155235928	G	C	intronic	rs41264925	c.1225-84C>G	.	Likely benign -1 points = 1 P - 2 B	.	.	.	.	.	.	.	.
155236050	G	A	intronic	rs1036605613	c.1224+195C>T	.	Uncertain significance 0 points = 1 P – 1 B	.	.	.	.	.	.	.	0.00003232
155236246	G	A	exonic	rs75548401	c.1223C>T	p.Thr408Met	Uncertain significance 1 points = 1 P – 0 B	Uncertain significance (2); Benign (4); Likely benign (3)	T	B	B	2.993	22.2	3.57	0.0076
155236294	C	T	exonic	rs11558184	c.1175G>A	p.Arg392Gln	Uncertain significance 1 points = 1 P – 0 B	.	T	P	B	3.323	22.9	3.67	0.00003233
155236331	C	T	exonic	rs781306264	c.1138G>A	p.Ala380Thr	Uncertain significance 1 points = 1 P – 0 B	.	D	D	D	6.843	33	3.67	.
155236366	C	T	exonic	rs1064648	c.1103G>A	p.Arg368His	Uncertain significance 1 points = 1 P – 0 B	.	T	B	B	3.16	22.6	2.75	.
155236367	G	A	exonic	rs374306700	c.1102C>T	p.Arg368Cys	Uncertain significance 1 points = 1 P – 0 B	Likely pathogenic (2); Uncertain significance (2)	T	D	P	6.006	27.9	3.67	.
155236376	C	T	exonic	rs2230288	c.1093G>A	p.Glu365Lys	Likely benign -1 points = 1 P – 2 B	Benign/Likely benign; risk factor	T	B	B	2.173	17.33	3.67	0.0128
155236459	T	C	exonic	rs1306645655	c.1010A>G	p.Asp337Gly	Uncertain significance 1 points = 1 P – 0 B	.	D	D	D	5.661	26.7	3.67	.
155236558	C	T	intronic	rs772645370	c.1000-89G>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0002
155236787	G	A	intronic	rs531447697	c.1000-318C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.00003263
155237162	G	A	intronic	rs547873878	c.999+179C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0003
155237222	A	C	intronic	rs946743963	c.999+119T>G	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.00003232
155237239	G	A	intronic	rs72704130	c.999+102C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0138
155237265	G	A	intronic	rs556277010	c.999+76C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.00003232
155237412	T	C	exonic	rs1057942	c.928A>G	p.Ser310Gly	Uncertain significance 3 points = 3 P – 0 B	Pathogenic/Likely pathogenic	T	P	B	1.77	14.81	3.51	0.00006461
155237438	C	T	exonic	rs140955685	c.902G>A	p.Arg301His	Uncertain significance 1 points = 1 P – 0 B	Uncertain significance	T	B	B	2.204	17.54	2.59	0.0003

155237596	A	T	intronic	rs140335079	c.762-18T>G	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.009
155237623	C	G	intronic	rs377217353	c.762-45G>C	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.00003236
155237914	A	G	intronic	rs549565365	c.761+220T>C	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0045
155238057	G	T	intronic	rs183540501	c.761+77C>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0031
155238175	G	A	exonic	rs376613535	c.720C>T	p.Pro240=	Likely benign -2 points = 1 P – 3 B	.	.	.	.	.	.	.	0.00006465
155238265	G	A	exonic	rs201615998	c.630C>T	p.Pro210=	Likely benign -2 points = 1 P – 3 B	.	.	.	.	.	.	.	.
155238570	C	G	exonic	rs147138516	c.535G>C	p.Asp179His	Uncertain significance 1 points = 1 P – 0 B	Likely pathogenic (1); Uncertain significance (2)	T	P	P	0.611	8.229	2.62	0.00009723
155238629	C	T	exonic	rs79653797	c.476G>A	p.Arg159Gln	Uncertain significance 3 points = 3 P – 0 B	Pathogenic/Likely pathogenic	D	D	D	6.336	29.3	3.55	.
155238630	G	A	exonic	rs439898	c.475C>T	p.Arg159Trp	Uncertain significance 3 points = 3 P – 0 B	Pathogenic	D	D	D	6.19	28.6	3.55	0.00003238
155238631	G	A	exonic	rs147411159	c.474C>T	p.Ile158=	Likely benign -2 points = 1 P – 3 B	Uncertain significance (1); Likely benign (3)	.	.	.	.	.	.	0.0006
155238833	A	G	intronic	rs188328778	c.455-183T>C	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0119
155238857	G	A	intronic	rs1042674060	c.455-207C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0001
155238927	A	G	intronic	rs778649863	c.455-277T>C	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155238984	C	G	intronic	rs752258174	c.455-334G>C	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155238985	G	A	intronic	rs951266434	c.455-335C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155239079	G	A	intronic	rs572108051	c.455-429C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0
155239287	C	T	intronic	rs1005434278	c.454+329G>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.00003262
155239509	C	T	intronic	rs570088632	c.454+107G>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155239633	G	T	exonic	rs758447515	c.437C>A	p.Ser146Ter	Likely pathogenic 9 points = 9 P – 0 B	.	.	.	.	10.665	36	3.25	.
155239858	C	A	intronic	rs369792423	c.307+28G>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.00006462
155239939	C	T	exonic	rs77829017	c.254G>A	p.Gly85Glu	Uncertain significance 1 points = 1 P – 0 B	Pathogenic	D	D	D	4.841	24.8	3.46	.
155239961	G	A	exonic	rs146774384	c.232C>T	p.Arg78Cys	Uncertain significance 1 points = 1 P – 0 B	.	T	D	P	4.822	24.8	3.46	0.00009697
155240072	G	C	exonic	.	c.121C>G	p.Arg41Gly	Uncertain significance 1 points = 1 P – 0 B	.	T	B	B	0.931	10.26	3.41	.
155240122	T	G	intronic	rs199565854	c.116-45A>C	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0003
155240171	C	T	intronic	rs114217696	c.116-94G>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.008

155240336	C	T	intronic	rs142348200	c.116-259G>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0012
155240779	T	C	intronic	rs2361534	c.28-62A>G	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.001
155240816	G	A	intronic	rs940168433	c.28-99C>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155241114	C	T	UTR5	rs1141801	c.-2G>T	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155241127	T	C	UTR5	rs41264927	c.-15A>G	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	0.0012
155241257	C	T	intronic	rs371157845	NM_001005742.3:c.-49-96G>A	.	Likely benign -1 points = 1 P – 2 B	.	.	.	.	.	.	.	.
155241315	T	C	intronic	rs188978150	NM_001005742.3:c.-49-154A>G	.	Likely benign -1 points = 1 P – 2 B	Uncertain Significance	.	.	.	.	.	.	0.0086

Variants with a gnomAD frequency >2% and without information on SNP ID or amino acid change were excluded. Pathogenicity scores were used from ACMG, Varsome, Clinvar, SIFT, Polphen2, CADD, and GERP++. Variants that were Sanger sequenced are highlighted in gray.

Supplementary Table S2: Rare *GBA1* variants (predicted as "pathogenic"/"likely pathogenic"/"uncertain significance") sequenced with the Oxford Nanopore and confirmed with Sanger sequencing

***GBA1* variant**	**cDNA transcript GBA(NM_000157.4)**	**Exon number**	**PD cases (*n*=462)**	**Controls (*n*=367)**
p.R78C	c.232C>T	3	2	0
p.S146X	c.437C>A	4	1	0
p.R159W	c.475C>T	5	12	1
p.R301H	c.902G>A	7	0	1
p.S310G	c.928A>G	7	1	1
p.D337G	c.1010A>G	8	2	1
p.E365K	c.1093G>A	8	33	10
p.R368C	c.1102C>T	8	2	0
p.A380T	c.1138G>A	8	1	0
p.T408M	c.1223C>T	8	10	13
p.N409S	c.1226A>G	9	13	4
p.L483P	c.1448T>C	10	6	0
p.G493D	c.1478G>A	10	1	0

Supplementary Table S3. Publications on *GBA1* variants and frequencies in patients with PD included in PubMed

Authors, Year (PMID)	*n* (PD/control)	Population/ Region	Method	*GBA1* variants screened*	*GBA1* variants found*	*n* mutation carriers/n total (frequency)*
Toft et al., 2006 (16476943)	311/474	Norwegian	Variant screening	L444P, N370S	L444P, N370S	L444P: PD: 3/311 (0.96%), Controls: 1/474 (0.21%) N370S: PD: 4/311 (1.29%), Controls: 7/474 (1.48%)
Lunde et al., 2018 (29792872)	442/419	Norwegian	Genotyping	N370S, T369M, E326K, V460L, Y135C, L444P	N370S, T369M, E326K, V460L, Y135C, L444P	All: PD: 53/442 (12.0%), Controls: 29/419 (6.9%) N370S: PD: 1/442 (0.2%), Controls: 1/419 (0.2%) T369M: PD: 7/442 (1.7%), Controls: 16/419 (3.6%) E326K: PD: 18/442 (4.3%), Controls: 29/419 (6.6%) V460L: PD: 1/442 (0.2%), Controls: 1/419 (0.2%) Y135C: PD: 0/442 (0%), Controls: 1/419 (0.2%) L444P: PD: 2/442 (0.5%), Controls: 6/419 (1.4%)
Berge-Seidl et al., 2017 (28830825)	1152/713	Scandinavian	Targeted deep sequencing; genotyping	All *GBA1* exons	E326K, T369M, N370S, R463C, IVS3+1G>A, V457A, G377D, W357R	E326K: PD: 20/330 (6.06%) T369M: PD: 13/366 (3.55%) N370S: PD: 1/339 (0.29%) R463C: PD: 1/366 (0.27%) IVS3+1G>A: PD: 1/366 (0.27%) V457A: PD: 1/366 (0.27%) G377D: PD: 1/366 (0.27%) W357R: PD: 1/366 (0.27%)
Ran et al., 2022 (35779693)	1131/1594	Swedish	Pyrosequencing for genotyping of T369M	T369M	T369M	T369M: PD: 47/1091 (4.31%), Controls: 50/1474 (3.39%)
Ran et al., 2016 (27255555)	1625/2025	Swedish	Genotyping by pyrosequencing of E326K, N370S, and L444P	E326K, N370S, L444P	E326K, N370S, L444P	E326K: PD: 90/1625 (5.54%), Controls: 65/2025 (3.21%) N370S: PD: 10/1625 (0.62%), Controls: 2/2025 (0.10%) L444P: PD: 35/1625 (2.15%), Controls: 65/2025 (0.15%)
Ylonen et al., 2017 (29029963)	852/403	Finnish	Variant screening for N370S and L444P; Whole exome sequencing in 225 EOPD cases	All *GBA1* exons; N370S, L444P	N370S, L444P	N370S: PD: 4/852 (0.5%), Controls: 1/403 (0.2%) L444P: PD: 17/852 (2.0%), Controls: 2/403 (0.5%)
Muldmaa et al., 2021 (32740907)	189/158	Estonian	Next-generation sequencing	NA	L444P, T369M, E326K, L276I, E10X	All: PD: 19/189 (10.1%), Controls: 6/158 (3.8%) *GBA1*-related risk variants: 18/189 (9.5%) L444P: PD: 1/189 (0.5%), Controls: 0/158 (0%) T369M: PD: 10/189 (5.3%), Controls: 3/158 (1.9%)

						E326K: PD: 6/189 (3.2%), Controls: 3/158 (1.9%) L276I: PD: 1/189 (0.5%), Controls: 0/158 (0%) E10X: PD: 1/189 (0.5%), Controls: 0/158 (0%)
Neumann et al., 2009 (19286695)	790/257	British	DNA sequencing of full *GBA1* gene	All *GBA1* exons and the flanking introns	L483P, D482N, R502C, RecNciI (L483P, A495P, V499V), RecA456P (L483P, A495P), N409S, D448H, D419A, N421PfsX4, R296Q, G232E, R170C, K46E, V497L	All: PD: 33/790 (4.18%), Controls: 3/257 (1.17%) L483P: PD: 11/790 (1.39%), Controls: 0/257 (0%) D482N: PD: 1/790 (0.13%), Controls: 0/257 (0%) R502C: PD: 3/790 (0.38%), Controls: 0/257 (0%) RecNciI (L483P, A495P, V499V): PD: 2/790 (0.25%), Controls: 0/257 (0%) RecA456P (L483P, A495P): PD: 1/790 (0.13%), Controls: 0/257 (0%) N409S: PD: 8/790 (1.01%), Controls: 1/257 (0.39%) D448H: PD: 1/790 (0.13%), Controls: 0/257 (0%) D419A: PD: 1/790 (0.13%), Controls: 0/257 (0%) N421PfsX4: PD: 1/790 (0.13%), Controls: 0/257 (0%) R296Q: PD: 1/790 (0.13%), Controls: 1/257 (0.39%) G232E: PD: 1/790 (0.13%), Controls: 0/257 (0%) R170C: PD: 1/790 (0.13%), Controls: 0/257 (0%) K46E: PD: 1/790 (0.13%), Controls: 0/257 (0%) V497L: PD: 0/790 (0%), Controls: 1/257 (0.39%)
Winder-Rhodes et al., 2013 (23413260)	259/0	British (2 South Asian)	DNA sequencing of all *GBA1* exons	All *GBA1* exons	L444P, N370S, N462K, R463C, R257Q, E326K, T369M, E388K, L119L	All: PD: 9/259 (3.5%) L444P: PD: 3/259 (1.2%) N370S: PD: 3/259 (1.2%) N462K: PD: 1/259 (0.4%) R463C: PD: 1/259 (0.4%) R257Q: PD: 1/259 (0.4%) E326K: PD: 8/259 (3.1%) T369M: PD: 5/259 (1.9%) E388K: PD: 1/259 (0.4%) L119L: PD: 1/259 (0.4%)
Duran et al., 2013 (23225227)	185/283	UK Caucasian	Sanger sequencing of the full *GBA1* gene	Full *GBA1* gene	N370S, L444P, RecNciI (L444P+A456P+V460V), R463C, E326K, IVS2+1, R131C, W184R, N188S, H255Q, R257Q, D409H, RecTL, E388K, G113A, T369M, S465P, L(−14)V, V172L, S177T, L217P, L317L, L354P, V375G, IVS10-4 C>T, IVS10-12 C>T, E340A, V458L	All: PD: 48/185 (25.94%), Controls: 12/283 (4.24%) All pathogenic: PD: 37/185 (20%), Controls: 9/283 (3.18%) N370S: PD: 5/185 (2.70%), Controls: 1/283 (0.35%) L444P: PD: 2/185 (1.08%), Controls: 0/283 (0%) RecNciI (L444P+A456P+V460V): PD: 3/185 (1.62%), Controls: 0/283 (0%) R463C: PD: 3/185 (1.62%), Controls: 0/283 (0%) E326K: PD: 14/185 (7.57%), Controls: 7/283 (2.47%), 6/202 (2.97%) IVS2+1: PD: 1/185 (0.54%), Controls: 0/283 (0%) R131C: PD: 2/185 (1.08%), Controls: 0/283 (0%) W184R: PD: 1/185 (0.54%), Controls: 0/283 (0%)

						N188S: PD: 1/185 (0.54%), Controls: 0/283 (0%) H255Q: PD: 1/185 (0.54%), Controls: 0/283 (0%) R257Q: PD: 1/185 (0.54%), Controls: 1/283 (0.35%) D409H: PD: 2/185 (1.08%), Controls: 0/283 (0%) RecTL: PD: 1/185 (0.54%), Controls: 0/283 (0%) E388K: PD: 1/185 (0.54%), Controls: 0/283 (0%) G113A: PD: 2/185 (1.08%), Controls: 0/283 (0%) T369M: PD: 1/185 (0.54%), Controls: 1/283 (0.35%) S465P: PD: 1/185 (0.54%), Controls: 0/283 (0%) L(−14)V: PD: 1/185 (0.54%), Controls: 0/283 (0%) V172L: PD: 2/185 (1.08%), Controls: 0/283 (0%) S177T: PD: 1/185 (0.54%), Controls: 0/283 (0%) L217P: PD: 1/185 (0.54%), Controls: 0/283 (0%) L317L: PD: 1/185 (0.54%), Controls: 0/283 (0%) L354P: PD: 1/185 (0.54%), Controls: 0/283 (0%) V375G: PD: 1/185 (0.54%), Controls: 0/283 (0%) IVS10-4 C>T: PD: 1/185 (0.54%), Controls: 0/283 (0%) IVS10-12 C>T: PD: 1/185 (0.54%), Controls: 0/283 (0%) E340A: PD: 0/185 (0%), Controls: 1/283 (0.35%) V458L: PD: 0/185 (0%), Controls: 1/283 (0.35%)
Olszewska et al., 2020 (32714263)	314/96 (friends or spouses)	Irish	DNA sequencing of all *GBA1* exons	All *GBA1* exons	T408M, E365K, F255Y, N409S, D448H, L483P, A495P, V499V, G416C, G234E, R301H, R368C	T408M: PD: 6/314 (1.91%), Controls: 4/96 (4.17%) E365K: PD: 13/314 (4.14%), Controls: 4/96 (4.17%) F255Y: PD: 1/314 (0.32%), Controls: 0/96 (0%) N409S: PD: 3/314 (0.96%), Controls: 0/96 (0%) D448H: PD: 1/314 (0.32%), Controls: 0/96 (0%) L483P: PD: 3/314 (0.96%), Controls: 0/96 (0%) A495P: PD: 3/314 (0.96%), Controls: 0/96 (0%) V499V: PD: 3/314 (0.96%), Controls: 0/96 (0%) G416C: PD: 1/314 (0.32%), Controls: 0/96 (0%) G234E: PD: 1/314 (0.32%), Controls: 0/96 (0%) R301H: PD: 1/314 (0.32%), Controls: 0/96 (0%) R368C: PD: 1/314 (0.32%), Controls: 0/96 (0%)
Crosiers et al., 2016 (27397011)	266/536	Flanders-Belgian	In-depth Sanger sequencing of all *GBA1* exons	All *GBA1* exons	D179H, Q256SfsX9, L363P, N409S, L483P, RecNcil (L483P-A495S-V499V), E365K, T408M, G39R, H529R	All rare: PD: 12/266 (4.5%), Controls: 2/536 (0.37%) D179H: PD: 1/266 (0.4%), Controls: 0/536 (0%) Q256SfsX9: PD: 1/266 (0.4%), Controls: 0/536 (0%) L363P: PD: 1/266 (0.4%), Controls: 0/536 (0%) N409S: PD: 3/266 (1.1%), Controls: 1/536 (0.2%) L483P: PD: 3/266 (1.1%), Controls: 1/536 (0.2%) RecNcil (L483P-A495S-V499V): PD: 1/266 (0.4%), Controls: 0/536 (0%) E365K: PD: 12/266 (4.5%), Controls: 15/536 (2.8%) T408M: PD: 3/266 (1.1%), Controls: 11/536 (2.0%)

						G39R: PD: 1/266 (0.4%), Controls: 0/536 (0%) H529R: PD: 1/266 (0.4%), Controls: 0/536 (0%)
den Heijer et al., 2020 (32618053)	3402/655	Dutch	Next-generation sequencing of full *GBA1* gene	Full *GBA1* gene	E-30Gfs*8, L-24S, L-24S+S23G, Q-7R, C18*, R39C, S45Rfs*15, R120W, D140H, R170H, A190T, G202R, F216Y, G250S, H255Q, I260T, L324P, G325R, E326K, R329C, W348G, Q350H, T369M, N370S, V375G, D380Y, E388K, N392S, D409H, L444P, D453L, V460M, R463P, S484L, S488T, H490R, L268=, S271G, A456P, V460=, S-1T, V459=, R496H, G390E, V17=, T61=, I119=, I130=, Q143=, G193=, G195=, G344=, T369=, P452=, c.762-5G>A, c.1000-4G>T	E-30Gfs*8: PD: 1/3402 (0.03%), Controls: 0/655 (0%) L-24S: PD: 1/3402 (0.03%), Controls: 0/655 (0%) L-24S+S23G: PD: 1/3402 (0.03%), Controls: 0/655 (0%) Q-7R: PD: 2/3402 (0.06%), Controls: 0/655 (0%) C18*: PD: 1/3402 (0.03%), Controls: 0/655 (0%) R39C: PD: 1/3402 (0.03%), Controls: 0/655 (0%) S45Rfs*15: PD: 1/3402 (0.03%), Controls: 0/655 (0%) R120W: PD: 5/3402 (0.15%), Controls: 0/655 (0%) D140H: PD: 84/3402 (2.47%), Controls: 6/655 (0.92%) R170H: PD: 2/3402 (0.06%), Controls: 0/655 (0%) A190T: PD: 1/3402 (0.03%), Controls: 0/655 (0%) G202R: PD: 1/3402 (0.03%), Controls: 0/655 (0%) F216Y: PD: 1/3402 (0.03%), Controls: 0/655 (0%) G250S: PD: 1/3402 (0.03%), Controls: 0/655 (0%) H255Q: PD: 2/3402 (0.06%), Controls: 0/655 (0%) I260T: PD: 2/3402 (0.06%), Controls: 0/655 (0%) L324P: PD: 2/3402 (0.06%), Controls: 1/655 (0.15%) G325R: PD: 1/3402 (0.03%), Controls: 0/655 (0%) E326K: PD: 314/3402 (9.23%), Controls: 18/655 (2.75%) R329C: PD: 2/3402 (0.06%), Controls: 0/655 (0%) W348G: PD: 1/3402 (0.03%), Controls: 0/655 (0%) Q350H: PD: 1/3402 (0.03%), Controls: 1/655 (0.15%) T369M: PD: 98/3402 (2.88%), Controls: 12/655 (1.83%) N370S: PD: 32/3402 (0.94%), Controls: 2/655 (0.31%) V375G: PD: 1/3402 (0.03%), Controls: 0/655 (0%) D380Y: PD: 1/3402 (0.03%), Controls: 0/655 (0%) E388K: PD: 3/3402 (0.09%), Controls: 0/655 (0%) N392S: PD: 1/3402 (0.03%), Controls: 0/655 (0%) D409H: PD: 1/3402 (0.03%), Controls: 0/655 (0%) L444P: PD: 26/3402 (0.76%), Controls: 0/655 (0%) D453L: PD: 5/3402 (0.15%), Controls: 0/655 (0%) V460M: PD: 1/3402 (0.03%), Controls: 0/655 (0%) R463P: PD: 2/3402 (0.06%), Controls: 1/655 (0.15%) S484L: PD: 1/3402 (0.03%), Controls: 0/655 (0%) S488T: PD: 1/3402 (0.03%), Controls: 0/655 (0%) H490R: PD: 1/3402 (0.03%), Controls: 0/655 (0%) L268=+S271G+D409H: PD: 1/3402 (0.03%), Controls: 0/655 (0%) RecTL (D409H+L444P+A456P+V460=): PD: 1/3402 (0.03%), Controls: 0/655 (0%) RecNciI (L444P+A456P+V460=): PD: 4/3402 (0.12%),

						Controls: 0/655 (0%) S-1T: PD: 1/3402 (0.03%), Controls: 0/655 (0%) V459=: PD: 5/3402 (0.15%), Controls: 0/655 (0%) R496H: PD: 1/3402 (0.03%), Controls: 0/655 (0%) G390E: PD: 1/3402 (0.03%), Controls: 1/655 (0.15%) V17=: PD: 0/3402 (0%), Controls: 1/655 (0.15%) T61=: PD: 1/3402 (0.03%), Controls: 0/655 (0%) I119=: PD: 5/3402 (0.15%), Controls: 0/655 (0%) I130=: PD: 1/3402 (0.03%), Controls: 0/655 (0%) Q143=: PD: 1/3402 (0.03%), Controls: 0/655 (0%) G193=: PD: 1/3402 (0.03%), Controls: 1/655 (0.15%) G195=: PD: 1/3402 (0.03%), Controls: 0/655 (0%) G344=: PD: 1/3402 (0.03%), Controls: 0/655 (0%) T369=: PD: 2/3402 (0.06%), Controls: 0/655 (0%) P452=: PD: 1/3402 (0.03%), Controls: 0/655 (0%) V460=: PD: 6/3402 (0.18%), Controls: 0/655 (0%) c.762-5G>A: PD: 1/3402 (0.03%), Controls: 0/655 (0%) c.1000-4G>T: PD: 0/3402 (0%), Controls: 1/655 (0.15%)
Anheim et al., 2012 (22282650)	525/71 (relatives)	French (88%)	DNA sequencing of all *GBA1* exons	All *GBA1* exons	N370S, L444P, F246L, G202R, R120W, R463C, R463H, S125N, S173SfsX50, S364N, T323I, Y304C, 1263-1217del55bp, E326K	All: PD (probands): 24/525 (4.6%) N370S: PD (probands): 10/525 (1.9%), Relatives with PD: 9/32 (28.1%), Relatives without PD: 12/71 (16.9%) L444P: PD (probands): 5/525 (1.0%), Relatives with PD: 6/32 (18.8%), Relatives without PD: 5/71 (7.0%) F246L: PD (probands): 1/525 (0.2%), Relatives with PD: 0/32 (0%), Relatives without PD: 0/71 (0%) G202R: PD (probands): 1/525 (0.2%), Relatives with PD: 0/32 (0%) Relatives without PD: 0/71 (0%) R120W: PD (probands): 1/525 (0.2%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 0/71 (0%) R463C: PD (probands): 1/525 (0.2%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 1/71 (1.4%) R463H: PD (probands): 1/525 (0.2%), Relatives with PD: 3/32 (9.4%), Relatives without PD: 3/71 (4.2%) S125N: PD (probands): 1/525 (0.2%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 0/71 (0%) S173SfsX50: PD (probands): 1/525 (0.2%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 3/71 (4.2%) S364N: PD (probands): 1/525 (0.2%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 0/71 (0%) T323I: PD (probands): 1/525 (0.2%), Relatives with PD:1/32 (3.1%), Relatives without PD: 1/71 (1.4%) Y304C: PD (probands): 1/525 (0.2%), Relatives with PD: 2/32 (6.3%), Relatives without PD: 6/71 (8.5%)

						1263-1217del55bp: PD (probands): 1/525 (0.2%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 0/71 (0%) E326K: PD (probands): 0/525 (0%), Relatives with PD: 1/32 (3.1%), Relatives without PD: 0/71 (0%)
Lesage et al., 2011a (20947659)	1130/391 (mainly spouses)	French (89%)	DNA sequencing of full *GBA1* gene	*GBA1* exons and flanking introns	K(227)R, K79M, G80R, I119L, R120W, S125N, R131C, S173SfsX50, G202R, P246L, Y304C, T323I, R329C, S364N, N370S, G377S, E388K, D409H, L444P, P452L, R463C, R463H, G113A, A446A, RecΔ5, RecNciI (L444P+A456P+V460V), RecA456P (L444P+A456P), c.1263del+RecTL (c.1263–1317del+D409H+L444P+A456P+V460V), D140H, E326K, A190A, Y313Y	K(227)R: PD: 1/1130 (0.09%), Controls: 0/391 (0%) K79M: PD: 0/1130 (0%), Controls: 1/391 (0.26%) G80R: PD: 1/1130 (0.09%), Controls: 0/391 (0%) I119L: PD: 1/1130 (0.09%), Controls: 0/391 (0%) R120W: PD: 1/1130 (0.09%), Controls: 0/391 (0%) S125N: PD: 1/1130 (0.09%), Controls: 0/391 (0%) R131C: PD: 1/1130 (0.09%), Controls: 0/391 (0%) S173SfsX50: PD: 1/1130 (0.09%), Controls: 0/391 (0%) G202R: PD: 2/1130 (0.18%), Controls: 0/391 (0%) P246L: PD: 1/1130 (0.09%), Controls: 0/391 (0%) Y304C: PD: 1/1130 (0.09%), Controls: 0/391 (0%) T323I: PD: 1/1130 (0.09%), Controls: 0/391 (0%) R329C: PD: 2/1130 (0.18%), Controls: 0/391 (0%) S364N: PD: 1/1130 (0.09%), Controls: 0/391 (0%) N370S: PD: 37/1130 (3.27%), Controls: 2/391 (0.51%) G377S: PD: 1/1130 (0.09%), Controls: 0/391 (0%) E388K: PD: 1/1130 (0.09%), Controls: 1/391 (0.26%) D409H: PD: 1/1130 (0.09%), Controls: 0/391 (0%) L444P: PD: 13/1130 (1.15%), Controls: 0/391 (0%) P452L: PD: 1/1130 (0.09%), Controls: 0/391 (0%) R463C: PD: 1/1130 (0.09%), Controls: 0/391 (0%) R463H: PD: 1/1130 (0.09%), Controls: 0/391 (0%) G113A/A446A: PD: 1/1130 (0.09%), Controls: 0/391 (0%) RecΔ5: PD: 2/1130 (0.18%), Controls: 0/391 (0%) RecNciI (L444P+A456P+V460V): PD: 2/1130 (0.18%), Controls: 0/391 (0%) RecA456P (L444P+A456P): PD: 1/1130 (0.09%), Controls: 0/391 (0%) c.1263del+RecTL (c.1263–1317del+D409H+L444P+A456P+V460V): PD: 1/1130 (0.09%), Controls: 0/391 (0%) A190A: PD: 1/1130 (0.09%), Controls: 0/391 (0%) Y313Y: PD: 0/1130 (0%), Controls: 1/391 (0.26%) E326K: PD: 49/1130 (4.34%), Controls: 8/391 (2.04%) T369M: PD: 17/1130 (1.50%), Controls: 1/391 (0.26%)

Spataro et al., 2017 (28124432)	249/145	Spanish	Targeted resequencing and CNF detection by eXome-Hidden Markov Model (XHMM) software	NA	N370S	N370S: PD: 1/249 (0.4%), Controls: 0/145 (0%)
Seto-Salvia et al., 2012 (22173904)	225/186	Spanish	Cycle sequencing of *GBA1* coding region	All *GBA1* exons	N370S, L444P, L144V, S488T, M123T, G202R, I260T, T369M, W393R, D409H, RecNcil	N370S: PD: 5/225 (2.22%), Controls: 0/186 (0%) L444P: PD: 6/225 (2.67%), Controls: 0/186 (0%) L144V: PD: 1/225 (0.44%), Controls: 0/186 (0%) S488T: PD: 1/225 (0.44%), Controls: 0/186 (0%) M123T: PD: 1/225 (0.44%), Controls: 0/186 (0%) G202R: PD: 1/225 (0.44%), Controls: 0/186 (0%) I260T: PD: 1/225 (0.44%), Controls: 0/186 (0%) T369M: PD: 2/225 (0.89%), Controls: 1/186 (0.5%) W393R: PD: 1/225 (0.44%), Controls: 0/186 (0%) D409H: PD: 2/225 (0.89%), Controls: 0/186 (0%) RecNcil: PD: 1/225 (0.44%), Controls: 0/186 (0%)
Jesus et al., 2016 (28030538)	532/542	Southern Spanish	High-resolution melting (HRM) analysis and direct DNA resequencing	Full *GBA1* gene	N370S, L444P, W312R, V457D, E326K, T369M, etc.	All variants: PD: (12.2%), Controls: (7.9%) N370S: PD: 5/532 (0.94%), Controls: 0/542 (0%) L444P: PD: 13/532 (2.44%), Controls: 6/542 (1.11%) W312R: PD: 6/532 (1.13%), Controls: 2/542 (0.37%) V457D: PD: 3/532 (0.56%), Controls: 4/542 (0.74%) c.116-8C>T: PD: 4/532 (0.75%), Controls: 7/542 (1.29%) E326K: PD: 16/532 (3.00%), Controls: 13/542 (2.40%) T369M: PD: 5/532 (0.94%), Controls: 2/542 (0.37%)
Bras et al., 2009 (18160183)	230/430	Portuguese	DNA sequencing of the complete open-reading frame, as well as intron/exon boundaries, of the *GBA1* gene	All coding exons and exon/intron boundaries of the *GBA1* gene	N409S, N435T, D448H, L483P, K13R, R41L, E365K, T408M, E427K	Pathogenic variants: PD: 14/230 (6.1%), Controls: 3/430 (0.7%) N409S: PD: 5/230 (2.2%), Controls: 3/430 (0.7%) N435T: PD: 5/230 (2.2%), Controls: 0/430 (0%) D448H: PD: 1/230 (0.4%), Controls: 0/430 (0%) L483P: PD: 3/230 (1.3%), Controls: 0/430 (0%) K13R: PD: 1/230 (0.4%), Controls: 0/430 (0%) R41L: PD: 0/230 (0%), Controls: 1/430 (0.2%) E365K: PD: 2/230 (0.9%), Controls: 3/430 (0.7%) T408M: PD: 2/230 (0.9%), Controls: 5/430 (1.2%) E427K: PD: 0/230 (0%), Controls: 2/430 (0.5%)
Petrucci et al., 2020 (32658388)	874/0	Italian	Whole exome sequencing	All *GBA1* exons	D24N, S107L, R120W, R131C, P182L, N188S, G202R, H255Q, D409H, L444P, R463C,	All: PD: 125/874 (14.3%) N370S: PD: 30/874 (3.43%) L444P: PD: 29/874 (3.32%) E326K: PD: 16/874 (1.83%)

					W209Gfs*6, R257*, E388K, S196P, G202R, H255Q, D409H, T369M, L444P, A456P, V460V, G46E, G193R, R329C, N370S, E326K, N188K, W184R, I161N, K(-27)R, M85V, E326D, T369T, V460L	
De Marco et al., 2008 (18074383)	395/483	Italian	Genotyping	L444P, N370S	L444P, N370S	All: PD: 11/395 (2.8%), Controls: 1/483 (0.2%) L444P: PD: 8/395 (2.0%), Controls: 1/483 (0.2%) N370S: PD: 3/395 (0.8%), Controls: 0/483 (0%)
Asselta et al., 2014 (25249066)	2350/1111	Italian	High-resolution melting (HRM) analysis (exon 9) and direct DNA sequencing (exon 10)	*GBA1* exons 9 and 10	IVS8-24T>G, N370S, E388K, IVS9+32C>T, IVS9-36C>G, IVS9-5T>A, D443N, L444P, IVS10+1G>T, IVS10+8C>A	N370S or D443N or L444P or IVS10+1G>T: PD: 106/2350 (4.5%), Controls: 7/1111 (0.63%) N370S: PD+DLB+MSA+PSP+CBD: 69/2766 (2.5%), Controls: 4/1111 (0.36%) L444P: PD+DLB+MSA+PSP+CBD: 47/2766 (1.7%), Controls: 3/1111 (0.27%)
Cilia et al., 2016 (27632223)	2843/0	Italian	Variant screening of *GBA1* exons 9 and 10	*GBA1* exons 9 and 10	N370S, L444P, G377S, IVS10+1G>T	N370S: PD: 70/2843 (2.46%) L444P: PD: 54/2843 (1.90%) G377S: PD: 1/2843 (0.04%) IVS10+1G>T: PD: 1/2843 (0.04%)
Straniero et al., 2020 (33209983)	3691/7757 (1625 partners and caregivers of patients with PD)	Italian	Variant screening	E326K, T369M, N370S, L444P	E326K, T369M, N370S, L444P	E326K: PD: 61/3691 (1.65%), Controls: 55/7755 (0.71%) T369M: PD: 49/3691 (1.33%), Controls: 61/7755 (0.79%) N370S: PD: 76/3691 (2.06%), Controls: 43/7755 (0.55%) L444P: PD: 62/3691 (1.68%), Controls: 11/7755 (0.14%)
Quadri et al., 2015 (25294124)	100/0	Sardinian	Whole exome sequencing	All *GBA1* exons	N370S, R131C	N370S: PD: 4/100 (4%) R131C: PD: 2/100 (2%)
Kalinderi et al., 2009 (19383421)	172/132	Greek	DNA sequencing of all *GBA1* exons	All *GBA1* exons	L444P, D409H, E326K, H255Q, R329H, L268L, S271G, T428K, V460L	L444P: PD: 2/172 (1.2%), Controls: 0/132 (0%) D409H: PD: 1/172 (0.6%), Controls: 0/132 (0%) E326K: PD: 1/172 (0.6%), Controls: 1/132 (0.8%) H255Q: PD: 4/172 (2.3%), Controls: 0/132 (0%) R329H: PD: 1/172 (0.6%), Controls: 0/132 (0%) L268L+S271G: PD: 1/172 (0.6%), Controls: 0/132 (0%) T428K: PD: 1/172 (0.6%), Controls: 1/132 (0.8%) V460L: PD: 0/172 (0%), Controls: 4/132 (3.0%)
Moraitou et al., 2011 (21745757)	205/206	Greek	Restriction enzyme	N370S, D409H, L444P,	N370S, D409H, L444P, H255Q, Y108C, IVS10-1G>A, IVS6-2A>G	N370S: PD: 6/205 (2.93%), Controls: 4/206 (1.94%) D409H: PD: 7/205 (3.41%), Controls: 0/206 (0%) L444P: PD: 6/205 (2.93%), Controls: 1/206 (0.49%)

			analysis for eight variants	H255Q, R120W, Y108C, IVS10-1G>A, IVS6-2A>G		H255Q: PD: 7/205 (3.41%), Controls: 1/206 (0.49%) Y108C: PD: 0/205 (0%), Controls: 1/206 (0.49%) IVS10-1G>A: PD: 1/205 (0.49%), Controls: 0/206 (0%)
Emekli et al., 2021 (34781237)	82/0	Turkish	Next-generation sequencing	All *GBA1* exons and intron/exon boundaries	R434P, H294Q, D448H, G241R, N227K	R434P: PD: 2/82 (2.4%) H294Q: PD: 1/82 (1.2%) D448H: PD: 1/82 (1.2%) G241R: PD: 1/82 (1.2%) N227K: PD: 1/82 (1.2%)
Kumar et al., 2013 (22812582)	360/348	Serbian	DNA sequencing of *GBA1* exons 8-11	*GBA1* exons 8-11	N370S, D409H, H255Q, L444P, A456P, R463C, RecNciI, T369M, E388K, D380V, N392S, V459V	All: PD: 21/360 (5.8%), Controls: 5/348 (1.4%) N370S: PD: 9/360 (2.5%), Controls: 0/348 (0%) D409H, H255Q: PD: 7/360 (1.9%), Controls: 2/348 (0.6%) L444P: PD: 2/360 (0.6%), Controls: 1/348 (0.3%) A456P: PD: 0/360 (0%), Controls: 1/348 (0.3%) R463C: PD: 1/360 (0.3%), Controls: 0/348 (0%) RecNciI (L444P+A456P+V460V): PD: 1/360 (0.3%), Controls: 0/348 (0%) T369M: PD: 8/360 (2.08%), Controls: 6/348 (1.72%) E388K: PD: 0/360 (0%), Controls: 1/348 (0.3%) D380V: PD: 1/360 (0.3%), Controls: 0/348 (0%) N329S: PD: 1/360 (0.3%), Controls: 0/348 (0%) V459V: PD: 0/360 (0%), Controls: 1/348 (0.3%)
Torok et al., 2016 (26547032)	124/122	Hungarian	Variant screening	L444P, N370S, R120W	L444P	L444P: PD: 3/124 (2.4%), Controls: 0/122 (0%)
Benitez et al., 2016 (27094865)	478/337	European-American	Deep-sequencing of all *GBA1* exons	All *GBA1* exons	R83C, H294Q, T336S, E365K, T408M, N409S, E427K, D448H, L483P, A495P	R83C: PD: 2/478 (0.41%), Controls: 0/337 (0%) H294Q: PD: 2/478 (0.41%), Controls: 0/337 (0%) T336S: PD: 1/478 (0.21%), Controls: 0/337 (0%) E365K: PD: 19/478 (3.97%), Controls: 11/337 (3.26%) T408M: PD: 17/478 (3.56%), Controls: 0/337 (0%) N409S: PD: 7/478 (1.46%), Controls: 1/337 (0.30%) E427K: PD: 1/478 (0.21%), Controls: 0/337 (0%) D448H: PD: 1/478 (0.21%), Controls: 1/337 (0.30%) L483P: PD: 7/478 (1.46%), Controls: 2/337 (0.59%) A495P: PD: 17/478 (3.56%), Controls: 10/337 (2.97%)
Noreau et al., 2011 (21856586)	212/189	French-Canadian	Sequencing of the entire coding region of *GBA1*	All *GBA1* exons	L197F, E326K, S339L, T369M, N370S, W378G, L444P	L197F: PD: 1/212 (0.47%), Controls: 0/189 (0%) E326K: PD: 5/212 (2.36%), Controls: 3/189 (1.59%) S339L: PD: 1/212 (0.47%), Controls: 0/189 (0%) T369M: PD: 9/212 (4.25%), Controls: 5/189 (2.65%) N370S: PD: 0/212 (0%), Controls: 2/189 (1.06%)

						W378G: PD: 1/212 (0.47%), Controls: 0/189 (0%) L444P: PD: 5/212 (2.36%), Controls: 1/189 (0.53%)
Han et al., 2016 (26000814)	225/110 (spouses)	Canadian	DNA sequencing of full *GBA1* gene	All *GBA1* exons and flanking introns	c.-119A/G, S(-35)N, R120W, N370S, L444P, RecNciI, RecTL (del55/D409H/RecNciI), E326K, T369M, S13L	All: PD: 25/225 (11.11%), Controls: 9/110 (8.19%) c.-119A/G: PD: 1/225 (0.44%), Controls: 0/110 (0%) S(-35)N: PD: 1/225 (0.44%), Controls: 0/110 (0%) R120W: PD: 1/225 (0.44%), Controls: 0/110 (0%) N370S: PD: 2/225 (0.89%), Controls: 0/110 (0%) L444P: PD: 4/225 (1.78%), Controls: 0/110 (0%) RecNciI (L444P-A456P-V460V): PD: 1/225 (0.44%), Controls: 0/110 (0%) RecTL (del55/D409H/RecNciI): PD: 2/225 (0.89%), Controls: 0/110 (0%) E326K: PD: 4/225 (1.78%), Controls: 4/110 (3.64%) T369M: PD: 11/225 (4.89%), Controls: 4/110 (3.64%) S13L: PD: 0/225 (0%), Controls: 1/110 (0.91%)
Sato et al., 2005 (15517592)	88/122	Canadian	Genotyping	N370S, L444P, IVS2+1, K198T, R329C, 84insGG, Rec	N370S, L444P, Rec	N370S: PD: 1/88 (1.14%), Controls: 1/122 (0.82%) L444P: PD: 1/88 (1.14%), Controls: 0/122 (0%) Rec: PD: 3/88 (3.41%), Controls: 0/122 (0.%)
Gonzalez-Del Rincon Mde et al., 2013 (23448517)	128/252 (128 sex and age matched, 124 (aged >60))	Mexican Mestizo	Variant screening	N370S, L444P	L444P	L444P: PD: 7/128 (5.47%), Controls: 0/252 (0%)
Tipton et al., 2020 (32197197)	209/58	Colombian and Hispanic American	Variant screening	K198E	K198E	Colombian: K198E: PD: 3 (2.1%), Controls: 1 (1.7%) Hispanic American: K198E: PD: 0 (0%)
Velez-Pardo et al., 2019 (30765263)	602/319	Colombian, Peruvian	DNA sequencing of all *GBA1* exons and intron/exon boundaries	All *GBA1* exons and intron/exon boundaries	R86X, R159W, R170C, G234W, K237E, N409S, L483P, L483P + RecG or L483P + Rec6b (L483P, +92G>A), Rec1 (L483P, A495P, V499V), RecD, E, or AZRecTL (L483P, A495P, V499V, +92G>A), D66H, R250K, R316H, M400I, E427K, D482N, I528V, R534H, K13R, E365K, T408M	Colombian: R86X: PD: 0/131 (0%), Controls: 0/164 (0%) R159W: PD: 0/131 (0%), Controls: 0/164 (0%) R170C: PD: 0/131 (0%), Controls: 0/164 (0%) G234W: PD: 0/131 (0%), Controls: 1/164 (0.6%) K237E: PD: 7/131 (5.3%), Controls: 2/164 (1.2%) N409S: PD: 3/131 (2.3%), Controls: 0/164 (0%) L483P: PD: 3/131 (2.3%), Controls: 0/164 (0%) L483P + RecG or L483P + Rec6b (L483P, +92G>A): PD: 0/131 (0%), Controls: 0/164 (0%) Rec1 (L483P, A495P, V499V): PD: 0/131 (0%), Controls: 0/164 (0%) RecD, E, or AZRecTL (L483P, A495P, V499V, +92G>A):

						PD: 0/131 (0%), Controls: 0/164 (0%) D66H: PD: 0/131 (0%), Controls: 0/164 (0%) R250K: PD: 0/131 (0%), Controls: 1/164 (0.6%) R316H: PD: 0/131 (0%), Controls: 0/164 (0%) M400I: PD: 0/131 (0%), Controls: 0/164 (0%) E427K: PD: 0/131 (0%), Controls: 1/164 (0.6%) D482N: PD: 0/131 (0%), Controls: 0/164 (0%) I528V: PD: 0/131 (0%), Controls: 0/164 (0%) R534H: PD: 0/131 (0%), Controls: 1/164 (0.6%) K13R: PD: 1/131 (0.8%), Controls: 1/164 (0.6%) E365K: PD: 2/131 (1.5%), Controls: 1/164 (0.6%) T408M: PD: 0/131 (0%), Controls: 0/164 (0%) Peruvian: R86X: PD: 1/471 (0.2%), Controls: 0/155 (0%) R159W: PD: 2/471 (0.4%), Controls: 0/155 (0%) R170C: PD: 3/471 (0.6%), Controls: 0/155 (0%) G234W: PD: 0/471 (0%), Controls: 0/155 (0%) K237E: PD: 0/471 (0%), Controls: 0/155 (0%) N409S: PD: 1/471 (0.2%), Controls: 1/155 (0.6%) L483P: PD: 7/471 (1.5%), Controls: 0/155 (0%) L483P + RecG or L483P + Rec6b (L483P, +92G>A): PD: 1/471 (0.2%), Controls: 0/155 (0%) Rec1 (L483P, A495P, V499V): PD: 4/471 (0.8%), Controls: 1/155 (0.6%) RecD, E, or AZRecTL (L483P, A495P, V499V, +92G>A): PD: 1/471 (0.2%), Controls: 0/155 (0%) D66H: PD: 1/471 (0.2%), Controls: 0/155 (0%) R250K: PD: 0/471 (0%), Controls: 0/155 (0%) R316H: PD: 1/471 (0.2%), Controls: 0/155 (0%) M400I: PD: 0/471 (0%), Controls: 1/155 (0.6%) E427K: PD: 0/471 (0%), Controls: 0/155 (0%) D482N: PD: 1/471 (0.2%), Controls: 0/155 (0%) I528V: PD: 1/471 (0.2%), Controls: 0/155 (0%) R534H: PD: 0/471 (0%), Controls: 0/155 (0%) K13R: PD: 3/471 (0.6%), Controls: 0/155 (0%) E365K: PD: 5/471 (1.1%), Controls: 0/155 (0%) T408M: PD: 3/471 (0.6%), Controls: 0/155 (0%)
Eblan et al., 2006 (16261622)	33/31	Venezuelan	DNA sequencing of all *GBA1* exons and	All *GBA1* exons and most flanking introns	N370S, L444P, RecNciI, D443N	N370S: PD: 1/33 (3.0%), Controls: 0/31 (0%) L444P: PD: 1/33 (3.0%), Controls: 0/31 (0%) RecNciI: PD: 2/33 (6.1%), Controls: 0/31 (0%) D443N: PD: 0/33 (0%), Controls: 1/31 (3.32%)

			most flanking introns			
Dos Santos et al., 2010 (20816920)	110/155	Brazilian	Variant screening	N370S, L444P, 84GG, IVS2+1G>A, G377S	N370S, L444P, D409H+L444P+A456P+V460V, IVS2+1G>A	N370S: PD: 2/110 (1.8%), Controls: 0/155 (0%) L444P: PD: 2/110 (1.8%), Controls: 0/155 (0%) D409H+L444P+A456P+V460V: PD: 1/110 (0.9%), Controls: 0/155 (0%) IVS2+1G>A: PD: 1/110 (0.9%), Controls: 0/155 (0%)
Spitz et al., 2008 (17703984)	65/267	Brazilian	Variant screening	L444P, N370S	L444P	L444P: PD: 2/65 (3.1%), Controls: 0/267 (0%)
Guimaraes Bde et al., 2012 (22192918)	237/186	Brazilian	Direct sequencing	N370S, L444P	N370S, L444P	L444P: PD: 3/237 (1.27%) N370S: PD: 6/237 (2.53%)
Socal et al., 2009 (18358758)	62/0	Brazilian (with mixed ethnic backgrounds)	Variant screening	L444P, N370S, IVS2+1, 84GG	L444P, N370S	L444P: PD: 1/62 (1.61%) N370S: PD: 1/62 (1.61%)
Barkhuizen et al., 2017 (28361101)	105/40	Caucasian/South African (82.7% Afrikaner)	Sanger sequencing of *GBA1* exons 8-11 in all participants; direct sequencing of all *GBA1* exons in 20 PD cases	All *GBA1* exons; *GBA1* exons 8-11	G35A, E326K, I368T, T369M, N370S, P387L, K441N	G35A: PD: 1/20 (5.0 %) E326K: PD: 5/105 (4.8%), Controls: 1/40 (2.5%) I368T: PD: 1/105 (1.0%), Controls: 0/40 (0%) T369M: PD: 2/105 (1.9%), Controls: 1/40 (2.5%) N370S: PD: 1/105 (1.0%), Controls: 0/40 (0%) P387L: PD: 2/105 (1.9%), Controls: 0/40 (0%) K441N: PD: 1/105 (1.0%), Controls: 0/40 (0%)
Mahungu et al., 2020 (32035846)	30/0	Black South African	Sanger sequencing of all *GBA1* exons	All *GBA1* exons	K13R, T75del, R159W, R170L, F255L, Q536, G517R, Q471Q	K13R: PD: 6/30 (20%) T75del: PD 2/30 (6.6%) R159W: PD: 1/30 (3.3%) R170L: PD: 1/30 (3.3%) F255L: PD: 6/30 (6.6%) Q536: PD 1/30 (together with F255L) (3.3%) G517R: PD 1/30 (3.3%) Q471Q: PD: 1/30 (3.3%)
Lesage et al., 2011b (21242499)	194/177	North African (PD: Algeria: n=147, Morocco: n=23, Tunisia: n=14, Libya: n=1, unknown:	DNA sequencing of *GBA1* coding regions	All *GBA1* exons	K(-27)R, R131C, N370S, L444P, E326K, RecNciI (A456P/V460V/L444P), D443N, T369M	K(-27)R: PD: 2/194 (1.03%), Controls: 0/177 (0%) R131C: PD: 2/194 (1.03%), Controls: 0/177 (0%) N370S: PD: 2/194 (1.03%), Controls: 0/177 (0%) L444P/E326K: PD: 1/194 (0.52%), Controls: 0/177 (0%) RecNciI (A456P/V460V/L444P): PD: 2/194 (1.03%), Controls: 0/177 (0%) D443N: PD: 0/194 (0%), Controls: 1/177 (0.56%) E326K: PD: 2/194 (1.03%), Controls: 1/177 (0.56%) T369M: PD: 2/194 (1.03%), Controls: 0/177 (0%)

		n=9; Controls: Algeria: n=95, Morocco: n=46)				
Nishioka et al., 2010 (19945510)	395/372	North African Arab-Berber	DNA sequencing of all *GBA1* exons	All *GBA1* exons	K13R, K225R, N370S	K13R: PD: 2/33 (3.0%) K225R: PD: 1/33 (6.1%) K13R: PD: 5/155 (familial, including the 33) (3.2%), 9/240 (sporadic) (3.8%), Controls: 16/372 (4.3%) K225R: PD: 2/155 (familial, including the 33) (1.3%), 0/240 (sporadic) (0%), Controls: 0/372 (0%) N370S: PD: 0/155 (familial, including the 33) (0%), 1/240 (sporadic) (0.4%), Controls: 3/372 (0.8%)
Emelyanov et al., 2018 (30146349)	762/400	Russian	Variant screening	L444P, N370S, E326K, T369M	L444P, N370S, E326K, T369M	L444P: PD: 9/762 (1.1%), Controls: 1/400 (0.1%) N370S: PD: 4/762 (0.5%), Controls: 0/400 (0%) E326K: PD: 14/762 (2.4%), Controls: 5/400 (1.3%) T369M: PD: 15/762 (2.5%), Controls: 4/400 (1.1%)
Emelyanov et al., 2012 (21915911)	330/240	Russian	Variant screening	L444P, N370S	L444P, N370S	L444P: PD: 6/330 (1.8%), Controls: 1/240 (0.5%) N370S: PD: 3/330 (0.9%), Controls: 0/240 (0%)
Mao et al., 2010 (20004703)	616/411	Han-Chinese	Variant screening	L444P	L444P	L444P: PD: 20/616 (3.2%), Controls: 1/411 (0.2%)
Hu et al., 2010 (20528910)	328/300	Han Chinese	Genotyping	N370S	N370S	N370S: PD: 6/328 (1.8%), Controls: 2/300 (0.7%)
Zhang et al., 2012 (23286447)	195/443	Han Chinese	Genotyping	L444P, N370S, R120W	L444P	L444P: PD: 6/195 (3.08%), Controls: 0/443 (0%)
Guo et al., 2015 (25623333)	1019/1030	Han-Chinese	Genotyping	L444P	L444P	L444P: PD: 26/1019 (2.7%), Controls: 1/1030 (0.1%)
Wang et al., 2014 (24095219)	1638/0	Han Chinese	Genotyping	L444P	L444P	L444P: PD: 49/1638 (2.99%)
Ren et al., 2022 (34951095)	737/0	Chinese	Next-generation sequencing	Full *GBA1* gene	N370S, E326K, T369M, R163Q, L444P, R120W, etc.	All variants: PD: 79/737 (10.72%) Mild (e.g. N370S): PD: 8/737 (1.09%) Severe (e.g. L444P): PD: 28/737 (3.80%) Risk (e.g. E326K): PD: 1/737 (0.14%) Complex (e.g. L444P-A456P-V460V): PD: 7/737 (0.95%) Unknown: PD: 35/737 (4.75%) R163Q: PD: 12/737 (1.63%) L444P: PD: 10/737 (1.36%) R120W: PD: 6/737 (0.81%)
Yu et al., 2015 (25518742)	184/130	Chinese	DNA sequencing of	All *GBA1* exons	R163Q, F213I, E326K, S364S, F347L, V375L,	All: PD: 16/184 (8.7%), Controls: 2/130 (1.54%) R163Q: PD: 1/184 (0.54%), Controls: 0/130 (0%)

			all *GBA1* exons		L444P, RecNciI (L444P-A456P-V460V), A456P Q497R, c.334_338delCAGAA, L264I, L314V	F213I: PD: 1/184 (0.54%), Controls: 0/130 (0%) E326K: PD: 1/184 (0.54%), Controls: 0/130 (0%) S364S: PD: 1/184 (0.54%), Controls: 0/130 (0%) F347L: PD: 1/184 (0.54%), Controls: 1/130 (0.77%) V375L: PD: 1/184 (0.54%), Controls: 0/130 (0%) L444P: PD: 2/184 (1.09%), Controls: 0/130 (0%) RecNciI (L444P-A456P-V460V): PD: 3/184 (1.63%), Controls: 0/130 (0%) A456P: PD: 0/184 (0%), Controls: 1/130 (0.77%) Q497R: PD: 1/184 (0.54%), Controls: 0/130 (0%) c.334_338delCAGAA: PD: 1/184 (0.54%), Controls: 0/130 (0%) L264I: PD: 2/184 (1.09%), Controls: 0/130 (0%) L314V: PD: 1/184 (0.54%), Controls: 0/130 (0%)
Sun et al., 2010 (20131388)	402/413	Chinese	Variant screening	L444P, F213I, R353W, N370S	L444P	L444P: PD: 11/402 (2.74%), Controls: 0/413 (0%)
Wang et al., 2012 (23227814)	208/298	Chinese	Variant screening	L444P, N370S, R120W	L444P	L444P: PD: 7/208 (3.4%), Controls: 1/298 (0.3%) N370S: PD: 0/208 (0%), Controls: 0/298 (0%) R120W: PD: 0/208 (0%), Controls: 0/298 (0%)
Tan et al., 2007 (17620502)	331/347	Chinese	Allelic discrimination using the 5' nuclease activity assay, adapted to detect the L444P and N370S variants	L444P, N370S	L444P	L444P: PD: 8/331 (2.4%), Controls: 0/347 (0%)
Li et al., 2020 (32171587)	240/0	Chinese	Whole-exome sequencing	All *GBA1* exons	IVS2+1, G202R, D409H, L444P, R163Q, Y205C, V499M	IVS2+1: PD: 1/240 (0.4%) G202R: PD: 1/240 (0.4%) D409H: PD: 2/240 (0.8%) L444P: PD: 1/240 (0.4%) R163Q: PD: 1/240 (0.4%) Y205C: PD: 1/240 (0.4%) V499M: PD: 1/240 (0.4%)
Huang et al., 2011 (21338444)	967/780 (spouses, patients with unrelated diseases,	Chinese	DNA sequencing of whole *GBA1* coding region (in 30 PD	All *GBA1* exons; L444P, D409H, R120W,	L444P, D409H, RecNciI (L444P-A456P-V460V)	All: PD: 36/967 (3.72%), Controls: 2/780 (0.26%) L444P: PD: 27/967 (2.79%), Controls: 1/780 (0.13%) D409H: PD: 2/967 (0.21%), Controls: 0/780 (0%) RecNciI (L444P-A456P-V460V): PD: 7/967 (0.72%), Controls: 1/780 (0.13%)

	healthy volunteers)		cases); Genotyping of L444P, D409H, R120W, L174P, Q497R in all participants	L174P, Q497R		
Zhang et al., 2015 (26421210)	1147/0	Chinese	Variant screening	L444P	L444P	L444P: PD: 34/1147 (2.96%)
Foo et al., 2014 (24565865)	1085/9445	Chinese, Korean	Sequencing of all *GBA1* exons in EOPD cases and matching controls; genotyping in LOPD cases and matching controls	All *GBA1* exons	Various, including S350R, V221I, P210fs, c.281+1G>A	Chinese EOPD: All rare and low frequency variants: PD: 4/195 (2.05%), Controls: 3/219 (1.37%) Korean EOPD: All rare and low frequency variants: PD: 18/180 (10.0%), Controls: 6/180 (3.33%) Chinese LOPD: All rare and low frequency variants: PD: 0/710 (0%), Controls: 3/9046 (0.03%)
Ziegler et al., 2007 (17462935)	92/92	Chinese from Taiwan	Direct sequencing of full *GBA1* gene	All *GBA1* exons and flanking introns	L444P, D409H, L174P, Q497R, V460M	L444P: PD: 1/92 (1.1%), Controls: 0/92 (0%) D409H: PD: 1/92 (1.1%), Controls: 0/92 (0%) L174P: PD: 1/92 (1.1%), Controls: 0/92 (0%) Q497R: PD: 1/92 (1.1%), Controls: 0/92 (0%) V460M: PD: 0/92 (0%), Controls: 1/92 (1.1%)
Gutti et al., 2008 (18541817)	184/0	Chinese from Taiwan	Sequencing of *GBA1* gene	Full *GBA1* gene	L444P, R131S, R163Q, L174P, S271G, D409H, Q497R	L444P: PD: 4/184 (2.17%) R131S: PD: 1/184 (0.54%) R163Q: PD: 1/184 (0.54%) L174P: PD: 1/184 (0.54%) S271G: PD: 1/184 (0.54%) D409H: PD: 1/184 (0.54%) Q497R: PD: 1/184 (0.54%)
Wu et al., 2007 (17702778)	518/339	Taiwanese	Variant screening	L444P, RecNciI, R120W	L444P, RecNciI, R120W	L444P: PD: 13/518 (2.5%), Controls: 2/339 (0.6%) RecNciI: PD: 2/518 (0.4%), Controls: 2/339 (0.6%) R120W: PD: 1/518 (0.2%), Controls: 0/339 (0%)
Choi et al., 2012 (22387070)	277/291	Korean	Direct DNA sequencing of all *GBA1* exons in 277 PD cases and 100 controls, only exon 2 and exons 5–	All *GBA1* exons; *GBA1* exons 2, 5-11	I-20V, R163Q, N188S, P201H, R257Q, L268L, S271G, R277C, F347L, L444P, K466K	I-20V: PD: 1/277 (0.36%), Controls: 4/291 (1.37%) R163Q: PD: 1/277 (0.36%), Controls: 0/291 (0%) N188S: PD: 1/277 (0.36%), Controls: 0/291 (0%) P201H: PD: 1/277 (0.36%), Controls: 0/291 (0%) R257Q: PD: 3/277 (1.08%), Controls: 0/291 (0%) L268L: PD: 1/277 (10.36%), Controls: 0/291 (0%) S271G: PD: 2/277 (0.72%), Controls: 0/291 (0%) R277C: PD: 1/277 (0.36%), Controls: 0/291 (0%)

			11 in 191 controls			F347L: PD: 1/277 (0.36%), Controls: 0/291 (0%) L444P: PD: 2/277 (0.72%), Controls: 0/291 (0%) K466K: PD: 4/277 (1.44%), Controls: 1/291 (0.34%)
Li et al., 2014 (24126159)	147/100	Japanese	DNA sequencing of all *GBA1* exons and intron/exon boundaries	All *GBA1* exons and exon/intron boundaries of *GBA1*	I(-20)V, G64V, R120W, D409H, L444P, I489V, W393X, K466K, c.1447-1466delTGins, RecNciI	I(-20)V: PD: 13/144 (9.0%), Controls: 10/100 (10.0%) G64V: PD: 1/144 (0.7%), Controls: 0/100 (0%) R120W: PD: 9/144 (6.3%), Controls: 0/100 (0%) D409H: PD: 4/144 (2.8%), Controls: 0/100 (0%) L444P: PD: 12/144 (8.3%), Controls: 0/100 (0%) I489V: PD: 2/144 (1.4%), Controls: 0/100 (0%) W393X: PD: 1/144 (0.7%), Controls: 0/100 (0%) c.1447-1466delTGins: PD: 1/144 (0.7%), Controls: 0/100 (0%) RecNciI: PD: 1/144 (0.7%), Controls: 1/100 (1.0%)
Mitsui et al., 2009 (19433656)	534/544	Japanese	DNA sequencing of all *GBA1* exons	All *GBA1* exons	R120W, R131C, N188S, R120W-N188R-V191G-S196P-F213I, G193W, F213I, R329C, L444P, L444P-A456P-V460V (RecNciI), A456P-V460V, R496C, I(-20), L(-15)F, L67Q, V121V, D153N, R163Q, P299T, G307S, T334I, L336L, G344G, F347L, R359L, V460V, K466K, I489V	R120W: PD: 15/534 (2.8%), Controls: 0/544 (0%) R131C: PD: 1/534 (0.2%), Controls: 0/544 (0%) N188S: PD: 4/534 (0.7%), Controls: 0/544 (0%) R120W-N188R-V191G-S196P-F213I: PD: 1/534 (0.2%), Controls: 0/544 (0%) G193W: PD: 1/534 (0.2%), Controls: 0/544 (0%) F213I: PD: 1/534 (0.2%), Controls: 0/544 (0%) R329C: PD: 2/534 (0.4%), Controls: 0/544 (0%) L444P: PD: 8/534 (1.5%), Controls: 0/544 (0%) L444P-A456P-V460V (RecNciI): PD: 14/534 (2.6%), Controls: 2/544 (0.4%) A456P-V460V: PD: 1/534 (0.2%), Controls: 0/544 (0%) R496C: PD: 2/534 (0.4%), Controls: 0/544 (0%) I(-20)V: PD: 77/534 (14.4%), Controls: 66/544 (12.1%) L(-15)F: PD: 1/534 (0.2%), Controls: 0/544 (0%) L67Q: PD: 0/534 (0%), Controls: 1/544 (0.2%) V121V: PD: 0/534 (0%), Controls: 1/544 (0.2%) D153N: PD: 1/534 (0.2%), Controls: 0/544 (0%) R163Q: PD: 4/534 (0.7%), Controls: 7/544 (1.3%) P299T: PD: 1/534 (0.2%), Controls: 0/544 (0%) G307S: PD: 1/534 (0.2%), Controls: 0/544 (0%) T334I: PD: 0/534 (0%), Controls: 1/544 (0.2%) L336L: PD: 1/534 (0.2%), Controls: 0/544 (0%) G344G: PD: 1/534 (0.2%), Controls: 0/544 (0%) F347L: PD: 0/534 (0%), Controls: 1/544 (0.2%) R359L: PD: 1/534 (0.2%), Controls: 0/544 (0%) V460V: PD: 2/534 (0.4%), Controls: 1/544 (0.2%) K466K: PD: 11/534 (2.1%), Controls: 8/544 (1.5%) I489V: PD: 4/534 (0.7%), Controls: 3/544 (0.6%)

Pulkes et al., 2014 (24997549)	480/395	Thai	Direct DNA sequencing in all EOPD and 100 patients with AAO>50; Variant screening in remaining patients with AAO>50 and controls	All *GBA1* exons and exon/intron boundaries of *GBA1*	L444P, N386K, P428S, IVS2+1G>A, IVS9+3G>C, IVS10-9_10GT>AG, V398fsX404	L444P: PD: 15/480 (3.1%), EOPD: 8/108 (7.4%), AAO>50: 7/372 (1.9%), Controls: 1/395 (0.3%) N386K: PD: 1/480 (0.2%), EOPD: 1/108 (0.9%), AAO>50: 0/372 (0%), Controls: 0/395 (0%) P428S: PD: 2/480 (0.4%), EOPD: 1/108 (0.9%), AAO>50: 1/372 (0.3%), Controls: 1/395 (0.3%) IVS2+1G>A: PD: 1/480 (0.2%), EOPD: 1/108 (0.9%), AAO>50: 0/372 (0%), Controls: 0/395 (0%) IVS9+3G>C: PD: 1/480 (0.2%), EOPD: 1/108 (0.9%), AAO>50: 0/372 (0%), Controls: 0/395 (0%) IVS10-9_10GT>AG: PD: 3/480 (0.6%), EOPD: 1/108 (0.9%), AAO>50: 2/372 (0.5%), Controls: 0/395 (0%) V398fsX404: PD: 1/480 (0.2%), EOPD: 1/108 (0.9%), AAO>50: 0/372 (0%), Controls: 0/395 (0%)
Yadav et al., 2018 (30504558)	100/0	Indian	DNA sequencing of all *GBA1* exons and intron/exon boundaries	All *GBA1* exons and exon/intron junctions	IVS1+191G>C, IVS4+47G>A, IVS6−86A>G, IVS9+141A>G, IVS10+3G>A	IVS1+191G>C: PD 1/100 (1%) IVS4+47G>A: PD 64/100 (64%) IVS6−86A>G: PD 65/100 (65%) IVS9+141A>G: PD 65/100 (65%) IVS10+3G>A: PD 1/100 (1%)
Biswas et al., 2021 (33711404)	198/241	Indian	Variant screening	IVS2+1A>G, R120W, H255Q, R257Q, E326K, N370S, D409H, L444P, RecNciI	L444P	L444P: PD: 2/198 (1.01%), Controls: 0/241 (0%)
Halder et al., 2016 (NA)	114/120	Indian	Variant screening	L444P, N370S	L444P	L444P: PD: 4/114 (3.51%), Controls: 0/120 (0%)
Goldstein et al., 2019 (31662221)	1200/378	Ashkenazi Jewish	Genotyping	E326K, T369M, R44C, N370S, R496H, L444P, 84GG, IVS2+1G->A, V394L, Rec370	E326K, T369M, R44C, N370S, R496H, L444P, 84GG, IVS2+1G->A, V394L, Rec370	E326K: PD: 17/1200 (1.4%), Controls: 5/378 (1.32%) R44C: PD: 1/1200 (0.08%), Controls: 0/378 (0%) T369M: PD: 11/1200 (0.92%), Controls: 1/378 (0.26%) N370S or R535H: PD: 140/1200 (11.67%), Controls: 14/378 (3.7%) 84GG or IVS2+1 or V349L or L444P or Rec370: PD: 46/1200 (3.83%), Controls: 2/378 (0.53%)

Gan-Or et al., 2008 (18434642)	420/4138	Ashkenazi Jewish	Variant screening	N370S, R496H, 84GG IVS2+1, V394L, D409H, L444P, RecTL	N370S, R496H, 84GG, IVS2+1, V394L, D409H, L444P, RecTL	All: PD: 75/420 (17.9%), Elderly controls: (4.2%), Young controls: (6.35%) N370S: PD: 46/420 (10.95%), Elderly controls: 11/333 (3.3%), Young controls: 224/3805 (5.89%) R496H: PD: 7/420 (1.67%), Elderly controls: 1/333 (0.3%) 84GG: PD: 8/420 (1.90%), Elderly controls: 1/333 (0.3%), Young controls: 6/3805 (0.16%) IVS2+1: PD: 4/420 (0.95%), Young controls: 4/3805 (0.11%) V394L: PD: 3/420 (0.71%), Young controls: 4/3805 (0.11%) L444P: PD: 2/420 (0.48%), Young controls: 2/3805 (0.05%)
Gan-Or et al., 2015 (25653295)	1000/3805	Ashkenazi Jewish	Variant screening	84GG, IVS2+1, N370S, L444P, V394L, R496H, 370Rec	84GG, IVS2+1, N370S, L444P, V394L, R496H, 370Rec	All: PD: 192/1000 (19.2%), Controls: 242/3805 (6.4%) N370S: PD: 131/1000 (13.1%), Controls: 225/3805 (5.9%) R496H: PD: 19/1000 (1.9%), Controls: NT 84GG: PD: 21/1000 (2.1%), Controls: 6/3805 (0.16%) IVS2+1G>A: PD: 5/1000 (0.5%), Controls: 1/3805 (0.03%) V394L: PD: 11/1000 (1.1%), Controls: 4/3805 (0.11%) L444P: PD: 3/1000 (0.3%), Controls: 4/3805 (0.11%) 370Rec: PD: 10/1000 (1.0%), Controls: 2/3805 (0.05%)
Dagan et al., 2015 (26169695)	287/400	Ashkenazi Jewish	Variant screening	N370S, L444P, c.84GG, c.115+1G>A (IVS2+1G>A), V394L, R496H	N370S, c.84GG, V394L, R496H	N370S: PD: 54/287 (18.8%), Controls: 14/400 (3.5%) c.84GG: PD: 9/287 (3.1%), Controls: 2/400 (0.5%) V394L: PD: 1/287 (0.3%) R496H: PD: 4/287 (1.4%), Controls: 3/400 (0.7%)
Liu et al., 2011 (21812969)	268/178	Ashkenazi Jewish	Genotyping	NA	N370S	N370S: PD: 28/268 (10.4%)
Ruskey et al., 2019 (29842932)	735/622	Ashkenazi Jewish	Targeted next-generation sequencing; Sanger sequencing of exons 10 and 11	Full *GBA1* gene	84GG, R44C, N188S, E326K, T369M, N370S, A384D, V394L, T410M, L444P, L461P, R496H	84GG: PD: 13/735 (1.77%), Controls: 1/622 (0.15%) R44C: PD: 2/735 (0.27%), Controls: 7/622 (1.06%) N188S: PD: 1/735 (0.14%), Controls: 0/622 (0%) E326K: PD: 13/735 (1.77%), Controls: 2/622 (0.3%) T369M: PD: 2/735 (0.27%), Controls: 0/622 (0%) N370S: PD: 92/735 (12.52%), Controls: 37/622 (5.58%) A384D: PD: 1/735 (0.14%), Controls: 0/622 (0%) V394L: PD: 1/735 (0.14%), Controls: 0/622 (0%) T410M: PD: 0/735 (0%), Controls: 1/622 (0.15%) L444P: PD: 3/735 (0.41%), Controls: 0/622 (0%)

						L461P: PD: 1/735 (0.14%), Controls: 0/622 (0%) R496H: PD: 9/735 (1.22%), Controls: 2/622 (0.3%)
Clark et al., 2005 (15517591)	160/92	Ashkenazi Jewish	Direct sequencing	N370S	N370S	N370S: PD: 17/160 (10.6%), Controls: 4/92 (4.3%)
Aharon-Peretz et al., 2004 (15525722)	99/1543	Ashkenazi Jewish	Variant screening	N370S, L444P, 84GG, IVS+1, V394L, R496H	N370S, 84GG, R496H	N370S: PD: 26/99 (26.3%), Controls: 92/1543 (5.96%) 84GG: PD: 4/99 (4.0%), Controls: 3/1543 (0.19%) R496H: PD: 1/99 (1.0%), Controls: 0/1543 (0%)
Aharon-Peretz et al., 2005 (16148263)	148/0	Ashkenazi Jewish	Digestion with appropriate enzymes to detect N370S, L444P, 84GG, IVS+1, V394L, R496H	N370S, L444P, 84GG, IVS+1, V394L, R496H	N370S,4GG, R496H	N370S: PD: 34/148 (22.97%) 84GG: PD: 4/148 (2.70%) R496H: 2/148 (1.35%)
Gan-Or et al., 2010 (19458969)	600/0	Ashkenazi Jewish	Variant screening	84GG, IVS2+1, N370S, V394L, D409H, L444P, R496H and RecTL	84GG, IVS2+1, N370S, V394L, D409H, L444P, R496H and RecTL	All: PD: 117/600 (19.5%)
Alcalay et al., 2015 (26117366)	517/252 (mostly spouses)	Mixed (PD: 231 with Ashkenazi Jewish Grandparent, Controls: 97 with Ashkenazi Jewish Grandparent)	DNA sequencing of the full *GBA1* gene	Full *GBA1* gene	N370S, L444P, 84GG, R496H, IVS2+1, K-27R, E326K, T369M, L461P, V294M, A456P, G241R, Rearrangement exon 8, Q-8H, R44C, N392S, S110A, T410M, F-36V, P387P, E349K	N370S: PD: 36/517 (7.0%), Controls: 4/252 (1.6%) L444P: PD: 7/517 (1.4%), Controls: 1/252 (0.4%) 84GG: PD: 4/517 (0.8%), Controls: 0/252 (0%) R496H: PD: 4/517 (0.8%), Controls: 0/252 (0%) IVS2+1: PD: 2/517 (0.4%), Controls: 0/252 (0%) K-27R: PD: 2/517 (0.4%), Controls: 0/252 (0%) E326K: PD: 13/517 (2.5%), Controls: 3/252 (1.2%) T369M: PD: 5/517 (1.0%), Controls: 4/252 (1.6%) L461P: PD: 1/517 (0.2%), Controls: 0/252 (0%) V294M: PD: 1/517 (0.2%), Controls: 0/252 (0%) A456P: PD: 1/517 (0.2%), Controls: 0/252 (0%) G241R: PD: 1/517 (0.2%), Controls: 0/252 (0%) Rearrangement exon 8: PD: 1/517 (0.2%), Controls: 0/252 (0%) Q-8H: PD: 1/517 (0.2%), Controls: 0/252 (0%) R44C: PD: 1/517 (0.2%), Controls: 0/252 (0%) N392S: PD: 1/517 (0.2%), Controls: 0/252 (0%) S110A: PD: 0/517 (0%), Controls: 1/252 (0.4%)

						T410M: PD: 0/517 (0%), Controls: 1/252 (0.4%) F-36V: PD: 0/517 (0%), Controls: 1/252 (0.4%) P387P: PD: 0/517 (0%), Controls: 1/252 (0.4%) E349K: PD: 0/517 (0%), Controls: 1/252 (0.4%)
Clark et al., 2007 (17875915)	278/179	Mixed (178 PD Jewish, 85 controls Jewish)	DNA sequencing of all *GBA1* exons	All *GBA1* exons	84insGG, E326K, T369M, N370S, D409H, R496H, L444P, RecNciI (L444P + A456P + V460V), P175P	All: PD: 38/278 (13.7%), Controls: 8/179 (4.5%) 84insGG: PD: 5/278 (1.8%), Controls: 0/179 (0%) E326K: PD: 1/278 (0.4%), Controls: 1/179 (0.6%) T369M: PD: 3/278 (1.1%), Controls: 3/179 (1.7%) N370S: PD: 23/278 (8.3%), Controls: 4/179 (2.2%) D409H: PD: 1/278 (0.4%), Controls: 0/179 (0%) R496H: PD: 1/278 (0.4%), Controls: 0/179 (0%) L444P: PD: 2/278 (0.7%), Controls: 0/179 (0%) RecNciI (L444P + A456P + V460V): PD: 1/278 (0.4%), Controls: 0/179 (0%) P175P: PD: 1/278 (0.4%), Controls: 0/179 (0%)
Alcalay et al., 2010 (20837857)	953/0	Mixed (77 Hispanics, 139 of Jewish ancestry)	DNA sequencing of full *GBA1* gene in 90 cases (previously reported); Genotyping for L444P and N370S in 515 cases; direct sequencing of L444P and N370S in 348 cases	Full *GBA1* gene; L444P, N370S	L444P, N370S	L444P: PD: 18/953 (1.9%) N370S: PD: 40/953 (4.2%)
Goker-Alpan et al., 2006 (16790605)	28/44	Mixed	DNA sequencing of full *GBA1* gene	All *GBA1* exons and flanking introns	N370S	N370S: PD: 1/28 (3.57%), Controls: 0/44 (0%)
Sidransky et al., 2009 (19846850)	5691/4898	Mixed	Variant screening for L444P, N370S or sequencing for all *GBA1* exons	L444P, N370S; all *GBA1* exons	L444P, N370S, E326K, T369M	Ashkenazi Jews: L444P/N370S: PD (15.3%), Controls (3.4%) Non-Ashkenazi Jews: L444P/N370S: PD (3.2%), Controls (0.6%) Center Brazil (N370S, L444P, G377S): PD: 4/65 (6.2%), Controls: 0/264 (0%) Center NYC, USA (Full sequencing): PD: 34/275 (177 AJ) (12.4%), Controls: 3/140. (65 AJ)

						(2.14%) Center France (N370S, L444P, D409H): PD: 12/297 (4.0%), Controls: 1/251 (0.39%) Center Haifa, IL (D409H, 84GG, V394L, IVS2+1, R496H): PD: 40/162 (162 AJ) (24.7%), Controls: NP Center Italy (L444P, N370S): PD: 11/395 (2.8%), Controls: 1/483 (0.21%) Center Norway (L444P, N370S): PD: 7/311 (2.3%), Controls: 8/473 (1.69%) Center NHGRI, USA (Full sequencing): PD: 29/539 (5.4%), Controls: 6/209 (1 AJ) (2.87%) Center Portugal (Full sequencing): PD: 15/231 (6.5%), Controls: 6/482 (1.24%) Center Rostock, DE (Full Sequencing): PD: 18/298 (6.0%), Controls: 5/212 (2.4%) Center Singapore (L444P, N370S): PD: 8/329 (2.4%), Controls: 0/201 (0%) Center Taiwan (L444P, recNciI, R120W, some full sequencing): PD: 22/559 (3.9%), Controls: 4/377 (1.06%) Center Tel Aviv, IL (84GG, IVS2+1, N370S, V394L, D409H, L444P, R496H, RecTL): PD: 81/420 (419 AJ) (19.3%), Controls: 13/321 (321 AJ) (4.05%) Center Japan (full sequencing): PD: 50/534 (9.4%), Controls: 2/546 (0.37%) Center Tübingen, DE (L444P, N370S): PD: 12/377 (3.2%), Controls: 0/325 (0%) Center Toronto, CA (N370S, K178T, L444P, 84GG, R329C, IVS2+1, recNciiI): PD: 5/88 (2 AJ) (5.7%), Controls: 1/96 (1.0%)
Mata et al., 2016 (26296077)	1369/0	Mixed	DNA sequencing of all *GBA1* exons and intron-exons boundaries	All *GBA1* exons and intron-exons boundaries	IVS2+1G>A, 84dupG, S125N, T134P, D140H, R163X, N188S, S196P, G202R, F216Y, 914delC, S271G, R359X, N370S, Rec3 (c1263-1317 del, D409H, L444P, A456P, V460V), D409H, L444P, Rec1 (L444P, A456P, V460V), Rec L444P + V460V, V460M, R463C,	IVS2+1G>A: PD: 2/1369 (0.15%) 84dupG: PD: 3/1369 (0.22%) S125N: PD: 1/1369 (0.07%) T134P: PD: 1/1369 (0.07%) D140H: PD: 2/1369 (0.22%) R163X: PD: 1/1369 (0.07%) N188S: PD: 1/1369 (0.07%) S196P: PD: 1/1369 (0.07%) G202R: PD: 1/1369 (0.07%) F216Y: PD: 1/1369 (0.07%) 914delC: PD: 1/1369 (0.07%)

					R496H, R(−32)T, P(−28)S, R44C, G193E, R262H, F316I, G344S, D443N, V460L, S488T, K(−27)R, E326K, T369M	S271G: PD: 1/1369 (0.07%) R359X: PD: 1/1369 (0.07%) N370S: PD: 18/1369 (1.31%) Rec3 (c1263-1317 del, D409H, L444P, A456P, V460V): PD: 1/1369 (0.07%) D409H: PD: 1/1369 (0.07%) L444P: PD: 16/1369 (1.17%) Rec1 (L444P, A456P, V460V): PD: 2/1369 (0.22%) Rec L444P + V460V: PD: 1/1369 (0.07%) V460M: PD: 1/1369 (0.07%) R463C: PD: 3/1369 (0.22%) R496H: PD: 2/1369 (0.22%) R(−32)T: PD: 1/1369 (0.07%) P(−28)S: PD: 1/1369 (0.07%) R44C: PD: 1/1369 (0.07%) G193E: PD: 1/1369 (0.07%) R262H: PD: 1/1369 (0.07%) F316I: PD: 1/1369 (0.07%) G344S: PD: 1/1369 (0.07%) D443N: PD: 1/1369 (0.07%) V460L: PD: 1/1369 (0.07%) S488T: PD: 1/1369 (0.07%) K(−27)R: PD: 6/1369 (0.44%) E326K: PD: 69/1369 (5.04%) T369M: PD: 30/1369 (2.19%)
Nichols et al., 2009 (18987351)	1325/359	International	DNA sequencing of all *GBA1* exons and corresponding intron/exon boundaries in 96 samples; Variant screening in 1325 cases and 359 controls	All *GBA1* exons and intron/exon boundaries	IVS6 589-2A>G, R262H, K303K, E326K, T369M, N370S, L444P, IVS10 1389-3C>G, RecNciI (L444P+A456P+V60V)	IVS6 589-2A>G/R262H/K303K/E326K/T369M/N370S/L444P/IVS10 1389-3C>G/RecNciI (L444P+A456P+V60V): PD: 21/96 (21.9%) All 9 *GBA1* variants: PD: 161/1325 (12.2%) All 5 previous *GBA1* variants (E326K, T369M, N370S, L444P, RecNcil) : PD: ?/450 (12.6%), Controls: ?/359 (5.3%) E326K: PD: ?/450 (6.2%), Controls: ?/359 (3.1%) T369M: PD: ?/450 (2.3%), Controls: ?/359 (1.1%) N370S: PD: ?/450 (1.4%), Controls: ?/359 (0.8%) L444P: PD: ?/450 (1.9%), Controls: ?/359 (0.0%) A456P/V460V/L444P: PD: ?/450 (0.8%), Controls: ?/359 (0.3%) IVS6 589-2A>G: Controls: 0/359 (0%) R262H: Controls: 0/359 (0%) IVS10 1389- 3C>G: Controls: 0/359 (0%)

Liu et al., 2016 (27717005)	2304/0	International	Depending on the study: Whole exome or targeted sequencing or genotyping of N370S, E326K, T369M	All *GBA1* exons; N370S, E326K, T369M	K(-27)R, 84GG, R120W, D140H, G195E, H255Q, R257Q, P266L, R359X, G377S, D409H, L444P, L444R, A456P, N462K, R463C, R463P, N370S, E326K, T369M, E388K	84GG: PD: 1/1921 (0.05%) R120W: PD: 1/1921 (0.05%) D140H: PD: 8/1921 (0.42%) G195E: PD: 1/1921 (0.05%) H255Q: PD: 1/1921 (0.05%) R257Q: PD: 2/1921 (0.10%) P266L: PD: 1/1921 (0.05%) R359X: PD: 1/1921 (0.05%) G377S: PD: 1/1921 (0.05%) D409H: PD: 1/1921 (0.05%) L444P: PD: 13/1921 (0.68%) L444R: PD: 1/1921 (0.05%) A456P: PD: 1/1921 (0.05%) N462K: PD: 1/1921 (0.05%) R463C: PD: 2/1921 (0.10%) N370S: PD: 28/1921 (1.5%) E326K: PD: 92/1921 (4.79%) T369M: PD: 48/1921 (2.50%) E388K: PD: 2/1921 (0.10%)
Stoker et al., 2020 (32303560)	250/0	NA	DNA sequencing of all *GBA1* exons in 250 patients; genotyping in 127 patients	All *GBA1* exons	N370S, L444P, R463C, G10S, N426K, R48W, R257Q, c.762 18T>A, E326K, T369M, E388K, L119L, c.589 86A>G)	All: PD: 36/250 (14.4%) N370S: PD: 3/250 (1.2%) L444P: PD: 3/250 (1.2%) R463C: PD: 1/250 (0.4%) G10S: PD: 1/250 (0.4%) N426K: PD: 1/250 (0.4%) R48W: PD: 1/250 (0.4%) R257Q: PD: 1/250 (0.4%) c.762 18T>A: PD: 8/250 (3.2%) E326K PD: 7/250 (2.8%) T369M: PD: 7/250 (2.8%) E388K: PD: 1/250 (0.4%) L119L: PD: 1/250 (0.4%) c.589 86A>G: PD: 1/250 (0.4%) N370S: PD: 4/127 (3.1%) R463C: PD: 1/127 (0.8%) G10S PD: 1/127 (0.8%) T369M PD: 3/127 (2.4%) E326K PD: 2/127 (1.6%) E388K PD: 1/127 (0.8%)
Gorostidi et al., 2016 (27294386)	92/0	NA	Targeted DNA sequencing	NA	K13R, Y244C, T408M	K13R: PD: 1/92 (1.1%) Y244C: PD: 1/92 (1.1%) T408M: PD: 3/92 (3.3%)

Mata et al., 2008 (18332251)	721/554 (310 spouses, 244 volunteers)	NA	Genotyping	N370S, L444P	N370S, L444P	N370S: PD: 11/721 (1.5%), Controls: 2/554 (0.4%) L444P: PD: 10/721 (1.4%), Controls: 0/554 (0%)
Lwin et al., 2004 (14728994)	57/44	NA	DNA sequencing of full *GBA1* gene	All *GBA1* exons and flanking introns	N370S, L444P, K198T, R329C, T369M, E326K	N370S: PD: 5/57 (8.77%), Controls: 0/44 (0%) L444P: PD: 1/57 (1.75%), Controls: 0/44 (0%) K198T: PD: 1/57 (1.75%), Controls: 0/44 (0%) R329C: PD: 1/57 (1.75%), Controls: 0/44 (0%) T369M PD: 3/57 (5.26%), Controls: 0/44 (0%) E326K PD: 1/57 (1.75%), Controls: 0/44 (0%)
Malek et al., 2018 (29378790)	1893/0	NA	DNA sequencing of all *GBA1* exons	All *GBA1* exons	L444P, N370S, R463C, G202R, R359S, E326K, T369M, D409H, F213I, G189V, G377S, K157Q, L383Xfs, L66P, M123T, N382Xfs, R163s, R257Q, S173s, E481Xfs, G10S, G325W, R170H, T323I, L175I, L324V, P55S, R262H, R329H, R395C, T267I, L268L, Asp315His, Exon 3 hemizygous deletion, A456P, V460V, D140H, I308T, Ex4 hemizygous deletion	L444P: PD: 30/1893 (1.6%) N370S: PD: 11/1893 (0.6%) R463C: PD: 5/1893 (0.3%) G202R: PD: 2/1893 (0.1%) R359S: PD: 2/1893 (0.1%) E326K: PD: 86/1893 (4.5%) T369M: PD: 35/1893 (1.8%) D409H: PD: 1/1893 (0.05%) F213I: PD: 1/1893 (0.05%) G189V: PD: 1/1893 (0.05%) G377S: PD: 1/1893 (0.05%) K157Q: PD: 1/1893 (0.05%) L383Xfs: PD: 1/1893 (0.05%) L66P: PD: 1/1893 (0.05%) M123T: PD: 1/1893 (0.05%) N382Xfs: PD: 1/1893 (0.05%) R163s: PD: 1/1893 (0.05%) R257Q: PD: 1/1893 (0.05%) S173s: PD: 1/1893 (0.05%) E481Xfs: PD: 1/1893 (0.05%) G10S: PD: 1/1893 (0.05%) G325W: PD: 1/1893 (0.05%) R170H: PD: 1/1893 (0.05%) T323I: PD: 1/1893 (0.05%) L175I: PD: 1/1893 (0.05%) L324V: PD: 1/1893 (0.05%) P55S: PD: 1/1893 (0.05%) R262H: PD: 1/1893 (0.05%) R329H: PD: 1/1893 (0.05%) R395C: PD: 1/1893 (0.05%) T267I: PD: 1/1893 (0.05%) L268L: PD: 1/1893 (0.05%) Asp315His: PD: 1/1893 (0.05%) Exon 3 hemizygous deletion: PD: 1/1893 (0.05%)

						A456P: PD: 6/1893 (0.3%) V460V: PD: 6/1893 (0.3%) D140H: PD: 2/1893 (0.1%) I308T: PD: 2/1893 (0.1%) Ex4 hemizygous deletion: PD: 2/1893 (0.1%)
Eblan et al., 2005 (15716572)	26/0	NA	DNA sequencing	NA	D140H, RecNciI	D140H: PD: 1/26 (3.85%) RecNciI: PD: 1/26 (3.85%)
Barber et al., 2017 (28472425)	106/283	NA	Variant screening	N370S, L444P	N370S	N370S: PD: 1/106 (0.9%), Controls: 1/283 (0.4%)
Malec-Litwinowicz et al., 2014 (25168325)	138/0	NA	DNA sequencing of *GBA1* exons 8 and 9	*GBA1* exons 8 and 9	N370S, T369M	N370S: PD: 5/138 (3.6%) T369M: PD: 11/138 (7.9%)
McNeill et al., 2012 (22577228)	220/0	NA	Sanger sequencing of the full *GBA1* gene	Full *GBA1* gene	N370S, L444P, Recombinant alleles, R496H, V460L, IVS2+1	N370S: PD: 5/220 (2.27%) L444P: PD: 2/220 (0.91%) Recombinant alleles: PD: 2/220 (0.91%) R496H: PD: 1/220 (0.45%) V460L: PD: 1/220 (0.45%) IVS2+1: PD: 1/220 (0.45%)
Graham et al., 2020 (31809948)	229/50	NA	Nanopore sequencing of full *GBA1* gene	Full *GBA1* gene	E365K, T408M, D179H, N409S, L335=, R78C	E365K: PD: 12/229 (5.24%), Controls: 0/50 (0%) T408M: PD: 7/229 (3.06%), Controls: 2/50 (4.0%) D179H: PD: 1/229 (0.44%), Controls: 0/50 (0%) N409S: PD: 1/229 (0.44%), Controls: 0/50 (0%) L335=: PD: 1/229 (0.44%), Controls: 0/50 (0%) R78C: PD: 1/229 (0.44%), Controls: 0/50 (0%)

Population	**Variants**	**n mutation carriers/n total (frequency) in PD**	**n mutation carriers/n total (frequency) in controls**	**n studies included**	**References**
White/Caucasian	p.L483P p.N409S p.T408M p.E365K p.R159W	329/21492 (1.53%) 345/21714 (1.59%) 310/15266 (2.03%) 668/16408 (4.07%) 8/10011 (0.08%)	101/17770 (0.57%) 76/17770 (0.43%) 182/14298 (1.27%) 239/14849 (1.61%) 0/4630 (0%)	34	(Toft et al., 2006), (Lunde et al., 2018), (Berge-Seidl et al., 2017), (Ran et al., 2022), (Ran et al., 2016), (Ylonen et al., 2017), (Muldmaa et al., 2021), (Neumann et al., 2009), (Winder-Rhodes et al., 2013), (Duran et al., 2013), (Olszewska et al., 2020), (Crosiers et al., 2016), (den Heijer et al., 2020), (Anheim et al., 2012), (Lesage et al., 2011a), (Spataro et al., 2017), (Seto-Salvia et al., 2012), (Jesus et al., 2016), (Bras et al., 2009), (Petrucci et al., 2020), (De Marco et al., 2008), (Asselta et al., 2014), (Cilia et al., 2016), (Straniero et al., 2020), (Quadri et al., 2015), (Kalinderi et al., 2009), (Moraitou et al., 2011), (Emekli et al., 2021), (Kumar et al., 2013), (Torok et al., 2016), (Benitez et al., 2016), (Noreau et al., 2011), (Han et al., 2016), (Sato et al., 2005)

South American	p.L483P p.N409S p.T408M p.E365K p.R159W	26/1237 (2.10%) 14/1237 (1.13%) 3/635 (0.47%) 7/635 (1.10%) 2/635 (0.31%)	0/1024 (0%) 1/1024 (0.10%) 0/350 (0%) 1/350 (0.29%) 0/350 (0%)	7	(Gonzalez-Del Rincon Mde et al., 2013), (Velez-Pardo et al., 2019), (Eblan et al., 2006), (Dos Santos et al., 2010), (Spitz et al., 2008), (Guimaraes Bde et al., 2012), (Socal et al., 2009)
North African	p.L483P p.N409S p.T408M p.E365K p.R159W	1/227 (0.44%) 3/589 (0.51%) 2/227 (0.88%) 3/227 (1.32%) 0/227 (0%)	0/177 (0%) 3/549 (0.55%) 0/177 (0%) 1/177 (0.56%) 0/177 (0%)	2	(Lesage et al., 2011b), (Nishioka et al., 2010)
Asian	p.L483P p.N409S p.T408M p.E365K p.R159W	261/10233 (2.55%) 6/3739 (0.16%) 0/1963 (0%) 1/2161 (0.05%) 31/4786 (0.65%)	7/5974 (0.12%) 2/3319 (0.06%) 0/1157 (0%) 0/1398 (0%) 0/3285 (0%)	23	(Mao et al., 2010), (Hu et al., 2010), (Zhang et al., 2012), (Guo et al., 2015), (Wang et al., 2014), (Ren et al., 2022), (Yu et al., 2015), (Sun et al., 2010), (Wang et al., 2012), (Tan et al., 2007), (Li et al., 2020), (Huang et al., 2011), (Zhang et al., 2015), (Ziegler et al., 2007), (Gutti et al., 2008), (Wu et al., 2007), (Choi et al., 2012), (Li et al., 2014), (Mitsui et al., 2009), (Pulkes et al., 2014), (Yadav et al., 2018), (Biswas et al., 2021), (Halder et al., 2016)
Ashkenazi Jewish	p.L483P p.N409S p.T408M p.E365K p.R159W	8/2689 (0.30%) 428/3117 (13.7%) 13/1935 (0.67%) 30/1935 (1.55%) 0/735 (0%)	6/10175 (0.06%) 382/6795 (5.62%) 1/1000 (0.01%) 7/1000 (0.07%) 0/622 (0%)	9	(Goldstein et al., 2019), (Gan-Or et al., 2008), (Gan-Or et al., 2015), (Dagan et al., 2015), (Liu et al., 2011), (Ruskey et al., 2019), (Clark et al., 2005), (Aharon-Peretz et al., 2004), (Aharon-Peretz et al., 2005)

* The reported studies use either the conventional nomenclature for *GBA1* alleles excluding the 39-residue signal peptide or refer to the processed protein that includes the 39-residue signal peptide following Human Genome Variation Society (HGVS) recommendation. Articles investigating Norwegian and Scandinavian cohorts are highlighted in gray.

References:

Aharon-Peretz, J., Badarny, S., Rosenbaum, H., and Gershoni-Baruch, R. (2005). Mutations in the glucocerebrosidase gene and Parkinson disease: phenotype-genotype correlation. Neurology *65*, 1460-1461.

Aharon-Peretz, J., Rosenbaum, H., and Gershoni-Baruch, R. (2004). Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med *351*, 1972-1977.

Alcalay, R.N., Caccappolo, E., Mejia-Santana, H., Tang, M.X., Rosado, L., Ross, B.M., Verbitsky, M., Kisselev, S., Louis, E.D., Comella, C.*, et al.* (2010). Frequency of known mutations in early-onset Parkinson disease: implication for genetic counseling: the consortium on risk for early onset Parkinson disease study. Arch Neurol *67*, 1116-1122.

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.*, et al.* (2015). Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. Brain *138*, 2648-2658.

Anheim, M., Elbaz, A., Lesage, S., Durr, A., Condroyer, C., Viallet, F., Pollak, P., Bonaiti, B., Bonaiti-Pellie, C., Brice, A.*, et al.* (2012). Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. Neurology *78*, 417-420.

Asselta, R., Rimoldi, V., Siri, C., Cilia, R., Guella, I., Tesei, S., Solda, G., Pezzoli, G., Duga, S., and Goldwurm, S. (2014). Glucocerebrosidase mutations in primary parkinsonism. Parkinsonism Relat Disord *20*, 1215-1220.

Barber, T.R., Lawton, M., Rolinski, M., Evetts, S., Baig, F., Ruffmann, C., Gornall, A., Klein, J.C., Lo, C., Dennis, G.*, et al.* (2017). Prodromal Parkinsonism and Neurodegenerative Risk Stratification in REM Sleep Behavior Disorder. Sleep *40*.

Barkhuizen, M., Anderson, D.G., van der Westhuizen, F.H., and Grobler, A.F. (2017). A molecular analysis of the GBA gene in Caucasian South Africans with Parkinson's disease. Mol Genet Genomic Med *5*, 147-156.

Benitez, B.A., Davis, A.A., Jin, S.C., Ibanez, L., Ortega-Cubero, S., Pastor, P., Choi, J., Cooper, B., Perlmutter, J.S., and Cruchaga, C. (2016). Resequencing analysis of five Mendelian genes and the top genes from genome-wide association studies in Parkinson's Disease. Mol Neurodegener *11*, 29.

Berge-Seidl, V., Pihlstrom, L., Maple-Grodem, J., Forsgren, L., Linder, J., Larsen, J.P., Tysnes, O.B., and Toft, M. (2017). The GBA variant E326K is associated with Parkinson's disease and explains a genome-wide association signal. Neurosci Lett *658*, 48-52.

Biswas, A., Sadhukhan, D., Biswas, A., Das, S.K., Banerjee, T.K., Bal, P.S., Pal, S., Ghosh, A., Ray, K., and Ray, J. (2021). Identification of GBA mutations among neurodegenerative disease patients from eastern India. Neurosci Lett *751*, 135816.

Bras, J., Paisan-Ruiz, C., Guerreiro, R., Ribeiro, M.H., Morgadinho, A., Januario, C., Sidransky, E., Oliveira, C., and Singleton, A. (2009). Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal. Neurobiol Aging *30*, 1515-1517.

Choi, J.M., Kim, W.C., Lyoo, C.H., Kang, S.Y., Lee, P.H., Baik, J.S., Koh, S.B., Ma, H.I., Sohn, Y.H., Lee, M.S.*, et al.* (2012). Association of mutations in the glucocerebrosidase gene with Parkinson disease in a Korean population. Neurosci Lett *514*, 12-15.

Cilia, R., Tunesi, S., Marotta, G., Cereda, E., Siri, C., Tesei, S., Zecchinelli, A.L., Canesi, M., Mariani, C.B., Meucci, N.*, et al.* (2016). Survival and dementia in GBA-associated Parkinson's disease: The mutation matters. Ann Neurol *80*, 662-673.

Clark, L.N., Nicolai, A., Afridi, S., Harris, J., Mejia-Santana, H., Strug, L., Cote, L.J., Louis, E.D., Andrews, H., Waters, C.*, et al.* (2005). Pilot association study of the beta-glucocerebrosidase N370S allele and Parkinson's disease in subjects of Jewish ethnicity. Mov Disord *20*, 100-103.

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.*, et al.* (2007). Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. Neurology *69*, 1270-1277.

Crosiers, D., Verstraeten, A., Wauters, E., Engelborghs, S., Peeters, K., Mattheijssens, M., De Deyn, P.P., Theuns, J., Van Broeckhoven, C., and Cras, P. (2016). Mutations in glucocerebrosidase are a major genetic risk factor for Parkinson's disease and increase susceptibility to dementia in a Flanders-Belgian cohort. Neurosci Lett *629*, 160-164.

Dagan, E., Schlesinger, I., Ayoub, M., Mory, A., Nassar, M., Kurolap, A., Peretz-Aharon, J., and Gershoni-Baruch, R. (2015). The contribution of Niemann-Pick SMPD1 mutations to Parkinson disease in Ashkenazi Jews. Parkinsonism Relat Disord *21*, 1067-1071.

De Marco, E.V., Annesi, G., Tarantino, P., Rocca, F.E., Provenzano, G., Civitelli, D., Ciro Candiano, I.C., Annesi, F., Carrideo, S., Condino, F.*, et al.* (2008). Glucocerebrosidase gene mutations are associated with Parkinson's disease in southern Italy. Mov Disord *23*, 460-463.

den Heijer, J.M., Cullen, V.C., Quadri, M., Schmitz, A., Hilt, D.C., Lansbury, P., Berendse, H.W., van de Berg, W.D.J., de Bie, R.M.A., Boertien, J.M.*, et al.* (2020). A Large-Scale Full GBA1 Gene Screening in Parkinson's Disease in the Netherlands. Mov Disord *35*, 1667-1674.

Dos Santos, A.V., Pestana, C.P., Diniz, K.R., Campos, M., Abdalla-Carvalho, C.B., de Rosso, A.L., Pereira, J.S., Nicaretta, D.H., de Carvalho, W.L., Dos Santos, J.M.*, et al.* (2010). Mutational analysis of GIGYF2, ATP13A2 and GBA genes in Brazilian patients with early-onset Parkinson's disease. Neurosci Lett *485*, 121-124.

Duran, R., Mencacci, N.E., Angeli, A.V., Shoai, M., Deas, E., Houlden, H., Mehta, A., Hughes, D., Cox, T.M., Deegan, P.*, et al.* (2013). The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. Mov Disord *28*, 232-236.

Eblan, M.J., Nguyen, J., Ziegler, S.G., Lwin, A., Hanson, M., Gallardo, M., Weiser, R., De Lucca, M., Singleton, A., and Sidransky, E. (2006). Glucocerebrosidase mutations are also found in subjects with early-onset parkinsonism from Venezuela. Mov Disord *21*, 282-283.

Eblan, M.J., Walker, J.M., and Sidransky, E. (2005). The glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med *352*, 728-731; author reply 728-731.

Emekli, I., Tepgec, F., Samanci, B., Toksoy, G., Hasanogullari Kina, G., Tufekcioglu, Z., Basaran, S., Bilgic, B., Gurvit, I.H., Emre, M.*, et al.* (2021). Clinical and molecular genetic findings of hereditary Parkinson's patients from Turkey. Parkinsonism Relat Disord *93*, 35-39.

Emelyanov, A., Boukina, T., Yakimovskii, A., Usenko, T., Drosdova, A., Zakharchuk, A., Andoskin, P., Dubina, M., Schwarzman, A., and Pchelina, S. (2012). Glucocerebrosidase gene mutations are associated with Parkinson's disease in Russia. Mov Disord *27*, 158-159.

Emelyanov, A.K., Usenko, T.S., Tesson, C., Senkevich, K.A., Nikolaev, M.A., Miliukhina, I.V., Kopytova, A.E., Timofeeva, A.A., Yakimovsky, A.F., Lesage, S.*, et al.* (2018). Mutation analysis of Parkinson's disease genes in a Russian data set. Neurobiol Aging *71*, 267 e267-267 e210.

Foo, J.N., Tan, L.C., Liany, H., Koh, T.H., Irwan, I.D., Ng, Y.Y., Ahmad-Annuar, A., Au, W.L., Aung, T., Chan, A.Y.*, et al.* (2014). Analysis of non-synonymous-coding variants of Parkinson's disease-related pathogenic and susceptibility genes in East Asian populations. Hum Mol Genet *23*, 3891-3897.

Gan-Or, Z., Amshalom, I., Kilarski, L.L., Bar-Shira, A., Gana-Weisz, M., Mirelman, A., Marder, K., Bressman, S., Giladi, N., and Orr-Urtreger, A. (2015). Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology *84*, 880-887.

Gan-Or, Z., Bar-Shira, A., Mirelman, A., Gurevich, T., Kedmi, M., Giladi, N., and Orr-Urtreger, A. (2010). LRRK2 and GBA mutations differentially affect the initial presentation of Parkinson disease. Neurogenetics *11*, 121-125.

Gan-Or, Z., Giladi, N., Rozovski, U., Shifrin, C., Rosner, S., Gurevich, T., Bar-Shira, A., and Orr-Urtreger, A. (2008). Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. Neurology *70*, 2277-2283.

Goker-Alpan, O., Giasson, B.I., Eblan, M.J., Nguyen, J., Hurtig, H.I., Lee, V.M., Trojanowski, J.Q., and Sidransky, E. (2006). Glucocerebrosidase mutations are an important risk factor for Lewy body disorders. Neurology *67*, 908-910.

Goldstein, O., Gana-Weisz, M., Cohen-Avinoam, D., Shiner, T., Thaler, A., Cedarbaum, J.M., John, S., Lalioti, M., Gurevich, T., Bar-Shira, A.*, et al.* (2019). Revisiting the non-Gaucher-GBA-E326K carrier state: Is it sufficient to increase Parkinson's disease risk? Mol Genet Metab *128*, 470-475.

Gonzalez-Del Rincon Mde, L., Monroy Jaramillo, N., Suarez Martinez, A.I., Yescas Gomez, P., Boll Woehrlen, M.C., Lopez Lopez, M., and Alonso Vilatela, M.E. (2013). The L444P GBA mutation is associated with early-onset Parkinson's disease in Mexican Mestizos. Clin Genet *84*, 386-387.

Gorostidi, A., Marti-Masso, J.F., Bergareche, A., Rodriguez-Oroz, M.C., Lopez de Munain, A., and Ruiz-Martinez, J. (2016). Genetic Mutation Analysis of Parkinson's Disease Patients Using Multigene Next-Generation Sequencing Panels. Mol Diagn Ther *20*, 481-491.

Graham, O.E.E., Pitcher, T.L., Liau, Y., Miller, A.L., Dalrymple-Alford, J.C., Anderson, T.J., and Kennedy, M.A. (2020). Nanopore sequencing of the glucocerebrosidase (GBA) gene in a New Zealand Parkinson's disease cohort. Parkinsonism Relat Disord *70*, 36-41.

Guimaraes Bde, C., Pereira, A.C., Rodrigues Fda, C., dos Santos, A.V., Campos, M., Jr., dos Santos, J.M., dos Santos, F.L., de Rosso, A.L., Nicaretta, D.H., Pereira, J.S.*, et al.* (2012). Glucocerebrosidase N370S and L444P mutations as risk factors for Parkinson's disease in Brazilian patients. Parkinsonism Relat Disord *18*, 688-689.

Guo, J.F., Li, K., Yu, R.L., Sun, Q.Y., Wang, L., Yao, L.Y., Hu, Y.C., Lv, Z.Y., Luo, L.Z., Shen, L.*, et al.* (2015). Polygenic determinants of Parkinson's disease in a Chinese population. Neurobiol Aging *36*, 1765 e1761-1765 e1766.

Gutti, U., Fung, H.C., Hruska, K.S., Lamarca, M.E., Chen, C.M., Wu, Y.R., and Sidransky, E. (2008). The need for appropriate genotyping strategies for glucocerebrosidase mutations in cohorts with Parkinson disease. Arch Neurol *65*, 850-851; author reply 851.

Halder, T., Raj, J., Pandey, S., Kumar, A., Kawale, S., Chaudhary, S., Sharma, V., Joshi, D., and Das, P. (2016). Screening of Genetic Mutations in GBA1, GIGYF2 and VPS35 in Parkinson Disease Patients from India. Journal of Genetic Disorders & Genetic Reports.

Han, F., Grimes, D.A., Li, F., Wang, T., Yu, Z., Song, N., Wu, S., Racacho, L., and Bulman, D.E. (2016). Mutations in the glucocerebrosidase gene are common in patients with Parkinson's disease from Eastern Canada. Int J Neurosci *126*, 415-421.

Hu, F.Y., Xi, J., Guo, J., Yu, L.H., Liu, L., He, X.H., Liu, Z.L., Zou, X.Y., and Xu, Y.M. (2010). Association of the glucocerebrosidase N370S allele with Parkinson's disease in two separate Chinese Han populations of mainland China. Eur J Neurol *17*, 1476-1478.

Huang, C.L., Wu-Chou, Y.H., Lai, S.C., Chang, H.C., Yeh, T.H., Weng, Y.H., Chen, R.S., Huang, Y.Z., and Lu, C.S. (2011). Contribution of glucocerebrosidase mutation in a large cohort of sporadic Parkinson's disease in Taiwan. Eur J Neurol *18*, 1227-1232.

Jesus, S., Huertas, I., Bernal-Bernal, I., Bonilla-Toribio, M., Caceres-Redondo, M.T., Vargas-Gonzalez, L., Gomez-Llamas, M., Carrillo, F., Calderon, E., Carballo, M.*, et al.* (2016). GBA Variants Influence Motor and Non-Motor Features of Parkinson's Disease. PLoS One *11*, e0167749.

Kalinderi, K., Bostantjopoulou, S., Paisan-Ruiz, C., Katsarou, Z., Hardy, J., and Fidani, L. (2009). Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Greece. Neurosci Lett *452*, 87-89.

Kumar, K.R., Ramirez, A., Gobel, A., Kresojevic, N., Svetel, M., Lohmann, K., C, M.S., Rolfs, A., Mazzulli, J.R., Alcalay, R.N.*, et al.* (2013). Glucocerebrosidase mutations in a Serbian Parkinson's disease population. Eur J Neurol *20*, 402-405.

Lesage, S., Anheim, M., Condroyer, C., Pollak, P., Durif, F., Dupuits, C., Viallet, F., Lohmann, E., Corvol, J.C., Honore, A.*, et al.* (2011a). Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. Hum Mol Genet *20*, 202-210.

Lesage, S., Condroyer, C., Hecham, N., Anheim, M., Belarbi, S., Lohman, E., Viallet, F., Pollak, P., Abada, M., Durr, A.*, et al.* (2011b). Mutations in the glucocerebrosidase gene confer a risk for Parkinson disease in North Africa. Neurology *76*, 301-303.

Li, N., Wang, L., Zhang, J., Tan, E.K., Li, J., Peng, J., Duan, L., Chen, C., Zhou, D., He, L.*, et al.* (2020). Whole-exome sequencing in early-onset Parkinson's disease among ethnic Chinese. Neurobiol Aging *90*, 150 e155-150 e111.

Li, Y., Sekine, T., Funayama, M., Li, L., Yoshino, H., Nishioka, K., Tomiyama, H., and Hattori, N. (2014). Clinicogenetic study of GBA mutations in patients with familial Parkinson's disease. Neurobiol Aging *35*, 935 e933-938.

Liu, G., Boot, B., Locascio, J.J., Jansen, I.E., Winder-Rhodes, S., Eberly, S., Elbaz, A., Brice, A., Ravina, B., van Hilten, J.J.*, et al.* (2016). Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. Ann Neurol *80*, 674-685.

Liu, X., Cheng, R., Verbitsky, M., Kisselev, S., Browne, A., Mejia-Sanatana, H., Louis, E.D., Cote, L.J., Andrews, H., Waters, C.*, et al.* (2011). Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. BMC Med Genet *12*, 104.

Lunde, K.A., Chung, J., Dalen, I., Pedersen, K.F., Linder, J., Domellof, M.E., Elgh, E., Macleod, A.D., Tzoulis, C., Larsen, J.P.*, et al.* (2018). Association of glucocerebrosidase polymorphisms and mutations with dementia in incident Parkinson's disease. Alzheimers Dement *14*, 1293-1301.

Lwin, A., Orvisky, E., Goker-Alpan, O., LaMarca, M.E., and Sidransky, E. (2004). Glucocerebrosidase mutations in subjects with parkinsonism. Mol Genet Metab *81*, 70-73.

Mahungu, A.C., Anderson, D.G., Rossouw, A.C., van Coller, R., Carr, J.A., Ross, O.A., and Bardien, S. (2020). Screening of the glucocerebrosidase (GBA) gene in South Africans of African ancestry with Parkinson's disease. Neurobiol Aging *88*, 156 e111-156 e114.

Malec-Litwinowicz, M., Rudzinska, M., Szubiga, M., Michalski, M., Tomaszewski, T., and Szczudlik, A. (2014). Cognitive impairment in carriers of glucocerebrosidase gene mutation in Parkinson disease patients. Neurol Neurochir Pol *48*, 258-261.

Malek, N., Weil, R.S., Bresner, C., Lawton, M.A., Grosset, K.A., Tan, M., Bajaj, N., Barker, R.A., Burn, D.J., Foltynie, T.*, et al.* (2018). Features of GBA-associated Parkinson's disease at presentation in the UK Tracking Parkinson's study. J Neurol Neurosurg Psychiatry *89*, 702-709.

Mao, X.Y., Burgunder, J.M., Zhang, Z.J., An, X.K., Zhang, J.H., Yang, Y., Li, T., Wang, Y.C., Chang, X.L., and Peng, R. (2010). Association between GBA L444P mutation and sporadic Parkinson's disease from Mainland China. Neurosci Lett *469*, 256-259.

Mata, I.F., Leverenz, J.B., Weintraub, D., Trojanowski, J.Q., Chen-Plotkin, A., Van Deerlin, V.M., Ritz, B., Rausch, R., Factor, S.A., Wood-Siverio, C.*, et al.* (2016). GBA Variants are associated with a distinct pattern of cognitive deficits in Parkinson's disease. Mov Disord *31*, 95-102.

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.*, et al.* (2008). Glucocerebrosidase gene mutations: a risk factor for Lewy body disorders. Arch Neurol *65*, 379-382.

McNeill, A., Duran, R., Hughes, D.A., Mehta, A., and Schapira, A.H. (2012). A clinical and family history study of Parkinson's disease in heterozygous glucocerebrosidase mutation carriers. J Neurol Neurosurg Psychiatry *83*, 853-854.

Mitsui, J., Mizuta, I., Toyoda, A., Ashida, R., Takahashi, Y., Goto, J., Fukuda, Y., Date, H., Iwata, A., Yamamoto, M.*, et al.* (2009). Mutations for Gaucher disease confer high susceptibility to Parkinson disease. Arch Neurol *66*, 571-576.

Moraitou, M., Hadjigeorgiou, G., Monopolis, I., Dardiotis, E., Bozi, M., Vassilatis, D., Vilageliu, L., Grinberg, D., Xiromerisiou, G., Stefanis, L.*, et al.* (2011). beta-Glucocerebrosidase gene mutations in two cohorts of Greek patients with sporadic Parkinson's disease. Mol Genet Metab *104*, 149-152.

Muldmaa, M., Mencacci, N.E., Pittman, A., Kadastik-Eerme, L., Sikk, K., Taba, P., Hardy, J., and Koks, S. (2021). Community-based genetic study of Parkinson's disease in Estonia. Acta Neurol Scand *143*, 89-95.

Neumann, J., Bras, J., Deas, E., O'Sullivan, S.S., Parkkinen, L., Lachmann, R.H., Li, A., Holton, J., Guerreiro, R., Paudel, R.*, et al.* (2009). Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. Brain *132*, 1783-1794.

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.*, et al.* (2009). Mutations in GBA are associated with familial Parkinson disease susceptibility and age at onset. Neurology *72*, 310-316.

Nishioka, K., Vilarino-Guell, C., Cobb, S.A., Kachergus, J.M., Ross, O.A., Wider, C., Gibson, R.A., Hentati, F., and Farrer, M.J. (2010). Glucocerebrosidase mutations are not a common risk factor for Parkinson disease in North Africa. Neurosci Lett *477*, 57-60.

Noreau, A., Riviere, J.B., Diab, S., Dion, P.A., Panisset, M., Soland, V., Jodoin, N., Langlois, M., Chouinard, S., Dupre, N.*, et al.* (2011). Glucocerebrosidase mutations in a French-Canadian Parkinson's disease cohort. Can J Neurol Sci *38*, 772-773.

Olszewska, D.A., McCarthy, A., Soto-Beasley, A.I., Walton, R.L., Magennis, B., McLaughlin, R.L., Hardiman, O., Ross, O.A., and Lynch, T. (2020). Association Between Glucocerebrosidase Mutations and Parkinson's Disease in Ireland. Front Neurol *11*, 527.

Petrucci, S., Ginevrino, M., Trezzi, I., Monfrini, E., Ricciardi, L., Albanese, A., Avenali, M., Barone, P., Bentivoglio, A.R., Bonifati, V.*, et al.* (2020). GBA-Related Parkinson's Disease: Dissection of Genotype-Phenotype Correlates in a Large Italian Cohort. Mov Disord *35*, 2106-2111.

Pulkes, T., Choubtum, L., Chitphuk, S., Thakkinstian, A., Pongpakdee, S., Kulkantrakorn, K., Hanchaiphiboolkul, S., Tiamkao, S., and Boonkongchuen, P. (2014). Glucocerebrosidase mutations in Thai patients with Parkinson's disease. Parkinsonism Relat Disord *20*, 986-991.

Quadri, M., Yang, X., Cossu, G., Olgiati, S., Saddi, V.M., Breedveld, G.J., Ouyang, L., Hu, J., Xu, N., Graafland, J.*, et al.* (2015). An exome study of Parkinson's disease in Sardinia, a Mediterranean genetic isolate. Neurogenetics *16*, 55-64.

Ran, C., Brodin, L., Forsgren, L., Westerlund, M., Ramezani, M., Gellhaar, S., Xiang, F., Fardell, C., Nissbrandt, H., Soderkvist, P.*, et al.* (2016). Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. Neurobiol Aging *45*, 212 e215-212 e211.

Ran, C., Brodin, L., Gellhaar, S., Westerlund, M., Fardell, C., Nissbrandt, H., Soderkvist, P., Sydow, O., Markaki, I., Hertz, E.*, et al.* (2022). Glucocerebrosidase variant T369M is not a risk factor for Parkinson's disease in Sweden. Neurosci Lett *784*, 136767.

Ren, J., Zhang, R., Pan, C., Xu, J., Sun, H., Hua, P., Zhang, L., Zhang, W., Xu, P., Ma, C.*, et al.* (2022). Prevalence and genotype-phenotype correlations of GBA-related Parkinson disease in a large Chinese cohort. Eur J Neurol *29*, 1017-1024.

Ruskey, J.A., Greenbaum, L., Ronciere, L., Alam, A., Spiegelman, D., Liong, C., Levy, O.A., Waters, C., Fahn, S., Marder, K.S.*, et al.* (2019). Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. Eur J Med Genet *62*, 65-69.

Sato, C., Morgan, A., Lang, A.E., Salehi-Rad, S., Kawarai, T., Meng, Y., Ray, P.N., Farrer, L.A., St George-Hyslop, P., and Rogaeva, E. (2005). Analysis of the glucocerebrosidase gene in Parkinson's disease. Mov Disord *20*, 367-370.

Seto-Salvia, N., Pagonabarraga, J., Houlden, H., Pascual-Sedano, B., Dols-Icardo, O., Tucci, A., Paisan-Ruiz, C., Campolongo, A., Anton-Aguirre, S., Martin, I.*, et al.* (2012). Glucocerebrosidase mutations confer a greater risk of dementia during Parkinson's disease course. Mov Disord *27*, 393-399.

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.*, et al.* (2009). Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med *361*, 1651-1661.

Socal, M.P., Bock, H., Michelin-Tirelli, K., Hilbig, A., Saraiva-Pereira, M.L., Rieder, C.R., and Jardim, L.B. (2009). Parkinson's disease and the heterozygous state for glucocerebrosidase mutations among Brazilians. Parkinsonism Relat Disord *15*, 76-78.

Spataro, N., Roca-Umbert, A., Cervera-Carles, L., Valles, M., Anglada, R., Pagonabarraga, J., Pascual-Sedano, B., Campolongo, A., Kulisevsky, J., Casals, F.*, et al.* (2017). Detection of genomic rearrangements from targeted resequencing data in Parkinson's disease patients. Mov Disord *32*, 165-169.

Spitz, M., Rozenberg, R., Pereira Lda, V., and Reis Barbosa, E. (2008). Association between Parkinson's disease and glucocerebrosidase mutations in Brazil. Parkinsonism Relat Disord *14*, 58-62.

Stoker, T.B., Camacho, M., Winder-Rhodes, S., Liu, G., Scherzer, C.R., Foltynie, T., Evans, J., Breen, D.P., Barker, R.A., and Williams-Gray, C.H. (2020). Impact of GBA1 variants on long-term clinical progression and mortality in incident Parkinson's disease. J Neurol Neurosurg Psychiatry *91*, 695-702.

Straniero, L., Asselta, R., Bonvegna, S., Rimoldi, V., Melistaccio, G., Solda, G., Aureli, M., Della Porta, M., Lucca, U., Di Fonzo, A.*, et al.* (2020). The SPID-GBA study: Sex distribution, Penetrance, Incidence, and Dementia in GBA-PD. Neurol Genet *6*, e523.

Sun, Q.Y., Guo, J.F., Wang, L., Yu, R.H., Zuo, X., Yao, L.Y., Pan, Q., Xia, K., and Tang, B.S. (2010). Glucocerebrosidase gene L444P mutation is a risk factor for Parkinson's disease in Chinese population. Mov Disord *25*, 1005-1011.

Tan, E.K., Tong, J., Fook-Chong, S., Yih, Y., Wong, M.C., Pavanni, R., and Zhao, Y. (2007). Glucocerebrosidase mutations and risk of Parkinson disease in Chinese patients. Arch Neurol *64*, 1056-1058.

Tipton, P.W., Soto-Beasley, A.I., Walton, R.L., Soler-Rangel, S., Romero-Osorio, O., Diaz, C., Moreno-Lopez, C.L., Ross, O.A., Wszolek, Z.K., and Cerquera-Cleves, C. (2020). Prevalence of GBA p.K198E mutation in Colombian and Hispanic populations. Parkinsonism Relat Disord *73*, 16-18.

Toft, M., Pielsticker, L., Ross, O.A., Aasly, J.O., and Farrer, M.J. (2006). Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population. Neurology *66*, 415-417.

Torok, R., Zadori, D., Torok, N., Csility, E., Vecsei, L., and Klivenyi, P. (2016). An assessment of the frequency of mutations in the GBA and VPS35 genes in Hungarian patients with sporadic Parkinson's disease. Neurosci Lett *610*, 135-138.

Velez-Pardo, C., Lorenzo-Betancor, O., Jimenez-Del-Rio, M., Moreno, S., Lopera, F., Cornejo-Olivas, M., Torres, L., Inca-Martinez, M., Mazzetti, P., Cosentino, C.*, et al.* (2019). The distribution and risk effect of GBA variants in a large cohort of PD patients from Colombia and Peru. Parkinsonism Relat Disord *63*, 204-208.

Wang, C., Cai, Y., Gu, Z., Ma, J., Zheng, Z., Tang, B.S., Xu, Y., Zhou, Y., Feng, T., Wang, T.*, et al.* (2014). Clinical profiles of Parkinson's disease associated with common leucine-rich repeat kinase 2 and glucocerebrosidase genetic variants in Chinese individuals. Neurobiol Aging *35*, 725 e721-726.

Wang, Y., Liu, L., Xiong, J., Zhang, X., Chen, Z., Yu, L., Chen, C., Huang, J., Zhang, Z., Mohmed, A.A.*, et al.* (2012). Glucocerebrosidase L444P mutation confers genetic risk for Parkinson's disease in central China. Behav Brain Funct *8*, 57.

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., and Barker, R.A. (2013). Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. Brain *136*, 392-399.

Wu, Y.R., Chen, C.M., Chao, C.Y., Ro, L.S., Lyu, R.K., Chang, K.H., and Lee-Chen, G.J. (2007). Glucocerebrosidase gene mutation is a risk factor for early onset of Parkinson disease among Taiwanese. J Neurol Neurosurg Psychiatry *78*, 977-979.

Yadav, R., Kapoor, S., Madhukar, M., Naduthota, R.M., Kumar, A., and Pal, P.K. (2018). Genetic analysis of the glucocerebrosidase gene in South Indian patients with Parkinson's disease. Neurol India *66*, 1649-1654.

Ylonen, S., Siitonen, A., Nalls, M.A., Ylikotila, P., Autere, J., Eerola-Rautio, J., Gibbs, R., Hiltunen, M., Tienari, P.J., Soininen, H.*, et al.* (2017). Genetic risk factors in Finnish patients with Parkinson's disease. Parkinsonism Relat Disord *45*, 39-43.

Yu, Z., Wang, T., Xu, J., Wang, W., Wang, G., Chen, C., Zheng, L., Pan, L., Gong, D., Li, X.*, et al.* (2015). Mutations in the glucocerebrosidase gene are responsible for Chinese patients with Parkinson's disease. J Hum Genet *60*, 85-90.

Zhang, X., Bao, Q.Q., Zhuang, X.S., Gan, S.R., Zhao, D., Liu, Y., Hu, Q., Chen, Y., Zhu, F., Wang, L.*, et al.* (2012). Association of Common Variants in the Glucocerebrosidase Gene with High Susceptibility to Parkinson's Disease among Chinese. Chin J Physiol *55*, 398-404.

Zhang, Y., Sun, Q.Y., Zhao, Y.W., Shu, L., Guo, J.F., Xu, Q., Yan, X.X., and Tang, B.S. (2015). Effect of GBA Mutations on Phenotype of Parkinson's Disease: A Study on Chinese Population and a Meta-Analysis. Parkinsons Dis *2015*, 916971.

Ziegler, S.G., Eblan, M.J., Gutti, U., Hruska, K.S., Stubblefield, B.K., Goker-Alpan, O., LaMarca, M.E., and Sidransky, E. (2007). Glucocerebrosidase mutations in Chinese subjects from Taiwan with sporadic Parkinson disease. Mol Genet Metab *91*, 195-200.

SUPPLEMENTARY MATERIAL FOR

The combined effect of lifestyle factors and polygenic scores on age at onset in Parkinson's disease

Carolin Gabbert[1], MSc, Leonie Blöbaum[1], BSc, Theresa Lüth[1], MSc, Prof. Inke R. König[2], Amke Caliebe[3], PhD, Sebastian Sendel[3], MSc, Björn-Hergen Laabs[2], PhD, Christine Klein[1], MD, Joanne Trinh[1*], PhD

Author affiliations:

[1]Institute of Neurogenetics, University of Lübeck, Lübeck, Germany

[2]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany

[3]Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel, Germany

[*]Correspondence to: Joanne Trinh

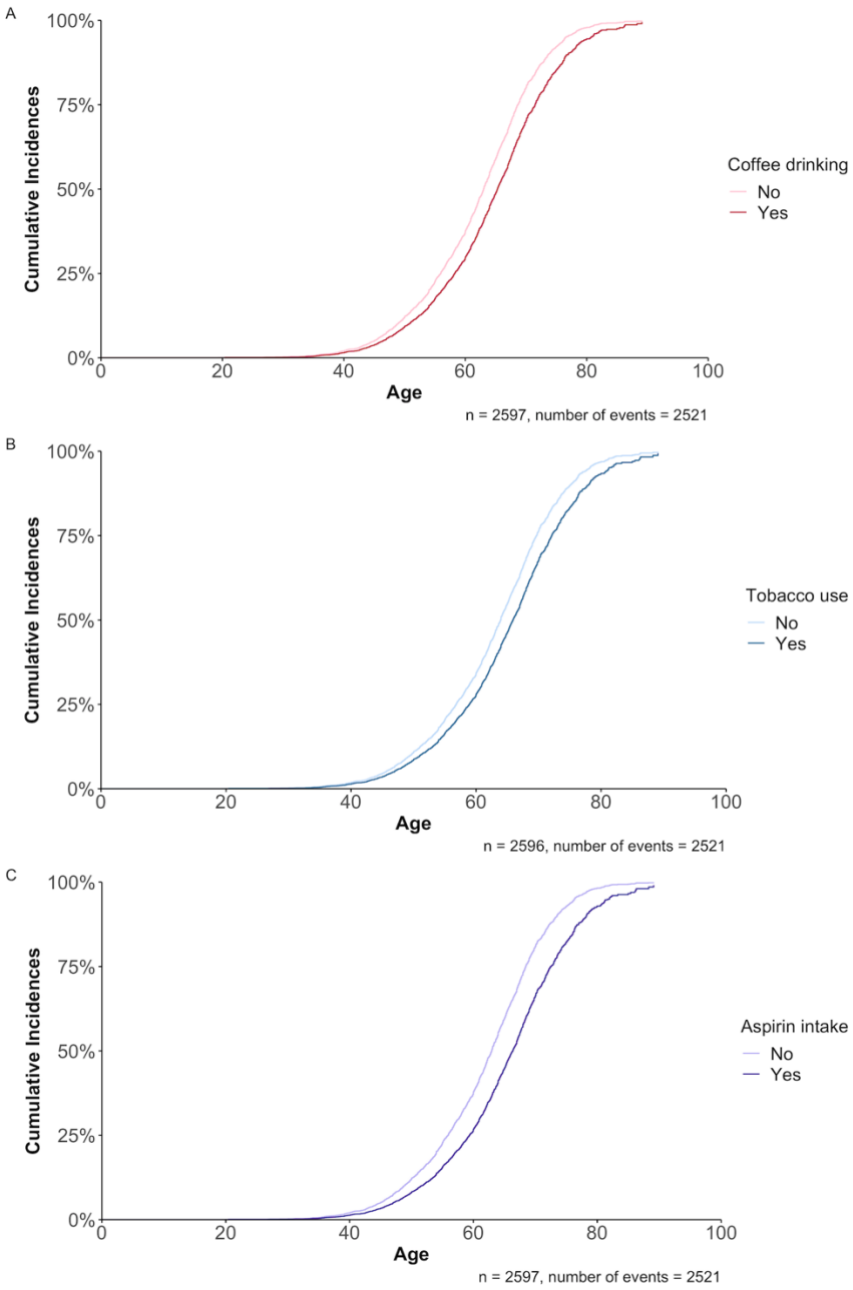
University of Lübeck

Ratzeburger Allee 160

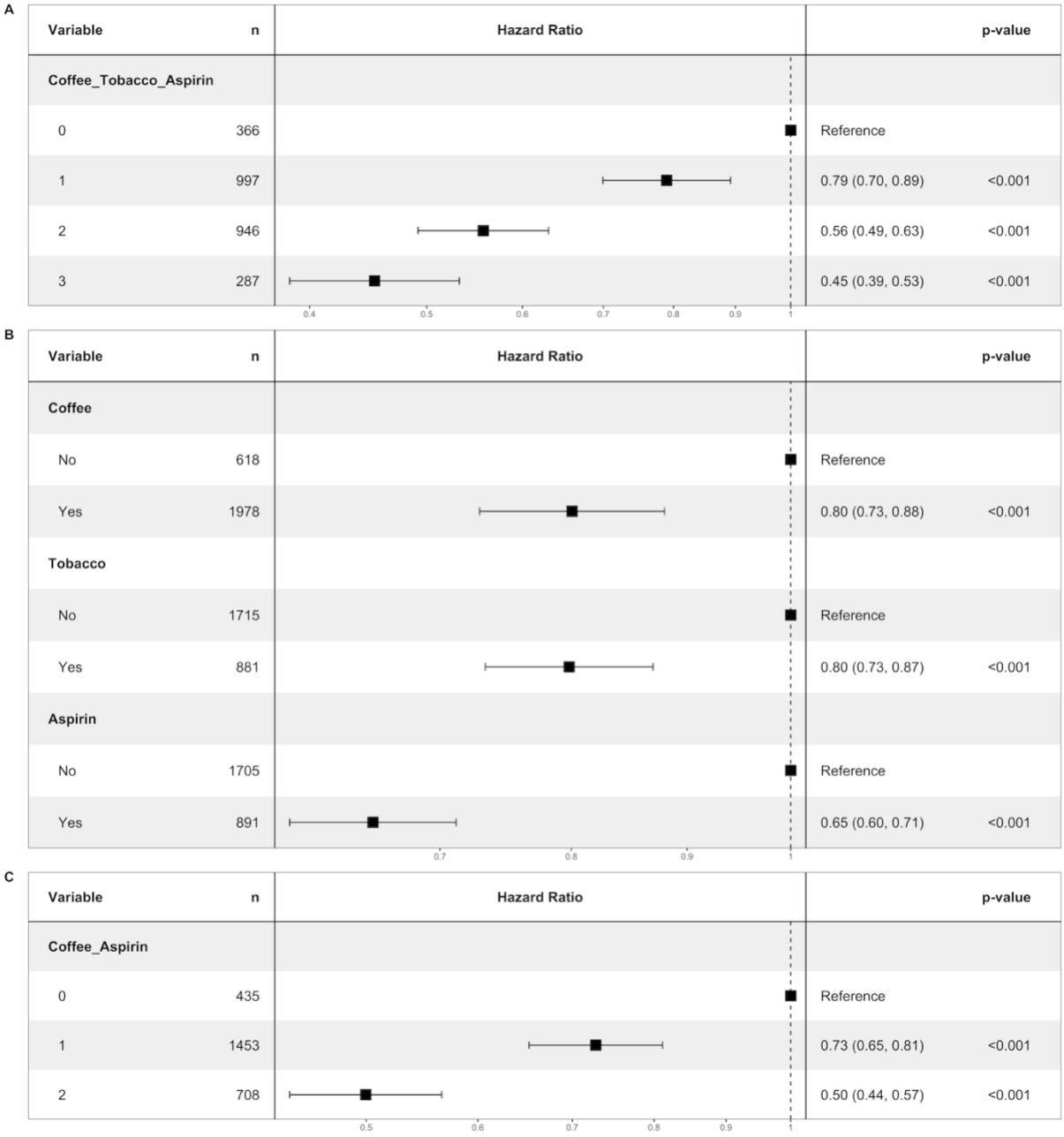
23538 Lübeck, Germany

Email: joanne.trinh@uni-luebeck.de

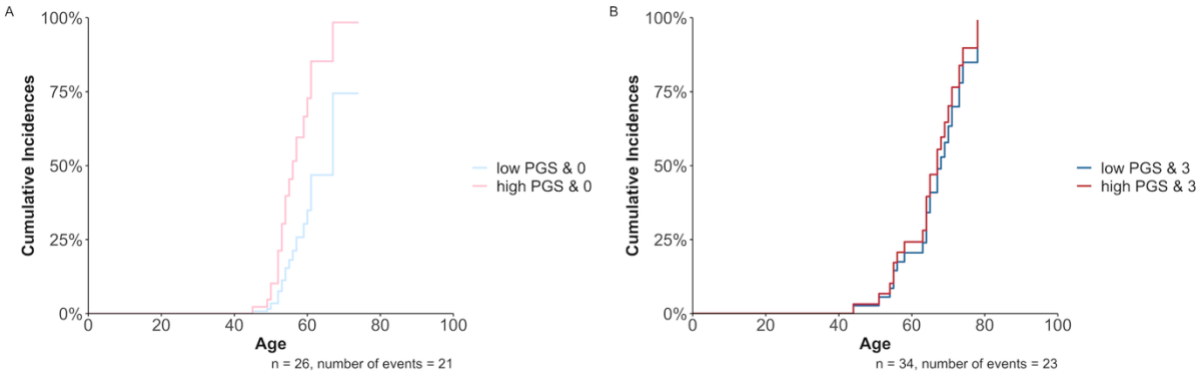
Tel.: +49-451-31018202



Supplementary Figure S1: Plot of the Cox proportional hazards models to investigate the association between the lifestyle factors coffee drinking, tobacco use, and aspirin intake on the AAO of PD patients, while censoring with the AAE of healthy controls. (A) The different curves describe the coffee drinkers and non-coffee drinkers. A Cox proportional hazards model was used to investigate the difference in AAO while censoring with the AAE of healthy controls. The sex and the study site were additionally included as covariates (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Coffee + Sex + Study, data = data)) **(B)** The different curves describe the tobacco users and non-tobacco users. A Cox proportional hazards model was used to investigate the difference in AAO while censoring with the AAE of healthy controls. The sex and the study site were additionally included as covariates (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Tobacco + Sex + Study, data = data)) **(C)** The different curves describe the aspirin users and non-aspirin users. A Cox proportional hazards model was used to investigate the difference in AAO while censoring with the AAE of healthy controls. The sex and the study site were additionally included as covariates (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Aspirin + Sex + Study, data = data))



Supplementary Figure S2: Plot of the Cox proportional hazards model to investigate the additive effects between the use of the lifestyle factors coffee drinking, tobacco use, and aspirin intake on the AAO of PD patients, while censoring with the AAE of healthy controls. (A) The use of the lifestyle factors coffee drinking, tobacco use, and aspirin intake was used as cumulative number (0-3). The sex and study site were additionally included as covariables but are not displayed. (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Coffee/Tobacco/Aspirin + Sex + Study, data = data)). **(B)** The lifestyle factors coffee drinking, tobacco use, and aspirin intake were included as separate covariables. The sex and study site were additionally included as covariables but are not displayed. (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Coffee + Tobacco + Aspirin + Sex + Study, data = data)). **(C)** The use of the lifestyle factors coffee drinking and aspirin intake was used as cumulative number (0-2). The sex and study site were additionally included as covariables but are not displayed. (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ Coffee/Aspirin + Sex + Study, data = data)).



Supplementary Figure S3: Additive effects of the PGS and lifestyle factors on the AAO of PD patients, while censoring with the AAE of healthy controls. (A) The different curves describe the PGS categorized into “low PGS” and “high PGS” according to the median PGS in the subgroup of participants that used no protective lifestyle factor. A Cox proportional hazards model was used to investigate the difference in AAO while censoring with the AAE of healthy controls. The sex and the first two PCs were additionally included (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ PGS low/high + Sex + PC1 + PC2, data = data_Coffee/Tobacco/Aspirin 0)). **(B)** The different curves describe the PGS categorized into “low PGS” and “high PGS” according to the median PGS in the subgroup of participants that used all three protective lifestyle factors. A Cox proportional hazards model was used to investigate the difference in AAO while censoring with the AAE of healthy controls. The sex and the first two PCs were additionally included (→ coxph(formula = Surv(AAO/AAE, Diagnosis) ~ PGS low/high + Sex + PC1 + PC2, data = data_Coffee/Tobacco/Aspirin 3)).

Supplementary Table S1: Linear model on the association of coffee drinking, tobacco use, and aspirin intake with AAO in the *GBA1*-PD study group.

	Estimate	Standard error	*p*-value	
Coffee drinking (binary) (*n* = 145)[1]				
Intercept	58.8245	1.9240	**$<2\times10^{-16}$**	*
Coffee drinking (binary)	3.3597	2.0773	0.1080	
Sex (Male)	0.5488	1.6176	0.7349	
Coffee drinking dosage (*n* = 113)[1]				
Intercept	60.7349	1.5662	**$<2\times10^{-16}$**	*
Coffee drinking dosage	0.0791	0.1137	0.4878	
Sex (Male)	0.7757	1.7817	0.6641	
Coffee drinking duration (*n* = 93)[1]				
Intercept	55.2783	1.7340	**$<2\times10^{-16}$**	*
Coffee drinking duration	0.2423	0.0455	**7×10^{-7}**	*
Sex (Male)	-0.4705	1.9440	0.8093	
Tobacco use (binary) (*n* = 149)[1]				
Intercept	60.2231	1.2026	**$<2\times10^{-16}$**	*
Tobacco use (binary)	3.6527	1.5808	**0.0223**	*
Sex (Male)	0.9111	1.5198	0.5498	
Tobacco use dosage (*n* = 124)[1]				
Intercept	60.8739	1.2053	**$<2\times10^{-16}$**	*
Tobacco use dosage	0.1386	0.0791	0.0823	
Sex (Male)	-0.2229	1.6733	0.8942	
Tobacco use duration (*n* = 121)[1]				
Intercept	60.7810	1.2286	**$<2\times10^{-16}$**	*
Tobacco use duration	0.2193	0.0882	**0.0143**	*
Sex (Male)	0.3614	1.6704	0.8291	
Aspirin intake (binary) (*n* = 92)[1]				
Intercept	59.3073	1.6076	**$<2\times10^{-16}$**	*
Aspirin intake (binary)	3.5143	2.2708	0.1253	
Sex (Male)	1.1559	2.0902	0.5816	
Aspirin intake dosage (*n* = 88)[1]				
Intercept	59.3123	1.4826	**$<2\times10^{-16}$**	*
Aspirin intake dosage	0.4355	0.2213	0.0524	
Sex (Male)	1.5660	2.0077	0.4376	
Aspirin intake duration (*n* = 83)[1]				
Intercept	58.985	1.5121	**$<2\times10^{-16}$**	*
Aspirin intake duration	0.4810	0.2066	**0.0224**	*
Sex (Male)	1.7848	2.0705	0.3913	
All lifestyle factors (binary) (*n* = 89)[2]				
Intercept	56.4789	2.7708	**$<2\times10^{-16}$**	*
Coffee drinking (binary)	0.2295	2.8641	0.9363	
Tobacco use (binary)	6.7844	2.1707	**0.0024**	*
Aspirin intake (binary)	2.1989	2.2613	0.3336	
Sex (Male)	1.8093	2.0733	0.3853	

[1]glm(formula = AAO ~ Lifestyle factor + Sex, family = gaussian, data=data).
[2]glm(formula = AAO ~ Coffee drinking (binary) + Tobacco use (binary) + Aspirin intake (binary) + Sex, family = gaussian, data=data).
* *p*-value < 0.05 are highlighted in bold.
Abbreviations: AAO, age at onset; PD, Parkinson's disease; glm, generalized linear model.

Supplementary Table S2: Linear model on the association of coffee drinking, tobacco use, and aspirin intake with AAO using *GBA1* variant carrier status (*GBA1*-PD vs. iPD) as another covariate.

	Estimate	**Standard error**	***p*-value**	
Coffee drinking (binary) (*n* = 2666)[1]				
Intercept	57.4021	0.4039	**$<2\times10^{-16}$**	*
Coffee drinking (binary)	2.9247	0.4293	**2×10^{-11}**	*
Sex (Male)	1.1880	0.3657	**0.0012**	*
GBA1 mutation status (*GBA1*-PD)	1.4501	0.8007	0.0702	
Coffee drinking dosage (*n* = 2438)[1]				
Intercept	58.7078	0.3216	**$<2\times10^{-16}$**	*
Coffee drinking dosage	0.0815	0.0193	**2×10^{-5}**	*
Sex (Male)	1.0858	0.3880	**0.0052**	*
GBA1 mutation status (*GBA1*-PD)	2.0984	0.8989	**0.0197**	*
Coffee drinking duration (*n* = 1816)[1]				
Intercept	55.5061	0.3723	**$<2\times10^{-16}$**	*
Coffee drinking duration	0.1791	0.0099	**$<2\times10^{-16}$**	*
Sex (Male)	-0.4503	0.4184	0.2821	
GBA1 mutation status (*GBA1*-PD)	1.4076	0.9306	0.1306	
Tobacco use (binary) (*n* = 2670)[1]				
Intercept	58.8585	0.2913	**$<2\times10^{-16}$**	*
Tobacco use (binary)	2.0021	0.3847	**2×10^{-7}**	*
Sex (Male)	1.3576	0.3645	**0.0002**	*
GBA1 mutation status (*GBA1*-PD)	1.7411	0.7919	**0.0280**	*
Tobacco use dosage (*n* = 2410)[1]				
Intercept	59.1459	0.2808	**$<2\times10^{-16}$**	*
Tobacco use dosage	0.0681	0.0125	**6×10^{-8}**	*
Sex (Male)	1.0795	0.3832	**0.0049**	*
GBA1 mutation status (*GBA1*-PD)	1.4818	0.8588	0.0846	
Tobacco use duration (*n* = 2180)[1]				
Intercept	59.9896	0.2977	**$<2\times10^{-16}$**	*
Tobacco use duration	0.1311	0.0221	**3×10^{-9}**	*
Sex (Male)	0.9518	0.4015	**0.0178**	*
GBA1 mutation status (*GBA1*-PD)	1.8386	0.8750	**0.0357**	*
Aspirin intake (binary) (*n* = 2612)[1]				
Intercept	58.2492	0.2813	**$<2\times10^{-16}$**	*
Aspirin intake (binary)	4.6349	0.3854	**$<2\times10^{-16}$**	*
Sex (Male)	0.7787	0.3656	**0.0333**	*
GBA1 mutation status (*GBA1*-PD)	0.9098	0.9778	0.3522	
Aspirin intake dosage (*n* = 2495)[1]				
Intercept	58.8474	0.2811	**$<2\times10^{-16}$**	*
Aspirin intake dosage	0.2684	0.0383	**3×10^{-12}**	*
Sex (Male)	1.0059	0.3757	**0.0075**	*
GBA1 mutation status (*GBA1*-PD)	1.1040	1.0053	0.2722	
Aspirin intake duration (*n* = 2166)[1]				
Intercept	58.6431	0.2888	**$<2\times10^{-16}$**	*
Aspirin intake duration	0.2948	0.0285	**$<2\times10^{-16}$**	*
Sex (Male)	0.5388	0.3982	0.1762	
GBA1 mutation status (*GBA1*-PD)	1.3935	1.0246	0.1739	

All lifestyle factors (binary) (*n* = 2610)[2]				
Intercept	56.1502	0.4099	**$<2\times10^{-16}$**	*
Coffee drinking (binary)	2.3116	0.4313	**9×10^{-8}**	*
Tobacco use (binary)	1.5787	0.3858	**4×10^{-5}**	*
Aspirin intake (binary)	4.5174	0.3820	**$<2\times10^{-16}$**	*
Sex (Male)	0.4930	0.3643	0.1761	
GBA1 mutation status (*GBA1*-PD)	0.5397	0.9842	0.5835	

[1]glm(formula = AAO ~ Lifestyle factor + Sex + *GBA1* mutation status, family = gaussian, data=data).
[2]glm(formula = AAO ~ Coffee drinking (binary) + Tobacco use (binary) + Aspirin intake (binary) + Sex + *GBA1* mutation status, family = gaussian, data=data).
* *p*-value < 0.05 are highlighted in bold.
Abbreviations: AAO, age at onset; PD, Parkinson's disease; glm, generalized linear model.

SUPPLEMENTARY MATERIAL FOR

Interaction of Mitochondrial Polygenic Score and Lifestyle Factors in LRRK2 p.Gly2019Ser Parkinsonism

Theresa Lüth[1], MSc, Carolin Gabbert[1], Sebastian Koch[2], MSc, MSc, Prof. Inke R. König[3], Amke Caliebe[2], PhD, Björn-Hergen Laabs[3], PhD, Faycel Hentati[4], MD, Samia Ben Sassi[4], MD, Rim Amouri[4], MD, Malte Spielmann[5], MD, Christine Klein[1], MD, Anne Grünewald[1,6], PhD, Matthew J. Farrer[7], PhD, Joanne Trinh[1*], PhD

Christine Klein[1], MD, Joanne Trinh[1*], PhD

Author affiliations:

[1]Institute of Neurogenetics, University of Lübeck, Lübeck, Germany

[2]Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel, Germany

[3]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany

[4]Neurology Department, National Institute of Neurology, Tunis, Tunisia

[5]Institute of Human Genetics, University of Lübeck, Lübeck, Germany

[6]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[7]Clinical Genomics, University of Florida, Gainesville, FL, USA

[*]Correspondence to: Joanne Trinh

University of Lübeck

Ratzeburger Allee 160

23538 Lübeck, Germany

Email: joanne.trinh@uni-luebeck.de

Tel.: +49-451-31018202

Supplementary Table S1: Demographics of patients with PD from the three investigated data sets.

		LRRK2-PD	iPD
Total	***N***	**486**	**9259**
	Number of men (%)	**250 (51.44%)**	**5262 (56.83%)**
	Mean AAE (SD)	**66.68 (12.36)**	**65.22 (9.64)**
	Mean AAO (SD)	**58.17 (11.07)**	**61.17 (10.18)**
AMP-PD	*N*	127	2077
	Race/Ethnicity (%)	European/White (94.4%), Arab (1.6%), Black/African American (1.6%), Hispanic/Latino (1.6%), Native Hawaiian/Other Pacific Islander (0.8%)	European/White (95.4%), Black/African American (1.2%), Hispanic/Latino (1.5%), American Indian/Alaska Native (>0.1%), Asian (0.9%), Multiethnic (0.8%)
	Number of men (%)	67 (52.76%)	1309 (63.02%)
	Mean AAE (SD)	67.97 (10.63)	64.39 (9.50)
	Mean AAO (SD)	58.49 (11.24)	60.73 (9.97)
Fox Insight	*N*	154	6949
	Race/Ethnicity (%)	European/White (100%)	European/White (100%)
	Number of men (%)	82 (53.25%)	3837 (55.22%)
	Mean AAE (SD)	66.95 (8.09)	65.36 (9.475)
	Mean AAO (SD)	62.12 (9.176)	61.51 (9.98)
Tunisian cohort	*N*	205	233
	Race/Ethnicity (%)	Tunisian/Arab (100%)	Tunisian/Arab (100%)
	Number of men (%)	101 (49.27%)	117 (50.21)
	Mean AAE (SD)	65.67 (15.56)	68.37 (14.10)
	Mean AAO (SD)	56.30 (11.63)	55.03 (14.75)

N=Number of individuals, AAE=Age at examination, AAO=Age at onset, iPD=idiopathic Parkinson's disease, *LRRK2*-PD= Patients with PD that carry the LRRK2 p.Gly2019Ser variant

Supplementary Text 1. Description of the genetic data from the study cohorts.

AMP-PD contains whole-genome sequencing (WGS) data from four harmonized cohorts. All samples of the AMP-PD dataset were processed by the TOPMed Freeze 9 Variant Calling Pipeline for joint genotyping(Iwaki et al., 2021). The Fox Insight dataset is a cohort within the Michael J. Fox Foundation (MJFF) (Smolensky et al., 2020). For Fox Insight, the genetic data (array-based genotyping) of the patients with PD were provided by 23andMe, as previously described (Smolensky et al., 2020). Three platforms were used to perform the genotyping within the Fox Insight data: V3 (Illumina OmniExpress + BeadChip), V4 (fully custom array) and V5 (customized Illumina Infinium Global Screening Array)(Smolensky et al., 2020). For the cohort recruited from the Tunisian Arab-Berber population, the array-based genotyping data were obtained from Affymetrix and Illumina MEGA arrays, as previously described (Trinh et al., 2016).

Supplementary Text 2. Description of the environment and lifestyle data from the study cohorts.

In the Fox Insight and Tunisian cohort, lifestyle and environmental information were assessed with the PD Risk Factor Questionnaire (PD-RFQ) for tobacco use, caffeine consumption, and pesticide exposure. The questionnaire was developed by the National Institute of Environmental Health Sciences, validated for interview and self-reporting settings and provides a tool for assessing lifestyle/environmental factors (https://www.commondataelements.ninds.nih.gov/report-viewer/23723/Risk%20Factor%20Questionnaire%20(RFQ-U)). According to the PD-RFQ, tobacco use is defined as smoking at least one cigarette per day for more than six months, smoking more than 100 cigarettes in your lifetime, or using smokeless tobacco at least once per day for more than six months. Caffeine consumption means drinking coffee, black tea, green tea or caffeinated soda at least once per week for more than six months

Supplementary Table S2: Prevalence of tobacco use, caffeine consumption and pesticide exposure from included cohorts (i.e., Fox Insight and Tunisian cohort) stratified by *LRRK2*-PD and iPD status.

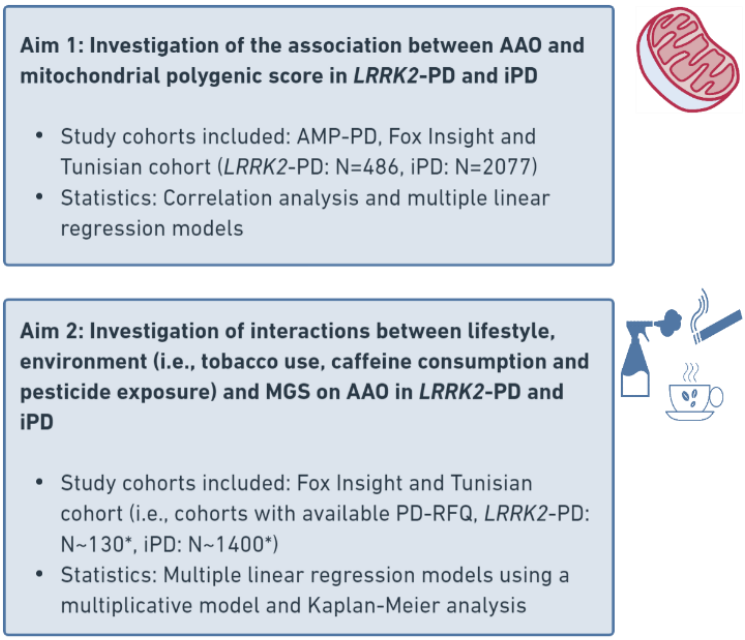
Lifestyle/environmental factors	**Fox Insight**		**Tunisian cohort**	
	***LRRK2*-PD**	**iPD**	***LRRK2*-PD**	**iPD**
Tobacco use *N* (%)	Yes: 20 (54.1%) No: 17 (45.9%)	Yes: 702 (40.5%) No: 1032 (59.5%)	Yes: 39 (36.4%) No: 68 (63.6%)	Yes: 8 (34.8%) No: 15 (65.2%)
Coffee consumption *N* (%)	Yes: 23 (82.1%) No: 5 (17.9%)	Yes: 1291 (78.2%) No: 360 (21.8%)	Yes: 55 (51.9%) No: 51 (48.1%)	Yes: 13 (56.5%) No: 10 (43.5%)
Black tea consumption *N* (%)	Yes: 7 (28.0%) No: 18 (72.%)	Yes: 557 (39.4%) No: 857 (60.6%)	Yes: 62 (60.2%) No: 41 (39.8%)	Yes: 10 (43.5%) No: 13 (56.5%)
Green tea consumption *N* (%)	Yes: 4 (15.4%) No: 22 (14.6%)	Yes: 227 (16.7%) No: 1130 (83.3%)	Yes: 30 (28.3%) No: 76 (71.7%)	Yes: 6 (26.1%) No: 17 (73.9%)
Caffeinated soda consumption *N* (%)	Yes: 14 (53.8%) No: 12 (46.2%)	Yes: 939 (69.7%) No: 408 (30.3%)	Yes: 33 (31.4%) No: 72 (68.6%)	Yes: 8 (38.1%) No: 13 (61.9%)
Pesticide exposure in a work setting *N* (%)	Yes: 4 (22.2%) No: 14 (77.8%)	Yes: 197 (20.9%) No: 746 (79.1%)	Yes: 18 (16.8%) No: 89 (83.2%)	Yes: 3 (13.6%) No: 19 (86.4%)
Pesticide exposure in a non-work setting *N* (%)	Yes: 13 (76.5%) No: 4 (23.5%)	Yes: 820 (90.3%) No: 88 (9.7%)	Yes: 36 (38.3%) No: 58 (61.7%)	Yes: 14 (66.7%) No: 7 (33.3%)

N=Number of individuals, iPD=idiopathic Parkinson's disease, *LRRK2*-PD=Patients with PD that carry the LRRK2 p.Gly2019Ser variant

Supplementary Text 3. Mitochondrial polygenic score generation.

The MGS was generated using samples collated within the framework of the Research Group 'ProtectMove' (FOR2488) funded by the German Research Foundation DFG. Those samples originated from five German cohorts comprising a total of 1914 Parkinson's disease cases and

4464 controls after quality control. A detailed description of the cohorts, genotyping and performed quality control is available in a previous publication (Koch et al., 2021).
Regarding summary statistics, the GWAS performed by Nalls *et al.*(Nalls et al., 2019) was used, and permission was granted by 23andMe.
We restricted our analysis to SNPs in gene regions associated with mitochondrial function as listed in the "secondary gene set" in the publication by Billingsley *et al.* (Billingsley et al., 2019). This left us with a total of 168,629 SNPs as the basis for generating the MGS. We removed all samples that had a mutation in at least one of the Parkinson-associated genes "CHCHD2", "LRRK2", "SNCA", "VPS35", "DJ-1", "Parkin", "PINK1", and "GBA". This left us with 6,000 samples (2,800 females, 3,161 males, 39 ambiguous) comprising 1,805 cases and 4,195 controls. We divided the samples into 5 equally sized sets. We stratified for case-control status, sex, the age for controls and age at onsets for cases with the R package "groupdata2" (version 2.0.2) (https://CRAN.R-project.org/package=groupdata2). Using these five datasets, we performed a five-fold cross-validation, where the MGS was developed on 4/5 of the data as a training dataset and validated on the remaining fifth. For each of these five training datasets, we generated multiple MGSs using the PRSice2, LDpred2, and lassosum2 methods. For PRSice2, we used all combinations that resulted from window sizes 250, 500, 750, and 1000 kB and r^2 thresholds 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, and 0.8. In addition, we had PRSice2 calculate the LD (linkage disequilibrium) once from the training data and once from the 1000 Genomes Project as a reference dataset. Regarding LDpred2 and lassosum2, we calculated the correlation matrix for all SNPs available to us. For LDpred2, the "inf", "auto", and "grid" models were then used, the latter via a grid consisting of weights for the SNP heritability h^2 with values 0.7, 1, and 1.4 as well as 17 values for the proportion of causal variants p between 0.0001 and 1.0 on a logarithmic scale, for both the sparse and non-sparse versions. The "auto" model varied over 30 p parameters logarithmically distributed between 0.0001 and 0.5. Both LDpred2 and lassosum2 were calculated using the R package "bigsnpr" (version 1.9.11) (Prive et al., 2018).
After calculation of the described MGS models, we selected per training dataset and method (PRSice2 (LD-reference own data), PRSice2 (LD-reference 1000 genomes), LDpred2 inf, LDpred2 auto, LDpred2 grid, lassosum2) the model with associated parameters that performed best on the respective training dataset with respect to AUC.
These models were then applied to the associated validation datasets and the mean AUC was determined across all 5 validation datasets per model. Here, the LDpred2 auto model performed best. Therefore, we subsequently computed an MGS with the whole dataset as a training dataset using LDpred2 auto. This final model achieved an AUC of 0.560 (95% CI: [0.544, 0.576]) with an odds ratio of 1.245 per standard deviation.



Supplementary Figure S1: Overview of the Study aims. The study aims, along with the included cohorts and used statistics, are displayed.
*The exact sample size depends on the specific lifestyle/environment factor investigated. Please see Table 2 and Supplementary Table S3 for exact sample sizes.

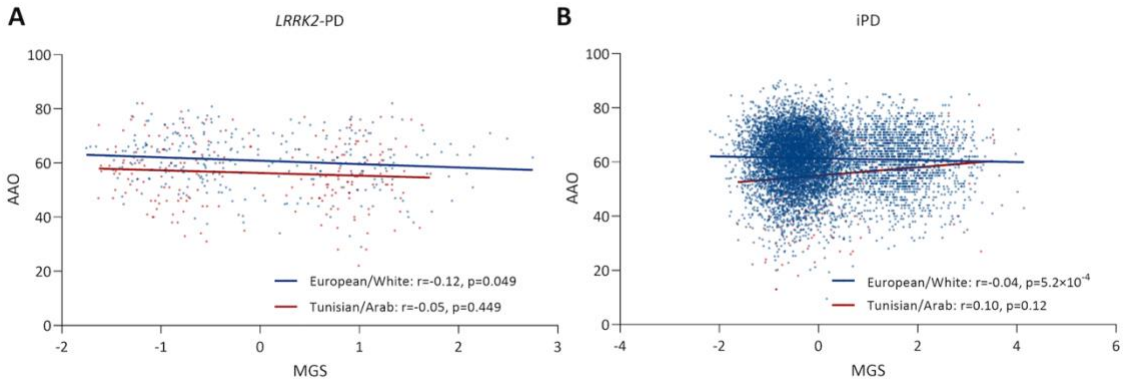
Supplementary Text 4. Principal component analysis.

Visualization of the race/ethnicity with 1000 Genomes Project

In order to visualize populations, we used PLINK to perform a principal component analysis (PCA) based on common SNPs (MAF > 0.3). Furthermore, we performed LD-pruning and filtered for a genotyping rate larger than 99%. As a reference, we used the publicly available 1000 Genomes Project data. After filtering for common SNPs, we excluded variants not present in both data sets (i.e, AMP-PD and 1000 Genomes Project, Fox Insight and 1000 Genomes Project or Tunisian cohort and 1000 Genomes Project) and then three individual PCAs were performed. Finally, the output was visualized with R and colored according to the study cohort and race/ethnicity.

Principal component analysis as covariates in the regression model

In order to utilize the first five principal components in the regression models, we performed one PCA with the merged PLINK file of AMP-PD, Fox Insight and the Tunisian cohort data sets. Before merging the three data sets, we filtered each individually as described above (i.e., MAF> 0.3, genotyping rate > 99% and LD-pruning using the PLINK parameter --indep-pairwise 50 10 0.3). Subsequently, the SNPs present in all three data sets were extracted and the merging was performed followed by the PCA analysis. PC1-5 from the PLINK output was utilized as covariates in the regression models of our study.



Supplementary Figure S2: Relationship between age at onset (AAO) and mitochondrial polygenic score (MGS). The correlation plots show the association between MGS and AAO in patients with Parkinson's disease, stratified by ethnicity (i.e., European/white and Tunisian/Arab), carrying the LRRK2 p.Gly2019Ser mutation (LRRK2-PD) (A) or patients with idiopathic PD (iPD) (B).
r = Spearman's rank correlation coefficient, *p* = Spearman's exploratory p-value.

Supplementary Table S3: Interaction between the mitochondrial polygenic score and tobacco use on the age at onset in patients with idiopathic PD.

iPD[1]			
	Estimate	**95% *CI***	***p*-value**
Tobacco users (*N*=710)			
MGS	**0.71**	**-0.14, 1.55**	**0.103**
Sex Male	2.14	0.66, 3.62	0.005*
PC1	-0.60	-1.58, 0.39	0.235
PC2	-0.32	-1.12, 0.48	0.432
PC3	1.70	0.86, 2.53	7.3×10^{-5}*
PC4	-0.24	-0.97, 0.50	0.524
PC5	0.15	-0.60, 0.89	0.696
Tobacco non-users (*N*=1049)			
MGS	**-0.79**	**-1.41, -0.17**	**0.013***
Sex Male	0.63	-0.47, 1.74	0.262
PC1	-1.26	-1.99, -0.52	0.001*
PC2	0.20	-0.32, 0.73	0.451
PC3	0.34	-0.29, 0.97	0.287
PC4	0.32	-0.23, 0.88	0.251
PC5	-0.28	-0.81, 0.25	0.298

The regression analysis was applied on idiopathic PD patients that used tobacco and those that did not, separately.
iPD=idiopathic Parkinson's disease, MGS=Mitochondrial polygenic score, * *p*-value < 0.05
[1]glm(formula = AAO ~ MGS + Sex + PC1 + PC2 + PC3 + PC4 + PC5 family = gaussian)
Baseline categories: Sex=Female

Supplementary Table S4: Interaction between the mitochondrial polygenic score and caffeinated soda consumption on the age at onset in patients with *LRRK2*-PD.

***LRRK2*-PD**[1]			
	Estimate	**95% *CI***	***p*-value**
Caffeinated soda consumption (*N*=49)			
MGS	**-2.30**	**-7.99, 3.39**	**0.432**
Sex Male	5.53	-0.66, 11.72	0.088
PC1	-4.04	-12.75, 4.67	0.369
PC2	-3.01	-7.31, 1.29	0.177
PC3	1.32	-4.79, 7.42	0.675
PC4	0.11	-3.15, 3.37	0.949
PC5	1.53	-1.96, 5.03	0.394
No caffeinated soda consumption (*N*=84)			
MGS	**3.86**	**0.36, 7.35**	**0.034***
Sex Male	3.14	-1.62, 7.91	0.200
PC1	-4.45	-10.23, 1.33	0.136
PC2	2.56	-0.27, 5.39	0.080
PC3	-0.90	-6.00, 4.20	0.731
PC4	1.17	-1.18, 3.52	0.332
PC5	0.03	-2.02, 2.09	0.975

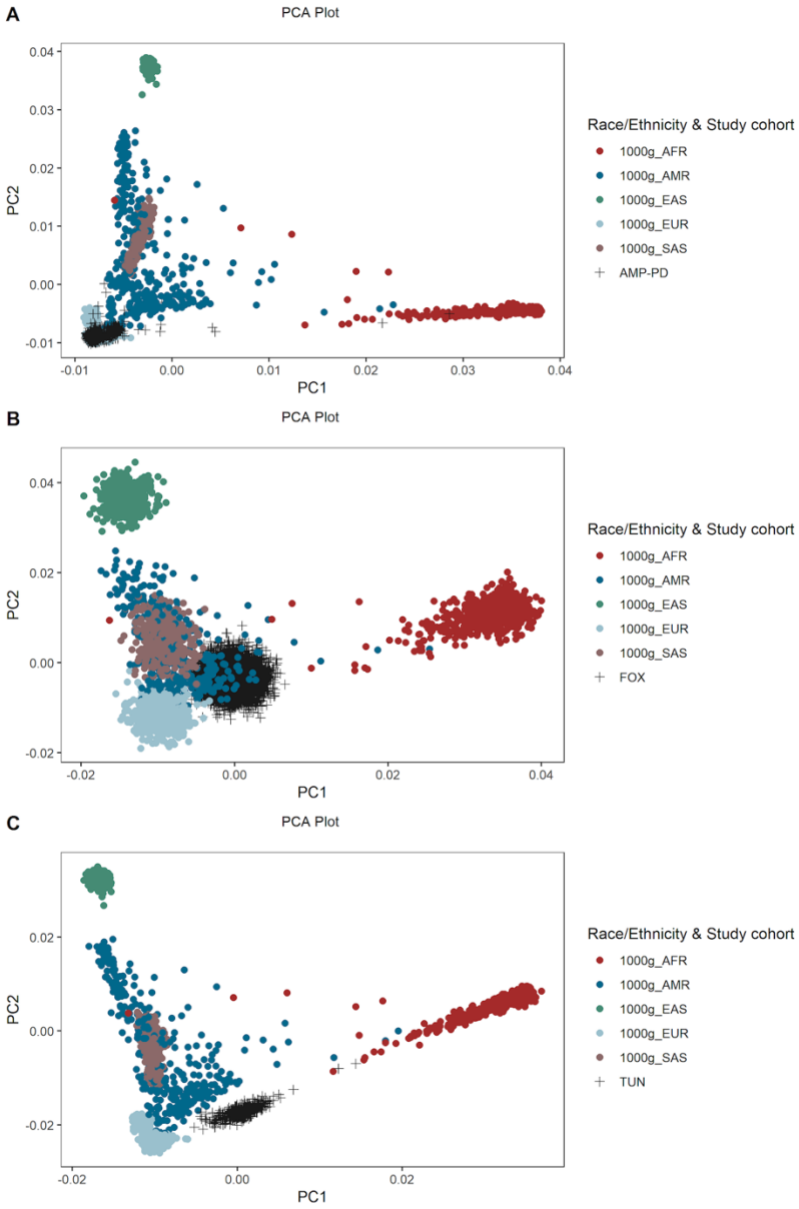
The regression analysis was applied on LRRK2-PD patients that consumed caffeinated soda and those that did not, separately. *LRRK2*-PD= Patients with PD that carry the LRRK2 p.Gly2019Ser variant, * p-value < 0.05
[1]glm(formula = AAO ~ MGS + Sex + PC1 + PC2 + PC3 + PC4 + PC5 family = gaussian)
Baseline categories: Sex=Female

Supplementary Table S5: Interaction between the mitochondrial polygenic score, caffeine consumption and pesticide exposure and the age at onset in patients with *LRRK2*-PD and idiopathic PD.

	***LRRK2*-PD**[1]			**iPD**[1]		
	Estimate	**95% *CI***	***p*-value**	**Estimate**	**95% *CI***	***p*-value**
	Black tea consumption (*N*=130)			**Black tea consumption (*N*=1437)**		
Lifestyle factor	4.09	-1.13, 9.30	0.127	0.74	-0.26, 1.75	0.148
MGS	0.13	-3.71, 3.97	0.946	-0.63	-1.33, 0.07	0.080
Sex Male	3.61	-0.23, 7.46	0.068	0.75	-0.24, 1.73	0.138
PC1	-4.30	-9.21, 0.60	0.088	-1.04	-1.68, -0.41	0.001*
PC2	0.36	-2.13, 2.85	0.777	0.06	-0.43, 0.55	0.812
PC3	0.25	-3.61, 4.10	0.901	0.68	0.13, 1.24	0.016*
PC4	0.83	-1.13, 2.80	0.408	0.17	-0.32, 0.66	0.492
PC5	0.29	-1.52, 2.10	0.756	-0.13	-0.61, 0.35	0.592
MGS:Lifestyle factor	**1.68**	**-2.52, 5.87**	**0.434**	**1.05**	**-0.02, 2.13**	**0.056**
	Green tea consumption (*N*=134)			**Green tea consumption (*N*=1382)**		
Lifestyle factor	-1.46	-7.71, 4.78	0.647	1.01	-0.33, 2.34	0.141
MGS	2.21	-1.07, 5.49	0.189	-0.25	-0.87, 0.36	0.425
Sex Male	3.17	-0.65, 7.00	0.107	0.60	-0.40, 1.60	0.242
PC1	-4.47	-9.36, 0.42	0.076	-1.04	-1.68, -0.41	0.001*
PC2	0.82	-1.57, 3.21	0.503	-0.02	-0.52, 0.48	0.926
PC3	0.29	-3.58, 4.16	0.882	0.75	0.18, 1.31	0.010*
PC4	1.05	-0.94, 3.04	0.304	0.23	-0.27, 0.73	0.367
PC5	0.24	-1.59, 2.08	0.795	-0.17	-0.66, 0.32	0.490
MGS:Lifestyle factor	**-1.03**	**-5.77, 3.71**	**0.671**	**0.25**	**-1.11, 1.62**	**0.714**
	Pesticide exposure in a work setting (*N*=126)			**Pesticide exposure in a work setting (*N*=967)**		
Exposure	2.63	-4.70, 9.97	0.483	-1.35	-2.88, 0.19	0.087
MGS	2.02	-1.21, 5.24	0.223	-0.28	-1.04, 0.49	0.478
Sex Male	3.44	-0.64, 7.52	0.101	1.28	0.04, 2.52	0.044*
PC1	-3.85	-9.06, 1.36	0.150	-1.03	-1.77, -0.29	0.006*
PC2	0.60	-2.07, 3.27	0.661	-0.02	-0.65, 0.62	0.959
PC3	0.31	-3.67, 4.28	0.880	1.02	0.33, 1.70	0.004*
PC4	1.40	-0.62, 3.43	0.177	0.17	-0.45, 0.80	0.585
PC5	0.05	-1.85, 1.96	0.957	0.00	-0.60, 0.60	0.996
MGS:Exposure	**-2.35**	**-7.75, 3.05**	**0.396**	**0.77**	**-0.88, 2.42**	**0.358**
	Pesticide exposure in a non-work setting (*N*=112)			**Pesticide exposure in a non-work setting (*N*=931)**		
Exposure	-3.76	-10.18, 2.65	0.253	1.27	-0.77, 3.30	0.224
MGS	-0.03	-4.30, 4.23	0.988	1.50	-0.38, 3.39	0.118
Sex Male	3.86	-0.41, 8.13	0.080	0.77	-0.47, 2.00	0.225
PC1	-5.18	-10.57, 0.21	0.062	-1.31	-2.07, -0.54	0.001*
PC2	0.43	-2.32, 3.19	0.758	0.08	-0.56, 0.72	0.813
PC3	0.18	-4.04, 4.40	0.934	0.87	0.17, 1.56	0.015*
PC4	0.66	-1.63, 2.94	0.574	0.12	-0.50, 0.74	0.703
PC5	0.29	-1.80, 2.38	0.788	0.16	-0.45, 0.77	0.615
MGS:Exposure	**3.24**	**-1.52, 8.00**	**0.185**	**-1.61**	**-3.58, 0.36**	**0.109**

iPD=idiopathic Parkinson's disease, *LRRK2*-PD= Patients with PD that carry the LRRK2 p.Gly2019Ser variant, MGS=Mitochondrial polygenic score , * *p*-value < 0.05

[1]glm(formula = AAO ~ MGS * Lifestyle factor/Exposure + Sex + PC1 + PC2 + PC3 + PC4 + PC5, family = gaussian)
Baseline categories: Sex=Female



Supplementary Figure S3: Principal component analysis (PCA). The PCA plot displays the clustering of samples included in this study. The PCA was derived from common SNPs overlapping in all data sets. The samples were colored by their study cohort and their ethnicity/race. In addition to our study cohorts (i.e., AMP-PD (A), Fox Insight (B) and Tunisian cohort (C)), we used the publicly available 1000 Genomes Project dataset for validation.
1000 Genomes Project super populations: AFR=African, EAS=East Asian, AMR=Ad Mixed American, EUR=European and SAS=South Asian

Supplementary Table S6: Association between MGS and AAO of patients with early-onset Parkinson's disease (EOPD: AAO<50 years) and late-onset Parkinson's disease (LOPD: AAO>50 years).

	***LRRK2*-PD**			**iPD**		
	Estimate	**95% *CI***	***p*-value**	**Estimate**	**95% *CI***	***p*-value**
EOPD (*LRRK2*-PD: *N*=105, iPD: *N*=1226)						
MGS	**-1.13**	**-2.61, 0.35**	**0.136**	**0.15**	**-0.39, 0.70**	**0.582**
Sex Male	0.38	-1.90, 2.66	0.747	0.06	-0.62, 0.74	0.865
PC1	-0.91	-2.41, 0.58	0.235	-0.36	-0.93, 0.22	0.223
PC2	-0.64	-1.93, 0.65	0.331	-0.05	-0.39, 0.29	0.768
PC3	-0.14	-1.42, 1.15	0.835	1.09	0.75, 1.43	3.0×10^{-10}*
PC4	0.32	-0.81, 1.46	0.578	-0.05	-0.39, 0.29	0.768
PC5	-0.60	-1.73, 0.52	0.295	0.14	-0.20, 0.48	0.419
LOPD (*LRRK2*-PD: *N*=378, iPD: *N*=7810)						
MGS	**-0.86**	**-1.99, 0.27**	**0.138**	**-0.25**	**-0.53, 0.03**	**0.082**
Sex Male	0.94	-0.63, 2.50	0.243	0.83	0.49, 1.17	1.3×10^{-6}*
PC1	0.19	-1.13, 1.50	0.781	-0.16	-0.44, 0.12	0.275
PC2	0.26	-0.70, 1.22	0.595	0.19	0.02, 0.35	0.026*
PC3	-0.18	-1.16, 0.79	0.716	-0.01	-0.17, 0.16	0.945
PC4	-0.96	-1.76, -0.17	0.018*	-0.02	-0.19, 0.15	0.818
PC5	0.18	-0.62, 0.98	0.664	0.13	-0.04, 0.29	0.133

iPD=idiopathic Parkinson's disease, *LRRK2*-PD= Patients with PD that carry the LRRK2 p.Gly2019Ser variant, MGS=Mitochondrial polygenic score, * *p*-value < 0.05
glm(formula = AAO ~ MGS + Sex + PC1 + PC2 + PC3 + PC4 + PC5 , family = gaussian)
Baseline categories: Sex=Female

Supplementary Table S7: Interaction between the mitochondrial polygenic score, tobacco use, caffeine consumption and pesticide exposure and the age at onset in patients with idiopathic PD from the Fox Insight cohort exclusively.

iPD[1]			
	Estimate	**95% *CI***	***p*-value**
Smoking (*N*=1734)			
Lifestyle factor	1.43	0.54, 2.32	0.002*
MGS	-0.37	-0.93, 0.19	0.194
Sex Male	1.23	0.36, 2.11	0.006*
PC1	0.07	-0.36, 0.50	0.751
PC2	0.11	-0.34, 0.55	0.640
PC3	0.40	-0.04, 0.84	0.075
PC4	0.14	-0.30, 0.58	0.533
PC5	-0.07	-0.50, 0.35	0.735
MGS:Lifestyle factor	**0.86**	**-0.02, 1.75**	**0.056**
Coffee consumption (*N*=1651)			
Lifestyle factor	2.15	1.06, 3.24	0.0001*
MGS	0.06	-0.92, 1.04	0.903
Sex Male	0.70	-0.20, 1.60	0.129
PC1	0.01	-0.42, 0.45	0.953
PC2	-0.01	-0.46, 0.44	0.962
PC3	0.30	-0.15, 0.75	0.192
PC4	0.09	-0.37, 0.54	0.711
PC5	-0.06	-0.50, 0.38	0.781
MGS:Lifestyle factor	**0.01**	**-1.09, 1.11**	**0.979**
Black tea consumption (*N*=1412)			
Lifestyle factor	0.78	-0.21, 1.77	0.124
MGS	-0.34	-0.95, 0.27	0.276
Sex Male	0.80	-0.17, 1.77	0.105
PC1	0.18	-0.29, 0.66	0.443
PC2	0.19	-0.29, 0.67	0.441
PC3	0.15	-0.33, 0.63	0.544
PC4	0.19	-0.29, 0.67	0.441
PC5	-0.11	-0.58, 0.36	0.651
MGS:Lifestyle factor	**1.06**	**0.10, 2.03**	**0.031***
Green tea consumption (*N*=1358)			
Lifestyle factor	0.89	-0.43, 2.20	0.186
MGS	0.002	-0.53, 0.53	0.994
Sex Male	0.58	-0.40, 1.56	0.244
PC1	0.18	-0.30, 0.65	0.470
PC2	0.12	-0.37, 0.61	0.639
PC3	0.20	-0.30, 0.69	0.433
PC4	0.25	-0.24, 0.75	0.322
PC5	-0.14	-0.62, 0.34	0.570
MGS:Lifestyle factor	**0.44**	**-0.82, 1.70**	**0.493**

Caffeinated soda consumption (*N*=1347)			
Lifestyle factor	-2.22	-3.29, -1.16	4.5×10^{-5}*
MGS	0.04	-0.81, 0.90	0.918
Sex Male	0.51	-0.47, 1.49	0.305
PC1	0.16	-0.32, 0.63	0.520
PC2	0.07	-0.42, 0.56	0.779
PC3	0.19	-0.30, 0.68	0.444
PC4	0.29	-0.20, 0.78	0.247
PC5	-0.22	-0.70, 0.25	0.355
MGS:Lifestyle factor	0.004	-1.03, 1.03	0.994
Pesticide exposure in a work setting (*N*=943)			
Exposure	-1.61	-3.10, -0.11	0.035*
MGS	0.22	-0.45, 0.88	0.528
Sex Male	1.27	0.05, 2.48	0.041*
PC1	0.36	-0.22, 0.94	0.219
PC2	0.19	-0.43, 0.81	0.554
PC3	0.35	-0.25, 0.94	0.259
PC4	0.21	-0.41, 0.82	0.511
PC5	0.02	-0.57, 0.60	0.957
MGS:Exposure	**0.10**	**-1.35, 1.55**	**0.892**
Pesticide exposure in a non-work setting (*N*=908)			
Exposure	1.35	-0.67, 3.38	0.190
MGS	1.47	-0.49, 3.42	0.143
Sex Male	0.79	-0.41, 1.99	0.197
PC1	0.36	-0.23, 0.95	0.230
PC2	0.30	-0.33, 0.92	0.353
PC3	0.16	-0.44, 0.76	0.604
PC4	0.17	-0.43, 0.77	0.578
PC5	0.19	-0.40, 0.78	0.536
MGS:Exposure	**-1.12**	**-3.17, 0.94**	**0.287**

iPD=idiopathic Parkinson's disease, MGS=Mitochondrial polygenic score, * *p*-value < 0.05
[1]glm(formula = AAO ~ MGS * Lifestyle factor/Exposure + Sex + PC1 + PC2 + PC3 + PC4 + PC5 family = gaussian)
Baseline categories: Sex=Female

References:

Billingsley, K.J., Barbosa, I.A., Bandres-Ciga, S., Quinn, J.P., Bubb, V.J., Deshpande, C., Botia, J.A., Reynolds, R.H., Zhang, D., Simpson, M.A.*, et al.* (2019). Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. NPJ Parkinsons Dis *5*, 8.

Iwaki, H., Leonard, H.L., Makarious, M.B., Bookman, M., Landin, B., Vismer, D., Casey, B., Gibbs, J.R., Hernandez, D.G., Blauwendraat, C.*, et al.* (2021). Accelerating Medicines Partnership: Parkinson's Disease. Genetic Resource. Mov Disord *36*, 1795-1804.

Koch, S., Laabs, B.H., Kasten, M., Vollstedt, E.J., Becktepe, J., Bruggemann, N., Franke, A., Kramer, U.M., Kuhlenbaumer, G., Lieb, W.*, et al.* (2021). Validity and Prognostic Value of a Polygenic Risk Score for Parkinson's Disease. Genes (Basel) *12*.

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.*, et al.* (2019). 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.

Prive, F., Aschard, H., Ziyatdinov, A., and Blum, M.G.B. (2018). Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. Bioinformatics *34*, 2781-2787.

Smolensky, L., Amondikar, N., Crawford, K., Neu, S., Kopil, C.M., Daeschler, M., Riley, L., andMe Research, T., Brown, E., Toga, A.W.*, et al.* (2020). Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Sci Data *7*, 67.

Trinh, J., Gustavsson, E.K., Vilarino-Guell, C., Bortnick, S., Latourelle, J., McKenzie, M.B., Tu, C.S., Nosova, E., Khinda, J., Milnerwood, A.*, et al.* (2016). DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. Lancet Neurol *15*, 1248-1256.