Title
Genetic risk factors for dyskinesia and hallucinations among Parkinson's disease patients

Author
Kusters, Cynthia Diana Johanna

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Genetic risk factors for dyskinesia and hallucinations among Parkinson's disease patients

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Cynthia Diana Johanna Kusters

2019

ABSTRACT OF THE DISSERTATION

Genetic risk factors for dyskinesia and hallucinations among Parkinson's disease patients

by

Cynthia Diana Johanna Kusters

Doctor of Philosophy of Epidemiology

University of California, Los Angeles, 2019

Professor Beate R. Ritz, Chair

Parkinson's disease (PD) will become more prevalent in the next decades as the world's population ages.[1] Although PD is diagnosed based on its motor symptoms, PD includes a wide range of non-motor symptoms. Two symptoms, in particular, significantly decrease quality-of-life: hallucinations and dyskinesia. Hallucinations are an important co-morbidity and dyskinesia is a common treatment-related complication. This dissertation analyzes the genetic risk factors for both dyskinesia and hallucinations, and identifies specific genetic variants and combinations of genetic variants that are associated with a higher risk of developing these symptoms.

In the first study, a large population-based study (PEG) was used to establish the association between four candidate genes (DRD1, DRD2, DRD3 and BDNF) and dyskinesia. This study consisted of 418 patients whose diagnosis was confirmed by a movement disorder specialist, who were using levodopa, and who had a minimum of three years disease duration at the time of

assessment. Applying Haploview and Phase, haploblocks for DRD1-3 and BDNF were created. Risk scores for DRD2 and DRD3 were generated. Risk ratios were calculated using Poisson regression with robust error variance. One haplotype in each DRD2 haploblock was associated with a 29% to 50% increase in dyskinesia risk. For each unit increase in risk score we observed a 16% increase in dyskinesia risk for DRD2 (95%CI: 1.05-1.29) and a 17% (95%CI: 0.99-1.40) increase for DRD3. The BDNF haploblock was not associated, but the minor allele of the rs6265 SNP was associated with dyskinesia (adjusted RR 1.31 (95%CI: 1.01-1.70)). Among the candidate genes for dyskinesia the following were genetic risk factors for dyskinesia: several haplotypes in DRD2, possibly some haplotypes in DRD3, and the minor allele of rs6265 in BDNF. Among PD patients, there is a constant tradeoff between increasing medication to address PD symptoms and increasing the risk of dyskinesia. Genetic information could help prevent or postpone this debilitating consequence of treatment and may improve patient-centered, personalized therapy. Future studies are needed to confirm our findings and quantify the benefits of implementing a personalized treatment based on a genetic risk score. PD patients with these genetic variants may be prime candidates for treatments aiming to prevent or delay the onset of dyskinesia.

The second and third study are based on three longitudinal PD cohorts: two population-based studies (ParkWest and PEG) and one international clinic-based study (PPMI). The population was restricted to Caucasians only (N=745). The second and third study in this dissertation analyze the association between polygenic risk scores (PRS) and hallucinations. In the second study, we describe the strengths and limitations of a PRS. In addition, a PRS for hallucinations based on PD candidate genes was generated and validation was attempted. The PEG and PW studies were used for the creation of the PRS, and the PPMI was used for validation.

Unfortunately, the PRS generated with the two population-based studies could not be replicated, most likely due to the sample size. A second PRS was created based on a large GWAS for PD. Based on the findings from a pooled analysis of all three studies, the hallucinations PRS, based on the GWAS, indicated that the following genes might contribute to increased risk of developing hallucinations in PD: LRRK2, APOE, SLC6A4, BDNF and MAPT.

In the third study, the overlap of the genetic architecture for Alzheimer's disease (AD) and schizophrenia (SZ) with Parkinson's hallucinations was assessed. For this purpose, two PRS were created. Both PRS were based on previously performed, large GWAS; one for SZ and one for AD. Various PRS were created using different p-value thresholds. The full PRS model, using all SNPs consisted of over 70,000 SNPs (AD and SZ). The genetic risk for hallucinations appears to differ by age at onset of PD. Stratifying by younger (<60 years) and older (60+ years) age at diagnosis, the SZ-PRS was associated with an increased risk for hallucinations among young PD patients (adjusted OR=1.18 (95%CI: 1.03-1.35, p-value 0.02). The AD-PRS was positively associated with hallucinations in older onset PD patients (adjusted OR=1.27 (95%CI: 1.08-1.50, p-value 0.005). The results suggest that the biological mechanisms for hallucinations may depend on age at diagnosis. Among young onset PD patients, SZ susceptibility factors may play a role in the development of hallucinations. In contrast, among older onset PD patients, hallucinations appear to be influenced by the genetic architecture seen in AD that contributes to cognitive decline.

The dissertation of Cynthia Diana Johanna Kusters is approved.

Onyebuchi Aniweta Arah

Janet S Sinsheimer

Stefan Horvath

Beate R. Ritz, Committee Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

List of tables

Supplemental Table 4.2b Logistic regression for the association between AD-PRS and hallucinations among subjects stratified by the studies (PEG, PW and PPMI).

List of figures

List of acronyms and abbreviations

AD	Alzheimer's disease
aOR	Adjusted odds ratio
aRR	Adjusted risk ratio
AUC	Area under the curve
BDNF	Brain derived neurotrophic factor
BMI	Body mass index
CAD	Coronary artery disease
CI	Confidence interval
COMT	Catechol O-MethylTransferase
CSF	Cerebrospinal fluid
DBS	Deep brain stimulation
DRD	Dopamine receptor
GSA	Global Screening Array
GWAS	Genome-wide association study
HR-QoL	Health Related Quality of Life
IBD	Identical by descent
LD	Linkage disequilibrium
L-DOPA	Levodopa
LEDD	levodopa-equivalent daily dosage
LID	Levodopa-induced dyskinesia
MAF	Minor allele frequency
MAO	MonoAmine Oxidase

MDS	Movement disorder society
MSA	Multiple system atrophy
NMDA	N-Methyl-D-Aspartate
NMS	Non-motor symptoms
OR	Odds ratio
PD	Parkinson's disease
PEG	Parkinson's Environment and Gene
PIGD	Postural-instability gait disorder
PPMI	Parkinson's Progression Markers Initiative
PRS	Polygenic risk score
PSP	Progressive supranuclear palsy
PW	ParkWest
RBD	REM sleep behavior disorder
RR	Risk ratio
SNP	Single nucleotide polymorphisms
SNpc	Substantia Nigra pars compacta
SZ	Schizophrenia
UBC	University of British Columbia
UPDRS	Unified Parkinson's disease rating scale

Acknowledgements

I would first and foremost like to thank my advisor, Prof. Beate Ritz, for her invaluable guidance and support throughout my degree. I am very grateful to have her as my mentor and look forward to many collaborations in the future. Thank you to my co-advisor, Prof. Janet Sinsheimer, for her regular advice and critical feedback during my dissertation. Finally, I would like to thank Profs. Horvath and Arah for serving on my dissertation committee and their insightful comments.

Thank you to my office mate, Kim Paul, for sharing her deep knowledge of genetic epidemiological methods.

My time as a student was profoundly enhanced by the support of the Department of Epidemiology, the Burroughs Wellcome Fund fellowship, and the Genomic Analysis Training program.

I would like to share my appreciation to my family. I would like to thank my husband, Benji, and mother-in-law, Joy, for their eagle-eyed proofreading. Last, but not least, to Esther, I am truly honored to be your mom, your energy and desire to explore new adventures is inspirational.

Chapter 2 was originally published in Parkinsonism and Related Disorders published by Elsevier at:

Kusters CDJ, Paul K, Bordelon Y, Bronstein J, Sinsheimer J, Farrer M, Ritz BR. Dopamine receptors and BDNF-haplotypes predict dyskinesia in Parkinson's disease. Parkinsonism Relat Disord 2018 Feb; 47: p39-44.

VITA

EDUCATION

2003 Bachelor of Science, Biomedical Science, University of Nijmegen, the Netherlands
2007 Doctor of Medicine, University of Nijmegen, the Netherlands
2013 Certification of Epidemiologist (equivalence for a Master of Science in epidemiology), Council of the Netherlands Epidemiological Society, Rotterdam, The Netherlands

CLINICAL, RESEARCH AND TEACHING EXPERIENCE

April 2019 – June 2019: Teaching Assistant (primary instructor), Department of Education Initiative, University of California Los Angeles, CA, USA

October 2018 – March 2019: Teaching Assistant, Department of Undergraduate education Initiative, University of California Los Angeles, CA, USA

Jan 2018 – March 2018: Teaching assistant, Department of Epidemiology, University of California Los Angeles, CA, USA

Jan 2015 – Sept 2016: Graduate Student Researcher, Department of Epidemiology, University of California Los Angeles, CA, USA

Jan 2014 – July 2014: Physician Researcher, EUROCAT, Department of Clinical Genetics, University Medical Center Groningen, The Netherlands

Sept 2012 – Sept 2013: Physician, Department of Clinical Genetics, University Medical Center Groningen, The Netherlands

Jan 2011 – July 2012: Physician, Department of Pediatrics, Rijnstate Ziekenhuis Arnhem, The Netherlands

Mar 2009 – Sept 2010: Physician, Neonatal Intensive Care Unit, Department of Pediatrics, Radboud University Medical Center, Nijmegen, the Netherlands

July 2007 – Dec 2008: Research Assistant, Department of Pediatrics, Radboud University Medical Center, Nijmegen, the Netherlands

Jan 2008 – July 2008: Research Intern, Department of Newborn Medicine, Tufts Medical Center, Floating Hospital for Children, Boston, MA, USA

July 2007 – Sept 2007: Physician, Department of Pediatrics, Radboud University Medical Center, Nijmegen, the Netherlands

PUBLICATIONS

Kusters CDJ, Paul K, Bordelon Y, Bronstein J, Sinsheimer J, Farrer M, Ritz BR. Dopamine receptors and BDNF-haplotypes predict dyskinesia in Parkinson's disease. Parkinsonism Relat Disord 2018 Feb; 47: p39-44.

Kheifets L, Swanson J, Yuan Y, Kusters C, Vergara X, "Comparative analyses of studies of childhood leukemia and magnetic fields, radon and gamma radiation," J. Radiol. Prot., 2017 June 26; 37(2), p. 459–491.

Levine ME, Chen BH, Lu A, Hernandez DG, Singleton AB, Ferruci L, Bandinelli S, Salfati E, Manson J., Quach A, Kusters CDJ, Kuh D, Wong A, Teschendorff AE, Widschwendter M, Ritz BR, Absher D, Themistocles A, Horvath S. DNA methylation age of blood is associated with age at menopause. Proc Natl Acad Sci 2016 Aug 16; 113 (33): 9327-32.

Veenendaal van NR, Kusters CDJ, Oostra R-J, Bergman JEH, Cobben J-M. When the right (Drug) should be left: Prenatal drug exposure and heterotaxy syndrome. Birth Defects Res A Clin Mol Teratol. 2016. doi:10.1002/bdra.23497

Kusters CDJ, Pal van der SM, Steenbrugge GJ van, Ouden den L, Kollee LAA. Impact van vroeggeboorte op het gezin. [The impact of a premature birth on the family; consequences are experienced even after 19 years.] NTVG. 2013; 157(25): A5449.

Kusters CDJ, Chen ML, Follett PL, Dammann O. "Intraventricular" hemorrhage and cystic periventricular leukomalacia in preterm infants: how are they related? Journal of Child Neurology. 2009; 24 (9): 1158-70.

Iersel v C, Velden vd HVN, Kusters CDJ, Spauwen PHM, Blokx WAM, Kiemeney LALM, Gerritsen MJP. Prognostic factors for a subsequent basal cell carcinoma: implications for follow-up. British Journal of Dermatology. 2005; 153: p1078-1079.

1. Introduction

1.1 Background Parkinson's disease

1.1.a. Parkinson's disease

James Parkinson was the first to describe a syndrome he named "Shaking Palsy" in 1817.[2] Later this syndrome became known as Parkinson's disease (PD). Dr. Parkinson described two main criteria: "involuntary tremulous motion, with lessened voluntary muscular power, in parts, not in action, and even, supported" and "a propensity to bend the trunk forwards, and to pass from a walking to a running pace."[2] The cardinal symptoms are tremor, rigidity, and bradykinesia. Postural instability is considered a major motor symptom and is sometimes considered a fourth cardinal symptom. Besides the presentation of these three cardinal symptoms, a PD diagnosis requires that there is no indication of other neurodegenerative disorders (e.g. progressive supranuclear palsy (PSP), multiple system atrophy (MSA) and corticobasal degeneration), or a history of specific drugs or medications.[3–6] While the only way to confirm a PD diagnosis is a post mortem examination, repeated antemortem medical examinations are the next best method.[3,7] Biomarker studies, including imaging studies, have been explored unsuccessfully and further research is required.[8–11]

It is well established that PD is caused by the destruction of dopaminergic neurons, specifically in the substantia nigra pars compacta (SNpc) and the accumulation of Lewy bodies in the midbrain. However, the exact mechanism of this destruction remains elusive.[12–15] The neuronal loss in PD is not restricted to dopaminergic neurons, nonetheless dopaminergic neurons are influenced the most.[16] Other neurotransmitter neurons that are affected include: serotonergic, noradrenergic and cholinergic systems. Dopamine is a common neurotransmitter in the brain involved in many processes. These processes include, but are not limited to, movement, cognition, and emotion-driven, the symptoms commonly seen in PD.[17,18] The Lewy bodies contain α-synuclein aggregates, which can be identified in the central and peripheral nervous system. These aggregates are shown to be related to PD, Lewy body dementia, and MSA.[19–21]

1.1.b. Treatment options for PD

The motor symptoms of PD are treated with a variety of medications. The most essential medication for PD is levodopa (L-3,4-dihydroxyphenylalanine). Levodopa increases the concentration of dopamine in the brain. Used in place of, or in combination with levodopa are dopamine agonists, which stimulate the dopamine receptors. Other categories of treatment are monoamine oxidase-B inhibitors (MAO-B inhibitors), catechol O-methyltransferase inhibitors (COMT inhibitors), N-methyl-D-aspartate receptor inhibitors (NMDA inhibitors) and anticholinergics. When either symptoms or side-effects of medications become too severe, deep brain stimulation is explored as an option. Although it has shown high efficacy, the risks of deep brain stimulation are considerable and it does not slow disease progression. [5,22,23]

1.1.c. Incidence/prevalence

Parkinson's disease (PD) is a devastating, progressive neurological disorder affecting more than a million Americans. It is the second most common neurodegenerative disease among the elderly in the U.S.[24] Incidence of PD increases with advancing age.[25] It is estimated to affect more than 8.5 million individuals worldwide by 2030.[1] The worldwide prevalence has been estimated from 41 per 100,000 among individuals between 40 to 49 years of age up to a prevalence of 1,903 per 100,000 among individuals over 80 years of age.[26] The prevalence appears to be lower in Africa, the Middle-East and Asia compared to North-America, South-America and Europe.[26] The reasons for these differences has been hypothesized to be a result of methodological issues, underreporting PD in countries with less-developed health care, differences in exposures, and population stratification. Another study showed that in the U.S. the prevalence appeared highest among Hispanics, followed by non-Hispanic Whites, Asians and then Blacks.[27]

1.1.d. Known risk factors for PD

Recent research points to a multi-factorial etiology of PD, involving genetic and environmental factors.[28–30] The strongest non-genetic risk factor for PD is increasing age. Other non-genetic risk factors include: pesticides; toxins (such as manganese and methamphetamines); traumatic brain injury; and a positive family history. Physical activity and caffeine have been

found to be protective.[26 31–33] Other variables that have been suggested to be protective include: dietary products, smoking, melanoma, reproductive variables, body mass index (BMI), cholesterol concentration, hypertension, alcohol abuse, vitamins, urate, and certain medications. Results for these variables are inconsistent and/or might be due to reverse causation.[32,34,35] Several genes have been identified to cause hereditary PD, although these mutations are rare. These include SNCA, LRRK2, VPS35, Parkin, PINK1, DNAJC6 and DJ1.[36–38] In addition, several more common genetic variants are associated with a slight increased risk for PD.[37]

1.1.e. Clinical features of PD

Although PD is diagnosed based on the motor symptoms, PD includes a wide range of symptoms including non-motor symptoms. Some of these non-motor symptoms are also prodromal symptoms. Prodromal symptoms include: constipation, anosmia, rapid eye movement sleep disorder, fatigue, excessive daytime sleepiness, orthostatic hypotension, erectile and urinary dysfunction, and depression.[5,39–42] Although these symptoms are associated with a higher risk of developing PD, they are non-specific and poor predictors. [5,39] Other non-motor symptoms include dysautonomia, sleep disturbances, mood disorders, cognitive disorders, pain, sensory disturbances and hallucinations.[23,31,43]

1.1.f. Progression

As the neurodegeneration of the dopamine neurons advance, the motor and non-motor symptoms of PD patients worsen over time. While initial treatment can be very effective, over time treatment-related complications become prevalent.[44–47] These progressive symptoms include, but are not limited to, motor- and non-motor fluctuation, dyskinesia, increased postural instability, freezing of gait, cognitive decline, and hallucinations.[37] The Hoehn and Yahr scale is used to measure motor progression. It was devised in 1967 and has been well established.[48]

The diagnosis of PD is associated with an increased risk of overall mortality compared to the general population. [49–51] Factors associated with this increased mortality and progression are: increasing age at PD onset, presence of dementia, gender, psychosis/hallucinations, and a postural-instability gait disorder (PIGD) phenotype.[49–53] Depression and hallucinations are the strongest predictors of nursing home placements among PD patients.[54] In early PD, when motor-

symptoms are under control with medication, the Health Related Quality of Life (HR-QoL) is mainly reduced by the non-motor symptoms.[55–58] Over time, PD symptoms contribute to significant disabilities and a greatly diminished HR-QoL.[56,59]

1.2. Dyskinesia

Dyskinesia refers to the involuntary movements that are caused by L-dopa treatment in patients with PD (It is commonly known as levodopa-induced dyskinesia (LID)). These involuntary movements include chorea, dystonia, myoclonus, tics, stereotypies and akathisia. There are three categories of dyskinesia: peak-dose dyskinesia, diphasic dyskinesia, and off-period dystonia. It is possible for patients to have overlapping types of dyskinesia. Peak-dose dyskinesia is the most common LID. It accounts for more than 80% of all LID and is most prevalent when the levodopa-concentration in the body is highest. Diphasic dyskinesia occurs in about 20% of LID. The dyskinesia symptoms generally begin soon after a levodopa dose is taken, subside for a few hours and then reoccur while the levodopa is wearing off. Off-period dystonia occurs in about 30% of patients when levodopa is wearing off (e.g. in early morning).[60,61] See also figure 1.1,[62] for a visualization of the various types of dyskinesia.

Dyskinesia has been reported to affect between 24-45% of all treated PD patients within the first five years of diagnosis.[44,63,64] This prevalence increases to 80% among patients 10 years into the disease.[65] Several risk factors have been identified for dyskinesia, these include PD disease duration, levodopa-equivalent daily dosage (LEDD), stress and anxiety, female gender, and young age of onset.[60,66–71]

Although LID has been shown to be a major component of the loss of HR-QoL,[72,73] in the early stages, patients are not always aware of the side-effect or confuse the dyskinesia with the PD tremor.[74–76] Dyskinesia is associated with depression in PD, even when motor fluctuations and disease stage are taken into account.[72]

Levodopa-sparing strategies could prevent or delay LID. However, treatment of PD with L-dopa is known to increase HR-QoL even during long-term follow-up.[77–79] In addition, L-dopa treatment has a much lower financial cost than dopamine agonists.[80] Several anti-dyskinetic medications have been approved for treatment, including amantadine.[81] However, the benefits from these treatments are limited. Several new treatments are in development.[72,82] A non-

medication treatment option is deep brain stimulation (DBS), which has demonstrated to have a robust effect. [83,84]

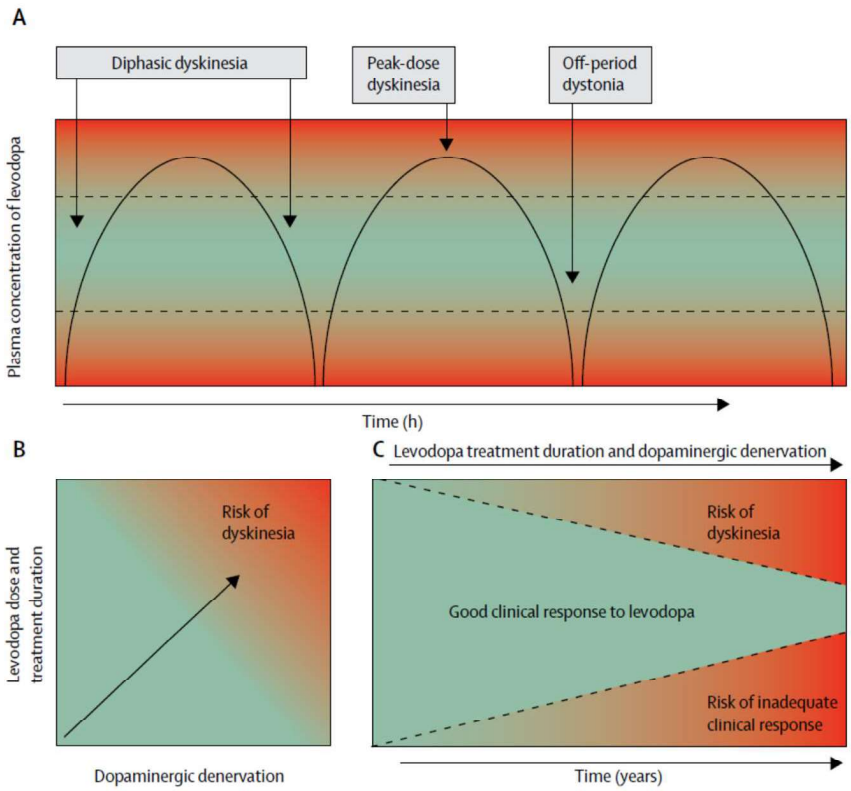


Figure 1.1: "Levodopa-induced dyskinesia: different types and risk factors (A) The different types of levodopa-induced dyskinesias are shown. Diphasic dyskinesias appear at the beginning of the effect of levodopa dosing, before full clinical benefit on motor symptoms is reached, and can reappear when the effect of levodopa starts to wear off. Peak-dose dyskinesias coincide with the full antiparkinsonian benefit of levodopa ("on" period), whereas off -period dystonia appears during "off " periods when levodopa is no longer effective. (B) The risk of developing levodopa-induced dyskinesias depends on both the progression of dopaminergic neurons degeneration and the exposure to levodopa (dose and treatment duration). (C) As the disease progresses, the probability of a good clinical response to levodopa decreases and the risk of both dyskinesias and unsatisfactory clinical response of motor symptoms rises."[62]

1.3 Pathobiology of dyskinesia

It is well established that PD is caused by the destruction of dopaminergic neurons and involves the accumulation of Lewy bodies in the midbrain. Many of the motor symptoms are due to dopamine imbalances in the basal ganglia circuitry.[85] The basal ganglia circuitry has two main pathways called direct and indirect. The direct pathway is associated with D1-like dopamine

receptors while the indirect pathway involves D2-like dopamine receptors. Adequate balance between the two pathways is necessary for the initiation of movement (see Figure 2.2). In PD, the loss of dopaminergic control leads to a hyperactivity of the inhibitory D2 pathway, which produces bradykinesia.[86–90] The restoration of dopamine levels with L-dopa treatment counteracts this imbalance, but can also produce hyperactivity through stimulation of the direct D1 pathway. This stimulation can lead to dyskinesia. DRD3 is a dopamine D2-like receptor. In animal studies, overexpression of the DRD3 receptor is associated with dyskinetic behavior, while D3 antagonists can diminish and/or prevent LID.[91–93] One PET-study found elevated D2/D3 binding in the globus pallidus in patients with LID.[94] Since the D1 and D2 pathways are controlled by dopamine receptors, common genetic variations in DRD loci are natural candidates to explain the risk of dyskinesia.[95,96] Brain derived neurotrophic factor (BDNF) influences neuroplasticity and survival of the dopaminergic neurons in the substantia nigra. BDNF may also affect dyskinesia due to modulation of the dopamine receptor expression.[97]

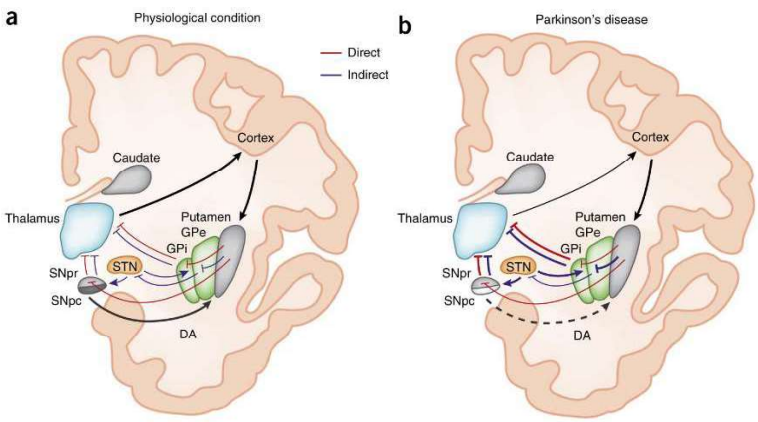


Figure 1.2 (**a**) "In the physiological condition, DA arising from the SNpc is thought to activate D1-expressing striatal MSNs of the direct pathway (red lines) and to inhibit D2-expressing striatal neurons of the indirect pathway (blue lines). The output nuclei GPi and SNpr project to the thalamus, which in turn sends efferents that complete the cortico-basal ganglia-thalamo-cortical loop. (**b**) In Parkinson's disease, degeneration of nigral neurons reduces DA receptor stimulation in striatal MSNs. The imbalance between direct and indirect pathways results in abnormal activation of output nuclei and over-inhibition of thalamic neurons projecting to the cortex."[98]

1.4 Genetic studies for dyskinesia

While other genes and loci (e.g. DAT, MAO-A/B, COMT) have been implicated, [65,90,95,96,99] the majority of studies have focused on DRD1-3 and BDNF. Many genetic studies of enzyme or

receptor function have concentrated on SNPs only, specifically the rs1800497 in DRD2,[66,69,99–104] the rs6280 in DRD3,[66,102,105]and the rs6265 in BDNF,[100,106,107] and have reported inconsistent results.[66,69,99–106] A relatively small clinic-based study using a Brazilian population implicated the DRD2 haplotypes (based on rs228365, rs1076560, rs6277, rs1800497 and rs2734849) in dyskinesia, but this result needs replication.[103]

1.5 Hallucinations

Hallucinations are a frequent non-motor symptom among PD patients.[108–110] Visual hallucinations are the most common, while auditory and tactile hallucinations are more unusual. Visual hallucinations most often include images of people, animals or inanimate objects. With longer disease duration, the prevalence of hallucinations increase and progress from minor hallucinations to delusions and psychosis. Generally, early in the disease, patients with visual hallucinations continue to have insight that the hallucinations are not real. Hallucinations are more common in dim light and at night and are often persistent. Over time, the visual hallucinations can change into more frightening images. In addition, delusions and loss of insight can occur with longer disease duration. Delusions among PD patients, false and/or irrational beliefs, tend to be of a paranoid nature.[108–110]

Many studies have attempted to establish the prevalence of hallucinations among PD patients. The majority of these studies were cross-sectional and clinic-based studies. There does not appear to be a large difference between clinic-based and population-based studies in the prevalence of hallucinations. Depending on the disease duration, the prevalence of hallucinations is estimated to be between 9 and 60% (cumulative).[111–117]

Non-genetic risk factors for hallucinations are REM sleep behavior disorder, female sex, later age at diagnosis, disease duration, cognitive decline, and motor subtype (postural instability and gait disorder).[118–120] Previously, medications, specifically dopamine agonists, were hypothesized

to cause hallucinations,[121] although recently several longitudinal studies were unable to identify an association.[122,123]

The underlying pathomechanism of hallucinations remains unresolved. One theory, similar to the psychosis theory among schizophrenia patients, is that there is an increased sensitivity in mesolimbic and meso-cortical areas that is causing the hallucinations.[124] A different hypothesis is that the serotonergic and acetylcholinergic systems are causing an imbalance in the serotonin/dopamine equilibrium[125] or a cholinergic deficit.[126] Other causal risk factors for hallucinations that have been proposed are visual-processing abnormalities, medications, sleep-wake cycle dysregulation and cognitive impairment.[111,127–132]

Imaging studies among PD patients indicate that those who report hallucinations have a reduced cortical volume in various areas of the brain.[133–144] Specifically, these studies have indicated that gray matter atrophy was detected in brain areas that are related to visuospatial processing, attention, and memory.[133,134,143,144,135–142] In addition, one study found reduced CSF amyloid $A\beta_{1-42}$ among PD patients with early onset hallucinations (within four years after diagnosis).[122] Post-mortem studies indicate that several neurodegeneration related protein aggregates i.e. Lewy bodies, amyloid and tau pathologies are more prevalent among those with hallucinations.[145]

Hallucinations among PD patients can increase their anxiety, and has been shown to increase the risk for dementia and depression.[114,146–149] In addition, hallucinations are a very strong predictor for placement in nursing homes[54,150,151] and increases a patient's mortality.[152,153] Finally, hallucinations have been shown to increase the caregiver burden substantially.[154]

Treatments for hallucinations among PD patients are limited. The atypical antipsychotic medications currently prescribed to PD patients to treat hallucinations include: pimavanserin,

olanzapine, clozapine, risperidone, and quetiapine. These atypical antipsychotic medications generally block the 5HT2A and DRD2 receptors. Some of these medications have shown to increase mortality and have severe side effects. Another reason why treatment is difficult, is that traditional antipsychotic medications are contraindicated and will worsen the PD symptoms.[155,156]

1.6 Genetic studies for hallucinations

Previous studies that researched the genetic risk factors for hallucinations reported conflicting results.[102,120,157–191] These studies were candidate-gene based and most had a small sample size (less than 250 subjects). The genes involved in these studies are dopamine transporter genes,[102,157,161,184,188] DRD1-4,[102,157,160,161,166,183] serotonin receptor,[158,159] serotonin transporter gene,[159,185] APOE,[160,175–179] COMT[161,177] HOMER1,[162,165] SNCA,[163,175] CCK,[164,182,187,191] CCKAR,[164,187,191] CCKBR,[164,187,191] GBA,[167,168,170–174] LRRK2,[167] PITX3,[169] and ACE.[180 181]

1.7 Dissertation objectives

The aim of this dissertation is to investigate genetic risk factors for dyskinesia and hallucinations, two major comorbidities among PD patients that have a significant influence on a patients' quality of life.

In chapter 2, a candidate-gene approach was used and haplotypes were created to analyze the association between the dopamine receptor 1-3 and BDNF genes and dyskinesia. The goal of this study is to determine the association between these candidate genes, as well to determine if there is a subgroup of patients, genetically, who are at increased risk for developing dyskinesia after treatment with L-Dopa medications.

In chapters 3 and 4, a genome-wide approach was utilized and polygenic risk scores were created to decipher the underlying pathomechanisms for hallucinations. In chapter 3, one PRS was created based on candidate genes in a training-set and validated in a target study, to decipher if previously identified genes are associated with hallucinations. A second PRS was created based on previous GWAS data for PD to decipher if the genetic architecture for PD is associated with hallucinations. The analysis of the PD-PRS would answer whether disease (progression) is affecting the risk for hallucinations. In chapter 4, two PRS were created, one based on Alzheimer's disease (AD) and one on schizophrenia, with the goal to decipher if the genetic architecture of AD or schizophrenia overlaps with the genetic risk factors for hallucinations.

2. STUDY 1: Dopamine receptors and BDNF-haplotypes predict dyskinesia in Parkinson's disease

2.1 Abstract

Objective: Dyskinesia is a known side-effect of the treatment of Parkinson's Disease (PD). We examined the influence of haplotypes in three dopamine receptors (DRD1, DRD2 and DRD3) and the Brain Derived Neurotrophic Factor (BDNF) on dyskinesia.

Methods: Patient data were drawn from a population-based case-control study. We included 418 patients with confirmed diagnoses by movement disorder specialists, using levodopa and a minimum three years disease duration at the time of assessment. Applying Haploview and Phase, we created haploblocks for DRD1-3 and BDNF. Risk scores for DRD2 and DRD3 were generated. We calculated risk ratios using Poisson regression with robust error variance.

Results: There was no difference in dyskinesia prevalence among carriers of various haplotypes in DRD1. However, one haplotype in each DRD2 haploblocks was associated with a 29 to 50% increase in dyskinesia risk. For each unit increase in risk score, we observed a 16% increase in dyskinesia risk for DRD2 (95%CI: 1.05-1.29) and a 17% (95%CI: 0.99-1.40) increase for DRD3. The BDNF haploblock was not associated, but the minor allele of the rs6265 SNP was associated with dyskinesia (adjusted RR 1.31 (95%CI: 1.01-1.70)).

Conclusion: Carriers of DRD2 risk haplotypes and possibly the BDNF variants rs6265 and DRD3 haplotypes, were at increased risk of dyskinesia, suggesting that these genes may be involved in dyskinesia related pathomechanisms. PD patients with these genetic variants might be prime candidates for treatments aiming to prevent or delay the onset of dyskinesia.

2.2 Introduction

Parkinson's disease (PD) leads to significant disability and loss of quality-of-life.[59] One important component to loss of quality-of-life is L-dopa-induced dyskinesia, a common side-effect of treatment. Dyskinesia affects 25% within the first five years [44] increasing up to 80% among patients ten years into disease.[65] Dyskinesia is associated with depression and increases health-related costs.[72] Interestingly, although some patients develop dyskinesia early in their disease, other patients never do.

The two basal ganglia circuitry pathways (D1 and D2) associated with movement control are regulated by dopamine receptors. Common genetic variations in DRD loci are natural candidates for dyskinesia risk.[95,96] In addition, Brain Derived Neurotrophic Factor (BDNF) may affect dyskinesia due to modulation of dopamine receptor expression.[97]

The majority of studies focused on DRD1-3 and BDNF. Many of these have concentrated on SNPs only, specifically the rs1800497 in DRD2,[66,69,99–104,192,193] the rs6280 in DRD3,[66,102,105,193] and the rs6265 in BDNF.[100,106,107] The results from these studies have reported inconsistent results.[66,69,99–107,192,193] One small clinic-based study using a Brazilian population, analyzed DRD2 haplotypes (based on rs228365, rs1076560, rs6277, rs1800497 and rs2734849) and implicated that DRD2 haplotypes are associated with dyskinesia.[103]

Here we are using a targeted approach for the three dopamine receptors (DRD1-3) and BDNF, to estimate risk of dyskinesia based on haplotypes' variants. Haplotypes are combinations of SNPs from a small region of a chromosome that commonly are inherited together. Hence, reviewing haplotypes instead of SNPs limits the number of tests and allows us to examine gene regions. Patients were enrolled in our large population-based study in California (N=747) and dyskinesia was assessed relatively early, on average five years after first diagnosis

based on cardinal motor symptoms. The goal of this study to determine whether a subgroup of patients can be identified genetically who are at increased risk for developing dyskinesia after treatment with L-Dopa medications.

2.3 Methods

2.3.a. Study population

We assembled a cohort from the population-based case control study Parkinson's Environment and Gene (PEG) study, which enrolled and followed patients from three Central California counties (Kern, Fresno, and Tulare) between 2001 and 2015. All patients were seen by movement disorder neurologists (JB, Dr. Bordelon) at least once at baseline, many on multiple follow-up occasions, and were confirmed as having probable idiopathic PD according to published criteria.[3] Recruitment occurred in two phases. Recruitment during the first phase (PEG1) in 2001-2007 has been described before.[194] Among 563 potential patients, 359 incident PD patients within the first three years after diagnosis were identified at baseline. Dyskinesia was assessed during the first follow-up appointment, where we saw 250 PEG1 patients at least once. The second recruitment strategy (PEG2) ensued during 2010-2015 and identified 388 idiopathic PD patients (from 589 potential patients). In PEG2, patients were allowed to have received a PD diagnosis after 2001.

Disease duration was measured from the diagnosis date to the time of the Unified Parkinson's disease Rating Scale (UPDRS) IV assessment. We restricted our population to those taking levodopa (N=459; 193 PEG1 and 266 PEG2 subjects). As dyskinesia is unlikely to develop within the first years after diagnosis, we further restricted to patients seen with a minimum of three calendar years of disease duration. Among patients who were seen too early in

disease (N=101) or were not taking levodopa medication (N=143) at the first assessment, we examined whether they met our criteria during a second assessment (34 PEG1; 26 PEG2). Figure 2.1 provides a flow diagram for the study population of 418 PD patients included in our final analyses.

All study protocols regarding human subjects have been approved by the local Institutional Review Board and all participants gave their written informed consent.

2.3.b. Definition of outcome and other variables

Dyskinesia was assessed during follow-up visits in PEG1, and starting with the first visit in PEG2 (for some patients dyskinesia was assessed shortly after their first contact) using the UPDRS part IV. Dyskinesia presence was noted (yes/no) and the severity of dyskinesia was measured in hours per day (no dyskinesia, ≤25%, 26-50%, 51-75% and >75% of the waking day).

Based on Parkinson's medication at the dyskinesia assessment, we calculated a levodopa-equivalent daily dosage (LEDD) converting all reported anti-Parkinsonian medications into a standardized equivalent dosage according to a previously described algorithm.[195] Race/ethnicity was established via genetic Ancestry Informative Markers if available (72.5% of those included here) or by self-report of ancestry if the Ancestry Informative Markers had not been measured (27.5%). We grouped race/ethnicity as European ancestry, Hispanic and other.

2.3.c. Genetic analysis

DNA was extracted at UCLA from whole blood and genotyping was conducted at University of British Columbia (UBC) for all patients with a confirmed diagnosis of PD. A total of 26 SNP

were selected to be genotyped across the three DRD genes: 5 SNPs for DRD1, 12 SNPS for DRD2, 8 for DRD3 loci (supplemental table 2.1). Markers were mainly selected for haplotype tagging. Genotyping was performed in duplicate at UBC using a Sequenom MassARRAY platform and custom iPLEX assay. Three SNPs in the BDNF locus were previously genotyped at UCLA for a subset of the study sample (N=312 patients with PD). All SNPs had a genotyping call rate >97%, MAF>5% and were in Hardy-Weinberg equilibrium (as defined by a p-value greater than 0.01 in Caucasians) except for rs11030104 (p-value: $1.7*10^{-9}$), which was subsequently excluded.

Using Haploview,[196] we created linkage disequilibrium (LD) plots based on the total study population (figure 2.2). A haploblock, consists of various haplotypes, i.e. SNP patterns. Haploblocks were created based on the confidence intervals.[197] The haplotypes with the highest probabilities for each subject were generated with Phase v2.1.1; rare haplotypes with a frequency of less than 30 patients (=3.7%) were combined into one subgroup.[198] We created one haploblock for the BDNF- and DRD1, and three haploblocks for DRD2 and DRD3 loci (Figure 2.2). The haplotypes in the three haploblocks are correlated. For the DRD2 haploblocks, the D' between the first and second haploblock is 1.0, and between the second and third 0.25. For DRD3, the D' between the first and second haploblock is 0.94 and between the second and third haploblock 0.61. The BDNF haplotype consisted of only two SNPs; the R-squared was 0.19 and D-prime was 1.0.

2.3.d. Statistical methods

All analyses were performed using SAS 9.4 (SAS Institute, Cary NC). We estimated risk ratios (RR) using Poisson regression models with robust error variance and a log link function[199]

to estimate effects for eight chromosomal regions in four genes on dyskinesia occurrence. The analyses were adjusted for gender, PEG recruitment period, disease duration, LEDD, race/ethnicity and age at diagnosis. Throughout, we adjust for LEDD, adjusting for L-DOPA daily dosage instead produced the same results. For analyses of dyskinesia severity, we distinguished three categories [no, mild (<25%) and moderate/severe (>25%)] to ensure that the proportional odds assumption of our model was met. When multiple haploblocks in one gene showed associations, we created a risk score summing across the three risk-haplotypes and estimated the effect according to the number of risk haplotypes a patient carried.

Finally, we assessed whether individual SNPs alone could explain the associations by analyzing each SNP independently. In sensitivity analyses, we examined whether results were consistent across race/ethnicity by restricting participants to their ancestry.

2.4. Results

2.4.a. Patient characteristics

The prevalence of dyskinesia in our patient sample was 24.6%, with the majority experiencing symptoms of dyskinesia less than 25% of their waking hours per day. The average age at diagnosis was 67 years, and on average 5.5 calendar years had passed since first diagnosis when we assessed dyskinesia in our study (see table 2.1). Haplotype frequencies ranged between 4 and 51% in our haploblocks (Figure 2.2). The BDNF haplotype had a minimum frequency of 18.9% and a maximum of 45.4%. Because haploblock 2 of the DRD3 loci consists almost exclusively of the GG haplotype, we lacked statistical power to examine associations for this block. This haploblock was therefore omitted from analyses.

Dyskinesia risk decreased with increasing age of PD diagnosis (with each year increase in age at diagnosis: RR=0.98 (95% CI: 0.96-0.99), increased with LEDD (per 1000 mg: RR=1.77 (95% CI: 1.29-2.44), increased with disease duration (per calendar year: RR=1.14 (95%CI: 1.07-1.21); and increased with UPDRS score (per five units increase: RR=1.07 (95%CI: 1.01-1.14)). Gender, years of schooling, smoking status, family history of PD, and ethnicity were not associated with dyskinesia in our population sample. The effect sizes and confidence intervals were very similar for the association among these variables with 'severity of dyskinesia'.

2.4.b. Association haplotypes with dyskinesia and 'severity of dyskinesia'

We did not find any association between dyskinesia or 'severity of dyskinesia' and haplotypes for DRD1. For DRD2, in both adjusted and unadjusted analyses, we found all three DRD2-haploblocks to be associated with dyskinesia. We adjusted for gender, PEG1 or 2 recruitment, disease duration, LEDD, race/ethnicity and age at diagnosis. In each haploblock there was one haplotype that accounted for about a 29 to 50% increased risk for dyskinesia. For 'severity of dyskinesia', the effect sizes were slightly higher (33-62% increase) and 95% confidence intervals (CIs) were narrower. For DRD3, the TGG-haplotype of the first haploblock and the CC-haplotype of the third haploblock suggested an increase in risk of developing dyskinesia, however the 95% CIs included the null value. Finally, for BDNF, the TA-haplotype was associated with 'severity of dyskinesia' with an increase of 43%, however the effect estimate for presence of dyskinesia did not reach statistical significance in our data (see tables 2.2 and 2.3).

Combining "risk haplotypes" (defined as the haplotypes in each haploblock associated with increased risk for dyskinesia) into a risk-score for the DRD2 (see supplemental table 2.2 for

prevalence of risk-scores), we estimated a 16% risk increase for dyskinesia and 21% risk increase for 'severity of dyskinesia' with each unit increase in risk score. Analyzing the risk score as a categorical variable, there is also a clear increase in odds with an increasing risk score. Compared to no risk haplotypes, having 1 to 3 risk haplotypes increases the risk by about 55%, while having 4 or more haplotypes increases the risk by 153%. The risk score for the DRD3 showed an increase of 17% with each additional "risk haplotype" and a 20% increase for 'severity of dyskinesia', though both confidence intervals included or were very close to the null. However, we observed a dose-response pattern when reviewing the association with the risk score as a categorical variable (see table 2.4).

Restricting the sample to patients of European ancestry did not change the results for dyskinesia nor for 'severity of dyskinesia'. None of the individual SNPs were responsible for the haplotype findings in the DRD loci. However, in BDNF, the rs6265 SNP was associated with dyskinesia (aRR: 1.31 (95%CI: 1.01-1.70; p:0.04)), while the rs11030101 was not (aRR: 0.96 (95%CI: 0.74-1.26); p:0.78). Risk among minor allele carriers of rs6265 was even more pronounced for 'severity of dyskinesia' (aRR: 1.43 (95%CI: 1.10-1.97); p:0.01).

2.5 Discussion

In our large community-based study of PD conducted in central California, we estimated an increased odds for developing dyskinesia after dopaminergic treatment in specific DRD2, DRD3, and BDNF but not DRD1 haplotype carriers. Our study finds that genetic variation in DRD2 influences the prevalence of dyskinesia and its severity in a dose response manner. Some variations in DRD3 and BDNF may further contribute to the risk and "severity of dyskinesia". Our findings proved robust in sensitivity analyses. The three DRD2 "risk haplotypes" are

frequently prevalent in the population (Figure 2.2 and table 2.2). Hence, a large percentage of L-dopa treated patients are at increased risk of developing dyskinesia. Overall, about 30% of subjects have no DRD2 risk haplotype, with about 60% having one to three risk haplotypes; and about 10% carrying four or more risk haplotypes. Especially patients with more than three risk haplotypes are at a much higher risk since the risk of dyskinesia in medicated patients increased by 153%.

Correction for multiple testing is not necessary in this study, because the four genes in this study were chosen based on previous findings and because the haplotypes within each gene are highly correlated and an adjustment for the number of tests would be overly restrictive. However, when using an excessively restrictive Bonferroni correction assuming 10 tests (8 haplotypes and 2 risk scores) leading to an alpha value of 0.005, our findings for the DRD2 risk score would still be considered statistically significant.

Only one previous study investigated an association between haplotypes in DRD2 and dyskinesia. The haplotypes in this relatively small Brazilian clinic-based population (N=199) were generated from different SNPs.[103] Our SNPs were chosen based on tagging characteristics rather than to replicate this previous study. Even though we cannot compare our results directly to these haplotypes, several SNPs are in LD with ours. Specifically, rs6277 and rs2734849 are associated with the rs1089154 and rs1554929 from our first haploblock (R-square is 1.0 for rs6277 and 0.77 for rs2734849) based on HapMap data. Furthermore, rs2283265, rs1076560 and rs1800497 are associated with the rs2471857 and rs2471854 from our second haploblock with R-squares of 0.98, 0.94 and 0.64, respectively. According to functional studies, the recessive allele of rs1800497 is associated with a decrease in DRD2 expression.[200] Hence, our haploblock may be associated with functional changes in DRD2 expression. Lower expression would diminish

the function of the inhibitory pathway and create an imbalance between the excitatory and inhibitory pathways leading to dyskinesia.[87]

Previous studies reviewed the association between DRD3 and dyskinesia, focusing mainly on the rs6280 SNP (Ser9Gly).[66,102,105] The particular Ser9Gly polymorphism in the dopamine 3 receptor has a higher binding affinity to dopamine,[201] and is in high LD with our third haploblock (R-square for rs226082 and rs324026 are 0.84 and 1.00, respectively). Although two studies found an association between the Ser9Gly polymorphism and dyskinesia, [66,193] one of which only found an association with a small subgroup of patients with diphasic dyskinesia,[66] other studies showed no association with dyskinesia.[102,105] One study has suggested that additional information on dyskinesia subtypes might improve the prediction of dyskinesia risk.[66] Unfortunately, we did not gather this information and cannot address this subtypes. The estimated effect size for DRD3 in our study is smaller than for DRD2 limiting our power to identify a statistically significant finding. However, our findings suggest that the DRD3 haplotypes may be relevant for dyskinesia, even though they are not formally statistically significant.

In our study, the rs6265 SNP of BDNF was associated with dyskinesia while haplotype associations were not evident. The minor allele of SNP rs6265 (Val66Met) has been previously associated with time to development of dyskinesia in a study of 315 patients from the Cambridge Centre for Brain repair.[107] Yet another study involving 285 Australian patients with PD did not find any difference in time to onset of dyskinesia.[106] The minor allele of rs6265 is associated with a change of valine to methionine at codon 66 causing decreased protein secretion, potentially changing the BDNF expression of the nigral dopaminergic neurons. This amino acid change has been used as a biologic argument for increased susceptibility for dyskinesia.[96,97]

In genetic studies, an important potential confounder is race/ethnicity. To avoid potential population stratification, we limited our analyses to patients from European ancestry in sensitivity analyses and our findings did not change. We adjusted our analysis for race/ethnicity according to three categories. Primary analyses were not adjusted for population substructure using fractional ancestry because Ancestry Informative Markers were not available on ~25% of the individuals. However, when we limited our analysis to those individuals with Ancestry Informative Markers and adjusted for population structure, we got essentially the same results (results not shown). Race/ethnicity in our study was based mostly on genetic Ancestry Informative Markers, or – if unavailable – on self-report of parental origin. The correlation between these two measures is high (95.4%) in our population. Although there may be some measurement error for race/ethnicity, based on self-report, there is no indication this would unduly influence our findings.

Information on dyskinesia on almost all subjects was available though we did not assess dyskinesia during the baseline examination early in disease for subjects enrolled in PEG1. A third of all patients were lost to follow up between baseline and follow-up exam in PEG1. The main reason for this was mortality and disease severity preventing participation. Thus, if both dyskinesia and genetic variants of interest influence mortality, we may have a survival bias in PEG1 subjects. However, as findings were consistent across both studies (PEG1 and PEG2) and dyskinesia was assessed in PEG2 subjects at baseline such a survival bias appears unlikely.

We only gathered information about dyskinesia prevalence at the time of motor function assessment during physical examinations and we did not record the exact start time of dyskinesia leading to left censoring. This limitation does not allow us to conduct a time-to-event analysis.

However, most of the PEG patients were early in their presentation (on average 5.1 years at first dyskinesia assessment) and were followed up within 2-5 years of the first dyskinesia assessment.

The presence of one of the risk haplotypes, or variant alleles, can have major implications for best practices in treatment of PD. Based on the findings of this study, we can hypothesize that patients who are at higher risk for developing dyskinesia based on their genetic risk profile may benefit from therapy with dopamine agonists along with a more strict L-dopa-sparing strategy. Treatment plans should reflect the difficulty in managing PD, to maximize optimal effects (L-dopa 'on' time) but also mitigate dyskinesia risk. Future individual treatment plans may start to consider a patient's genetic profile.

2.6 Conclusion

Several haplotypes in DRD2, possibly haplotypes in DRD3 and the minor allele of rs6265 in BDNF, increased the risk of dyskinesia in our study. Levodopa induced dyskinesia and PD symptoms must be approached as a tradeoff. Nevertheless, genetic information may help prevent or postpone this debilitating consequence of treatment and may improve patient-centered, personalized therapy. Association studies require confirmation and the health care economy of implementing more personalized treatment should also be quantified before decisions are made. PD patients with these specific risk haplotypes may be prime candidates for testing approaches to prevent or delay the development of dyskinesia.

2.7 Tables and Figures

Table 2.1 Overview of characteristics of the PEG study population

	Dyskinesia		No Dyskinesia		Total sample population	
	Total N	**N=103 %**	**Total N**	**N=315 %**	**Total N**	**N=418 %**
Race						
European ancestry	87	84.5	244	77.5	331	79.2
Hispanic	14	13.6	60	19.0	74	17.7
Other	2	1.9	11	3.5	13	3.1
Smoker						
Non-smoker	58	56.3	171	54.3	229	54.8
Former smoker	43	41.7	138	43.8	181	43.3
Current smoker	2	1.9	6	1.9	8	1.9
Gender						
Male	62	60.2	207	65.7	269	64.4
Female	41	39.8	108	34.3	149	35.6
Family history of PD						
No	87	84.5	259	83.5	346	83.8
Yes	16	15.5	51	16.5	67	16.2
Dyskinesia in hours						
no dyskinesia	0	0.0	315	100.0	315	75.4
< 25% of waking day	70	68.0	0	0.0	70	16.7
>26 of waking day	33	32.0	0	0.0	33	7.9
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Years of education	14.2	4.2	14.0	4.7	14.1	4.5
PD duration at time of measuring dyskinesia	6.2	2.3	5.3	2.1	5.5	2.1
Levodopa equivalent daily dosage	739.3	488.9	601.4	340.4	635.4	386.3
Age of diagnosis PD	64.1	10.9	67.6	10.0	66.7	10.3
UPDRS score	26.7	9.9	24.0	12.0	24.7	11.6

N: number. %: percentage. SD: Standard deviation

Table 2.2 Poisson regression analyses for the association between dyskinesia and the haplotypes for the DRD1, DRD2, DRD3 and the BDNF loci in PD patients (N=418 patients) with a minimum of 3 years PD disease duration.

	N	Crude	Adj		
	Dysk / No Dysk	**RR**	**RR**	**95%CI**	**p-value**
DRD1 *(*rs4867798|rs5326|rs265981)					
TCA	67 / 212	Ref			
TCG	64 / 211	0.97	1.00	0.75 – 1.34	0.98
CCG	34 / 109	1.01	1.06	0.74 – 1.52	0.75
CTG	30 / 87	1.07	1.14	0.78 – 1.66	0.50
Other	1 / 2	NA			
DRD2**, block 1** (rs10891549|rs1554929|rs1124493|rs2242592|rs2245805)					
CTAGG	78 / 293	Ref			
TCGTT	60 / 183	1.17	1.20	0.90 – 1.60	0.22
TCAGG	48 / 103	1.51	1.50	1.09 – 2.07	0.01
TCGTG	5 / 30	0.68	0.63	0.26 – 1.52	0.30
Other	5 / 9	NA			
DRD2**, block 2** (rs6265| rs11030101|rs1076563|rs1116313|rs2471857|rs2471854)					
CGCG	87 / 328	Ref			
AACG	62 / 183	1.21	1.25	0.94 – 1.65	0.12
AATC	47 / 105	1.48	1.50	1.10 – 2.03	0.01
Other	0 / 2	NA			
DRD2**, block 3** (rs4245146|rs7122454|rs11214611)					
TGA	79 / 287	Ref			
CGA	86 / 215	1.33	1.29	1.00 – 1.66	0.05
CCG	31 / 113	1.00	1.01	0.69 – 1.48	0.95
Other	0 / 3	NA			
DRD3**, block 1** (rs3732790|rs963468|rs324036)					
TGC	91 / 311	Ref			
AAC	73 / 219	1.10	1.08	0.85 – 1.38	0.52
TGG	30 / 75	1.26	1.26	0.90 – 1.74	0.17
Other	2 / 13	NA			
DRD3**, block 3** (rs226082|rs324026)					
TT	116 / 398	Ref			
CC	72 / 194	1.20	1.23	0.97 – 1.56	0.09
TC	8 / 26	1.04	1.08	0.61 – 1.91	0.80
BDNF (rs6265|rs11030101)					
CT	77 / 202	Ref			
CA	55 / 164	0.91	0.94	0.70 – 1.26	0.67
TA	40 / 76	1.25	1.27	0.93 – 1.72	0.13

Dysk / No Dysk: number of Parkinson's patients with dyskinesia / without dyskinesia. Ref: Reference. 95%CI: 95% confidence interval. NA: not available due to small number of subjects in cells. Adj. = adjusted analysis. Adjusted analysis was adjusted for PEG study, gender, disease duration, levodopa equivalent daily dosage, ethnicity, and age at diagnosis.

Table 2.3 Poisson regression analyses for the association between 'severity of dyskinesia' and the haplotypes for the DRD1, DRD2, DRD3 and the BDNF loci in PD patients (N=418 patients) with a minimum of 3 years PD disease duration.

	N	**Crude**	**Adj**		
	No / Mild / Severe	**RR**	**RR**	**95%CI**	**p-value**
DRD1 *(*rs4867798|rs5326|rs265981)					
TCA	212 / 46 / 21	Ref			
TCG	211 / 46 / 18	0.95	0.97	0.71 – 1.33	0.85
CCG	106 / 19 / 15	1.11	1.15	0.78 – 1.70	0.48
CTG	87 / 21 / 9	1.06	1.14	0.77 – 1.69	0.52
Other	2 / 0 / 1	NA	NA		
DRD2**, block 1** (rs10891549|rs1554929|rs1124493|rs2242592|rs2245805)					
CTAGG	293 / 55 / 23	Ref			
TCGTT	183 / 44 / 16	1.15	1.17	0.87 – 1.57	0.29
TCAGG	103 / 27 / 21	1.68	1.62	1.14 – 2.30	0.01
TCGTG	30 / 2 / 3	0.84	0.74	0.28 – 2.01	0.56
Other	9 / 4 / 1	NA	NA		
DRD2**, block 2** (rs6265| rs11030101rs1076563|rs1116313|rs2471857|rs2471854)					
CGCG	328 / 60 / 27	Ref			
AACG	183 / 46 / 16	1.16	1.20	0.90 – 1.60	0.22
AATC	105 / 26 / 21	1.63	1.62	1.15 – 2.28	0.01
Other	2 / 0 / 0	NA	NA		
DRD2**, block 3** (rs4245146|rs7122454|rs11214611)					
TGA	287 / 56 / 23	Ref			
CGA	215 / 57 / 29	1.37	1.33	1.01 – 1.74	0.04
CCG	113 / 19 / 12	1.07	1.08	0.72 – 1.63	0.71
Other	3 / 0 / 0	NA	NA		
DRD3**, block 1** (rs3732790|rs963468|rs324036)					
TGC	311 / 64 / 27	Ref			
AAC	219 / 49 / 24	1.13	1.11	0.86 – 1.43	0.42
TGG	75 / 18 / 12	1.36	1.35	0.95 – 1.92	0.10
Other	13 / 1 / 1	NA	NA		
DRD3**, block 3** (rs226082|rs324026)					
TT	398 / 82 / 34	Ref			
CC	194 / 46 / 26	1.26	1.29	1.00 – 1.65	0.05
TC	26 / 4 / 4	1.21	1.26	0.71 – 2.25	0.43
BDNF (rs6265|rs11030101)					
CT	202 / 54 / 23	Ref			
CA	164 / 40 / 15	0.89	0.92	0.68 – 1.25	0.60

TA	76 / 22 / 18	1.40	1.43	1.01 – 2.03	0.04

No/Mild/Severe: Number of Patients with respectively no, mild or severe dyskinesia. N: Number. Ref: Reference. 95%CI: 95% confidence interval. NA: not available due to small number of subjects in cells. Adj. = adjusted analysis. Adjusted analysis was adjusted for PEG study, gender, disease duration, levodopa equivalent daily dosage, ethnicity, and age at diagnosis.

Table 2.4 Poisson regression analyses with robust error variance and a log link function for the association between the DRD2 and DRD3 risk scores and dyskinesia.

Number of Risk Haplotypes	Dyskinesia					Severity of dyskinesia				
		Crude	Adj.				Crude	Adj.		
		RR	RR	95%CI	P		RR	RR	95%CI	P
***DRD2*, risk score**										
Continuous		1.17	1.16	1.05 – 1.29	0.004		1.20	1.21	1.08 – 1.35	0.001
	N Dysk / No Dysk					N No / Mild / Severe				
0	20 / 103	Ref				103 / 15 / 5	Ref			
1 to 3	61 / 181	1.55	1.54	0.98 – 2.42	0.06	181 / 41 / 20	1.76	1.61	1.00 – 2.60	0.05
4 to 6	17 / 25	2.49	2.53	1.48 – 4.34	0.001	25 / 10 / 7	3.62	2.80	1.58 – 4.93	<0.001
***DRD3*, risk score**										
Continuous		1.16	1.17	0.99 – 1.40	0.07		1.22	1.20	1.00 – 1.45	0.04
	N Dysk / No Dysk					N No / Mild / Severe				
0	30 / 130	Ref				130 / 22 / 8	Ref			
1	62 / 164	1.46	1.45	0.99 – 2.12	0.06	164 / 40 / 22	1.57	1.53	1.02 – 2.30	0.04
2 to 4	6 / 15	1.52	1.72	0.78 – 3.81	0.18	15 / 4 / 2	1.60	1.84	0.79 – 4.31	0.16

N Dysk / No Dysk: number of Parkinson's patients with dyskinesia / without dyskinesia. No/Mild/Severe: Number of Patients with respectively no, mild or severe dyskinesia. Ref: Reference category. 95%CI: 95% confidence interval. Adj.: adjusted analysis. P: P-value Adjusted analysis was adjusted for PEG study, gender, disease duration, levodopa equivalent daily dosage, ethnicity, and age at diagnosis.

Figure 2.1 Patient recruitment, in- and exclusion-criteria, flow diagram of the two PEG studies (N=418)

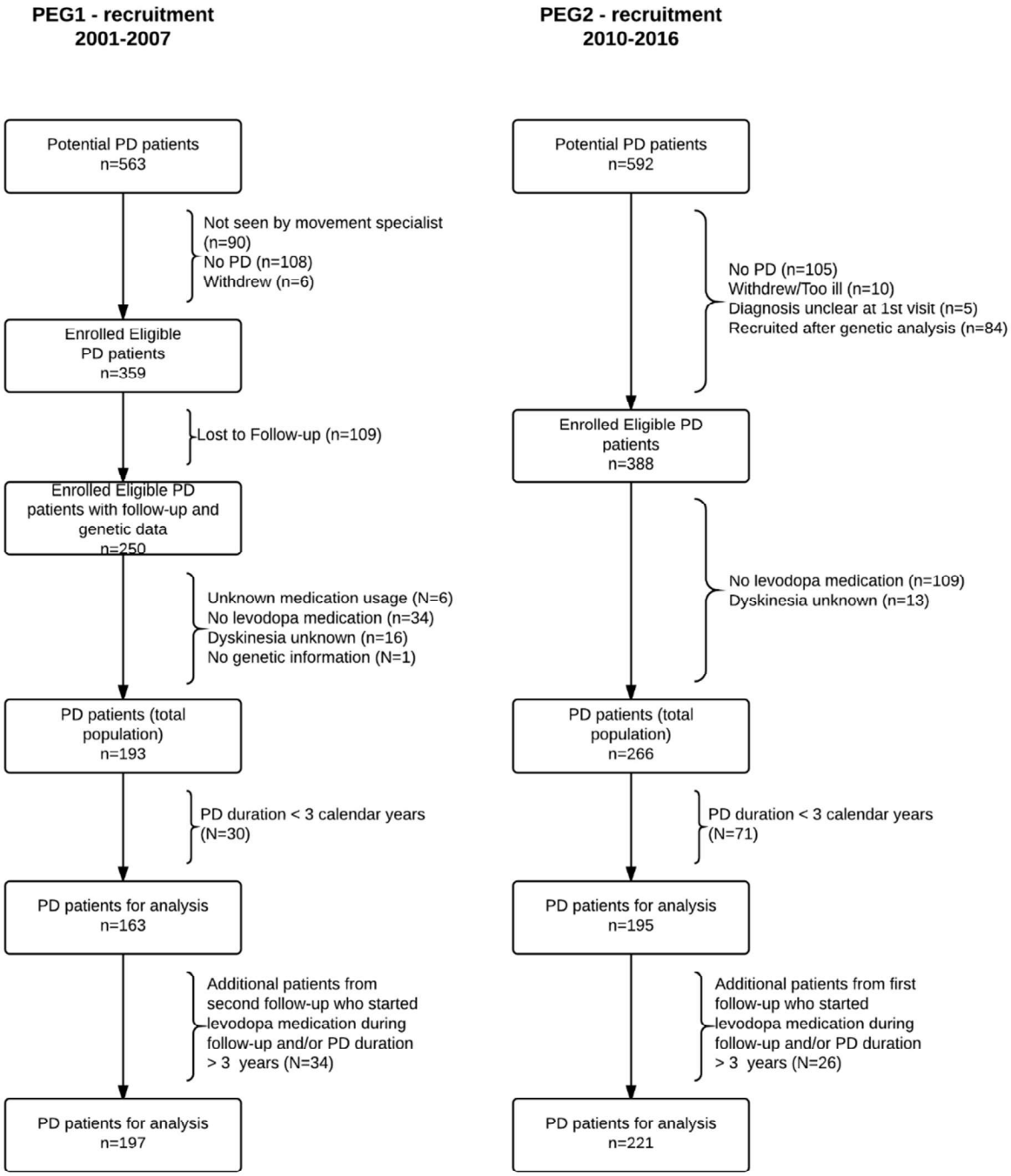
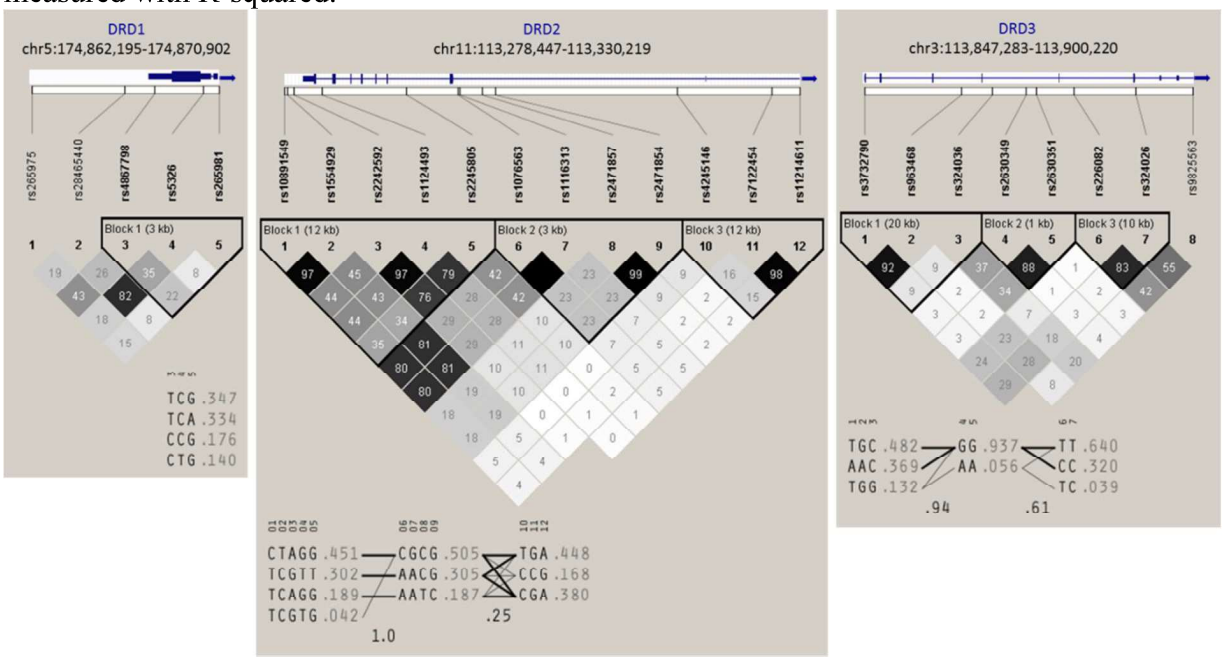


Figure 2.2 Linkage Disequilibrium (LD) plots for the *DRD*-loci in our study population, LD measured with R-squared.



Each LD plot shows the various haplotypes in a haploblock with its population frequency and connections. In the crossing areas, the value of multiallelic D' is shown. This represents the level of recombination between the two blocks.[196]

Supplemental table 2.1. Overview of SNPs genotyped for this study.

Gene	SNP	Position	Location	A1	A2	MAF	Function
DRD1	rs265975	chr5:174862195	5.5 kb downstream	T	C	0.350	Tagging SNPs
DRD1	rs28465440	chr5:174866541		A	G	0.142	Tagging SNPs
DRD1	rs4867798	chr5:174867899	utr variant 3 prime	C	T	0.320	Tagging SNPs
DRD1	rs5326	chr5:174870196	utr variant 5 prime	T	C	0.140	Tagging SNPs
DRD1	rs265981	chr5:174870902	UTR-variant-5prime	A	G	0.345	Tagging SNPs
DRD2	rs10891549	chr11:113278447		C	T	0.465	Tagging SNPs
DRD2	rs1554929	chr11:113278764		T	C	0.461	Tagging SNPs
DRD2	rs2242592	chr11:113283477		G	A	0.340	Tagging SNPs
DRD2	rs1124493	chr11:113282295	intron variant	T	G	0.340	Tagging SNPs
DRD2	rs2245805	chr11:113290699	intron variant	T	G	0.300	Tagging SNPs
DRD2	rs1076563	chr11:113295909	intron variant	A	C	0.489	Tagging SNPs
DRD2	rs1116313	chr11:113296107	intron variant	A	G	0.489	Tagging SNPs
DRD2	rs2471857	chr11:113298339	intron variant	T	C	0.198	Tagging SNPs
DRD2	rs2471854	chr11:113299639	intron variant	C	G	0.197	Tagging SNPs
DRD2	rs4245146	chr11:113317973	intron variant	T	C	0.450	Tagging SNPs
DRD2	rs7122454	chr11:113327468	intron variant	C	G	0.181	Tagging SNPs
DRD2	rs11214611	chr11:113330219	intron variant	G	A	0.180	Tagging SNPs
DRD3	rs3732790	chr3:113847283	downstream variant 500B,utr variant 3 prime	A	T	0.368	Tagging SNPs
DRD3	rs963468	chr3:113862887	intron 4	A	G	0.377	Literature
DRD3	rs324036	chr3:113870078	intron variant	G	C	0.138	Tagging SNPs
DRD3	rs2630349	chr3:113873372	intron variant	A	G	0.064	Tagging SNPs
DRD3	rs2630351	chr3:113875059	intron variant	A	G	0.057	Tagging SNPs
DRD3	rs226082	chr3:113890815	intron 2	C	T	0.316	Tagging SNPs

DRD3	rs324026	chr3:113891042	intron variant	C	T	0.351	Tagging SNPs
DRD3	rs9825563	chr3:113900220	5'-flanking / intron variant	G	A	0.312	Tagging SNPs
BDNF	rs6265	chr11:27658369	Val66Met	G	A	0.229	Functional variant
BDNF	rs11030101	chr11:27659197	intron variant,nc transcript variant,utr variant 5 prime	A	T	0.274	Tagging SNPs

A1 = major allele; A2 = minor allele; MAF = minor allele frequency, measured in the entire genotyped population

Supplemental table 2.2 Prevalence of the DRD2 and DRD3 risk scores, based on summation of risk haplotypes of the DRD2 and DRD3, respectively.

Number of Risk Haplotypes	**Dyskinesia**		**No dyskinesia**		**Total sample population**	
	Total	**N=98**	**Total**	**N=309**	**Total**	**N=407**
***DRD2*, risk score**						
	N	**%**	**N**	**%**	**N**	**%**
0	20	20.4	103	33.3	123	30.2
1	35	35.7	93	30.1	128	31.5
2	11	11.2	46	14.9	57	14.0
3	15	15.3	42	13.6	57	14.0
4	10	10.2	15	4.9	25	6.1
5	3	3.1	8	2.6	11	2.7
6	4	4.1	2	0.7	6	1.5
***DRD3*, risk score**						
	N	**%**	**N**	**%**	**N**	**%**
0	30	30.6	130	42.1	160	39.3
1	41	41.8	107	34.6	148	36.4
2	21	21.4	57	18.5	78	19.2
3	5	5.1	12	3.9	17	4.2
4	1	1.0	3	1.0	4	1.0

N= Number of PD patients with the number “risk haplotypes”. %=percentage of the risk score prevalence

3. STUDY 2: Polygenic risk scores in Parkinson's disease: opportunities and challenges in predicting hallucinations

3.1 Abstract

Creating polygenic risk scores (PRS) is an increasingly popular approach to uncover underlying disease pathomechanisms, to predict disease risk prior to onset, and to establish causality through a Mendelian Randomization approach. PRS can be a powerful tool and has demonstrated success in many diseases, including in Parkinson's disease (PD). To derive a valid and accurate PRS requires understanding its strengths and limitations. In this study, we attempted to generate and validate a PRS for hallucinations in PD based on three longitudinal patient cohorts. An approach to identify potential genes/pathways involved in hallucinations using a candidate-gene based PRS failed, most likely due to the restricted sample size.

We also developed a PRS for Parkinson's disease status based on previous GWAS results, and demonstrated a strong association with disease status in our independent dataset. This PD-status PRS is possibly weakly associated with hallucinations, although our analyses did not reach formal statistical significance. Based on the findings from the pooled analysis, the generated hallucinations PRS, and the PD-status PRS, the following genes of interest to explain an increased risk of developing hallucinations in PD are LRRK2, APOE, SLC6A4, BDNF and MAPT.

3.2 Introduction

The diagnosis of Parkinson's Disease (PD) is based on motor symptoms.[202] The heritability of PD is estimated to be 40%,[203] and large GWAS studies have been performed in the past. [204,205] In recent years, more attention has been given to common and debilitating non-motor symptoms of PD.[118,119] One important symptom influencing patients' quality of life and of vital concern for caregivers is hallucinations. Often PD patients retain insight during visual hallucinations. However, loss of insight and paranoid psychosis, although rare, may occur during disease progression. The prevalence of hallucinations increases with disease duration and risk factors, including female sex, older age at diagnosis, and REM sleep behavior disorder (RBD). Hallucinations often coincide with cognitive decline, [118–120] and it is well established that hallucinations are associated with an increase in mortality.[153]

A number of previous studies attempted to identify potential associations between hallucinations and genetic variants in various candidate genes with conflicting findings.[120,157–169,171–191,206,207] Twenty-one genes have been implicated or hypothesized as being associated with the development of hallucinations among PD patients. The majority of these studies focused on a few single-nucleotide polymorphisms (SNPs) within each gene to identify an association with hallucinations or psychosis. Generally, these studies used small sample sizes (less than 250 subjects). Associations found in one study were often not replicated in others or were not even targeted for replication.

In this study, we combined data from three longitudinal Parkinson's disease studies to attempt to decipher the role of these genes. We used the new promising tool, PRS, to explain or predict the onset of hallucinations in PD. Previously, PRS created for PD were associated with age at onset, cognitive decline and motor progression among PD patients.[208–210] We first created

a PRS based on genes previously implicated in the risk of developing hallucinations. This PRS improved upon single gene based studies in, at least, two ways. Whereas the single-gene studies were each restricted by statistical power to a few SNPs, the PRS increased the statistical power to analyze many SNPs. Further, PRS assessed the combined influence of a large number of biologically plausible candidate genes, an analysis that was unfeasible in the single-gene studies. In order to maximize the power of our study, we combined three studies and restricted to the number of candidate genes to develop the PRS. In addition, we created and validated a PRS associated with PD status based on previous GWAS studies. This allowed us to assess whether the underlying pathomechanisms of PD are also associated with the development of hallucinations, and to assess whether the progression of PD pathology is causing hallucinations among PD patients.

3.2.A Polygenic risk scores (PRS)

Complex diseases are thought to be caused by a combination of environmental exposures, lifestyle choices and genetics variations. Large genome-wide studies have been performed to identify potential genetic risk factors for these complex diseases. Unfortunately, genetic research has only been able to identify a fraction of the heritability among complex diseases. This phenomenon has been called the missing heritability.[211] One potential explanation for this missing heritability is the polygenic theory that states that a plethora of common genetic variants together influence the risk of disease, each contributing only a small effect.

Single SNPs may not act alone but in concert with many other genetic loci.[212] A PRS represents an aggregated score that sums up the effects of a multitude of protective and risk alleles (SNPs). A PRS commonly uses genome-wide association analysis, however, it does not

require individual SNPs to reach statistical significance and takes synergism between SNPs into account. Most commonly, the weights assigned to the risk alleles are derived from effect estimates established in genome wide approaches. Synonyms for PRS are genetic score, genome-wide score, multi-genic score or a multi-polygenic score. With the expansion of and increasing affordability of genomic arrays, PRS are gaining interest as a tool to identify pathomechanisms beyond simply predicting outcomes.

This multi-locus theory was first invoked as an explanation for the unsuccessful search for genetic contributors to schizophrenia.[213] Accordingly, the first polygenic score was developed for schizophrenia and has been very successful in predicting disease risk and deciphering underlying pathomechanisms.[214,215] After the development of the PRS for schizophrenia, other successful polygenic scores have been developed for a wide range of disorders and traits, including Alzheimer's disease, obesity, cardiovascular diseases, breast cancer, educational attainment, and height.[216–224]

By using an aggregated score, a PRS has more power to identify an association with the phenotype than identifying an association for each individual SNP. Many consortia that originally came together to conduct GWAS analyses in studies of tens of thousands of individuals can now be employed to generate PRS that include many SNPs with small effect estimates, even for rare diseases and sub-phenotypes of common diseases.

Below, we used two approaches to investigate the uses of a PRS for hallucinations in PD and to explain the creation and validation process as well as the strengths and limitations of the PRS approach: i.e. a candidate-gene based RPS and a PRS generated from previous GWAS analyses for PD status. We used this PD-status score to assess any overlap between genetic risk factors for PD and hallucinations among PD patients.

3.2.b Predicting risk of complex diseases.

One possible use of a PRS is to identify people who are at higher risk of developing a clinical outcome, and to provide a simple screening tool that can potentially enable earlier treatment. In order to fully measure the genetic risk associated with a disease, this score generally requires a very large number of risk alleles, in the order of tens-of-thousands or hundreds-of-thousands of risk alleles.[225] Although many PRS are associated with a disease, selecting SNPs to include in the PRS can be difficult when a large GWA study is not available and the disease has a moderate or low genetic heritability.[225,226] An example of a successful, predictive PRS is a score created for coronary artery disease (CAD).[227] This score uses 6.6 million genetic variants and has an Area Under the Curve (AUC) of 0.81. The success of this PRS is likely due to the large number of genetic variants, the secondary optimization in a large second dataset before validation, and the high heritability of CAD (50-60%).[228,229]

3.2.c. Identifying the underlying genetic architecture of a disease or phenotype.

Instead of using a large scale GWAS approach it is possible to create a PRS using pathway-related genes or candidate-genes. If an association is identified, this can facilitate the understanding of underlying mechanisms or biological functions.[230,231] An example of a pathway-specific approach is the Alzheimer's disease (AD) study that developed and tested PRS for amyloid-β clearance, cholesterol metabolism and immune response. This study attempted to identify an association between the PRS and cognitive functions, as well as an association between the PRS and biomarkers measured in cerebrospinal fluid (CSF) and PET scan. The study generated the PRS using pathway-specific genes selected for their biological function.

Interestingly, there was no indication that any of the genetic risk scores generated were associated with cognitive function, except for the immune-response PRS and the measured concentration of T-Tau in CSF.[230]

Another method to identify the underlying genetic architecture is to create a PRS based on a genome-wide array and then determine the biologic function and/or pathways the genetic variants or genes included in the PRS are representing.[212,232] A study that used this approach recently provided support for immune-related genetic susceptibility in PD. [232]

3.2.d. Determining causality by developing a PRS as an instrumental variable (Mendelian randomization approach)

Causality can be assessed when a PRS is developed to serve as a genetic predictor of a (modifiable) risk factor for a disease. This method can decrease the need to address confounding risk factors.[233] There are however important aspects to consider when using a PRS as an instrument (proxy), namely that the genetic variants do not have a direct effect on the outcome. The variants should only affect the outcome through the risk factor of interest for which the PRS is a proxy measure. This assumption can become problematic when the PRS contains thousands of genetic variants.[234,235] Also, instrumental variable analysis generally requires large sample sizes, especially when the PRS is weakly associated with the risk factor of interest. In this study, since our goal was to identify genetic risk factors of hallucinations, a Mendelian randomization approach was not appropriate.

3.2.e. Identifying potential overlap of genetic risk factors between various disorders.

Underlying pathomechanisms can be identified when a well-established PRS for one disease is utilized to explain another disorder. The goal of this approach is to analyze the overlapping pathobiology. Examples include studies identifying associations between a PRS developed based on SNPs associated with Schizophrenia that is also associated with a diagnosis of bipolar disorders, anxiety disorders, and AD-related psychosis.[214,236,237] In this study, we did not use this approach as it is beyond the scope of this manuscript.

3.2.f. Creation of a polygenic risk score

To create a PRS requires, choosing the sample size for the training data, and choosing whether to use a whole genome or candidate gene approach; i.e. an untargeted (or 'unbiased') approach that requires a very large dataset, or a targeted approach that can be performed with a smaller sample size. The latter is based on prior knowledge and is useful to determine whether a specific (targeted) pathway is related to the outcome.

It is important to develop the PRS on a training set and validate it on an independent dataset. Although it is technically possible to validate a PRS on the same dataset by using cross-validation, where a percentage of the study subjects are used for the training and the remaining subjects for validation, this cross-validation approach tends to overestimate the association between the PRS and the phenotype. This is mainly due to the homogeneity of the subjects within one study. The lack of generalizability to a different, separate study population tends to reduce the effect estimates and predictive qualities of the association between the PRS and the phenotype.[238]

Several methods have been proposed to create a PRS. The first method, the simplest and most commonly used, is to select SNPs based on a p-value threshold for statistical significance and to use the effect estimates as weights. This can be done using summary statistics from large GWA studies.[214] Another option is to use a second independent dataset to optimize the SNP selection threshold.[227] Newer techniques involve joint analysis of all SNPs using mixed-effect modeling, LD-based prediction and Bayesian techniques. These techniques have been shown to increase some of the PRS' predictive performance.[239–241] Another method that has been proposed is to test only functional SNPs; or to divide all genome-wide derived SNPs into categories based on their known or suggested function, and then calculating a p-value threshold dependent on the category.[242]

When using the weighted sum method (the first method described above), two criteria are important for the selection of the SNPs in the PRS: linkage disequilibrium (LD) clumping and p-value thresholds. As SNPs that are in close proximity are associated with each other (in LD), it is important to remove SNPs that are highly correlated to avoid double-counting the genetic location for the PRS. LD clumping is the process by which the SNP with the strongest association, based on the p-value, is selected first and the SNPs in close proximity, with a LD above a certain threshold, are removed. A high threshold runs the risk of including SNPs that are not independent, while a low threshold could be too restrictive. There is no gold standard for LD thresholds, however thresholds between 0.1 and 0.5 R-square are commonly used. [243]

For the creation of the PRS, either a p-value threshold for the SNP selection needs to be preset, or one can choose the number of SNPs based on a variance explained (R-square) increase. There are no preset guidelines to selecting the optimal threshold. The optimal threshold depends on the validity of the analysis and the polygenic nature of the phenotype. If the training set only

includes a small sample size, as is the case in this study's PRS for hallucination, it is prudent to restrict the number of false-positive effect estimates by using a low p-value threshold. However, in situations where the results of the training dataset were very precise or when the PRS is based on large GWAS results, and the phenotype is thought to be very polygenic (infinitesimal model; a very large number of SNPs with a very small effect size) than choosing a large p-value threshold could generate a better predictor of the outcome.

3.3 Methods

3.3.a. Study population

Three separate research populations were used for this study. See an overview in figure 3.1.

The Parkinson's Environment and Gene study (PEG) is a population-based case control study among residents of the Central Valley of California. The specifics of this study have been described previously. [194] In this paper, we used information from the PD patients' cohort that was followed between 2001 and 2019. These PD patients were classified by movement disorder neurologists as probable idiopathic PD according to published criteria. [3] If a PD diagnosis was deemed only 'possible' at baseline, a diagnostic re-assessment was scheduled. In addition, if necessary, disease classifications were changed after further follow-up visits. Recruitment occurred in two phases; during 2001-2007 (PEG1) and 2010-2018 (PEG2). At PEG1 baseline, 359 incident PD patients were seen within the first three years after their diagnosis, but the presence of hallucinations was first assessed during a follow-up appointment (N=251) on average 6 years into their disease. The second recruitment strategy (PEG2) allowed patients to have received a PD diagnosis after 2001 but they were first seen by UCLA specialists after 2010. PEG2 identified 388 idiopathic PD patients and at the time of this study, 183 had completed a

follow-up visit. Thus, the PEG study has 434 PD patients with information about their hallucinations. In addition, population-based controls were included during the PEG1 phase and between 2007 and 2010. A total of 741 patients and 840 controls were included in the validation for the PRS for PD status.

The ParkWest study (PW) is a longitudinal, population-based cohort of Parkinson's disease patients in South- and West-Norway.[244] All patients with incident Parkinson's disease between November 2004 and August 2006 were approached for this study. After informed consent, subjects were classified by certainty of the disease diagnosis, and if deemed necessary re-assessments were conducted before the participant was included in the study. Final disease classification occurred at a 28-months follow-up meeting according to acknowledged criteria.[202] The ParkWest study collected data for 191 patients at baseline, and 159 patients at a 5-year follow-up visit.

The Parkinson's Progression Markers Initiative study (PPMI) is a longitudinal study following incident Parkinson's disease patients and healthy controls. This study is an international multicenter study (US, Europe, Israel and Australia) and unlike the previous studies, it is clinic-based. Newly diagnosed, untreated PD patients were enrolled if they had a dopamine transporter deficit on 123-I Ioflupane dopamine transporter (DatScan) imaging, no dementia at baseline-visit, and consented to the extraction of biological samples (including serum, plasma, urine and cerebrospinal fluid), genetic analysis, extensive imaging (MRI and DatScan) as well as neuropsychiatric/cognitive testing. PPMI gathered information from 317 PD patients during follow-up.

For the development of the PRS for hallucinations in PD, we used two population-based studies (PEG and PW) and attempted to validate the score among a third independent study

(PPMI). We used PD cases and controls from the PEG study for the validation of the PRS associated with PD-status, and used PD patients from the three studies (PEG, PW, and PMI) when analyzing this validated PD-status PRS to determine the association with hallucinations.

3.3.b. Definition of outcome and other variables

The presence of hallucinations was assessed using the Movement Disorder Society - Unified Parkinson Disease Rating Scale (MDS-UPDRS) in PEG and PPMI, while PW used the UPDRS.[245,246] The specific sub-item is answered positively when a patient or caregiver indicates that there were illusions, hallucinations and/or psychosis. In this study, for simplicity, we use the term 'hallucinations' to include all sub-items. For the PEG study, we used the information provided at the first follow-up visit, about 5 years after disease diagnosis. To increase comparability between studies, we chose to use the information about the presence of hallucinations at a five-year disease duration in the ParkWest and the 48-months follow-up visit after baseline in the PPMI study (~5 years' disease duration).

3.3.c. Genotyping data

All three studies performed a genome-wide array. The PEG study used the Global Screening Array (GSA) chip, the PPMI study used the NeuroX chip and the ParkWest study the Illumina Infinium OmniExpress. To homogenize the strands and impute the whole genome and to increase coverage of the 21 genes, we imputed data by means of the Michigan Imputation Server.[247] The Michigan Imputation Server uses Minimac3 for imputation. Phasing was performed using ShapeIT v2.r790,[248] the reference panel was HRC.r1.2016. [249]

Standard data quality control was performed. Subjects with less than 95% of the SNPs genotyped were removed. SNPs with a minor allele frequency (MAF) less than 0.02 or with a Hardy Weinberg Equilibrium p-value of less than $1*10^{-7}$ were also excluded. To avoid any bias due to family structures, genetic relatedness between subjects was estimated by calculating the Identical by Descent (IBD) percentage. Among pairs or families with a genetic overlap of 20% or more, one subject was randomly selected to remain in the study.

For the hallucination PRS, as the sample sizes are relatively small for a genomics based approach, we limited our PRS creation to the 21 genes previously identified or hypothesized as being associated with hallucinations in PD. [120,157–169,171–191,206,207] This restricted selection of genes aims to reduce inclusion of false-positive findings in the PRS. SNPs within a 20Kb border of each gene were selected. We, furthermore selected only those SNPs that were available in the genomics data from all three studies (N=3,994 SNPs). The top genetic variants (N=9,830) of the combined GWAS of PDGene and PDBWS were used for the PRS creation to estimate PD association.[205] Of these top variants, there were 8,233 variants available in our PEG study.

After quality control, 841 PD patients provided genotyped data for the analysis (390 in PEG, 134 in PW, and 317 in PPMI) of hallucinations. For PD status, 741 patients and 840 controls from the PEG study had whole genome data. All study protocols regarding human subjects were approved by their local Institutional Review Board and written consent was given by all participants.

3.3.d. Ethnicity

In the three studies, ethnicity was estimated slightly differently. In the PW and the PEG study, principal component analysis was performed and compared to the 1000 Genomes Super-

populations. All subjects in PW were considered to be of European ancestry. In the PEG study, the subjects were divided into the 5 main Super-populations (African, American, East-Asian, European or South-Asian). In the PPMI study, ethnicity was based on questionnaire data and later confirmed with a hidden Markov model using Structure 2.3.4.[250]

For subjects in all three studies, fractional ancestry was estimated using hidden Markov modeling and clustering (Structure 2.3.4).[250] The ideal number of clusters was assessed using the Evanno method with Structure Harvester package in R.[251] The final model used 4 clusters, with 10,000 burnin periods and 10,000 Markov chain Monte Carlo repetition.

As population stratification and genetic ancestry are two main confounders in genetic studies, we restricted our analysis to subjects classified as representing the European Super-population. In analyses, we also adjusted for fractional ancestry identified by Structure.[250] In a sensitivity analysis, we also ran a PRS for the total population. However, as the majority of our populations are Caucasian, results are not necessarily generalizable to other ethnic groups.

3.3.e. Confounders

All analyses were adjusted for sex, fractional ancestry, age at diagnosis, disease duration at time of assessment for hallucinations and study as necessary. In order to increase the number of degrees of freedom we adjusted the logistic regression analyses using propensity scores derived from these variables.

3.3.f. Polygenic risk score generation

We chose to use a targeted approach to build a hallucination PRS from our own data, while the PRS for PD status was based on findings from a previous meta-GWAS.[205] First, the primary

creation of the hallucination-PRS was based on the PW and PEG study, since they represent population-based samples, while the replication was attempted with an independent and clinic-based study (the PPMI). Second, we showed the influence of alternating between studies in the creation and validation of the PRS. Third, we demonstrated the results of using cross-validation as a method to create and validate a PRS based on all three studies (using 70% of the subject for the training, and 30% for validation).

The weights for the hallucinations-PRS we created were based on the effects estimate from a logistic regression adjusted for the confounders mentioned above using propensity scores. The weights for the PD status-PRS were based on the effect estimates from the previously performed GWAS analysis. For both PRS we used a threshold r-square LD of 0.1 in order to reduce potential false-positives as much as possible.

In this study, the program 'PRSice' was used to create the PRS and in the replication study to calculate a weighted sum PRS.[252] The final score is the sum of the number of risk alleles per individual weighted by each β coefficient. As it is desirable to limit the number of false-positive alleles in a PRS, especially given a relatively small dataset, we limited the number of alleles using a p-value threshold of 0.05. In the sensitivity analysis, we used various p-value thresholds ($P<0.01$, $P<0.10$) as well as the "optimal" number of SNPs based on the explained variance (R-square). The latter tends to include almost all SNPs used for the creation of the score (in our case, it included all SNPs after clumping). While this method may be an effective approach with a large sample size, given our smaller sample size, selecting the PRS based on the R-square is likely to be ineffective as we will show below.

In summary, the PRS for hallucinations was calculated based on the effect estimates from the training dataset (PW and PEG) by summing the number of risk alleles times the weights for each subject. A p-value threshold of 0.05 and a LD-threshold of 0.1 were used for clumping. Using the weights and the SNPs chosen based on this protocol, the score was calculated in our target dataset (PPMI), and subsequently a logistic regression was performed to analyze the association between the PRS and hallucinations adjusted for the confounders using a propensity score.

For the PRS associated with PD status, we used the effect estimates from the previous GWA study [205] as the weights and calculated the PRS scores using various p-value thresholds among all subjects. After calculating the PRS, we performed logistic regression analysis to assess the association with PD status among PEG patients and controls, and with the presence of hallucinations among PD patients from PEG, PW and PPMI combined.

3.3.g. Software

Analyses were performed using a combination of Plink 1.9, RStudio 1.1.453, SAS 9.4 (SAS Institute, Cary NC), Structure 2.3.4, and PRSice 2.1.8.[250,252–256]

3.3.h. Alternative analysis

Instead of a PRS analysis, a gene set-based analysis could be performed to identify whether specific genes are associated with hallucinations among PD patients. A gene set-based analysis (GSA) analyzes the number of independent (clumped) SNPs within a gene that reaches a p-value threshold (often 0.05) and analyzes the average test-statistic of these SNPs. In this study, we show the results of the GSA for hallucinations based on the 21 candidate genes previously implicated, using a p-value threshold of 0.05, LD threshold of 0.5 and 25,000 permutations.

Finally, we also provided a pooled analysis of the association between the SNPs in the PRS and hallucinations.

Although the analyses above can be performed in lieu of PRS analyses, they generally require a larger sample size to identify statistically significant findings. The pooled analysis requires the largest sample size to identify single SNPs that are statistically significantly associated with the outcome. We provided these alternative analyses for comparison purposes and as they can contain additional insights.

3.4 Results

The average age of subjects with PD in this study was 64 years old, with a range of 23 to 87 years. The controls in this study were on average 66 years during the interview (range: 35 to 99). The main difference between subjects in the PPMI and the PEG/PW study was age at diagnosis. Subjects in the PPMI study were younger at diagnosis (61 years old vs 66-67). The PEG study was unable to examine patients at set time points, thus variation in disease duration was larger. Interestingly, though, the overall prevalence of hallucinations at, on average, 5 years of disease duration was similar in all three studies. On average 12.6% reported experiencing these phenomena, almost exclusively without loss of insight. See also Table 3.1.

3.4.a. PRS created for hallucinations based on a training dataset

First, we generated a PRS for hallucinations based on 21 candidate genes. The PRS based on our original criteria included 21 SNPs and was of moderate strength in the training set (PEG/PW). The adjusted OR was 2.69 (95%CI: 2.10 – 3.45), p-value was $2*10^{-13}$ and the R-square was 0.16. Unfortunately, this PRS did not replicate in the validation data with an adjusted

OR of 1.00 and p-value of 0.98 (see Table 3.2). When reviewing the associations of the individual SNPs with hallucinations, the effect estimates seen in the training dataset were not reproduced in the validation dataset, and several associations even changed the direction of the effect estimates explaining the overall null result. (See Supplemental Table 3.2)

Using different criteria to select SNPs did not improve the performance of the score, i.e. using more and less restrictive p-values (0.01 or 0.10), and selecting SNPs based on optimizing the R-square. None of the scores created (whether based on a minimum of 4 or a maximum of 180 SNPs) validated in the replication datasets. The adjusted ORs in the target datasets hovered around 1.0 (null value) and had high p-values (see Table 3.2a). Other variations in creating the PRS, where we exchanged training and target datasets or where we included other ethnicities, provided similar results. Even when using a cross-validation approach that resulted in a higher R-square in the training dataset (70% of subjects), the PRS effect was not replicated in the validation dataset (30% of subjects).

Results based on alternative methods to explore the associations between these 21 genes and hallucinations are presented in Table 3.3 and Table 3.4a. In a set-based analysis using all three studies we also did not identify associations with one specific gene at a formal statistical significant level, adjusted for multiple comparisons. Interestingly, a pooled analysis for all three studies showed a number of consistent findings across the three studies, even though it did not reach formal statistical significance after multiple comparison corrections were applied. Our pooled analysis suggests that genetic variants in APOE, LRRK2, SLC6A4, and BDNF might be important to review again in a larger dataset.

Prior to the development, we were unable to estimate the power to detect an association between the PRS and hallucinations. The heritability of hallucinations among PD patients is

currently unknown and our dataset was not large enough to calculate the heritability. However, it is likely we have only limited statistical power and therefore can only identify moderate to strong effect size SNPs in these datasets. A power calculation, after the PRS was created indicated that there was 80% power to detect an odds ratio of 1.3 or above. Assumptions were based on the distribution of the PRS in the training dataset and included a mean of -1.607 and a standard deviation of 2.272, in combination with an alpha of 0.05, sample size of 297 (PPMI) and two-sided testing.

3.4.b. PRS created for PD status based on previous GWAS results

We employed a PRS for PD status that was created based on the top SNPs from a previous meta-GWAS and included 442 genetic variants after clumping. This PD-status PRS validated very well in the PEG study where genetic data from controls was available. The association of the PRS with PD status, independent of the P-value threshold chosen, was highly associated with disease status. The adjusted OR for PD ranged between 1.34 to 1.50, with confidence intervals from 1.23 (lower) to 1.70 (upper) and p-values in the order of 10^{-10} to 10^{-12}. This PD status PRS was able to explain about 8% of the variance (R-square between 7.7 to 8.8%) in our combined datasets. (Table 3.2b) While we initially restricted to Caucasian subjects, the PRS appears to have an even slightly stronger association in the Hispanic population. The ORs ranged between 1.77 and 2.32, with confidence intervals from 1.43 to 3.2, p-values in the range of 10^{-7} to 10^{-8} and R-squares around 22 to 25% (see Supplemental Table 3.3). Evidence for the power of a PRS approach is that the individual SNPs of the PD-status PRS (p-value threshold $<1*10^{-8}$), are not statistically significant after adjustment for multiple testing (Table 3.5), while the overall PRS is highly significant.

However, while the PD-status PRS created using the meta-GWAS data showed a formally statistically significant (while weak) association with disease status, it did not show any association with hallucination status. The PRS generated using a rather low p-value threshold did suggest that there might be a small increase for hallucinations with this score based on 58 SNPs in 51 genes (aOR: 1.19 (95%CI: 0.92-1.52)), however the score did not reach formal statistical significance levels. (Table 3.2b) After reviewing the individual SNPs from the PD-status PRS and their association with hallucinations, the majority of the effect estimates had wide confidence intervals that overlapped with the effect estimates of the previous meta-GWAS. However, several effect estimates were very different and changed direction (Table 3.5). Genes that appeared to validate and especially would benefit from a larger study are LRRK2 and MAPT.

Based on the distribution of the created PRS for the PD-status with a p-value threshold $1*10^{-8}$, we had 80% power to detect an odds ratio of 1.65 or more. As we identified a very weak effect estimate around 1.2, our study was clearly not large enough to identify such a small PRS effect estimate.

3.5. Discussion and future directions

Using a meta-GWAS based PRS we found, as expected, a statistically formal significant association between PD-status and the PRS in our PEG study confirming that the sites identified as predictors in large GWAS can be widely applied to population-based PD studies, even across various ethnicities. However, we were unsuccessful when analyzing associations between this PD specific PRS and hallucinations, suggesting that whatever the influence is of these SNPs on PD risk does not necessarily translate to explaining these progression symptoms.

Alternatively, we attempted to generate a hallucination specific PRS from our data but the derived results did not replicate and while certain SNPs were consistently associated with hallucinations across three datasets, it is impossible to determine whether these genes (APOE, LRRK2, SLC6A4, or BDNF) are truly implicated in the development of hallucinations at the moment. Unfortunately, the lack of replication is most likely due to the limited sample size i.e. the small percentage of subjects with information about hallucination symptoms during follow-up and a lack of a large GWA studies on this subject.

Still, it appears that among the top genetic variants associated with hallucinations in our 21 candidate genes, there is consistency across the three studies. Unfortunately, the weights for the PRS we generated from the three datasets in the training phase, all with a relatively small sample size, were thus necessarily based on effect estimates with wide confidence intervals. This uncertainty of effect estimate is not taken into account during the creation of the score and this type of measurement error would lead to a bias towards the null and could explain our lack of replication.

Another potential type of measurement bias occurs when the outcome is misclassified. This is especially important for discrete phenotypes (e.g. PD status or hallucinations). Hence, using cases that are well characterized and have a very specific phenotype would increase validity. It is difficult to measure hallucinations accurately, and while the three studies used similar methods, the UPDRS only includes one question concerning hallucinations. It might be more accurate to measure hallucinations with a specialized questionnaire or instrument.[257] However, this instrument would have to be used in all studies consistently in order to pool these data. Thus, we expect some outcome misclassification independent of the genetic risk factor, which is likely to cause a further bias towards the null, making it more difficult to identify any association.

Instead of analyzing hallucinations at one specific time point, another option would be to use a time-to-event analysis. However, to get the timing right may depend on the ability of patients to recall accurately the first onset of hallucinations symptoms and to schedule patient follow-up assessments in tight enough intervals to facilitate accurate recall. Hence, in population-based studies time-to-event analysis may be challenging. In this research, we chose as the outcome ever-versus-never occurrence of hallucinations after on average ~5 years from a first PD diagnosis, a time period of great relevance for patients. Five years after diagnosis is also a reasonable chance of sufficient compliance with follow-up assessments, as both mortality and loss-to-follow-up increase substantially after five years from diagnosis.

In genetic studies, the most important potential confounder that influences results are genetic ancestry (population stratification). In this study, we restricted the analysis to subjects that were genetically identified as Caucasian, and additionally adjusted for fractional ancestry. For our PD-status PRS, we also analyzed the Hispanic population and were able to establish a stronger association between the PRS and PD status than in the Caucasian population. Commonly, a PRS is based on either a restricted Caucasian population or a mixed population with a majority of Caucasians. In these scenarios, it is important to recognize the likelihood that the polygenic score is invalid or less effective in a population of a different ethnicity or race. [258,259] In fact, extrapolations of GWAS results from one population to predict risks in a different population have often been inaccurate.[260,261] In addition, due to the misclassification and inaccuracy of PRS in different populations, health disparities might increase when such screening tools are implemented without taking these factors into account. [262]

3.6 Conclusion

PRS are gaining popularity and have the potential to decipher underlying pathomechanisms, predict outcomes, and even infer causality when using a Mendelian Randomization approach. While our PRS were not able to identify a strong association with hallucinations, there is some indication that genetic variants are associated with these symptoms. The PD-status PRS with a strict p-value threshold and the pooled analysis suggest genes such as LRRK2, APOE, SLC6A4, BDNF and MAPT are potentially involved in the occurrence of hallucinations among PD patients. We suggest that this outcome should be further explored in a larger dataset to identify subjects higher risk for developing hallucinations.

3.7 Tables and Figures

Figure 3.1 Inclusion and exclusion criteria for the study subjects in this study, stratified by the two separate PRS

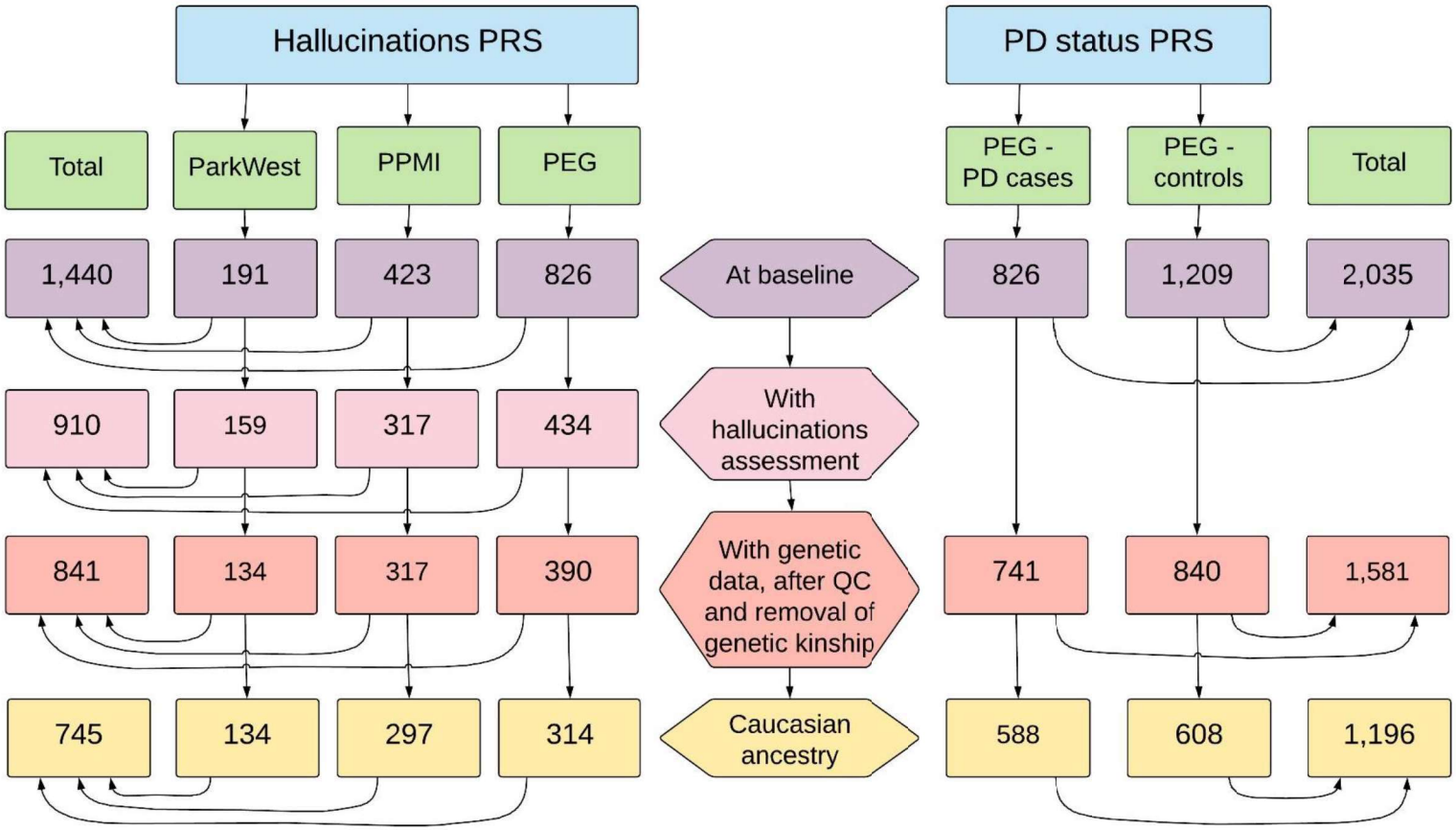


Table 3.1 Characteristics for the subjects in this study.

	Hallucinations						PD status	
	PEG		PW		PPMI		PEG	
	Nr	%/SD	Nr	%/SD	Nr	%/SD	Nr	%/SD
Nr. of subjects	390		134		317		1,581	
Age at diagnosis[1]	65.8	10	66.6	9.2	60.8	9.8	66.7	11.2
Dis. Duration[2]	6.2	2.8	5	0.1	4.5	0.5	NA	
Male sex	239	61.3	84	62.7	214	67.5	862	54.5
Race European	314	80.5	134	100	297	93.7	1,196	75.7
Hispanic	72	18.5	0	0	10	3.2	340	21.5
African	0	0	0	0	1	0.3	30	1.9
Asian	4	1	0	0	9	2.8	15	1.0
Hallucinations	51	13.1	15	10	40	12.6	NA	
UPDRS score at baseline	19.6	10.2	21.1	9.3	20.3	8.6	NA	

[1] Or age at interview for controls; [2] at time of hallucination assessment.
Abbreviations: Nr: Number; SD: Standard Deviation; Dis: disease; NA: Not Applicable

Table 3.2a Creation and validation of the PRS for hallucinations based on the 21 genes with varying criteria. Logistic regression to analyze the association between the PRS and hallucination at ≈5 years' disease duration.

	Nr. of	Training dataset					Validation dataset			
	SNPs	aOR	95%CI		P	R^2	aOR	95%CI		P
model 1	21	2.69	2.10	3.45	$2*10^{-13}$	0.16	1.00	0.75	1.33	0.98
model 2a	4	2.81	1.88	4.21	$4*10^{-6}$	0.06	1.04	0.80	1.35	0.76
model 2b	31	2.70	2.11	3.45	$8*10^{-14}$	0.17	1.03	0.59	1.81	0.91
model 2c	180	1.45	1.33	1.59	$2*10^{-20}$	0.28	0.92	0.75	1.14	0.45
model 3	19	3.01	2.30	3.96	$5*10^{-13}$	0.13	1.00	0.93	1.07	1.00
model 4	28	2.54	2.01	3.19	$3*10^{-12}$	0.14	0.95	0.80	1.13	0.56
model 5a	20	2.66	2.11	3.36	$3*10^{-15}$	0.18	0.89	0.57	1.39	0.62
model 5b	21	2.39	1.90	3.01	$6*10^{-10}$	0.10	0.99	0.82	1.19	0.89
model 5c	17	3.01	2.16	4.20	$7*10^{-7}$	0.06	1.01	0.97	1.05	0.59
model 5d	23	2.76	2.16	3.51	$1*10^{-16}$	0.38	1.03	0.91	1.18	0.64

Model 1: original creation of the PRS; training data is PEG/PW, LD threshold 0.1, P-value threshold 0.05, target data is PPMI – Caucasian population only

Model 2: variation of different p-value thresholds and optimizing R-squares. a) p-value threshold 0.01; b) p-value threshold 0.10; c) optimizing R-square

Model 3: Using the total population, including other ethnicities/race

Model 4: variation on the LD-threshold, instead of R-square of 0.1 using R-square of 0.5.

Model 5: different training and target datasets. a) PEG/PPMI as training dataset, PW as validation; b) PW/PPMI as training dataset, PEG as validation; c) PPMI as training dataset, PW/PEG as validation; d) using cross-validation, 70% of PEG/PPMI/PW as training dataset, remaining as validation

Abbreviation: Nr: Number; aOR: adjusted Odds Ratio; CI: Confidence Interval; P: P-value; R^2: R-square

Adjustment using a propensity score for sex, fractional ethnicity, age at diagnosis, disease duration, and study.

Table 3.2b Validation of the PRS was based on the top genetic variants from a previous GWAS, of which 8,233 variants are available for our subjects. [205] Logistic regression to analyze the association between the PRS with various P-value threshold and disease status as well as hallucination at ≈5 years' disease duration.

	Disease status (PD vs controls)						Hallucination status (among PD cases)					
Threshold P-value	Nr of SNPs[1]	aOR	95%CI		P	R^2	Nr of SNPs[1]	aOR	95%CI		P	R^2
$1*10^{-8}$	61	1.50	1.31	1.70	$9*10^{-10}$	0.08	58	1.19	0.92	1.52	0.18	0.03
$1*10^{-6}$	117	1.44	1.29	1.61	$1*10^{-10}$	0.08	119	1.13	0.92	1.40	0.24	0.03
$1*10^{-5}$	163	1.40	1.26	1.54	$7*10^{-11}$	0.08	161	1.14	0.95	1.39	0.17	0.03
$1*10^{-4}$	213	1.38	1.26	1.52	$3*10^{-11}$	0.08	206	1.12	0.93	1.34	0.23	0.03
$1*10^{-3}$	281	1.37	1.25	1.50	$2*10^{-11}$	0.08	268	1.06	0.90	1.27	0.48	0.03
0.01	358	1.36	1.25	1.49	$6*10^{-12}$	0.09	336	1.07	0.90	1.27	0.44	0.03
0.05	401	1.34	1.23	1.46	$4*10^{-11}$	0.08	372	1.06	0.90	1.26	0.49	0.03
0.10	413	1.34	1.23	1.46	$4*10^{-11}$	0.08	382	1.07	0.90	1.26	0.45	0.03
0.50	433	1.34	1.23	1.46	$4*10^{-11}$	0.08	399	1.06	0.90	1.26	0.48	0.03
All	442	1.50	1.31	1.70	$3*10^{-11}$	0.08	407	1.06	0.90	1.26	0.48	0.03

[1] differences in number of SNPs are caused by differences after LD-clumping in different populations

Abbreviation: Nr: Number; aOR: adjusted Odds Ratio; CI: Confidence Interval; P: P-value; R^2: R-square

Adjustment for the logistic regression analyzing PD status used a propensity score for sex, fractional ethnicity, and age at diagnosis or interview.

Adjustment for the logistic regression analyzing hallucination was done using a propensity score for sex, fractional ethnicity, age at diagnosis, disease duration, and study.

Table 3.3 Gene set-based analysis for the 21 genes and the association with hallucinations among PD patients in all three studies, with an LD threshold of 0.5, a cut-off p-value of 0.05, 25,000 permutations, and logistic regression adjusted for confounders using a propensity score

Gene	Nr of SNPs	Nr of indep SNPs	Nr of sign. SNP	Nr of indep sign. SNP	P Perm test	Independent significant SNPs
ACE	94	11	0	0	1.00	NA
APOE	87	22	6	4	0.22	19:45394211 - 19:45422160 - 19:45407788 - 19:45429543
BDNF	115	12	7	2	0.14	11:27751390 - 11:27702406
CCK	63	12	1	1	0.39	3:42293204
CCKAR	58	9	8	3	0.17	4:26511303 - 4:26473621 - 4:26508161
CCKBR	126	32	8	4	0.56	11:6311877 - 11:6292065 - 11:6312769 - 11:6291659
COMT	230	30	0	0	1.00	NA
DRD1	22	6	0	0	1.00	NA
DRD2	194	20	8	1	0.23	11:113321796
DRD3	154	19	0	0	1.00	NA
DRD4	130	31	4	3	0.56	11:637294 - 11:647395 - 11:628245
DRD5	67	5	0	0	1.00	NA
GBA	57	10	19	3	0.16	1:155205331 - 1:155198771 - 1:155198222
HOMER1	269	22	1	1	0.53	5:78665487
HTR2A	160	32	0	0	1.00	NA
LRRK2	497	47	9	5	0.25	12:40660487 - 12:40695776 - 12:40666760 - 12:40703878 - 12:40670998
MAPT	851	41	3	1	0.75	17:43969897
PITX3	59	15	1	1	0.45	10:104021085
SLC6A3	213	41	1	1	0.61	5:1461167
SLC6A4	90	13	1	1	0.07	17:28543389
SNCA	458	29	0	0	1.00	NA

Abbreviations: Nr: Number; sign.: significant, defined here as a p-value < 0.05; indep: independent; P: P-value; Perm: Permutation; NA: Not Applicable

Table 3.4 Pooled analysis for logistic regression to assess the association between the SNPs and hallucinations. The genetic variants included in this table are the top 20 genetic variants. A logistic regression adjusted for confounders was performed on all subjects from the three studies (PEG, PW and PPMI) combined, as well as stratified.

								Stratified by study		
								PEG	PW	PPMI
CHR	SNP	Gene	A1	aOR	95% CI		P	aOR	aOR	aOR
19	19:45394211	APOE	T	2.24	1.32	3.79	0.003	2.92	2.19	2.06
12	12:40660487	LRRK2	G	2.40	1.31	4.42	0.005	2.41	2.14	2.65
12	12:40695776	LRRK2	C	0.63	0.45	0.88	0.01	0.75	0.27	0.65
17	17:28543389	SLC6A4	T	2.06	1.20	3.54	0.01	1.34	2.66	4.16
11	11:27751390	BDNF	A	0.66	0.49	0.90	0.01	0.64	0.68	0.73
4	4:26511303	CCKAR	T	0.50	0.29	0.84	0.01	0.50	0.41	NA
11	11:113321796	DRD2	C	0.38	0.17	0.85	0.02	0.25	NA	1.26
11	11:6311877	CCKBR	A	0.58	0.37	0.91	0.02	0.64	1.03	0.21
12	12:40666760	LRRK2	T	1.67	1.09	2.57	0.02	1.49	1.71	2.29
11	11:637294	DRD4	T	2.23	1.13	4.39	0.02	2.34	1.33	2.77
1	1:155198771	GBA	G	0.56	0.34	0.92	0.02	0.56	1.23	0.39
19	19:45422160	APOE	G	1.60	1.07	2.38	0.02	2.12	1.64	1.30
11	11:27702406	BDNF	T	1.62	1.07	2.46	0.02	1.42	2.17	1.74
1	1:155198222	GBA	G	0.31	0.11	0.86	0.02	0.55	NA	0.16
5	5:1461167	SLC6A3	G	0.66	0.45	0.96	0.03	0.74	0.40	0.45
5	5:78665487	HOMER1	G	2.02	1.05	3.89	0.03	2.44	1.74	1.56
3	3:42293204	CCK	C	1.70	1.03	2.82	0.04	1.89	2.06	1.15
4	4:26473621	CCKAR	T	0.71	0.51	0.98	0.04	0.60	0.72	0.81
19	19:45429543	APOE	T	0.70	0.49	0.99	0.04	0.90	0.56	0.58
17	17:43969897	MAPT	C	1.55	1.01	2.35	0.04	1.14	3.24	1.53

Abbreviation: CHR: Chromosome; A1: Major Allele; aOR: adjusted Odds Ratio; CI: Confidence Interval; P: P-value
Adjustment for the logistic regression analyzing hallucination was done using a propensity score for sex, fractional ethnicity, age at diagnosis, disease duration, and study.

Table 3.5 Information for the genetic variants that are included in the PRS for PD using a P-value threshold of $1*10^{-8}$ (N=61), including the previously described results in the meta-analysis (the weights for the PRS) and a logistic regression for PD status in the PEG study.

					Meta-GWAS data				PD status				Hallucination			
Chr	SNP	(Closest) Gene	A	MAF[1]	OR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
1	1:156030037	RAB25	T	2.6	1.42	1.31	1.54	$4*10^{-17}$	1.76	1.03	3.04	0.04	0.87	0.33	2.27	0.78
1	1:205723572	NUCKS1	C	43.9	0.89	0.87	0.91	$1*10^{-23}$	0.92	0.78	1.09	0.33	0.93	0.68	1.27	0.64
1	1:226916078	ITPKB	C	27.3	0.92	0.89	0.94	$2*10^{-10}$	0.94	0.78	1.13	0.52	1.12	0.79	1.57	0.53
1	1:232664611	SIPA1L2	T	13.6	1.12	1.08	1.15	$8*10^{-13}$	1.00	0.79	1.27	1.00	0.93	0.58	1.49	0.76
2	2:135539967	R3HDM1	T	40.5	0.89	0.87	0.91	$8*10^{-24}$	0.96	0.81	1.14	0.64	1.06	0.78	1.44	0.71
2	2:169129145	STK39	T	13.6	1.21	1.17	1.25	$1*10^{-30}$	1.34	1.05	1.70	0.02	0.94	0.60	1.47	0.77
3	3:18277488	TBC1D5	G	11.8	1.11	1.07	1.15	$3*10^{-9}$	1.41	1.09	1.82	0.01	1.16	0.75	1.78	0.51
3	3:182762437	MCCC1	A	18.9	0.85	0.82	0.87	$2*10^{-30}$	0.99	0.80	1.22	0.93	0.67	0.43	1.04	0.07
4	4:893712	GAK	T	16.5	1.11	1.08	1.15	$8*10^{-11}$	0.90	0.72	1.12	0.34	1.36	0.93	2.00	0.11
4	4:897428	GAK	G	27.6	0.89	0.86	0.91	$3*10^{-18}$	0.95	0.79	1.14	0.57	1.13	0.79	1.60	0.50
4	4:951947	TMEM175	C	20.7	1.23	1.20	1.27	$1*10^{-50}$	1.29	1.06	1.58	0.01	1.38	0.97	1.96	0.07
4	4:15737101	BST1	C	45.2	0.90	0.88	0.92	$1*10^{-19}$	0.99	0.85	1.17	0.95	0.73	0.53	1.01	0.06
4	4:77176897	FAM47E	T	8.7	1.15	1.11	1.19	$7*10^{-13}$	1.38	1.02	1.86	0.04	1.25	0.77	2.02	0.36
4	4:77198986	FAM47E	T	37.5	0.92	0.90	0.94	$1*10^{-14}$	0.99	0.84	1.17	0.94	1.02	0.75	1.40	0.89
4	4:90551543	CCSER1	C	24.2	0.90	0.87	0.93	$4*10^{-11}$	NA				1.08	0.77	1.53	0.65
4	4:90562801	CCSER1	G	20.6	0.88	0.85	0.91	$2*10^{-12}$	0.98	0.81	1.20	0.88	NA			
4	4:90626111	SNCA	G	36.8	1.33	1.30	1.36	$5*10^{-123}$	1.33	1.12	1.57	0.001	0.92	0.66	1.28	0.62
4	4:90632025	SNCA	C	4.8	1.38	1.30	1.46	$1*10^{-29}$	1.64	1.12	2.42	0.01	0.95	0.47	1.90	0.88
4	4:90642531	SNCA	G	7.9	0.83	0.79	0.88	$2*10^{-12}$	1.04	0.77	1.41	0.78	1.00	0.55	1.80	0.99
4	4:90653134	SNCA	T	8.2	1.21	1.15	1.28	$3*10^{-11}$	1.30	0.98	1.74	0.07	NA			
4	4:90757294	SNCA	A	19.8	0.82	0.80	0.85	$4*10^{-38}$	0.78	0.63	0.96	0.02	0.89	0.59	1.35	0.59
4	4:90788943	MMRN1	A	6.3	1.26	1.19	1.34	$2*10^{-14}$	1.27	0.91	1.78	0.16	0.79	0.40	1.53	0.48
4	4:90882543	CCSER1	T	2.8	0.80	0.74	0.86	$1*10^{-9}$	0.81	0.49	1.34	0.41	NA			
4	4:91164040	CCSER1	T	6.5	1.25	1.19	1.31	$5*10^{-19}$	1.38	0.99	1.93	0.06	1.09	0.60	2.01	0.77
4	4:114360372	CAMK2D	C	12.1	1.14	1.09	1.18	$2*10^{-9}$	1.36	1.06	1.75	0.02	1.02	0.65	1.62	0.92
5	5:60273923	NDUFAF2	C	8.7	1.15	1.10	1.19	$2*10^{-11}$	1.30	0.97	1.73	0.08	1.36	0.84	2.20	0.21
6	6:27299584	TRNAM18	C	24.6	1.08	1.05	1.11	$9*10^{-9}$	1.08	0.90	1.31	0.41	1.30	0.92	1.84	0.14
6	6:27681215	Linc01012	A	19.8	1.13	1.09	1.16	$3*10^{-13}$	1.17	0.96	1.44	0.12	1.06	0.73	1.53	0.78
6	6:31875712	SLC44A4	T	33.2	1.08	1.06	1.11	$2*10^{-10}$	1.32	1.11	1.57	0.002	1.03	0.75	1.40	0.88
6	6:32395036	HLA-DRA	C	30.9	1.10	1.07	1.13	$1*10^{-11}$	1.06	0.90	1.26	0.49	1.07	0.77	1.48	0.68
6	6:32577973	HLA-DQA1	G	16.9	0.85	0.82	0.88	$5*10^{-17}$	0.75	0.60	0.94	0.01	0.65	0.40	1.06	0.08
7	7:23293746	GPNMB	G	38.4	0.91	0.89	0.93	$4*10^{-18}$	0.98	0.83	1.15	0.78	0.97	0.70	1.33	0.83

8	8:11707174	GATA4	G	26.3	0.91	0.88	0.94	$1*10^{-10}$	1.08	0.90	1.29	0.43	0.93	0.65	1.32	0.67
8	8:16697091	LOC 101929028	A	27.7	0.91	0.89	0.94	$2*10^{-11}$	1.11	0.92	1.33	0.28	1.19	0.84	1.69	0.34
9	9:17579690	SH3GL2	T	36.0	0.91	0.88	0.93	$2*10^{-12}$	0.91	0.78	1.08	0.27	1.33	0.97	1.84	0.08
9	9:17712754	SH3GL2	G	26.8	1.10	1.07	1.13	$2*10^{-11}$	1.05	0.87	1.26	0.61	1.38	0.99	1.92	0.06
10	10:121842595	LOC651144	A	4.7	1.32	1.21	1.44	$1*10^{-10}$	0.89	0.61	1.30	0.55	1.37	0.71	2.61	0.34
11	11:83544472	DLG2	A	41.9	0.93	0.90	0.95	$4*10^{-9}$	0.96	0.81	1.13	0.60	1.30	0.95	1.78	0.10
11	11:133765367	MIR4697	T	34.5	1.09	1.07	1.12	$1*10^{-13}$	1.02	0.86	1.20	0.85	0.81	0.58	1.12	0.20
12	12:40388109	SLC2A13	T	1.8	1.53	1.41	1.67	$2*10^{-24}$	2.09	1.11	3.93	0.02	NA			
12	12:40614434	LRRK2	T	15.7	1.15	1.12	1.19	$1*10^{-19}$	NA				1.13	0.73	1.75	0.59
12	12:40614656	LRRK2	A	34.5	0.92	0.90	0.94	$6*10^{-11}$	0.86	0.72	1.02	0.08	0.93	0.67	1.29	0.68
12	12:40617202	LRRK2	C	11.3	1.14	1.10	1.18	$9*10^{-13}$	1.33	1.02	1.72	0.03	NA			
12	12:40824739	MUC19	A	20.5	0.90	0.87	0.93	$8*10^{-13}$	1.15	0.94	1.41	0.17	NA			
12	12:41251554	CNTN1	G	2.2	1.32	1.22	1.43	$2*10^{-12}$	1.21	0.69	2.09	0.51	0.82	0.29	2.32	0.71
12	12:123303586	CCDC62	G	43.7	0.90	0.88	0.92	$2*10^{-20}$	0.91	0.77	1.07	0.24	0.98	0.72	1.34	0.90
14	14:55348869	GCH1	T	33.1	0.91	0.89	0.93	$4*10^{-16}$	0.88	0.73	1.04	0.14	0.84	0.60	1.19	0.33
14	14:88472612	GPR65	T	45.0	1.08	1.05	1.10	$1*10^{-9}$	1.03	0.87	1.21	0.74	0.92	0.67	1.26	0.60
15	15:61994134	LOC 107984782	G	24.3	0.91	0.89	0.93	$4*10^{-14}$	0.86	0.71	1.04	0.12	1.03	0.72	1.48	0.88
16	16:19279464	SYT17	T	45.9	1.07	1.05	1.10	$1*10^{-9}$	1.08	0.92	1.28	0.35	0.96	0.70	1.31	0.79
16	16:31121793	BCKDK	A	37.4	1.08	1.06	1.10	$5*10^{-12}$	1.27	1.07	1.51	0.01	0.94	0.68	1.30	0.70
17	17:40600319	ATP6V0A1	A	26.8	1.08	1.05	1.11	$9*10^{-9}$	1.21	1.01	1.45	0.04	1.01	0.72	1.42	0.95
17	17:43370481	SPATA32	T	24.3	0.90	0.87	0.92	$9*10^{-15}$	1.00	0.83	1.20	1.00	0.88	0.61	1.27	0.50
17	17:43460891	ARHGAP27	G	26.9	0.84	0.82	0.87	$1*10^{-27}$	0.84	0.69	1.01	0.06	1.02	0.71	1.45	0.92
17	17:43472507	ARHGAP27	G	27.7	1.09	1.06	1.12	$3*10^{-10}$	1.02	0.84	1.22	0.87	0.97	0.68	1.38	0.87
17	17:43731896	CRHR1	T	20.9	0.78	0.76	0.81	$4*10^{-57}$	0.83	0.67	1.01	0.07	0.92	0.61	1.40	0.70
17	17:43892520	CRHR1	A	32.1	1.09	1.06	1.12	$1*10^{-10}$	NA				1.19	0.86	1.66	0.30
17	17:43994648	MAPT	T	21.0	0.78	0.76	0.80	$1*10^{-68}$	0.83	0.67	1.01	0.06	0.92	0.61	1.40	0.70
17	17:44021960	MAPT	A	24.5	1.10	1.07	1.14	$4*10^{-10}$	1.02	0.84	1.23	0.85	NA			
17	17:44031596	MAPT	G	31.3	1.10	1.07	1.13	$1*10^{-11}$	NA				1.16	0.84	1.59	0.37
17	17:44246624	KANSL1	A	21.0	0.79	0.76	0.81	$2*10^{-52}$	0.84	0.68	1.02	0.08	0.91	0.60	1.38	0.66
17	17:44801784	NSF	A	20.0	0.79	0.77	0.82	$5*10^{-50}$	0.86	0.70	1.06	0.16	0.79	0.51	1.23	0.30
17	17:44852612	WNT3	C	25.5	1.11	1.07	1.15	$3*10^{-9}$	1.09	0.91	1.32	0.35	0.98	0.69	1.40	0.92
18	18:40673380	RIT2	G	32.9	1.10	1.07	1.12	$6*10^{-16}$	1.05	0.88	1.25	0.60	0.98	0.70	1.36	0.90
19	19:2358485	LSM7	C	39.4	1.08	1.05	1.10	$6*10^{9}$	0.97	0.82	1.15	0.74	1.16	0.86	1.59	0.33

Abbreviation: CHR: Chromosome; A: Major Allele; MAF: minor allele frequency; aOR: adjusted Odds Ratio; CI: Confidence Interval; P: P-value; R^2: R-square

[1]MAF is based on the subjects in the PEG study. However, if variant was only available in the PRS for hallucinations, than the MAF represents the minor allele frequency among PD patients in the three studies (PEG/PW/PPMI)

Adjustment for the logistic regression analyzing PD status was done using a propensity score for sex, fractional ethnicity, and age at diagnosis or interview.
Adjustment for the logistic regression analyzing hallucination was done using a propensity score for sex, fractional ethnicity, age at diagnosis, disease duration, and study.

Supplemental Table 3.1 Candidate genes for the association with hallucinations. These genes were chosen for the generation of a PRS based on previous research. (7,9-44) Basepair locations are based on the UCSC Genome Browser GRCh37/hg19 assembly

Chromosome	Basepair location		Gene
1	155204239	155214653	GBA
3	42299315	42307662	CCK
3	113847499	113918254	DRD3
4	9783258	9785633	DRD5
4	26483018	26492042	CCKAR
4	90645250	90759447	SNCA
5	1392905	1445543	SLC6A3
5	78669647	78809659	HOMER1
5	174867675	174871163	DRD1
10	103989946	104001231	PITX3
11	637305	640706	DRD4
11	6280841	6293363	CCKBR
11	27676440	27743605	BDNF
11	113280317	113346413	DRD2
12	40618813	40763087	LRRK2
13	47405677	47471211	HTR2A
17	28521337	28562986	SLC6A4
17	43971702	44105700	MAPT
17	61554422	61575741	ACE
19	45409039	45412650	APOE
22	19929263	19957498	COMT

Supplemental Table 3.2 Logistic regression analysis analyzing association between hallucinations and the 21 genetic variants that were included in the original PRS, based on the PEG/PW study, LD-clumping threshold 0.1, p-value threshold 0.05 among Caucasian PD patients, stratified by the training data (PEG and PW) and the validation dataset (PPMI)

					Training data (PEG & PW)				Validation data (PPMI)			
Chr	SNP	Gene	A1	MAF[1]	aOR	95% CI		P	aOR	95% CI		P
3	3:42293204	CCK	C	7.9	1.98	1.11	3.55	0.02	1.15	0.34	3.86	0.83
4	4:26473621	CCKAR	T	41.5	0.64	0.42	0.99	0.04	0.81	0.48	1.39	0.45
4	4:26511303	CCKAR	T	14.8	0.50	0.28	0.87	0.01	NA			
5	5:1374103	SLC6A3	C	14.7	1.62	1.03	2.56	0.04	0.63	0.25	1.62	0.34
5	5:1379490	SLC6A3	C	21.3	0.60	0.37	0.99	0.05	1.18	0.58	2.42	0.65
5	5:1454004	SLC6A3	T	9.8	1.61	1.02	2.54	0.04	1.17	0.38	3.58	0.79
5	5:1461568	SLC6A3	C	19.9	0.59	0.37	0.95	0.03	1.12	0.31	3.97	0.86
5	5:78665487	HOMER1	G	3.8	2.40	1.05	5.47	0.04	1.56	0.53	4.60	0.42
10	10:103984390	PITX3	C	3.7	2.69	1.12	6.45	0.03	1.00	0.30	3.33	0.99
11	11:636784	DRD4	C	41.5	1.55	1.03	2.33	0.04	0.93	0.56	1.53	0.76
11	11:6292065	CCKBR	A	13.6	1.90	1.20	3.01	0.01	0.80	0.30	2.13	0.65
11	11:27751390	BDNF	A	48.8	0.62	0.42	0.91	0.02	0.73	0.44	1.21	0.22
11	11:113321796	DRD2	C	8.2	0.19	0.06	0.62	0.01	1.26	0.41	3.87	0.69
12	12:40656164	LRRK2	G	2.7	3.17	1.32	7.63	0.01	1.21	0.26	5.69	0.81
12	12:40665454	LRRK2	G	3.3	2.35	1.03	5.37	0.04	1.44	0.35	5.98	0.62
12	12:40695776	LRRK2	C	40.2	0.62	0.40	0.96	0.03	0.65	0.38	1.11	0.11
13	13:47482083	HTR2A	C	5.4	0.21	0.05	0.93	0.04	1.80	0.63	5.10	0.27
17	17:28576856	SLC6A4	C	7.2	0.35	0.13	0.92	0.03	2.71	1.06	6.93	0.04
19	19:45422160	APOE	G	15.8	1.82	1.09	3.02	0.02	1.30	0.65	2.56	0.46
19	19:45394211	APOE	T	6.6	2.32	1.25	4.31	0.01	2.06	0.71	5.99	0.18
22	22:19921825	COMT	A	49.6	1.61	1.07	2.40	0.02	0.95	0.59	1.53	0.84

[1]Minor Allele Frequency is based on all PD patients in the three studies (PEG, PW and PPMI)

Abbreviation: CHR: Chromosome; A1: Major Allele; aOR: adjusted Odds Ratio; CI: Confidence Interval; P: P-value; NA: Not available due to cells containing zero

Adjustment for the logistic regression in the training data set was done using a propensity score for sex, fractional ethnicity, age at diagnosis, disease duration, and study.

Adjustment for the logistic regression in the validation data set was done using a propensity score for sex, fractional ethnicity, age at diagnosis, and disease duration.

Supplemental Table 3.3 Validation of the PRS was based on the top genetic variants from a previous GWAS, of which 8,233 variants are available for our subjects. [205] Logistic regression to analyze the association between the PRS with various P-value threshold and disease status, stratified by ethnicity. The total population includes subjects that were identified as Caucasians, Hispanics, Asians and Africans.

		Caucasians					Hispanics					Total population				
Threshold P-value	Nr of SNPs	aOR	95%CI		P	R^2	aOR	95%CI		P	R^2	aOR	95%CI		P	R^2
$1*10^{-8}$	61	1.50	1.31	1.70	$9*10^{-10}$	0.08	2.32	1.68	3.20	$3*10^{-7}$	0.22	1.60	1.42	1.80	$5*10^{-15}$	0.11
$1*10^{-6}$	117	1.44	1.29	1.61	$1*10^{-10}$	0.08	1.98	1.53	2.58	$3*10^{-7}$	0.22	1.51	1.37	1.67	$7*10^{-16}$	0.11
$1*10^{-5}$	163	1.40	1.26	1.54	$7*10^{-11}$	0.08	2.06	1.60	2.65	$2*10^{-8}$	0.24	1.47	1.34	1.61	$9*10^{-17}$	0.12
$1*10^{-4}$	213	1.38	1.26	1.52	$3*10^{-11}$	0.08	2.03	1.60	2.58	$8*10^{-8}$	0.25	1.46	1.33	1.59	$3*10^{-17}$	0.12
$1*10^{-3}$	281	1.37	1.25	1.50	$2*10^{-11}$	0.08	1.88	1.51	2.35	$3*10^{-8}$	0.24	1.42	1.31	1.54	$6*10^{-17}$	0.12
0.01	358	1.36	1.25	1.49	$6*10^{-12}$	0.09	1.78	1.44	2.20	$1*10^{-7}$	0.23	1.41	1.30	1.53	$3*10^{-17}$	0.12
0.05	401	1.34	1.23	1.46	$4*10^{-11}$	0.08	1.77	1.43	2.19	$1*10^{-7}$	0.23	1.39	1.28	1.50	$3*10^{-16}$	0.12
0.1	413	1.34	1.23	1.46	$4*10^{-11}$	0.08	1.78	1.44	2.20	$9*10^{-8}$	0.23	1.39	1.28	1.50	$3*10^{-16}$	0.12
0.5	433	1.34	1.23	1.46	$4*10^{-11}$	0.08	1.78	1.44	2.20	$9*10^{-8}$	0.23	1.39	1.28	1.50	$3*10^{-16}$	0.12
All	442	1.50	1.31	1.70	$3*10^{-11}$	0.08	1.78	1.44	2.20	$9*10^{-8}$	0.23	1.39	1.28	1.50	$2*10^{-16}$	0.12

Abbreviation: Nr: Number; aOR: adjusted Odds Ratio; CI: Confidence Interval; P: P-value; R^2: R-square
Adjustment for the logistic regression analyzing PD status used a propensity score for sex, fractional ethnicity, and age at diagnosis or interview.

4.1 Abstract

Objective: Although the diagnosis of Parkinson's disease (PD) is established based on motor symptoms, non-motor symptoms such as hallucinations are very common. The prevalence of hallucinations increases with disease duration. However, genetic risk factors for hallucinations remain elusive. In this study, we examined the overlap of genetic architecture for Alzheimer's disease (AD) and schizophrenia (SZ) with hallucinations in PD.

Methods: Three longitudinal PD studies were combined to analyze the association of two polygenic risk scores (PRS) with the prevalence of hallucinations after, on average, five years of disease duration. One international clinic-based study, and two population-based studies (one from Norway and one from the US). When combined, this included 745 Caucasian subjects. PRS were based on previously performed large GWA studies for AD and SZ.

Results: Stratifying by younger (<60 years) and older (60+ years) age at diagnosis, the SZ-PRS was associated with an increased risk for hallucinations among young PD patients (adjusted OR=1.18 (95%CI: 1.03-1.35, p-value 0.02). The AD-PRS had a positive association for hallucinations in older onset PD patients (adjusted OR=1.27 (95%CI: 1.08-1.50, p-value 0.005). The results were very similar when hallucination severity was used as the outcome.

Conclusions: Our results suggest that the specific biological mechanisms for hallucinations may depend on age at diagnosis. In young onset PD patients, SZ susceptibility factors may play a role in the development of hallucinations. While in older onset PD, hallucinations appear to be influenced by the genetic architecture seen in AD that contributes to cognitive decline.

4.2 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease in the United State and is estimated to affect 1.2 million individuals in the US by 2030.[263] A combination of genetic and environmental factors contribute to the loss of dopaminergic neurons and PD's characteristic motor symptoms, as well as to its peripheral non-motor symptoms.[264]

While for decades the clinical focus has been on the motor symptoms of PD,[202] in recent years more attention has been given to its non-motor symptoms (NMS) that considerably affect health related quality of life. NMS consist of a wide range of symptoms, from pain, depression and cognitive decline to sleep disorders.[265]

One important non-motor symptom that negatively affect quality of life is the occurrence of hallucinations among PD patients.[266] Depending on the type of measurement and when during disease progression hallucinations are assessed, the prevalence of hallucinations has been reported as low as 9% early in the disease to a cumulative occurrence in as many as 60% of patients after 12 years of disease.[111–114] Hallucinations among PD patients commonly manifest as visual hallucinations representing people, animals or inanimate objects. Patients generally retain insight into these phenomena not being 'real'. As the disease advances, more severe psychosis can manifest, including visual and non-visual hallucinations; (paranoid) delusions and a loss of insight.[109]

Risk factors for hallucinations include REM sleep behavior disorder, female sex, later age at diagnosis, disease duration, and motor subtype (postural instability and gait disorder).[118–120] Dopamine agonists were hypothesized to cause hallucinations,[121] although recently several longitudinal studies were unable to identify an association.[122,123] Hallucinations are not only

major contributors to a decrease in quality of life,[266] but have been associated with the need for nursing home care [150,151] and an increased mortality.[152,153]

Visual hallucinations experienced by PD patients resemble those of schizophrenia patients. Both auditory and visual hallucinations in patients with either Alzheimer disease (AD) or schizophrenia (SZ), have previously been attributed to a shared pathophysiologic mechanism.[267] Interestingly, in PD hallucinations often co-occur with cognitive decline; but it remains unclear whether one phenotype is causing the other. That is, some studies have shown that dementia increases the risk for subsequent development of hallucinations,[122,268,269] while other studies found the occurrence of hallucinations to increase the risk of subsequently developing dementia. [114,146–149] However, both phenomena may also have a common underlying pathobiology and whether one or the other occurs first might be random.[270,271]

Imaging studies among PD patients indicate that those who report hallucinations have a reduced cortical volume in various areas of the brain.[133,134,137–140] In addition, one study found reduced CSF amyloid $A\beta_{1\text{-}42}$ among PD patients with early onset hallucinations (within four years after diagnosis).[122] Post-mortem studies indicate that several neurodegeneration related protein aggregates i.e. Lewy bodies, amyloid and tau pathologies are more prevalent among those with hallucinations.[145] One study, however, reported that this pathology was limited to PD patients with hallucinations and co-existing dementia, while PD patients experiencing hallucinations without cognitive decline did not show these pathological changes.[272] This might indicate that patients who hallucinate with and without dementia are exhibiting separate subtypes of pathology leading to their hallucinations.

Previous studies that tried to identify genetic risk factors for hallucinations reported conflicting results.[102,120,157–191] These studies were candidate-gene based and most had a small

sample size (less than 250 subjects). Several studies focused on genes known to be related to cognitive decline and dementia, (e.g. APOE, MAPT), or to schizophrenia risk (e.g. COMT, DRD2). However, no study has previously taken a comprehensive look at the genetic risk factors for schizophrenia or AD using polygenic risk scores (PRS) and their association with hallucinations among PD patients.

In this study, we utilized GWAS results from large consortia to create PRS in order to identify potential genetic overlap between genes involved in AD or schizophrenia pathogenesis and hallucinations in PD. PRS are a very powerful tool, as the PRS combines the underlying genetic architecture into one score that can then be used to predict the occurrence of hallucinations among PD patients. We created PRS for schizophrenia and Alzheimer's disease separately and then estimated its effects based on three well characterized longitudinal studies of PD.

4.3 Methods

For a detailed description of the study population, outcome measurements and genotyping, refer to chapter 3. Briefly, this study combines three separate research populations (see also Figure 4.1). First, the Parkinson's Environment and Gene study (PEG) is a population-based case control study in the Central Valley of California. Second, the ParkWest study (PW) is a longitudinal, population-based cohort of Parkinson's disease patients in South and West-Norway. And the third study is the Parkinson's Progression Markers Initiative study (PPMI), a longitudinal, clinic-based study following incident Parkinson's disease patients and healthy controls. This study recruited internationally and had a multicenter design (US, Europe, Israel and Australia).

The presence of hallucinations was assessed using the Movement Disorder Society - Unified Parkinson Disease Rating Scale (MDS-UPDRS) in PEG and PPMI, while PW used the UPDRS.[245,246] The presence of hallucinations was considered positive when a patient or caregiver indicated the presence of illusions, hallucinations and/or psychosis. For this study, we used information at the first follow-up visit among PEG patients (on average 6.2 years after diagnosis), at five-year disease duration among patients from the PW, and 48-months after baseline-visit among the PPMI subjects.

The severity of hallucinations was estimated based on the score of the question. For the MDS-UPDRS this means 0 represents no hallucinations; 1 - slight illusions or non-formed hallucinations; 2 - formed hallucinations independent of environmental stimuli; 3 - formed hallucinations with loss of insight; and 4 - the patients is experiencing delusions or paranoia. The UPDRS has a slightly different scoring and unfortunately, it combined scores 1 and 2 of the MDS-UPDRS. Therefore, it does not differentiate between formed and non-formed hallucinations among subjects that have hallucinations without loss of insight. To take this uncertainty into account, we scored those subjects as 1.5 (N=9).

In all three studies a genome-wide array was performed and SNPs for the whole genome were imputed using the Michigan Imputation Server.[247] The Michigan Imputation Server uses Minimac3 for imputation, phasing using ShapeIT v2.r790,[248] and the reference panel was HRC.r1.2016.[249] Standard data quality control was performed. Subjects with less than 95% of the SNPs genotyped were removed, and SNPs with a minor allele frequency (MAF) less than 0.02 and or with a Hardy Weinberg Equilibrium p-value of less than $1*10^{-7}$ were excluded. Among family members with a genetic overlap of 20% or more (Identical by Descent), one subject was randomly selected to remain in the study.

After quality control, 841 PD patients, followed longitudinally, had provided both genotyping and hallucination information (390 in PEG, 134 in PW, and 317 in PPMI). All study protocols regarding human subjects have been approved by their local Institutional Review Board and written consent was given by all participants.

4.3.a. Confounders

Fractional ancestry among all subjects was estimated using hidden Markov modeling and clustering (Structure 2.3.4).[250] We used four clusters (based on the Evanno method),[251] 10,000 burnin periods and 10,000 Markov chain Monte Carlo repetition. We restricted the analysis to subjects classified as from the European Super-population and in the analysis we adjusted using fractional ancestry identified by Structure. As a sensitivity analysis, we also used the PRS on the total population, as well as the Hispanic subpopulation, and men and women separately.

All analyses were adjusted for sex, fractional ancestry, age at diagnosis, disease duration at time of assessment for hallucinations and study when necessary using propensity scores.

4.3.b. Creation of polygenic risk scores

Two weighted sum PRS were created, one for Alzheimer's disease and one for schizophrenia. The weights were based on the effect estimate from previous large scale GWAS data,[273,274] the final score being the sum of the number of risk alleles per individual weighted by each β coefficient (log of the odds ratio).

Alzheimer's disease (AD) – PRS: The PRS for AD was based on the IGAP GWAS and theAPOE status. The IGAP GWAS analysis was performed on 74,046 Caucasian subjects and consisted of over 7 million SNPs.[273] After excluding ambiguous variants and restricting to those

that were available in all three studies (PEG, PW, and PPMI), more than 3 million variants remained. Subsequent clumping of the data, to take linkage disequilibrium into account, using a R-square threshold of 0.1, led to a total of 77,881 SNPs remaining in the score.

As APOE allele status is a well-known genetic risk factor for AD, and allele status is best estimated using specific combinations of two SNPs (rs429358 and rs7412), we chose to use the APOE allele status instead of using the GWAS results for this specific gene region. The GWAS results have been known to be highly associated with APOE allele status. However, some measurement bias would still occur when using GWAS data instead of estimating the allele status based on the two SNPs. Among our subjects, there were no ambiguous APOE alleles; among the Caucasian population, 93 subjects had at least one APOE-e2 allele and 178 subjects had at least one APOE-e4 allele. The APOE risk was calculated using the odds ratio estimated from the meta-analysis of 28 case-control studies from Alzgene.org. The odds ratio among the Caucasian population has been estimated to be 0.55 (β: -0.598) for each APOE-e2 allele; and 3.77 (β: 1.327) for each APOE-e4 allele compared to the APOE-e3 allele.[275]

Schizophrenia (SZ) – PRS: The PRS for schizophrenia was based on the schizophrenia working group of the Psychiatric Genomics consortium. The GWAS analysis was performed on 77,096 subjects of mixed ethnicity, although the majority of the subjects were Caucasian (24 out of 28 studies) and the remaining studies were done among East-Asian populations.[274] After excluding ambiguous variants and restricting to those that were available in all three studies (PEG, PW, and PPMI), there were more than 3.5 million variants. Subsequent clumping of the data, using a R-square threshold of 0.1, led to a total of 70,966 SNPs.

4.3.c Analysis

After the creation of the two PRS, we performed logistic regression analysis to analyze the association between each PRS and the presence of hallucinations. In addition, we also performed an ordinal regression analysis analyzing hallucination severity. All analyses were adjusted for confounders using a propensity score.

In addition to our initial analysis, we performed stratified analysis by sex, age at diagnosis and ethnicity (European ancestry and Hispanic). Age at diagnosis was stratified by early onset (less than 60 years) versus those diagnosed at 60 years or older. As a sensitivity analysis, we also stratified by using the median age at diagnosis (65 years). Analyses were performed using a combination of Plink 1.9, RStudio 1.1.453, SAS 9.4 (SAS Institute, Cary NC), Structure 2.3.4, and PRSice 2.1.8.[250,252–256]

4.3.d. Power calculation

A power calculation, based on the distribution of the schizophrenia-PRS indicated that there was 80% power to detect an odds ratio of 1.18 or above. Assumptions were based on the distribution of the PRS (mean of -9.33 and a standard deviation of 3.63), in combination with an alpha of 0.05, sample size of 745 subjects and two-sided testing. Based on the distribution of the AD-PRS with a p-value threshold $1*10^{-8}$, we had 80% power to detect an odds ratio of 1.25 or more.

4.4. Results

The characteristics of the subjects in this study are shown in Table 4.1. On average, the age of diagnosis was 64 years (range: 23 to 87 years), 64% were male and our population consisted

mainly of subjects of European descent (89%). Hallucinations were prevalent in 12.6% of the subjects. Among the patients with hallucinations, 84% only had mild hallucinations without loss of insight, while 16% had severe hallucinations with loss of insight and/or psychosis or delusions.

The schizophrenia-PRS had between 106 to 70,966 SNPs depending on the P-value threshold. There did not appear to be any association between the PRS and the prevalence of hallucinations (Table 4.2a). The AD-PRS included 39 to 77,885 SNPs depending on the P-value threshold. Overall, there did not appear to be a strong association between hallucinations and the AD-PRS, although there was some indication of a small effect when using the AD-PRS with a smaller P-value threshold. For example, for the PRS with a p-value threshold of $1*10^{-3}$, using 810 SNPs, the adjusted OR (aOR) was 1.11 (95%CI: 1.00-1.22, p-value 0.05). The strongest genetic risk factor among the AD-PRS appeared to be the APOE-allele. When calculating the APOE risk based on a subject's APOE alleles, the adjusted OR was 1.30 (95%CI: 0.96-1.76, p-value 0.09). When analyzing according to hallucination severity, the effect estimates were very similar. (see Table 4.2b)

Interestingly, when stratifying by age at diagnosis, the schizophrenia-PRS is associated with an increase in hallucinations among the patients who were diagnosed before the age of 60 years (aOR for model with all SNPs: 1.18 (95%CI: 1.03-1.35, p-value 0.02), Table 4.3a), while no association was observed among those diagnosed at or after age 60 (aOR: 1.01 (95%CI: 0.95-1.08, p-value 0.76)). The opposite was found for the stratified analysis with the AD-PRS. The AD-PRS was associated with an increase of hallucinations among those who were diagnosed at a later age (aOR for model with p-value threshold $1*10^{-8}$: 1.27 (95%CI: 1.08-1.50, p-value 0.005), Table 4.4a), while no association or maybe even a protective effect was found among those

diagnosed before 60 years of age (aOR: 0.77 (95%CI: 0.57-1.05, p-value 0.10) Again, the effect of the APOE-allele appeared to be a strong driver for this increased risk. The adjusted OR for the APOE risk was 1.57 (95%CI: 1.11-2.23, p-value 0.01). Results for the ordinal analysis of hallucination severity were consistent with the logistic regression of the binary outcome (Table 4.3b and 4.4b).

Using different categories for age at diagnosis, for example using age 65 (the median for age at diagnosis) as the cut-off for stratification did not change the overall results. Effect estimates were very similar and the strength of the associations (p-values) were comparable (Supplemental Table 4.1).

The association seen among the Caucasian subpopulation for the schizophrenia-PRS did not replicate well in our Hispanic subpopulation. Among Hispanics, the more restrictive (lower p-value threshold) models using the schizophrenia-PRS appeared to have an association with hallucinations (see Supplemental Table 4.2); and the model using the p-value threshold of $1*10^{-4}$ with 1090 SNPs generates an aOR of 2.21 (95%CI: 1.12-4.34, p-value 0.02). This model also had an association independent of the age of diagnosis in Hispanics (aOR among those diagnosed before 60: 2.21; aOR among those diagnosed at age 60 or later: 2.29). However, there did not appear to be an association with the full 70,966 SNP schizophrenia-PRS among Hispanics, with an early diagnosis (aOR: 1.00 (95%CI: 0.67 – 1.47, p-value 0.98).

This inconsistency between the Caucasian and Hispanic subpopulations was also seen among the AD-PRS. While the AD-PRS appeared to have a small, positive association with hallucinations among the Caucasian subpopulation, there was no indication of an association among the Hispanic subpopulation, where the aOR is either below or near one. This same inconsistency was also found among those diagnosed at or after age 60. While the aOR among

Caucasians indicates a positive association in the smallest PRS on 39 SNPs (p-value threshold $1*10^{-8}$; aOR: 1.27 (95%CI: 1.08-1.50, p-value 0.004)), there was no association among those identified as Hispanic (aOR: 0.72 (95%CI: 0.33-1.53, p-value 0.39).

Even though associations were identified between the PRS and hallucinations, the R-square for these associations were rather small, i.e. less than 1% in the models using the total population and a maximum of 3.5% in the SZ-PRS using all SNPs among young PD patients. This suggests that the PRS only explains a small fraction of the risk for hallucinations overall.

4.5 Discussion

This study is the first to use a PRS to analyze the genetic risk factors for AD and schizophrenia with hallucinations among PD patients. Using a PRS, based on previously performed GWAS data is a very powerful method to determine the possible overlapping pathobiology between various phenotypes. The use of previously performed GWAS data for the creation of a PRS has several benefits. One benefit was that when the GWAS was performed on very large datasets, the estimates for the associations were precise. This ensured that the measurement errors for the weights in the PRS were minimalized. Another benefit was that there was no need for validation of the PRS in a separate, independent dataset. Previous studies have provided validation for both the AD-PRS and SZ-PRS.[223,276–278]

This study identified some age-specific associations between both PRS and hallucinations. Specifically, the AD-PRS is positively associated with hallucinations among patients diagnosed after age 60 and the SZ-PRS is positively associated with hallucinations among patients diagnosed before age 60. This suggests that there are different biological pathways for patients diagnosed at an early age vs. at a later age. This is in line with a prior study that suggested that

hallucinatory psychosis in young onset PD patients without dementia could be due to an underlying psychiatric illness.[279] Hypothesized in this study was that genetic susceptibility for psychiatric illnesses are represented by the SZ-PRS among young onset PD patients.

The results in this study indicate that the PRS using the full model (all SNPs) for schizophrenia performed optimally, while the PRS for AD had stronger associations (narrower confidence intervals) in much more restricted models. This is not surprising as schizophrenia is thought to have an infinitesimal genetic model, while for AD several genes and genetic variants with larger effect sizes have been indicated (e.g. APOE). As such, a more restrictive model for AD is justifiable.

The genetic architecture of the PRS for schizophrenia has been researched extensively. The SZ-PRS has been associated with many phenotypes, both psychiatric (such as bipolar disease, autism and anxiety disorders), cognition, and non-psychiatric (e.g. hypertension, diabetes). This shows the pleiotropy of this PRS.[280] Contrary to the large number of studies that identified correlations between PRS and other phenotypes, attempts to identify underlying pathomechanisms have been less successful.[281] Some biological functions, such as neurogenesis, mitochondrial disruption, RNA splicing and phosphoprotein gene pathways have been considered promising.[282,283] A reason for this lack of finding biological pathways is likely the infinitesimal genetic model, where the effect estimates are very small and a very large number of SNPs contribute to the genetic susceptibility. This might also indicate potential interactions between multiple networks. [284]

The effects of the AD-PRS seemed to be mainly driven by the APOE status of patients. APOE regulates amyloid-beta aggregation and clearance in the brain, as well as brain lipid transport, glucose metabolism, neuronal signaling, neuro-inflammation and mitochondrial

function.[285] An association between APOE and cognitive decline among PD patients has been established previously. [286] Prior studies that tried to identify an association with hallucinations among PD patients have been inconsistent, with some studies identifying an association,[176,178,186] while others did not.[160,175,177,179] However, none of these studies stratified by age at diagnosis, a majority only had a small sample size (less than 250 subjects) and disease duration varied by study.

We previously assessed the association between a PRS developed for PD, based on large GWAS data for PD risk factors (Chapter 3). This study did not find an association with hallucinations. This was independent of age at diagnosis. Combined with the findings in this study, these results suggest that the development of hallucinations is based on a pathobiology that is different from the progression of the disease itself, aka it is not the Parkinson's disease itself that is causing hallucinations among PD patients. Among older PD patients, the pathobiology of hallucinations appears more similar to that of cognitive decline and Alzheimer's disease among the older PD patients. While among younger PD patients hallucinations appear to be caused by an underlying genetic susceptibility for hallucinations, similar to the genetic risk factors for schizophrenia.

When identifying an association with hallucinations, it is important to take disease duration into account. Longer disease duration is associated with an increase in hallucinations and a time-to-event analysis would be optimal. However, in this study, we decided to choose one time point instead of a time-to-event analysis. The PEG study only gathered information on hallucinations at follow-up visits, and the disease duration varied by subjects but averaged around 6 years into their disease. The PPMI study and PW study had set time points where they estimated the prevalence of hallucinations. The PW had visits scheduled every 2 years, while the PPMI had

visits every 3 to 6 months in the first five years, so there was no consistency in visit frequency between the three studies and the accuracy of an accurate recall for the start date of the hallucinations would be biased according to study. We hypothesize that there is an increased risk for hallucinations at each time point. Therefore, choosing a single time point was appropriate for this study. The prevalence of hallucinations is around 10 to 15% at around 5 years' disease duration. In addition, the mortality and loss-to-follow-up in longitudinal PD studies has been shown to increase substantially after five years, which was the main reason for choosing this specific time point.

By choosing to use the estimate at around 5 years' disease duration, there is still a possibility of survival bias. This would be the case if loss-to-follow-up was associated with the PRS. A Cox-survival-regression analysis to estimate the association between the PRS and mortality was performed in the PEG study and no such association was identified. Therefore, survival bias appeared to be unlikely.

It is difficult to assess the prevalence of hallucinations accurately. Estimates were based on the UPDRS or MDS-UPDRS, that was available across the three studies. More specialized and validated instruments to measure hallucinations have been developed,[257] but were unavailable in these studies. It is likely that the number of patients with hallucinations was underestimated, especially considering that patients that did not report hallucinations at the specific time point might develop hallucinations in the near future or that patients were giving socially desirable answers and were afraid to acknowledge the presence of hallucinations. There is no indication that the genetic score would influence the mismeasurement differentially. The use of the UPDRS/MDS-UPDRS likely caused non-differential misclassification, leading to a bias towards null. This suggests that the true associations were likely to be larger in magnitude.

In genetic studies, the most important potential confounder that influences results are genetic ancestry (population stratification). In this study, the analysis was restricted to subjects that were genetically identified as Caucasian, and additionally adjusted for fractional ancestry. The SZ-PRS was based on a mixed population, while the AD-PRS was based on Caucasian subjects. However, when the association with the PRS among the Hispanic subpopulation was analyzed, the results differed from those among the Caucasians. Part of these differences could be the relatively small sample size of the Hispanic subpopulation, however it is also very likely that there is population stratification and that these PRS are not valid for this subpopulation. Previous studies have indicated that the APOE e4 allele is less prevalent among Hispanics and that the risk for dementia associated with APOE e4 among Hispanics is lower or even absent.[287–289] Even though the SZ-PRS was based on a mixed population, it consisted of mainly Caucasian subjects, with the remainder being mainly East-Asian. Unfortunately, the Hispanic subpopulation was not well-represented in the meta-GWAS and it appears that the PRS is not generalizable.

4.6 Conclusion

Pathomechanisms for hallucinations appear to differ by age at onset as suggested by the genetic architecture differences based on the association with the AD- and the SZ-PRS. Among subjects with young onset PD (before age 60), genetic susceptibility for schizophrenia appears important, while for older patients the genetic architecture associated with Alzheimer's disease (especially APOE) is significant for the development of hallucinations. These results suggest that the overlapping pathobiology and the genetic architecture for AD is an underlying cause for both cognitive decline and hallucinations among older-onset PD patients.

4.7 Tables and Figures

Figure 4.1 Inclusion and exclusion criteria for the subjects in this study

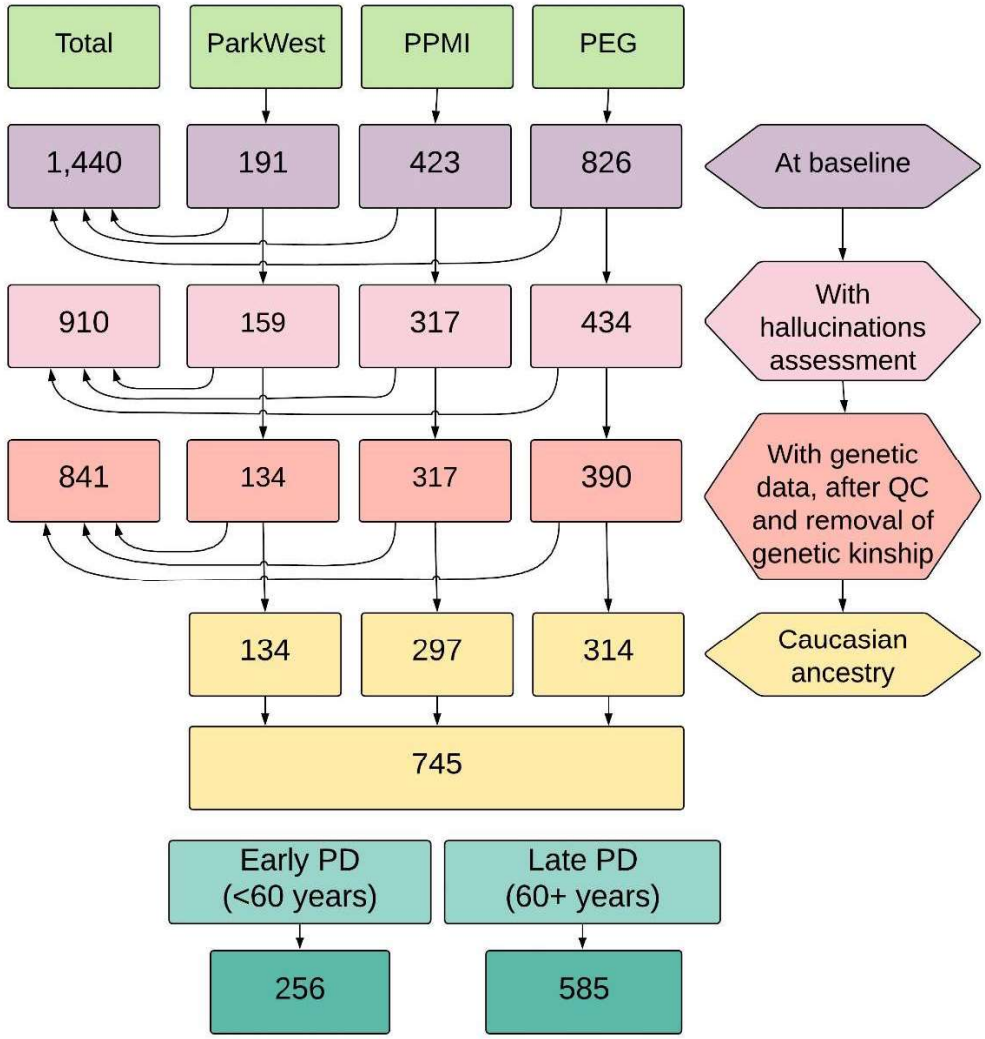


Table 4.1 Characteristics for the subjects in this study, stratified by study and stratified by age at diagnosis (less than 60 years of age vs. 60 years and older at diagnosis)

	Stratified by study						Stratified by age at diagnosis				Total population	
	PEG		PW		PPMI		Diagnosis <60		Diagnosis ≥ 60			
	No.	%/ SD	No.	%/ SD	No.	%/ SD	No.	%/ SD	No.	%/ SD	No.	%/ SD
No. of subjects	390		134		317		256		585		841	
Age at diagnosis	65.8	10	66.6	9.2	60.8	9.8	52	6.3	69.3	6.1	64.1	5.4
Dis. Duration	6.2	2.8	5	0.1	4.5	0.5	5.7	2.6	5.2	1.7	5.4	2.1
Male sex	239	61.3	84	62.7	214	67.5	160	62.5	377	64.4	537	63.9
Race												
European	314	80.5	134	100	297	93.7	222	86.7	523	89.4	745	88.6
Hispanic	72	18.5	0	0	10	3.2	29	11.3	53	9.1	82	9.8
African	0	0	0	0	1	0.3	0	0	1	0.2	1	0.1
Asian	4	1	0	0	9	2.8	5	2	8	1.4	13	1.6
Hallucinations	51	13.1	15	10	40	12.6	27	10.6	79	13.5	106	12.6
UPDRS score at baseline	19.6	10.2	21.1	9.3	20.3	8.6	19.4	9.2	20.4	9.7	20.1	9.5

Abbreviations: No.: Number; %: percentage; SD: Standard Deviation; Dis.: disease

Table 4.2a Logistic regression for the association between Schizophrenia-PRS or the AD-PRS and hallucinations among subjects from all three cohorts combined (PEG, PW and PPMI).

	Schizophrenia - PRS					Alzheimer's disease – PRS				
Threshold P-value	No. of SNPs	aOR	95% CI		P	No. of SNPs	aOR	95% CI		P
$1*10^{-8}$	106	0.94	0.60	1.46	0.77	39	1.11	0.97	1.28	0.13
$1*10^{-6}$	281	1.09	0.79	1.49	0.60	52	1.13	0.98	1.29	0.08
$1*10^{-5}$	509	1.06	0.83	1.36	0.63	84	1.12	0.98	1.27	0.10
$1*10^{-4}$	1,090	1.12	0.92	1.37	0.27	199	1.09	0.96	1.23	0.17
$1*10^{-3}$	2,760	0.94	0.82	1.09	0.42	810	1.11	1.00	1.22	0.05
0.01	8,097	1.02	0.92	1.12	0.75	4,523	1.04	0.98	1.11	0.19
0.05	18,452	1.05	0.97	1.14	0.25	14,670	1.03	0.98	1.07	0.25
0.10	26,621	1.03	0.96	1.10	0.46	23,504	1.02	0.98	1.05	0.37
0.20	37,560	1.04	0.98	1.11	0.23	36,706	1.01	0.98	1.04	0.72
0.50	57,013	1.04	0.98	1.11	0.19	60,898	1.00	0.98	1.03	0.87
All	70,966	1.04	0.98	1.11	0.18	77,885	1.00	0.98	1.03	0.90
APOE risk						2	1.30	0.96	1.76	0.09

The APOE risk is based rs429358 and rs7412, where each apoe4-allele increases risk (β: 1.327), while each apoe2-allele decreases risk (β:-0.598) compared to apoe3.
The logistic regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Table 4.2b Ordinal regression for the association between Schizophrenia-PRS or the AD-PRS and hallucinations severity among subjects from all three cohorts combined (PEG, PW and PPMI).

	Schizophrenia - PRS					Alzheimer's disease – PRS				
Threshold P-value	No. of SNPs	aOR	95% CI		P	No. of SNPs	aOR	95% CI		P
$1*10^{-8}$	106	0.90	0.58	1.39	0.63	39	1.15	0.93	1.42	0.20
$1*10^{-6}$	281	1.04	0.76	1.43	0.79	52	1.17	0.96	1.42	0.11
$1*10^{-5}$	509	1.02	0.80	1.31	0.85	84	1.15	0.95	1.38	0.14
$1*10^{-4}$	1,090	1.09	0.90	1.33	0.38	199	1.09	0.93	1.29	0.29
$1*10^{-3}$	2,760	0.92	0.80	1.05	0.23	810	1.12	0.99	1.26	0.07
0.01	8,097	1.00	0.91	1.11	0.97	4,523	1.05	0.98	1.12	0.19
0.05	18,452	1.04	0.96	1.13	0.33	14,670	1.03	0.98	1.08	0.21
0.10	26,621	1.02	0.95	1.09	0.59	23,504	1.02	0.98	1.06	0.32
0.20	37,560	1.03	0.97	1.10	0.34	36,706	1.01	0.98	1.04	0.60
0.50	57,013	1.03	0.97	1.10	0.27	60,898	1.01	0.98	1.03	0.71
All	70,966	1.03	0.98	1.10	0.26	77,885	1.00	0.98	1.03	0.73
APOE risk						2	1.33	0.98	1.79	0.06

The APOE risk is based rs429358 and rs7412, where each apoe4-allele increases risk (β: 1.327), while each apoe2-allele decreases risk (β: -0.598) compared to apoe3.
The ordinal regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.

Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Table 4.3a Logistic regression for the association between Schizophrenia-PRS and hallucinations among subjects from all three cohorts combined (PEG, PW and PPMI), stratified by age at diagnosis.

		Diagnosis PD before age 60				Diagnosis PD at or after age 60			
Threshold P-value	No. of SNPs	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	106	0.84	0.35	2.01	0.69	0.99	0.59	1.66	0.97
$1*10^{-6}$	281	0.87	0.47	1.63	0.67	1.20	0.83	1.74	0.33
$1*10^{-5}$	509	1.26	0.73	2.18	0.40	1.03	0.77	1.36	0.86
$1*10^{-4}$	1,090	1.20	0.78	1.84	0.40	1.10	0.88	1.38	0.39
$1*10^{-3}$	2,760	0.93	0.68	1.26	0.63	0.95	0.81	1.11	0.53
0.01	8,097	1.09	0.88	1.34	0.43	1.00	0.89	1.12	0.98
0.05	18,452	1.17	0.99	1.38	0.07	1.01	0.93	1.11	0.76
0.10	26,621	1.20	1.03	1.40	0.02	0.98	0.91	1.06	0.65
0.20	37,560	1.18	1.02	1.37	0.02	1.01	0.94	1.08	0.82
0.50	57,013	1.17	1.02	1.34	0.02	1.01	0.95	1.08	0.72
All	70,966	1.18	1.03	1.35	0.02	1.01	0.95	1.08	0.76

The logistic regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Table 4.3b ordinal regression for the association between Schizophrenia-PRS and hallucinations severity among subjects from all three cohorts combined (PEG, PW and PPMI), stratified by age at diagnosis.

		Diagnosis PD before age 60				Diagnosis PD at or after age 60			
Threshold P-value	No. of SNPs	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	106	0.84	0.35	2.02	0.70	0.94	0.56	1.57	0.82
$1*10^{-6}$	281	0.87	0.47	1.63	0.67	1.16	0.80	1.67	0.43
$1*10^{-5}$	509	1.26	0.73	2.18	0.40	0.99	0.75	1.31	0.96
$1*10^{-4}$	1,090	1.19	0.78	1.81	0.43	1.09	0.87	1.36	0.48
$1*10^{-3}$	2,760	0.92	0.67	1.25	0.58	0.93	0.79	1.09	0.35
0.01	8,097	1.08	0.88	1.33	0.44	0.99	0.88	1.11	0.85
0.05	18,452	1.18	1.00	1.40	0.05	1.01	0.92	1.10	0.90
0.10	26,621	1.22	1.04	1.42	0.01	0.97	0.90	1.05	0.52
0.20	37,560	1.18	1.02	1.37	0.02	1.00	0.93	1.08	0.96
0.50	57,013	1.17	1.03	1.34	0.02	1.01	0.94	1.08	0.85
All	70,966	1.18	1.04	1.36	0.01	1.00	0.94	1.07	0.89

The ordinal regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Table 4.4a Logistic regression for the association between AD-PRS and hallucinations among subjects from all three studies (PEG, PW and PPMI), stratified by age at diagnosis.

		Diagnosis PD before age 60				Diagnosis PD at or after age 60			
Threshold P-value	No. of SNPs	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	39	0.77	0.57	1.05	0.10	1.27	1.08	1.50	0.005
$1*10^{-6}$	52	0.80	0.60	1.07	0.13	1.26	1.08	1.48	0.003
$1*10^{-5}$	84	0.77	0.58	1.04	0.09	1.26	1.08	1.47	0.003
$1*10^{-4}$	199	0.77	0.58	1.02	0.07	1.21	1.05	1.40	0.01
$1*10^{-3}$	810	0.89	0.71	1.10	0.28	1.18	1.05	1.33	0.005
0.01	4,523	0.92	0.80	1.05	0.23	1.08	1.00	1.17	0.04
0.05	14,670	0.93	0.84	1.02	0.12	1.05	1.00	1.11	0.04
0.10	23,504	0.94	0.86	1.03	0.18	1.03	0.99	1.08	0.13
0.20	36,706	0.96	0.90	1.03	0.30	1.01	0.98	1.05	0.45
0.50	60,898	0.96	0.90	1.02	0.17	1.01	0.98	1.04	0.50
All	77,885	0.95	0.89	1.02	0.15	1.01	0.98	1.04	0.51
APOE risk	2	0.83	0.45	1.54	0.55	1.57	1.11	2.23	0.01

The APOE risk is based rs429358 and rs7412, where each apoe4-allele increases risk (β: 1.327), while each apoe2-allele decreases risk (β:-0.598) compared to apoe3.
The ordinal regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Table 4.4b Ordinal regression for the association between AD-PRS and hallucinations severity among subjects from all three studies (PEG, PW and PPMI), stratified by age at diagnosis.

		Diagnosis PD before age 60				Diagnosis PD at or after age 60			
Threshold P-value	No. of SNPs	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	39	0.60	0.38	0.95	0.03	1.40	1.09	1.80	0.01
$1*10^{-6}$	52	0.66	0.43	1.02	0.06	1.38	1.10	1.72	0.01
$1*10^{-5}$	84	0.64	0.42	0.97	0.03	1.34	1.08	1.66	0.01
$1*10^{-4}$	199	0.64	0.44	0.94	0.02	1.25	1.04	1.51	0.02
$1*10^{-3}$	810	0.87	0.67	1.13	0.29	1.19	1.04	1.37	0.01
0.01	4,523	0.93	0.80	1.08	0.32	1.08	1.00	1.17	0.07
0.05	14,670	0.93	0.85	1.03	0.18	1.05	1.00	1.11	0.05
0.10	23,504	0.95	0.87	1.04	0.27	1.03	0.99	1.08	0.16
0.20	36,706	0.97	0.91	1.04	0.45	1.01	0.98	1.05	0.48
0.50	60,898	0.97	0.91	1.03	0.27	1.01	0.98	1.04	0.50
All	77,885	0.96	0.90	1.03	0.25	1.01	0.98	1.04	0.51
APOE risk	2	0.82	0.44	1.53	0.53	1.60	1.13	2.27	0.01

The APOE risk is based rs429358 and rs7412, where each apoe4-allele increases risk (β: 1.327), while each apoe2-allele decreases risk (β:-0.598) compared to apoe3.

The ordinal regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Supplemental Table 4.1a Logistic regression for the association between Schizophrenia-PRS and hallucinations among subjects from all three cohorts combined (PEG, PW and PPMI), stratified by median age at diagnosis (65 years)

		Diagnosis PD before age 65				Diagnosis PD at or after age 65			
Threshold P-value	No. of SNPs	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	106	1.10	0.55	2.20	0.79	0.84	0.47	1.49	0.55
$1*10^{-6}$	281	1.20	0.74	1.93	0.46	1.03	0.67	1.57	0.90
$1*10^{-5}$	509	1.30	0.87	1.95	0.20	0.94	0.68	1.30	0.71
$1*10^{-4}$	1,090	1.22	0.89	1.67	0.22	1.06	0.82	1.37	0.67
$1*10^{-3}$	2,760	1.11	0.88	1.40	0.39	0.86	0.72	1.03	0.10
0.01	8,097	1.11	0.95	1.31	0.18	0.96	0.84	1.09	0.52
0.05	18,452	1.17	1.03	1.33	0.02	0.97	0.87	1.08	0.56
0.10	26,621	1.14	1.02	1.28	0.03	0.96	0.88	1.05	0.36
0.20	37,560	1.15	1.03	1.28	0.01	0.98	0.90	1.06	0.63
0.50	57,013	1.13	1.02	1.26	0.02	0.99	0.92	1.07	0.81
All	70,966	1.14	1.03	1.26	0.01	0.99	0.92	1.07	0.80

The logistic regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Supplemental Table 4.1b Logistic regression for the association between AD-PRS and hallucinations among subjects from all three studies (PEG, PW and PPMI), stratified by median age at diagnosis (65 years)

		Diagnosis PD before age 65				Diagnosis PD at or after age 65			
Threshold P-value	No. of SNPs	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	39	0.95	0.77	1.18	0.67	1.29	1.06	1.58	0.01
$1*10^{-6}$	52	0.97	0.79	1.18	0.73	1.29	1.07	1.55	0.01
$1*10^{-5}$	84	0.95	0.77	1.16	0.61	1.29	1.07	1.55	0.01
$1*10^{-4}$	199	0.93	0.77	1.13	0.46	1.25	1.05	1.48	0.01
$1*10^{-3}$	810	0.99	0.85	1.16	0.90	1.20	1.05	1.38	0.01
0.01	4,523	1.01	0.91	1.11	0.91	1.07	0.98	1.16	0.14
0.05	14,670	0.99	0.92	1.06	0.73	1.05	0.99	1.11	0.11
0.10	23,504	0.99	0.93	1.05	0.72	1.03	0.98	1.08	0.24
0.20	36,706	1.00	0.95	1.05	0.93	1.01	0.97	1.05	0.77
0.50	60,898	0.99	0.95	1.04	0.79	1.00	0.97	1.04	0.87
All	77,885	0.99	0.95	1.04	0.78	1.00	0.97	1.04	0.91
APOE risk	2	1.12	0.72	1.74	0.61	1.54	1.01	2.36	0.04

The APOE risk is based rs429358 and rs7412, where each apoe4-allele increases risk (β: 1.327), while each apoe2-allele decreases risk (β: -0.598) compared to apoe3.
The ordinal regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

Supplemental Table 4.2a Logistic regression for the association between schizophrenia-PRS and hallucinations among subjects stratified by ethnicity

		Caucasian subpopulation				Hispanic subpopulation				Total population			
P-value thresh.	No. of SNPS	aOR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	106	0.92	0.59	1.42	0.69	2.95	0.71	12.3	0.14	1.02	0.67	1.54	0.93
$1*10^{-6}$	281	1.06	0.78	1.46	0.70	1.73	0.64	4.67	0.28	1.11	0.82	1.49	0.49
$1*10^{-5}$	509	1.04	0.81	1.34	0.74	1.66	0.73	3.78	0.23	1.09	0.86	1.39	0.45
$1*10^{-4}$	1,090	1.10	0.90	1.34	0.35	2.21	1.12	4.34	0.02	1.19	0.99	1.44	0.06
$1*10^{-3}$	2,760	0.93	0.81	1.07	0.33	1.52	0.99	2.35	0.06	1.00	0.88	1.14	0.98
0.01	8,097	1.01	0.91	1.11	0.88	1.31	0.92	1.87	0.14	1.04	0.95	1.14	0.42
0.05	18,452	1.04	0.96	1.13	0.29	1.07	0.82	1.40	0.62	1.05	0.98	1.14	0.18
0.1	26,621	1.02	0.95	1.10	0.53	1.04	0.81	1.33	0.77	1.03	0.96	1.10	0.40
0.2	37,560	1.04	0.97	1.10	0.29	1.03	0.81	1.30	0.83	1.04	0.98	1.11	0.21
0.5	57,013	1.04	0.98	1.10	0.23	1.00	0.81	1.25	0.98	1.04	0.98	1.10	0.19
All	70,966	1.04	0.98	1.10	0.23	1.01	0.82	1.25	0.93	1.04	0.98	1.10	0.18
Less than 60 years old at diagnosis		Caucasian subpopulation				Hispanic subpopulation				Total population			
P-value thresh.	No. of SNPS	aOR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	106	0.82	0.34	1.97	0.66	1.15	0.08	16.1	0.92	0.80	0.35	1.82	0.59
$1*10^{-6}$	281	0.86	0.46	1.62	0.65	1.65	0.27	10.1	0.59	0.89	0.50	1.59	0.69
$1*10^{-5}$	509	1.26	0.73	2.18	0.41	1.43	0.34	5.92	0.63	1.29	0.77	2.14	0.33
$1*10^{-4}$	1,090	1.18	0.77	1.80	0.45	2.21	0.67	7.30	0.19	1.31	0.88	1.94	0.18
$1*10^{-3}$	2,760	0.92	0.67	1.25	0.58	2.20	0.68	7.10	0.19	0.99	0.74	1.32	0.93
0.01	8,097	1.08	0.88	1.32	0.48	1.32	0.61	2.85	0.48	1.10	0.90	1.34	0.36
0.05	18,452	1.16	0.98	1.38	0.08	0.97	0.58	1.62	0.91	1.14	0.97	1.34	0.10
0.1	26,621	1.20	1.03	1.40	0.02	1.07	0.66	1.73	0.78	1.19	1.03	1.38	0.02
0.2	37,560	1.17	1.01	1.36	0.03	1.05	0.66	1.67	0.84	1.17	1.02	1.34	0.03
0.5	57,013	1.16	1.01	1.33	0.03	1.01	0.69	1.49	0.96	1.15	1.02	1.31	0.03
All	70,966	1.17	1.02	1.34	0.02	1.00	0.67	1.47	0.98	1.16	1.02	1.32	0.02
60 years or older at diagnosis		Caucasian subpopulation				Hispanic subpopulation				Total population			
P-value thresh.	No. of SNPS	aOR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	106	0.98	0.58	1.64	0.93	4.89	0.75	31.7	0.10	1.13	0.70	1.83	0.62
$1*10^{-6}$	281	1.19	0.82	1.72	0.35	1.71	0.49	5.93	0.40	1.23	0.87	1.74	0.24
$1*10^{-5}$	509	1.01	0.76	1.34	0.93	1.76	0.60	5.10	0.30	1.06	0.81	1.39	0.68
$1*10^{-4}$	1,090	1.09	0.87	1.37	0.44	2.29	0.97	5.41	0.06	1.17	0.94	1.44	0.15
$1*10^{-3}$	2,760	0.95	0.81	1.11	0.48	1.40	0.88	2.23	0.16	1.01	0.87	1.17	0.92

0.01	8,097	1.00	0.89	1.12	0.96	1.31	0.87	1.96	0.20	1.03	0.92	1.15	0.60
0.05	18,452	1.01	0.92	1.11	0.79	1.11	0.81	1.53	0.51	1.03	0.94	1.12	0.54
0.1	26,621	0.98	0.90	1.06	0.61	1.03	0.76	1.39	0.86	0.99	0.92	1.07	0.77
0.2	37,560	1.01	0.94	1.08	0.87	1.03	0.77	1.37	0.86	1.01	0.94	1.08	0.72
0.5	57,013	1.01	0.94	1.08	0.76	1.00	0.75	1.33	1.00	1.01	0.95	1.08	0.68
All	70,966	1.01	0.94	1.08	0.81	1.02	0.78	1.33	0.90	1.01	0.95	1.08	0.72

The APOE risk is based rs429358 and rs7412, where each apoe4-allele increases risk (β: 1.327), while each apoe2-allele decreases risk (β:-0.598) compared to apoe3.
The logistic regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.
Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value; thresh.: threshold

Supplemental Table 4.2b Logistic regression for the association between AD-PRS and hallucinations among subjects stratified by the studies (PEG, PW and PPMI).

		Caucasian subpopulation				Hispanic subpopulation				Total population			
P-value thresh.	No. of SNPS	aOR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	39	1.12	0.97	1.29	0.12	0.81	0.44	1.49	0.50	1.11	0.97	1.27	0.14
$1*10^{-6}$	52	1.13	0.99	1.29	0.08	0.91	0.53	1.56	0.73	1.12	0.99	1.28	0.08
$1*10^{-5}$	84	1.12	0.98	1.28	0.09	0.96	0.58	1.59	0.87	1.12	0.99	1.27	0.08
$1*10^{-4}$	199	1.09	0.97	1.24	0.16	0.94	0.58	1.52	0.79	1.09	0.97	1.23	0.16
$1*10^{-3}$	810	1.11	1.00	1.23	0.04	0.93	0.66	1.30	0.66	1.09	0.99	1.20	0.08
0.01	4,523	1.05	0.98	1.12	0.16	0.94	0.76	1.16	0.56	1.03	0.97	1.09	0.36
0.05	14,670	1.03	0.98	1.07	0.21	0.99	0.87	1.12	0.82	1.02	0.98	1.06	0.40
0.1	23,504	1.02	0.98	1.06	0.31	0.99	0.90	1.10	0.85	1.01	0.98	1.04	0.60
0.2	36,706	1.01	0.98	1.04	0.60	0.98	0.90	1.07	0.63	1.00	0.98	1.03	0.92
0.5	60,898	1.00	0.98	1.03	0.73	0.97	0.90	1.05	0.43	1.00	0.97	1.02	0.84
All	77,885	1.00	0.98	1.03	0.76	0.97	0.89	1.04	0.38	1.00	0.97	1.02	0.78
APOE risk	2	1.31	0.97	1.77	0.08	0.97	0.29	3.27	0.97	1.29	0.97	1.72	0.08
Less than 60 years at diagnosis		Caucasian subpopulation				Hispanic subpopulation				Total population			
P-value thresh.	No. of SNPS	aOR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	39	0.77	0.57	1.05	0.10	1.42	0.41	4.91	0.58	0.78	0.58	1.05	0.10
$1*10^{-6}$	52	0.80	0.59	1.07	0.14	1.54	0.48	4.89	0.47	0.82	0.62	1.08	0.15
$1*10^{-5}$	84	0.77	0.58	1.04	0.09	1.67	0.53	5.21	0.38	0.80	0.61	1.06	0.12
$1*10^{-4}$	199	0.77	0.58	1.02	0.07	1.05	0.42	2.64	0.92	0.77	0.59	1.00	0.05
$1*10^{-3}$	810	0.89	0.71	1.11	0.30	1.04	0.56	1.94	0.89	0.87	0.71	1.07	0.20
0.01	4,523	0.92	0.80	1.06	0.25	1.09	0.74	1.62	0.65	0.93	0.82	1.05	0.23
0.05	14,670	0.93	0.84	1.02	0.14	1.04	0.79	1.36	0.78	0.94	0.86	1.02	0.12
0.1	23,504	0.95	0.87	1.03	0.23	1.02	0.84	1.25	0.82	0.95	0.89	1.02	0.19
0.2	36,706	0.97	0.91	1.04	0.38	1.04	0.87	1.25	0.67	0.97	0.92	1.03	0.34
0.5	60,898	0.96	0.90	1.02	0.23	1.03	0.88	1.20	0.71	0.97	0.92	1.02	0.23
All	77,885	0.96	0.90	1.02	0.21	1.03	0.89	1.20	0.68	0.97	0.92	1.02	0.21
APOE risk	2	0.83	0.45	1.55	0.56	2.65	0.30	23.1	0.38	0.87	0.48	1.57	0.64

60 years or older at diagnosis		Caucasian subpopulation				Hispanic subpopulation				Total population			
P-value thresh.	No. of SNPS	aOR	95% CI		P	aOR	95% CI		P	aOR	95% CI		P
$1*10^{-8}$	39	1.27	1.08	1.50	0.004	0.72	0.33	1.53	0.39	1.25	1.06	1.46	0.01
$1*10^{-6}$	52	1.27	1.08	1.48	0.003	0.81	0.40	1.61	0.54	1.25	1.08	1.45	0.004
$1*10^{-5}$	84	1.26	1.08	1.47	0.003	0.86	0.45	1.64	0.64	1.25	1.07	1.44	0.004
$1*10^{-4}$	199	1.21	1.05	1.40	0.01	0.96	0.52	1.75	0.88	1.21	1.06	1.39	0.01
$1*10^{-3}$	810	1.18	1.05	1.33	0.005	0.92	0.59	1.43	0.71	1.17	1.04	1.30	0.01
0.01	4,523	1.08	1.00	1.17	0.04	0.89	0.69	1.15	0.37	1.06	0.99	1.14	0.10
0.05	14,670	1.05	1.00	1.11	0.04	0.97	0.84	1.14	0.74	1.04	0.99	1.09	0.08
0.1	23,504	1.03	0.99	1.08	0.12	0.98	0.87	1.11	0.78	1.02	0.99	1.06	0.24
0.2	36,706	1.01	0.98	1.05	0.41	0.96	0.85	1.07	0.43	1.01	0.98	1.04	0.65
0.5	60,898	1.01	0.98	1.04	0.46	0.94	0.85	1.05	0.26	1.00	0.98	1.03	0.81
All	77,885	1.01	0.98	1.04	0.47	0.94	0.84	1.04	0.22	1.00	0.98	1.03	0.85
APOE risk	2	1.58	1.11	2.25	0.01	0.71	0.15	3.45	0.67	1.52	1.09	2.13	0.01

The logistic regression analyses were adjusted for sex, fractional ethnicity, age at diagnosis, disease duration, and study through the use of a propensity score.

Abbreviations: No.: Number; aOR: adjusted Odds Ratio; %: percentage; CI: Confidence Interval; P: P-value

4 Public Health Relevance

Parkinson's disease (PD) will become more prevalent in the next decade as the world's population ages.[1] Two symptoms in particular significantly decrease quality-of-life: hallucinations and dyskinesia. Hallucinations are an important co-morbidity and dyskinesia is a common treatment-related complication. This dissertation analyzed the genetic risk factors for both dyskinesia and hallucinations and identified specific genetic variants or combinations of genetic variants that are associated with an increase of these comorbidities.

Among the candidate genes (DRD1-3 and BDNF) for dyskinesia, the following were identified as genetic risk factors: several haplotypes in DRD2, possibly some haplotypes in DRD3, and the minor allele of rs6265 in BDNF. Among PD patients, there is a constant tradeoff between medication increases needed to address PD symptoms and increasing the risk of dyskinesia. Genetic information could help prevent or postpone this debilitating consequence of L-Dopa treatment and may improve patient-centered, personalized therapy. Future studies are needed to confirm our findings and quantify the health care benefits and risks of personalized treatments based on genetics.

Combining the results of the second and third study led to establishing that the genetic risk factors for hallucinations differ between young and older PD patients. Among young PD patients, the genetic architecture for schizophrenia appears to influence the risk for hallucinations, while for older PD patients, the genetic architecture for AD (especially the APOE-gene) is associated with an increased risk. Further research to decipher the difference between these two subpopulations would be highly recommend.

These results also suggest that the genetic risk factors for hallucinations are not the same as those for PD, although further study of certain genes (LRRK2, APOE, SLC6A4, BDNF and

MAPT) is recommended. The effort to create a PRS based on the candidate-genes was unsuccessful, potentially due to the limited sample size. Research into this possible PRS would benefit from adding additional larger studies with both genetic as well as phenotypic data.

The analysis for hallucinations were based on a Caucasian subpopulation and further analysis, including large GWAS of other ethnicities would be necessary to be able to generalize the data to other subpopulations. While sometimes a PRS can be used to predict the risk of developing a phenotype. Based on this research, the developed PRS's predictive qualities are suboptimal and preclude its use as a screening tool.

Overall, in this dissertation, it is demonstrated that certain genetic risk factors increased the chance of dyskinesia and hallucinations. These findings suggest certain underlying pathomechanisms and could identify subjects that are at an increased risk. Potentially in the future, when creating a personalized medical plan based on a person's genetic risk scores, this research will be able to help prevent or delay these serious comorbidities of hallucinations and dyskinesia in PD.

5 References

1. Dorsey, E. R. *et al.* Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology* **68**, 384–6 (2007).
2. Parkinson, J. An essay on the shaking palsy. 1817. *Sherwood, Neely and Jones, London* **14**, 223–36; discussion 222 (1817).
3. Hughes, A. J., Ben-Shlomo, Y., Daniel, S. E. & Lees, a J. What features improve the accuracy of clinical diagnosis in Parkinson's disease: a clinicopathologic study. *Neurology* **42**, 1142–1146 (1992).
4. Tysnes, O.-B. B. & Storstein, A. Epidemiology of Parkinson's disease. *J. Neural Transm.* **124**, 901–905 (2017).
5. Gazewood, J. D., Richards, D. R. & Clebak, K. Parkinson disease: an update. *Am. Fam. Physician* **87**, 267–73 (2013).
6. Postuma, R. B. *et al.* MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601 (2015).
7. Rizzo, G. *et al.* Accuracy of clinical diagnosis of Parkinson disease. *Neurology* **86**, 566–576 (2016).
8. Chahine, L. M., Stern, M. B. & Chen-Plotkin, A. Blood-based biomarkers for Parkinson's disease. *Parkinsonism Relat. Disord.* **20 Suppl 1**, S99-103 (2014).
9. Alonso-Navarro, H., Jimenez-Jimenez, F. J., Garcia-Martin, E. & Agundez, J. A. G. Genomic and pharmacogenomic biomarkers of Parkinson's disease. *Curr. Drug Metab.* **15**, 129–81 (2014).
10. Wu, X., Cai, H., Ge, R., Li, L. & Jia, Z. Recent Progress of Imaging Agents for Parkinson's Disease. *Curr. Neuropharmacol.* **12**, 551–563 (2015).

11. Delenclos, M., Jones, D. R., McLean, P. J. & Uitti, R. J. Biomarkers in Parkinson's disease: Advances and strategies. *Park. Relat. Disord.* **22**, S106–S110 (2016).

12. Beitz, J. M. Parkinson's disease: a review. *Front. Biosci. (Schol. Ed).* **6**, 65–74 (2014).

13. Varma, D. & Sen, D. Role of the unfolded protein response in the pathogenesis of Parkinson's disease. *Acta Neurobiol. Exp. (Wars).* **75**, 1–26 (2015).

14. Schapira, A. H. & Jenner, P. Etiology and pathogenesis of Parkinson's disease. *Mov. Disord.* **26**, 1049–1055 (2011).

15. Fujita, K. a. *et al.* Integrating pathways of parkinson's disease in a molecular interaction map. *Mol. Neurobiol.* **49**, 88–102 (2014).

16. Blesa, J. & Przedborski, S. Parkinsonâ€TMs disease: animal models and dopaminergic cell vulnerability. *Front. Neuroanat.* **8**, 155 (2014).

17. Missale, C., Nash, S. R., Robinson, S. W., Jaber, M. & Caron, M. G. Dopamine receptors: from structure to function. *Physiol. Rev.* **78**, 189–225 (1998).

18. Magrinelli, F. *et al.* Pathophysiology of Motor Dysfunction in Parkinson's Disease as the Rationale for Drug Treatment and Rehabilitation. *Parkinsons. Dis.* **2016**, 1–18 (2016).

19. Surmeier, D. J., Obeso, J. A. & Halliday, G. M. Selective neuronal vulnerability in Parkinson disease. *Nat. Rev. Neurosci.* **18**, 101–113 (2017).

20. Sulzer, D. & Surmeier, D. J. Neuronal vulnerability, pathogenesis, and Parkinson's disease. *Mov. Disord.* **28**, 715–724 (2013).

21. Goedert, M., Spillantini, M. G., Del Tredici, K. & Braak, H. 100 years of Lewy pathology. *Nat Rev Neurol* **9**, 13–24 (2013).

22. Moustafa, A. A. *et al.* Motor symptoms in Parkinson's disease: A unified framework. *Neurosci. Biobehav. Rev.* **68**, 727–740 (2016).

23. Sveinbjornsdottir, S. The clinical symptoms of Parkinson's disease. *J. Neurochem.* **139**, 318–324 (2016).

24. Sherer, T. B., Chowdhury, S., Peabody, K. & Brooks, D. W. Overcoming obstacles in Parkinson's disease. *Mov. Disord.* **27**, 1606–11 (2012).

25. Hirsch, L., Jette, N., Frolkis, A., Steeves, T. & Pringsheim, T. The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Neuroepidemiology* **46**, 292–300 (2016).

26. Pringsheim, T., Jette, N., Frolkis, A. & Steeves, T. D. L. The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Mov. Disord.* **29**, 1583–1590 (2014).

27. Van Den Eeden, S. K. S. *et al.* Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity. *Am. J. Epidemiol.* **157**, 1015–1022 (2003).

28. Dexter, D. T. & Jenner, P. Parkinson disease: from pathology to molecular disease mechanisms. *Free Radic. Biol. Med.* **62**, 132–44 (2013).

29. De Rosa, P., Marini, E. S., Gelmetti, V. & Valente, E. M. Candidate genes for Parkinson disease: Lessons from pathogenesis. *Clin. Chim. Acta.* **449**, 68–76 (2015).

30. Ritz, B. R., Paul, K. C. & Bronstein, J. M. Of Pesticides and Men: a California Story of Genes and Environment in Parkinson's Disease. *Curr. Environ. Heal. reports* **3**, 40–52 (2016).

31. Lee, A. & Gilbert, R. M. Epidemiology of Parkinson Disease. *Neurol. Clin.* **34**, 955–965 (2016).

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

33. Nandipati, S. & Litvan, I. Environmental Exposures and Parkinson's Disease. *Int. J.*

Environ. Res. Public Health **13**, 881 (2016).

34. Ritz, B., Lee, P.-C., Lassen, C. F. & Arah, O. A. Parkinson disease and smoking revisited: ease of quitting is an early sign of the disease. *Neurology* **83**, 1396–402 (2014).

35. Bellou, V., Belbasis, L., Tzoulaki, I., Evangelou, E. & Ioannidis, J. P. A. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Park. Relat. Disord.* **23**, 1–9 (2016).

36. Kalinderi, K., Bostantjopoulou, S. & Fidani, L. The genetic background of Parkinson's disease: Current progress and future prospects. *Acta Neurol. Scand.* **134**, 314–326 (2016).

37. Kalia, L. V. & Lang, A. E. Parkinson's disease. *Lancet* **386**, 896–912 (2015).

38. Lill, C. M. Genetics of Parkinson's disease. *Mol. Cell. Probes* **30**, 386–396 (2016).

39. Berg, D. *et al.* MDS research criteria for prodromal Parkinson's disease. *Mov. Disord.* **30**, 1600–1611 (2015).

40. Walter, U. *et al.* Potential impact of self-perceived prodromal symptoms on the early diagnosis of Parkinson's disease. *J. Neurol.* **260**, 3077–85 (2013).

41. Palma, J.-A. & Kaufmann, H. Autonomic disorders predicting Parkinson's disease. *Parkinsonism Relat. Disord.* **20**, S94–S98 (2014).

42. Lin, C.-H., Lin, J.-W., Liu, Y.-C., Chang, C.-H. & Wu, R.-M. Risk of Parkinson's disease following severe constipation: A nationwide population-based cohort study. *Parkinsonism Relat. Disord.* **20**, 1371–1375 (2014).

43. Jankovic, J. Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry* **79**, 368–376 (2008).

44. Bjornestad, A. *et al.* Risk and course of motor complications in a population-based incident Parkinson's disease cohort. *Park. Relat. Disord.* **22**, 48–53 (2016).

45. Poewe, W. The natural history of Parkinson's disease. *J. Neurol.* **253 Suppl**, VII2-6 (2006).

46. Poewe, W. & Mahlknecht, P. The clinical progression of Parkinson's disease. *Parkinsonism Relat. Disord.* **15 Suppl 4**, S28-32 (2009).

47. Evans, J. R. *et al.* The natural history of treated Parkinson's disease in an incident, community based cohort. *J. Neurol. Neurosurg. Psychiatry* **82**, 1112–8 (2011).

48. Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression and mortality. *Neurology* **17**, 427–42 (1967).

49. Macleod, A. D., Taylor, K. S. M. & Counsell, C. E. Mortality in Parkinson's disease: A systematic review and meta-analysis. *Mov. Disord.* **29**, 1615–1622 (2014).

50. Xu, J., Gong, D. D. D., Man, C. F. F. & Fan, Y. Parkinson's disease and risk of mortality: Meta-analysis and systematic review. *Acta Neurol. Scand.* **129**, 71–79 (2014).

51. Morgan, J. C. *et al.* Mortality in Levodopa-Treated Parkinson's Disease. *Parkinsons. Dis.* **2014**, 1–8 (2014).

52. De Lau, L. M. L., Verbaan, D., van Rooden, S. M., Marinus, J. & van Hilten, J. J. Relation of clinical subtypes in Parkinson's disease with survival. *Mov. Disord.* **29**, 150–151 (2014).

53. De Lau, L. M. L., Verbaan, D., Marinus, J. & van Hilten, J. J. Survival in Parkinson's disease. Relation with motor and non-motor features. *Park. Relat. Disord.* **20**, 613–616 (2014).

54. Pfeiffer, R. F. Non-motor symptoms in Parkinson's disease. *Park. Relat. Disord.* **22**, S119–S122 (2016).

55. Müller, B., Assmus, J., Herlofson, K., Larsen, J. P. & Tysnes, O.-B. Importance of motor

vs. non-motor symptoms for health-related quality of life in early Parkinson's disease. *Parkinsonism Relat. Disord.* **19**, 1027–32 (2013).

56. Duncan, G. W. *et al.* Health-related quality of life in early Parkinson's disease: the impact of nonmotor symptoms. *Mov. Disord.* **29**, 195–202 (2014).

57. Santos-García, D. & de la Fuente-Fernández, R. Impact of non-motor symptoms on health-related and perceived quality of life in Parkinson's disease. *J. Neurol. Sci.* **332**, 136–40 (2013).

58. Shearer, J., Green, C., Counsell, C. E. & Zajicek, J. P. The impact of motor and non motor symptoms on health state values in newly diagnosed idiopathic Parkinson's disease. *J. Neurol.* **259**, 462–468 (2012).

59. Global Parkinson's Disease Survey (GPDS) Steering Committee. Factors impacting on quality of life in Parkinson's disease: results from an international survey. *Mov. Disord.* **17**, 60–7 (2002).

60. Vijayakumar, D. & Jankovic, J. Drug-Induced Dyskinesia, Part 1: Treatment of Levodopa-Induced Dyskinesia. *Drugs* **76**, 759–777 (2016).

61. Zesiewicz, T. A., Sullivan, K. L. & Hauser, R. A. Levodopa-induced dyskinesia in Parkinson's disease: epidemiology, etiology, and treatment. *Curr. Neurol. Neurosci. Rep.* **7**, 302–10 (2007).

62. Calabresi, P., Filippo, M. Di, Ghiglieri, V., Tambasco, N. & Picconi, B. Levodopa-induced dyskinesias in patients with Parkinson's disease: Filling the bench-to-bedside gap. *Lancet Neurol.* **9**, 1106–1117 (2010).

63. Scott, N. W., Macleod, A. D. & Counsell, C. E. Motor complications in an incident Parkinson's disease cohort. *Eur. J. Neurol.* **23**, 304–12 (2016).

64. Rascol, O. *et al.* A five-year study of the incidence of dyskinesia in patients with early Parkinson's disease who were treated with ropinirole or levodopa. *N. Engl. J. Med.* **342**, 1484–91 (2000).

65. Fabbrini, G., Brotchie, J. M., Grandas, F., Nomoto, M. & Goetz, C. G. Levodopa-induced dyskinesias. *Mov. Disord.* **22**, 1379–89 (2007).

66. Lee, J.-Y., Cho, J., Lee, E.-K., Park, S.-S. & Jeon, B. S. Differential genetic susceptibility in diphasic and peak-dose dyskinesias in Parkinson's disease. *Mov. Disord.* **26**, 73–9 (2011).

67. Sharma, J. C., Bachmann, C. G. & Linazasoro, G. Classifying risk factors for dyskinesia in Parkinson's disease. *Parkinsonism Relat. Disord.* **16**, 490–7 (2010).

68. Papapetropoulos, S. & Mash, D. C. Motor fluctuations and dyskinesias in advanced/end stage Parkinson's disease: a study from a population of brain donors. *J. Neural Transm.* **114**, 341–345 (2007).

69. Zappia, M. *et al.* Sex differences in clinical and genetic determinants of levodopa peak-dose dyskinesias in Parkinson disease: an exploratory study. *Arch. Neurol.* **62**, 601–5 (2005).

70. Hauser, R. A., McDermott, M. P. & Messing, S. Factors Associated With the Development of Motor Fluctuations and Dyskinesias in Parkinson Disease. *Arch. Neurol.* **63**, 1756 (2006).

71. Manson, A., Stirpe, P. & Schrag, A. Levodopa-induced-dyskinesias clinical features, incidence, risk factors, management and impact on quality of life. *J. Parkinsons. Dis.* **2**, 189–98 (2012).

72. Péchevis, M. *et al.* Effects of dyskinesias in Parkinson's disease on quality of life and

health-related costs: a prospective European study. *Eur. J. Neurol.* **12**, 956–63 (2005).

73. Chapuis, S., Ouchchane, L., Metz, O., Gerbaud, L. & Durif, F. Impact of the motor complications of Parkinson's disease on the quality of life. *Mov. Disord.* **20**, 224–30 (2005).

74. Jenkinson, P. M., Edelstyn, N. M. J., Stephens, R. & Ellis, S. J. Why Are Some Parkinson Disease Patients Unaware of Their Dyskinesias? *Cogn. Behav. Neurol.* **22**, 117–121 (2009).

75. Pietracupa, S., Latorre, A., Berardelli, A. & Fabbrini, G. Parkinsonian Patients and Poor Awareness of Dyskinesias. *Front. Neurol.* **5**, 32 (2014).

76. Friedman, J. H. Misperceptions and Parkinson's disease. *J. Neurol. Sci.* **374**, 42–46 (2017).

77. Haaxma, C. A. *et al.* Risk of disabling response fluctuations and Dyskinesias for Dopamine Agonists Versus Levodopa in Parkinson's Disease1. *J. Parkinsons. Dis.* **5**, 847–853 (2015).

78. PD Med Collaborative Group *et al.* Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): a large, open-label, pragmatic randomised trial. *Lancet (London, England)* **384**, 1196–205 (2014).

79. Xie, C. *et al.* Levodopa alone compared with levodopa-sparing therapy as initial treatment for Parkinson's disease: a meta-analysis. *Neurol. Sci.* **36**, 1319–1329 (2015).

80. Nice. PARKINSON'S DISEASE National clinical guideline for diagnosis and management in primary and secondary care.

81. Metman, L. V *et al.* Amantadine for levodopa-induced dyskinesias: a 1-year follow-up

study. *Arch. Neurol.* **56**, 1383–6 (1999).

82. Rascol, O., Perez-Lloret, S. & Ferreira, J. J. New treatments for levodopa-induced motor complications. *Mov. Disord.* **30**, 1451–1460 (2015).

83. Munhoz, R. P., Cerasa, A. & Okun, M. S. Surgical treatment of dyskinesia in Parkinson's disease. *Front. Neurol.* **5**, 65 (2014).

84. Krack, P. *et al.* From off-period dystonia to peak-dose chorea. The clinical spectrum of varying subthalamic nucleus activity. *Brain* 1133–46 (1999).

85. DeLong, M. R. & Wichmann, T. Basal Ganglia Circuits as Targets for Neuromodulation in Parkinson Disease. *JAMA Neurol.* **72**, 1354–60 (2015).

86. Darmopil, S., Martín, A. B., De Diego, I. R., Ares, S. & Moratalla, R. Genetic inactivation of dopamine D1 but not D2 receptors inhibits L-DOPA-induced dyskinesia and histone activation. *Biol. Psychiatry* **66**, 603–13 (2009).

87. Bravo, S. A., Rangel-barajas, C. & Garduno, B. F. Pathophysiology of L-Dopa Induced Dyskinesia — Changes in D1 / D3 Receptors and Their Signaling Pathway. *A Synopsis Park. Dis.* **46425**, 1–28 (2014).

88. Obeso, J. A. *et al.* Pathophysiology of the basal ganglia in Parkinson's disease. *Trends Neurosci.* **23**, S8-19 (2000).

89. Olanow, C. W., Obeso, J. A. & Stocchi, F. Continuous dopamine-receptor treatment of Parkinson's disease: scientific rationale and clinical implications. *Lancet Neurol.* **5**, 677–687 (2006).

90. Jiménez-Urbieta, H. *et al.* Dyskinesias and impulse control disorders in Parkinson's disease: From pathogenesis to potential therapeutic approaches. *Neurosci. Biobehav. Rev.* **56**, 294–314 (2015).

91. Cote, S. R., Chitravanshi, V. C., Bleickardt, C., Sapru, H. N. & Kuzhikandathil, E. V. Overexpression of the dopamine D3 receptor in the rat dorsal striatum induces dyskinetic behaviors. *Behav. Brain Res.* **263**, 46–50 (2014).

92. Visanji, N. P. *et al.* Dopamine D3 receptor stimulation underlies the development of L-DOPA-induced dyskinesia in animal models of Parkinson's disease. *Neurobiol. Dis.* **35**, 184–192 (2009).

93. Bézard, E. *et al.* Attenuation of levodopa-induced dyskinesia by normalizing dopamine D3 receptor function. *Nat. Med.* **9**, 762–767 (2003).

94. Payer, D. E. *et al.* D3 dopamine receptor-preferring [11C]PHNO PET imaging in Parkinson patients with dyskinesia. *Neurology* **86**, 224–230 (2016).

95. Kalinderi, K., Fidani, L., Katsarou, Z. & Bostantjopoulou, S. Pharmacological treatment and the prospect of pharmacogenetics in Parkinson's disease. *Int. J. Clin. Pract.* **65**, 1289–94 (2011).

96. Droździk, M., Białecka, M. & Kurzawski, M. Pharmacogenetics of Parkinson's disease - through mechanisms of drug actions. *Curr. Genomics* **14**, 568–77 (2013).

97. Fumagalli, F., Racagni, G. & Riva, M. A. Shedding light into the role of BDNF in the pharmacotherapy of Parkinson's disease. *Pharmacogenomics J.* **6**, 95–104 (2006).

98. Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V. & Di Filippo, M. Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nat. Neurosci.* **17**, 1022–1030 (2014).

99. Greenbaum, L. *et al.* Do tardive dyskinesia and L-dopa induced dyskinesia share common genetic risk factors? An exploratory study. *J. Mol. Neurosci.* **51**, 380–8 (2013).

100. Kaplan, N. *et al.* Sequence variants in SLC6A3, DRD2, and BDNF genes and time to levodopa-induced dyskinesias in Parkinson's disease. *J. Mol. Neurosci.* **53**, 183–188

(2014).

101. Oliveri, R. L. *et al.* Dopamine D2 receptor gene polymorphism and the risk of levodopa-induced dyskinesias in PD. *Neurology* **53**, 1425–30 (1999).

102. Kaiser, R. *et al.* L-Dopa-induced adverse effects in PD and dopamine transporter gene polymorphism. *Neurology* **60**, 1750–1755 (2003).

103. Rieck, M. *et al.* DRD2 haplotype is associated with dyskinesia induced by levodopa therapy in Parkinson's disease patients. *Pharmacogenomics* **13**, 1701–10 (2012).

104. Strong, J. A. *et al.* Genotype and smoking history affect risk of levodopa-induced dyskinesias in Parkinson's disease. *Mov. Disord.* **21**, 654–9 (2006).

105. Paus, S. *et al.* Motor complications in patients form the German Competence Network on Parkinson's disease and the DRD3 Ser9Gly polymorphism. *Mov. Disord.* **24**, 1080–4 (2009).

106. Cheshire, P. *et al.* Influence of single nucleotide polymorphisms in COMT, MAO-A and BDNF genes on dyskinesias and levodopa use in Parkinson's disease. *Neurodegener. Dis.* **13**, 24–28 (2013).

107. Foltynie, T. *et al.* BDNF val66met influences time to onset of levodopa induced dyskinesia in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **80**, 141–4 (2009).

108. Fenelon, G., Mahieux, F., Huon, R. & Ziégler, M. Hallucinations in Parkinson's disease: Prevalence, phenomenology and risk factors. *Brain* **123**, 733–745 (2000).

109. Aarsland, D. & Kramberger, M. G. Neuropsychiatric Symptoms in Parkinson's Disease. *J. Parkinsons. Dis.* **5**, 659–667 (2015).

110. Levin, J., Hasan, A. & Höglinger, G. U. Psychosis in Parkinson's disease: identification, prevention and treatment. *J. Neural Transm.* **123**, 45–50 (2016).

111. Forsaa, E. B. *et al.* A 12-Year Population-Based Study of Psychosis in Parkinson Disease. *Arch. Neurol.* **67**, 996–1001 (2010).

112. Friedman, A. & Sienkiewicz, J. Psychotic complications of long-term levodopa treatment of Parkinson's disease. *Acta Neurol. Scand.* **84**, 111–3 (1991).

113. Meco, G., Bonifati, V., Cusimano, G., Fabrizio, E. & Vanacore, N. Hallucinations in Parkinson disease: neuropsychological study. *Ital. J. Neurol. Sci.* **11**, 373–9 (1990).

114. Bjornestad, A., Pedersen, K. F., Tysnes, O.-B. & Alves, G. Clinical milestones in Parkinson's disease: A 7-year population-based incident cohort study. *Parkinsonism Relat. Disord.* **42**, 28–33 (2017).

115. Graham, J. M., Grünewald, R. A. & Sagar, H. J. Hallucinosis in idiopathic Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **63**, 434–40 (1997).

116. Sanchez-Ramos, J. R., Ortoll, R. & Paulson, G. W. Visual hallucinations associated with Parkinson disease. *Arch. Neurol.* **53**, 1265–8 (1996).

117. Haeske-Dewick, H. C. Hallucinations in Parkinson's disease: Characteristics and associated clinical features. *Int. J. Geriatr. Psychiatry* **10**, 487–495 (1995).

118. Prange, S. *et al.* Age and time course of long-term motor and nonmotor complications in Parkinson disease. *Neurology* **92**, e148–e160 (2019).

119. Khan, A. U., Akram, M., Daniyal, M. & Zainab, R. Awareness and current knowledge of Parkinson's disease: a neurodegenerative disorder. *Int. J. Neurosci.* **129**, 55–93 (2019).

120. Ffytche, D. H. *et al.* The psychosis spectrum in Parkinson disease. *Nat. Rev. Neurol.* **13**, 81–95 (2017).

121. Borovac, J. A. Side effects of a dopamine agonist therapy for Parkinson's disease: a mini-review of clinical pharmacology. *Yale J. Biol. Med.* **89**, 37–47 (2016).

122. Pagonabarraga, J. *et al.* Minor hallucinations occur in drug-naive Parkinson's disease patients, even from the premotor phase. *Mov. Disord.* **31**, 45–52 (2016).

123. Ffytche, D. H. *et al.* Risk factors for early psychosis in PD: Insights from the Parkinson's Progression markers initiative. *J. Neurol. Neurosurg. Psychiatry* **88**, 325–331 (2017).

124. Wolters, E. C. Dopaminomimetic psychosis in Parkinson's disease patients: diagnosis and treatment. *Neurology* **52**, S10-3 (1999).

125. Perry, E. K. *et al.* Evidence of a monoaminergic-cholinergic imbalance related to visual hallucinations in Lewy body dementia. *J. Neurochem.* **55**, 1454–6 (1990).

126. Whitehouse, P. J. Clinical and neurochemical consequences of neuronal loss in the nucleus basalis of Meynert in Parkinson's disease and Alzheimer's disease. *Adv. Neurol.* **45**, 393–7 (1987).

127. Gallagher, D. A. *et al.* Testing an aetiological model of visual hallucinations in Parkinson's disease. *Brain* **134**, 3299–3309 (2011).

128. Goldman, J. G. *et al.* Visuoperceptive region atrophy independent of cognitive status in patients with Parkinson's disease with hallucinations. *Brain* **137**, 849–859 (2014).

129. Bowman, A. R., Bruce, V., Colbourn, C. J. & Collerton, D. Compensatory shifts in visual perception are associated with hallucinations in Lewy body disorders. *Cogn. Res. Princ. Implic.* **2**, 26 (2017).

130. Shine, J. M. *et al.* Abnormal connectivity between the default mode and the visual system underlies the manifestation of visual hallucinations in Parkinson's disease: a task-based fMRI study. *npj Park. Dis.* **1**, 15003 (2015).

131. Shine, J. M., Halliday, G. M., Naismith, S. L. & Lewis, S. J. G. Visual misperceptions and hallucinations in Parkinson's disease: dysfunction of attentional control networks? *Mov.*

Disord. **26**, 2154–9 (2011).

132. Damasceno dos Santos, E. U. *et al.* Pharmacogenetic Profile and the Occurrence of Visual Hallucinations in Patients With Sporadic Parkinson's Disease. *J. Clin. Pharmacol.* (2019). doi:10.1002/jcph.1394

133. Brockmann, K. *et al.* SNPs in Aβ clearance proteins. *Neurology* **89**, 2335–2340 (2017).

134. Bäckström, D. C. *et al.* Cerebrospinal Fluid Patterns and the Risk of Future Dementia in Early, Incident Parkinson Disease. *JAMA Neurol.* **72**, 1175 (2015).

135. Watanabe, H. *et al.* Cortical and subcortical brain atrophy in Parkinson's disease with visual hallucination. *Mov. Disord.* **28**, 1732–1736 (2013).

136. Gama, R. L. *et al.* Structural brain abnormalities in patients with Parkinson's disease with visual hallucinations: A comparative voxel-based analysis. *Brain Cogn.* **87**, 97–103 (2014).

137. Alves, G. *et al.* CSF A 42 predicts early-onset dementia in Parkinson disease. *Neurology* **82**, 1784–1790 (2014).

138. Taddei, R. N., Cankaya, S., Dhaliwal, S. & Chaudhuri, K. R. Management of Psychosis in Parkinson's Disease: Emphasizing Clinical Subtypes and Pathophysiological Mechanisms of the Condition. *Parkinsons. Dis.* **2017**, 1–18 (2017).

139. Braak, H., Rub, U., Jansen Steur, E. N. H., Del Tredici, K. & de Vos, R. A. I. Cognitive status correlates with neuropathologic stage in Parkinson disease. *Neurology* **64**, 1404–1410 (2005).

140. Aarsland, D., Perry, R., Brown, A., Larsen, J. P. & Ballard, C. Neuropathology of dementia in Parkinson's disease: A prospective, community-based study. *Ann. Neurol.* **58**, 773–776 (2005).

141. Ibarretxe-Bilbao, N. *et al.* Hippocampal head atrophy predominance in Parkinson's disease with hallucinations and with dementia. *J. Neurol.* **255**, 1324–1331 (2008).

142. Ramírez-Ruiz, B. *et al.* Cerebral atrophy in Parkinson's disease patients with visual hallucinations. *Eur. J. Neurol.* **14**, 750–756 (2007).

143. Meppelink, A. M., de Jong, B. M., Teune, L. K. & van Laar, T. Regional cortical grey matter loss in Parkinson's disease without dementia is independent from visual hallucinations. *Mov. Disord.* **26**, 142–147 (2011).

144. Janzen, J. *et al.* The pedunculopontine nucleus is related to visual hallucinations in Parkinson's disease: preliminary results of a voxel-based morphometry study. *J. Neurol.* **259**, 147–54 (2012).

145. Compta, Y. *et al.* Cerebrospinal tau, phospho-tau, and beta-amyloid and neuropsychological functions in Parkinson's disease. *Mov. Disord.* **24**, 2203–2210 (2009).

146. Xu, Y., Yang, J. & Shang, H. Meta-analysis of risk factors for Parkinson's disease dementia. *Transl. Neurodegener.* **5**, 11 (2016).

147. Anang, J. B. M. *et al.* Predictors of dementia in Parkinson disease: a prospective cohort study. *Neurology* **83**, 1253–60 (2014).

148. Aarsland, D., Andersen, K., Larsen, J. P., Lolk, A. & Kragh-Sørensen, P. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Arch. Neurol.* **60**, 387–92 (2003).

149. Uc, E. Y. *et al.* Incidence of and risk factors for cognitive impairment in an early parkinson disease clinical trial cohort. *Neurology* **73**, 1469–1477 (2009).

150. Aarsland, D., Larsen, J. P., Tandberg, E. & Laake, K. Predictors of nursing home placement in Parkinson's disease: a population-based, prospective study. *J. Am. Geriatr.*

Soc. **48**, 938–42 (2000).

151. Goetz, C. G. & Stebbins, G. T. Risk factors for nursing home placement in advanced Parkinson's disease. *Neurology* **43**, 2227–9 (1993).

152. Goetz, C. G. & Stebbins, G. T. Mortality and hallucinations in nursing home patients with advanced Parkinson ' s disease. *Int. Classif.* 669–671 (1995).

153. Forsaa, E. B., Larsen, J. P. & Alves, G. What predicts mortality in Parkinson disease ? (2010).

154. Schrag, A., Hovris, A., Morley, D., Quinn, N. & Jahanshahi, M. Caregiver-burden in parkinson's disease is closely associated with psychiatric symptoms, falls, and disability. *Parkinsonism Relat. Disord.* **12**, 35–41 (2006).

155. Goldman, J. G. & Holden, S. Treatment of psychosis and dementia in Parkinson's disease. *Curr. Treat. Options Neurol.* **16**, 281 (2014).

156. Gustaw Rothenberg, K. & Rajaram, R. Advances in Management of Psychosis in Neurodegenerative Diseases. *Curr Treat Options Neurol* **21**, 3 (1940).

157. Wang, J., Zhao, C., Chen, B. & Liu, Z.-L. Polymorphisms of dopamine receptor and transporter genes and hallucinations in Parkinson's disease. *Neurosci. Lett.* **355**, 193–6 (2004).

158. Ballanger, B. *et al.* Serotonin 2A Receptors and Visual Hallucinations in Parkinson Disease. *Arch. Neurol.* **67**, 416–21 (2010).

159. Kiferle, L. *et al.* Visual hallucinations in Parkinson's disease are not influenced by polymorphisms of serotonin 5-HT2A receptor and transporter genes. *Neurosci. Lett.* **422**, 228–231 (2007).

160. Goetz, C. G. *et al.* Genetic variation analysis in parkinson disease patients with and

without hallucinations: case-control study. *Arch. Neurol.* **58**, 209–13 (2001).

161. Damasceno dos Santos, E. U. *et al.* Pharmacogenetic Profile and the Occurrence of Visual Hallucinations in Patients With Sporadic Parkinson's Disease. *J. Clin. Pharmacol.* (2019). doi:10.1002/jcph.1394

162. De Luca, V. *et al.* HOMER1 Promoter Analysis in Parkinson’s Disease: Association Study with Psychotic Symptoms. *Neuropsychobiology* **59**, 239–245 (2009).

163. Kasten, M. & Klein, C. The many faces of alpha-synuclein mutations. *Mov. Disord.* **28**, 697–701 (2013).

164. Goldman, J. G. *et al.* Racial differences may influence the role of cholecystokinin polymorphisms in Parkinson's disease hallucinations. *Mov. Disord.* **26**, 1781–1783 (2011).

165. Schumacher-Schuh, a F. *et al.* Association of common genetic variants of HOMER1 gene with levodopa adverse effects in Parkinson's disease patients. *Pharmacogenomics J.* **14**, 289–94 (2014).

166. Ferrari, M. *et al.* Polymorphisms of dopamine receptor genes and risk of visual hallucinations in Parkinson's patients. *Eur. J. Clin. Pharmacol.* **72**, 1335–1341 (2016).

167. Yahalom, G. *et al.* Carriers of both GBA and LRRK2 mutations, compared to carriers of either, in Parkinson's disease: Risk estimates and genotype-phenotype correlations. *Park. Relat. Disord.* 0–1 (2018). doi:10.1016/j.parkreldis.2018.12.014

168. Thaler, A. *et al.* A "dose" effect of mutations in the GBA gene on Parkinson's disease phenotype. *Parkinsonism Relat. Disord.* **36**, 47–51 (2017).

169. Bäckström, D. *et al.* PITX3 genotype and risk of dementia in Parkinson's disease: A population-based study. *J. Neurol. Sci.* **381**, 278–284 (2017).

170. Thaler, A. *et al.* Parkinson's disease phenotype is influenced by the severity of the mutations in the GBA gene. *Park. Relat. Disord.* **55**, 45–49 (2018).

171. De Marco, E. V. *et al.* Glucocerebrosidase gene mutations are associated with Parkinson's disease in southern Italy. *Mov. Disord.* **23**, 460–463 (2008).

172. Li, Y. *et al.* Clinicogenetic study of GBA mutations in patients with familial Parkinson's disease. *Neurobiol. Aging* **35**, 935.e3-935.e8 (2014).

173. Oeda, T. *et al.* Impact of glucocerebrosidase mutations on motor and nonmotor complications in Parkinson's disease. *Neurobiol. Aging* **36**, 3306–3313 (2015).

174. Neumann, J. *et al.* Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. *Brain* **132**, 1783–94 (2009).

175. Factor, S. A. *et al.* Disease-related and genetic correlates of psychotic symptoms in Parkinson's disease. *Mov. Disord.* **26**, 2190–2195 (2011).

176. Monsell, S. E. *et al.* Clinical and pathologic presentation in Parkinson's disease by apolipoprotein e4 allele status. *Parkinsonism Relat. Disord.* **20**, 503–507 (2014).

177. Camicioli, R. *et al.* Apolipoprotein E ε4 and catechol-O-methyltransferase alleles in autopsy-proven Parkinson's disease: Relationship to dementia and hallucinations. *Mov. Disord.* **20**, 989–994 (2005).

178. de la Fuente-Fernández, R., Núñez, M. A. & López, E. The apolipoprotein E epsilon 4 allele increases the risk of drug-induced hallucinations in Parkinson's disease. *Clin. Neuropharmacol.* **22**, 226–30

179. Papapetropoulos, S. *et al.* Phenotypic associations of tau and ApoE in Parkinson's disease. *Neurosci. Lett.* **414**, 141–144 (2007).

180. Pascale, E. *et al.* Genetic polymorphism of Angiotensin-Converting Enzyme is not

associated with the development of Parkinson's disease and of l-dopa-induced adverse effects. *J. Neurol. Sci.* **276**, 18–21 (2009).

181. Lin, J.-J., Yueh, K.-C., Lin, S.-Z., Harn, H.-J. & Liu, J.-T. Genetic polymorphism of the angiotensin converting enzyme and l-dopa-induced adverse effects in Parkinson's disease. *J. Neurol. Sci.* **252**, 130–134 (2007).

182. Fujii, C. *et al.* Association between polymorphism of the cholecystokinin gene and idiopathic Parkinson's disease. *Clin. Genet.* **56**, 394–9 (1999).

183. Makoff, A. J. *et al.* Association study of dopamine receptor gene polymorphisms with drug-induced hallucinations in patients with idiopathic Parkinson's disease. *Pharmacogenetics* **10**, 43–8 (2000).

184. Schumacher-Schuh, A. F. *et al.* Polymorphisms in the dopamine transporter gene are associated with visual hallucinations and levodopa equivalent dose in Brazilians with Parkinson's disease. *Int. J. Neuropsychopharmacol.* **16**, 1251–1258 (2013).

185. Creese, B. *et al.* Determining the Association of the 5HTTLPR Polymorphism with Delusions and Hallucinations in Lewy Body Dementias. *Am. J. Geriatr. Psychiatry* **22**, 580–586 (2014).

186. Feldman, B., Chapman, J. & Korczyn, A. D. Apolipoprotein e4 advances appearance of psychosis in patients with Parkinson's disease. *Acta Neurol. Scand.* **113**, 14–17 (2006).

187. Goldman, J. G., Goetz, C. G., Berry-Kravis, E., Leurgans, S. & Zhou, L. Genetic Polymorphisms in Parkinson Disease Subjects With and Without Hallucinations. *Arch. Neurol.* **61**, 1280–4 (2004).

188. Redenšek, S. *et al.* Dopaminergic Pathway Genes Influence Adverse Events Related to Dopaminergic Treatment in Parkinson's Disease. *Front. Pharmacol.* **10**, 8 (2019).

189. Lenka, A., Arumugham, S. S., Christopher, R. & Pal, P. K. Genetic substrates of psychosis in patients with Parkinson's disease: A critical review. *J. Neurol. Sci.* **364**, 33–41 (2016).

190. Corvol, J.-C. & Poewe, W. Pharmacogenetics of Parkinson's Disease in Clinical Practice. *Mov. Disord. Clin. Pract.* **4**, 173–180 (2017).

191. Wang, J., Si, Y.-M., Liu, Z.-L. & Yu, L. Cholecystokinin, cholecystokinin-A receptor and cholecystokinin-B receptor gene polymorphisms in Parkinson's disease. *Pharmacogenetics* **13**, 365–9 (2003).

192. Wang, J., Liu, Z. L. & Chen, B. Association study of dopamine D2, D3 receptor gene polymorphisms with motor fluctuations in PD. *Neurology* **56**, 1757–9 (2001).

193. Comi, C. *et al.* Polymorphisms of Dopamine Receptor Genes and Risk of L-Dopa–Induced Dyskinesia in Parkinson's Disease. *Int. J. Mol. Sci.* **18**, 242 (2017).

194. Ritz, B., Rhodes, S. L., Bordelon, Y. & Bronstein, J. α-Synuclein genetic variants predict faster motor symptom progression in idiopathic Parkinson disease. *PLoS One* **7**, e36199 (2012).

195. Tomlinson, C. L. *et al.* Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov. Disord.* **25**, 2649–53 (2010).

196. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–5 (2005).

197. Gabriel, S. B. The Structure of Haplotype Blocks in the Human Genome. *Science (80-.).* **296**, 2225–2229 (2002).

198. Stephens, M., Smith, N. J. & Donnelly, P. A New Statistical Method for Haplotype Reconstruction from Population Data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).

199. Fang, J. Using SAS ® Procedures FREQ , GENMOD , LOGISTIC , and PHREG to Estimate Adjusted Relative Risks – A Case Study. SAS Global Forum 2011. *SAS Global Forum 2011* 345 (2011).

200. Eisenstein, S. A. *et al.* Prediction of striatal D2 receptor binding by DRD2/ANKK1 TaqIA allele status. *Synapse* **70**, 418–31 (2016).

201. Lundstrom, K. & Turpin, M. P. Proposed Schizophrenia-Related Gene Polymorphism : Expression of the Ser9Gly Mutant Human Dopamine D 3 Receptor with the Semliki Forest Virus System. *Biochem. Biophys. Res. Commun.* **225**, 1068–1072 (1996).

202. Gelb, D. J., Oliver, E. & Gilman, S. Diagnostic criteria for Parkinson disease. *Arch. Neurol.* **56**, 33–9 (1999).

203. Hamza, T. H. & Payami, H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. *J. Hum. Genet.* **55**, 241 (2010).

204. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).

205. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516 (2017).

206. Thaler, A. *et al.* Parkinson's disease phenotype is influenced by the severity of the mutations in the GBA gene. *Parkinsonism Relat. Disord.* **55**, 45–49 (2018).

207. Kaiser, R. *et al.* L -dopa-induced adverse effects in PD and dopamine transporter gene polymorphism. *Neurology* **60**, 1750–5 (2003).

208. Escott-Price, V. *et al.* Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann. Neurol.* **77**, 582–91 (2015).

209. Ibanez, L. *et al.* Parkinson disease polygenic risk score is associated with Parkinson

disease status and age at onset but not with alpha-synuclein cerebrospinal fluid levels. *BMC Neurol.* **17**, 198 (2017).

210. Paul, K. C., Schulz, J., Bronstein, J. M., Lill, C. M. & Ritz, B. R. Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. *JAMA Neurol.* **75**, 360–366 (2018).

211. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

212. International Schizophrenia Consortium, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–52 (2009).

213. Gottesman, I. I. & Shields, J. A polygenic theory of schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 199–205 (1967).

214. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–52 (2009).

215. Huo, Y., Li, S., Liu, J., Li, X. & Luo, X. J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat. Commun.* **10**, (2019).

216. Escott-Price, V., Myers, A. J., Huentelman, M. & Hardy, J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. *Ann. Neurol.* **82**, 311–314 (2017).

217. Khera, A. V. *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* **177**, 587-596.e9 (2019).

218. Dron, J. S. & Hegele, R. A. The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr. Opin. Lipidol.* **30**, 71–81 (2019).

219. Spiliopoulou, A. *et al.* Genomic prediction of complex human traits: relatedness, trait architecture and predictive meta-models. *Hum. Mol. Genet.* **24**, 4167–4182 (2015).

220. Elliott, M. L. *et al.* A Polygenic Score for Higher Educational Attainment is Associated with Larger Brains. *Cereb. Cortex* (2018). doi:10.1093/cercor/bhy219

221. Stapleton, C. P. *et al.* Polygenic risk score as a determinant of risk of non-melanoma skin cancer in a European-descent renal transplant cohort. *Am. J. Transplant.* **19**, 801–810 (2019).

222. Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).

223. Leonenko, G. *et al.* Polygenic risk and hazard scores for Alzheimer's disease prediction. *Ann. Clin. Transl. Neurol.* **6**, 456–465 (2019).

224. Escott-Price, V. *et al.* Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* **138**, 3673–84 (2015).

225. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genet.* **9**, e1003348 (2013).

226. Dudbridge, F., Pashayan, N. & Yang, J. Predictive accuracy of combined genetic and environmental risk scores. *Genet. Epidemiol.* **42**, 4–19 (2018).

227. Khera, A. V *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* doi:10.1038/s41588-018-0183-z

228. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

229. Dai, X., Wiernek, S., Evans, J. P. & Runge, M. S. Genetics of coronary artery disease and myocardial infarction. *World J. Cardiol.* **8**, 1–23 (2016).

230. Darst, B. F. *et al.* Pathway-specific polygenic risk scores as predictors of β-amyloid

deposition and cognitive function in a sample at increased risk for Alzheimer's disease HHS Public Access. *J Alzheimers Dis* **55**, 473–484 (2017).

231. Docherty, A. R. *et al.* Pathway-based polygene risk for severe depression implicates drug metabolism in CONVERGE. *Psychol. Med.* 1–6 (2019). doi:10.1017/S0033291719000618

232. Holmans, P. *et al.* A pathway-based analysis provides additional support for an immune-related genetic susceptibility to Parkinson's disease. *Hum. Mol. Genet.* **22**, 1039–1049 (2013).

233. Richardson, T. G., Harrison, S., Hemani, G. & Davey Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* **8**, 1–24 (2019).

234. Palmer, T. M. *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* **21**, 223–242 (2012).

235. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genet. Epidemiol.* **37**, 658–665 (2013).

236. Richards, A. *et al.* Associations between schizophrenia genetic risk, anxiety disorders and manic/hypomanic episode in a longitudinal population cohort study. *Br. J. Psychiatry* **214**, 96–102 (2019).

237. DeMichele-Sweet, M. A. A. *et al.* Genetic risk for schizophrenia and psychosis in Alzheimer disease. *Mol. Psychiatry* **23**, 963–972 (2018).

238. Spiliopoulou, A. *et al.* Genomic prediction of complex human traits: relatedness, trait architecture and predictive meta-models. *Hum. Mol. Genet.* **24**, 4167–4182 (2015).

239. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors. *bioRxiv* 416859 (2018). doi:10.1101/416859

240. Krapohl, E. *et al.* Multi-polygenic score approach to trait prediction. *Mol. Psychiatry* **23**, 1368–1374 (2018).

241. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).

242. Shi, J. *et al.* Winner's Curse Correction and Variable Thresholding Improve Performance of Polygenic Risk Modeling Based on Genome-Wide Association Study Summary-Level Data. *PLoS Genet.* **12**, 1–24 (2016).

243. Choi, S. W., Shin, T., Mak, H. & O'reilly, P. F. A guide to performing Polygenic Risk Score analyses. doi:10.1101/416545

244. Alves, G. *et al.* Incidence of Parkinson's disease in Norway: The Norwegian ParkWest study. *J. Neurol. Neurosurg. Psychiatry* **80**, 851–857 (2009).

245. Goetz, C. G. *et al.* Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.* **23**, 2129–70 (2008).

246. Goetz, C. G. Unified Parkinson's Disease Rating Scale (UPDRS) and Movement Disorder Society Revision of the UPDRS (MDS-UPDRS). in *Rating Scales in Parkinson's Disease* 62–83 (Oxford University Press, 2012). doi:10.1093/med/9780199783106.003.0112

247. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

248. Delaneau, O. *et al.* Integrating sequence and array data to create an improved 1000

Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).

249. Consortium, the H. R. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

250. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–59 (2000).

251. EVANNO, G., REGNAUT, S. & GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

252. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).

253. SAS institute Inc. 2012. Base SAS® 9.4 Procedures Guide. Cary, NC: SAS Institute Inc.

254. RStudio Team. RStudio: Integrated Development for R. (2015).

255. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).

256. Purcell, S. Plink.

257. Aynsworth, C., Collerton, D. & Dudley, R. Measures of visual hallucinations: Review and recommendations. *Clin. Psychol. Rev.* **57**, 164–182 (2017).

258. Grinde, K. E. *et al.* Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet. Epidemiol.* **43**, 50–62 (2019).

259. Márquez-Luna, C., Loh, P. R. & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).

260. Otto, J. M., Gizer, I. R., Bizon, C., Wilhelmsen, K. C. & Ehlers, C. L. Polygenic risk scores for cigarettes smoked per day do not generalize to a Native American population.

Drug Alcohol Depend. **167**, 95–102 (2016).

261. Marden, J. R., Walter, S., Tchetgen Tchetgen, E. J., Kawachi, I. & Glymour, M. M. Validation of a polygenic risk score for dementia in black and white individuals. *Brain Behav.* **4**, 687–697 (2014).

262. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).

263. Marras, C. *et al.* Prevalence of Parkinson's disease across North America. *npj Park. Dis.* **4**, 21 (2018).

264. Rietdijk, C. D., Perez-Pardo, P., Garssen, J., van Wezel, R. J. A. & Kraneveld, A. D. Exploring Braak's Hypothesis of Parkinson's Disease. *Front. Neurol.* **8**, 37 (2017).

265. Schapira, A. H. V., Chaudhuri, K. R. & Jenner, P. Non-motor features of Parkinson disease. *Nat. Rev. Neurosci.* **18**, 435–450 (2017).

266. McKinlay, A. *et al.* A profile of neuropsychiatric problems and their relationship to quality of life for Parkinson's disease patients without dementia. *Parkinsonism Relat. Disord.* **14**, 37–42 (2008).

267. Waters, F. *et al.* Visual hallucinations in the psychosis spectrum and comparative information from neurodegenerative disorders and eye disease. *Schizophr. Bull.* **40 Suppl 4**, S233-45 (2014).

268. Morgante, L. *et al.* Psychosis associated to Parkinson's disease in the early stages: relevance of cognitive decline and depression. *J Neurol Neurosurg Psychiatry* **83**, 76–82 (2012).

269. Zhu, K., van Hilten, J. J., Putter, H. & Marinus, J. Risk factors for hallucinations in Parkinson's disease: Results from a large prospective cohort study. *Mov. Disord.* **28**, 755–

762 (2013).

270. Alzahrani, H. & Venneri, A. Cognitive and neuroanatomical correlates of neuropsychiatric symptoms in Parkinson's disease: A systematic review. *J. Neurol. Sci.* **356**, 32–44 (2015).

271. Marinus, J., Zhu, K., Marras, C., Aarsland, D. & van Hilten, J. J. Risk factors for non-motor symptoms in Parkinson's disease. *Lancet Neurol.* **17**, 559–568 (2018).

272. Compta, Y. *et al.* High cerebrospinal tau levels are associated with the rs242557 tau gene variant and low cerebrospinal β-amyloid in Parkinson disease. *Neurosci. Lett.* **487**, 169–173 (2011).

273. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–8 (2013).

274. Schizophrenia Working Group of the Psychiatric Genomics Consortium *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

275. AlzGene. Available at: http://www.alzgene.org/geneoverview.asp?geneid=83.

276. Desikan, R. S. *et al.* Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. *PLOS Med.* **14**, e1002258 (2017).

277. Song, L., Liu, A., Consortium, M. G. of S. & Shi, J. SummaryAUC: a tool for evaluating the performance of polygenic risk prediction models in validation datasets with only summary level statistics. *bioRxiv* 359463 (2018). doi:10.1101/359463

278. Guloksuz, S. *et al.* Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry* **18**, 173–182 (2019).

279. Cannas, A. *et al.* Chronic delusional hallucinatory psychosis in early-onset Parkinson's

disease: drug-induced complication or sign of an idiopathic psychiatric illness? *Neurol. Sci.* **22**, 53–4 (2001).

280. Mistry, S., Harrison, J. R., Smith, D. J., Escott-Price, V. & Zammit, S. The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: Systematic review. *Schizophr. Res.* **197**, 2–8 (2018).

281. Curtis, D. Polygenic risk score for schizophrenia is not strongly associated with the expression of specific genes or gene sets. *Psychiatr. Genet.* **28**, 1 (2018).

282. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).

283. Sabunciyan, S. Gene Expression Profiles Associated with Brain Aging are Altered in Schizophrenia. *Sci. Rep.* **9**, 5896 (2019).

284. Rammos, A., Gonzalez, L. A. N., Weinberger, D. R., Mitchell, K. J. & Nicodemus, K. K. The role of polygenic risk score gene-set analysis in the context of the omnigenic model of schizophrenia. *Neuropsychopharmacology* 1 (2019). doi:10.1038/s41386-019-0410-z

285. Liu, C.-C., Liu, C.-C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.* **9**, 106–18 (2013).

286. Paul, K. C. *et al.* APOE, MAPT, and COMT and Parkinson's Disease Susceptibility and Cognitive Symptom Progression. *J. Parkinsons. Dis.* **6**, 349–59 (2016).

287. Gartenberg Livney, M. *et al.* Ethnoracial Differences in the Clinical Characteristics of Alzheimer Disease at Initial Presentation at an Urban Alzheimer's Disease Center. *Am J Geriatr Psychiatry* **19**, 430–439 (2011).

288. Campos, M., Edland, S. D. & Peavy, G. M. Exploratory study of apolipoprotein E ε4 genotype and risk of Alzheimer's disease in Mexican Hispanics. *J. Am. Geriatr. Soc.* **61**,

1038–40 (2013).

289. Haan, M. N. *et al. Prevalence of Dementia in Older Latinos: The Influence of Type 2 Diabetes Mellitus, Stroke and Genetic Factors*. *JAGS* **51**, (2003).