The genetics of sporadic Parkinson's disease

Refining the insights from genome-wide association studies

Lasse Pihlstrøm

Department of Neurology, Oslo University Hospital
Faculty of Medicine, University of Oslo

2015

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 2102

ISBN 978-82-8333-116-5

Cover: Hanne Baadsgaard Utigard.
Print production: John Grieg AS, Bergen.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

In examining disease, we gain wisdom about anatomy and physiology and biology. In examining the person with disease, we gain wisdom about life.

- Oliver Sacks

Table of Contents

ACKNOWLEDGEMENTS

The work presented in this thesis was carried out at Oslo University Hospital, Department of Neurology, and the University of Oslo, Faculty of Medicine, Institute for Clinical Medicine, between 2010 and 2015. The PhD project was funded by the South-Eastern Norway Regional Health Authority, and included studies were also supported by grants from the Norwegian Parkinson Association Research Fund and Reberg's Legacy. I am grateful to these institutions.

I thank all study participants for making this research possible by giving consent and donating blood samples, and also for the many moving and inspiring encounters in the clinic where individual patients have shared their unique personal experiences of life with Parkinson's disease. Those lives represent the importance of all our efforts.

I wish to express my deepest, most heartfelt gratitude to my supervisor Mathias Toft for your invaluable guidance, always coloured by your enthusiasm, empathy, creativity, intelligence and humour. Developing our close collaboration over this PhD period has been one of the great experiences of my life. Your thoughts and attitudes towards both science and clinical work are truly inspiring. Thank you for giving me trust, recognition and great freedom to develop my own ideas, while at the same time maintaining a close involvement in the projects, generously sharing your thoughts and time.

I am sincerely grateful to my co-supervisor and head of department Espen Dietrichs. Thank you for taking the initiative to get me involved in movement disorders research, and for all your wise and experienced support. Your interest in culture, art and history, putting neurological knowledge into a broader context, inspires me greatly. Thank you for generously taking me with you into an atmosphere where neurology is social, international, interesting and fun.

Over the time period when the studies presented in this thesis were conducted, Mathias Toft's project group went from having two pipettes in a plastic box, to modern laboratory facilities in a brand new building. It has been a privilege to take part in this development, and most of all, to enjoy the collaboration with an increasing number of good colleagues. Warm thanks to Kari Anne Bjørnarå, Silje Bjerknes, Aina Rengmark,

Zafar Iqbal, Victoria Berge, Sandra Pilar Henriksen, Camilla Sætehaug and Anne Fosby.

I also wish to thank several other people for important contributions and valuable help. I am indebted to all co-authors who contributed sample sets from ParkWest, Gothenburg, Linköping and Umeå. Special thanks to research nurse Lena Pedersen for invaluable help with sample collection and handling. I am grateful to Rotary clubs in Oslo and Bærum, as well as personnel at Markveien legesenter, Oslo, Fürst laboratories, Oslo and Sørumsand legesenter, who helped me in the recruitment of control subjects. Many thanks also to Gregor Gilfillan and Tim Hughes at the Norwegian Sequencing Centre, Ullevål, and Kristil Kindem Sundsaasen at Cigene.

During the course of my PhD work, I have also enjoyed interesting discussions and an inspiring shared research environment with many good colleagues at the Neuroscience Research Unit, Oslo University Hospital. In particular, there have been many fruitful interactions with the research groups led by Hanne Harbo, John Anker Zwart and Chantal Tallaksen, for which I am very grateful. Special thanks also to Kaja Kristine Selmer and Magnus Dehli Vigeland at the Department of Medical Genetics, Oslo University Hospital for enlightening discussions and inspiring enthusiasm.

While conducting the studies presented in this thesis, I have also periodically served as a resident physician and University teacher in the Department of Neurology at Oslo University Hospital, Rikshospitalet. Ever since I first came to the department, I have felt generously included and intellectually at home. I am truly grateful to all clinical colleagues for this, especially the consultant neurologists whom I have had the privilege of working closely with over the years.

Finally, I would like to thank my friends and family, in particular my parents Anne-Marie and Øystein Pihlstrøm, and also my sister Hege, for continuous care and support and genuine interest in my thoughts and work.

Most of all, thanks to my wife Karianne Tøsse, for sharing with me life itself, where work is but one part of the great conjoined adventure. Our daughters Julia and Maria have my deepest gratitude simply for being who they are; challenging, teaching, inspiring and rewarding me the way that only children can.

LIST OF ABBREVIATIONS

cDNA*	Complementary DNA
CpG*	Cytosine-phosphate-guanine
DBS	Deep brain stimulation
DLB	Dementia with Lewy bodies
DNA*	Deoxyribonucleic acid
GCTA	Genome-wide complex trait analysis
gDNA*	Genomic DNA
GWAS	Genome-wide association study
HTS	High throughput sequencing
HWE	Hardy-Weinberg equillibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
MPTP	1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine
mRNA*	Messenger RNA
MRSE	Methylation-sensitive restriction enzyme
OR	Odds ratio
PCR*	Polymerase chain reaction
PD	Parkinson's disease
PD-D	Parkinson's disease dementia
qPCR	Quantitative PCR

RIN	RNA integrity number
RNA*	Ribonucleic acid
SNP	Single-nucleotide polymorphism
SN	Substantia nigra
UKPDS	United Kingdom Parkinson's Disease Society
ΔCT*	Delta threshold cycle

*On the first appearance in the text, the terms are spelled out in full, followed by the abbreviated form in parentheses, except for the common abbreviations marked here by an asterix.

Note that for some terms used only once in the text, abbreviations are not listed here, but may still be indicated in parentheses if they are frequent in the literature. Genes are throughout the text referred to by their gene symbols in *italicized* uppercase letters.

PUBLICATIONS INCLUDED

The thesis is based on the following original articles:

Paper 1

Pihlstrøm L, Axelsson G, Bjørnaråa KA, Dizdar N, Fardell C, Forsgren L, Holmberg B, Larsen JP, Linder J, Nissbrandt H, Tysnes OB, Ohman E, Dietrichs E, Toft M. Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease. **Neurobiology of Aging. 2013;34(6):1708 e7-13.**

Paper 2

Pihlstrøm L, Rengmark A, Bjørnaråa KA, Toft M. Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease. **Annals of Human Genetics. 2014;78(3):243-52.**

Paper 3

Pihlstrøm L, Rengmark A, Bjørnaråa KA, Dizdar N, Fardell C, Forsgren L, Holmberg B, Larsen JP, Linder J, Nissbrandt H, Tysnes OB, Dietrichs E, Toft M. Fine mapping and resequencing of the PARK16 locus in Parkinson's disease. **Journal of Human Genetics. 2015;doi: 10.1038/jhg.2015.34. [Epub ahead of print]**

Paper 4

Pihlstrøm L, Berge V, Rengmark A, Toft M. Parkinson's disease correlates with promoter methylation in the α-synuclein gene. **Movement Disorders. 2015;30(4):577-80.**

SUMMARY IN ENGLISH

Parkinson's disease is a common neurological disorder where the prevalence increases with age. The diagnosis is based on motor symptoms, namely slowness of movements, rigidity, tremor and balance problems, yet patients also suffer a range of other complaints, including cognitive decline and dementia. We may currently offer therapy that effectively alleviates some of the symptoms of Parkinson's disease, but without available treatment to target the cause of the disease, the condition typically follows a progressive course, severely affecting the quality of life of patients and their families. An improved understanding of the disease mechanisms on a molecular level is crucial for the development of new therapeutic strategies in the future. Mapping the genetic causes of the disease represents a key aspect of this effort. In recent years, a number of gene regions affecting the risk of Parkinson's disease have been identified through large-scale genetic investigations known as genome-wide association studies. This has represented a significant breakthrough, yet many open questions remain to be addressed before clinical medicine can take advantage of the results from these studies. The work presented in this thesis aims to shed further light on the genetic causes of Parkinson's disease by building upon the findings from genome-wide association studies. We first examined genetic risk-variants in patients and healthy controls in a material of more than 2500 participants from Norway and Sweden and found supportive evidence for an association with disease risk for 11 different gene regions. Next, we presented an effective study design for investigating such gene regions in detail in large sample sets by modern sequencing technologies. We then used this method as part of a study where we explored in depth a susceptibility region on chromosome 1, called PARK16. In the final study we investigated the relation between genetic risk variants in the *SNCA* gene and methylation, an epigenetic modification of the DNA molecule that regulates gene expression. We found evidence indicating that genetic susceptibility may be mediated through epigenetic mechanisms. Taken together, the studies presented in this thesis involve a broad range of methods and topics that are expected to become increasingly important in the genetic study of Parkinson disease, as well as other common diseases, in the years to come.

SAMMENDRAG PÅ NORSK (SUMMARY IN NORWEGIAN)

Parkinsons sykdom er en nevrologisk lidelse, der forekomsten øker med alderen. Diagnosen stilles ut ifra motoriske symptomer med langsomme bevegelser, stivhet, skjelving og balanseproblemer, men sykdommen medfører også en rekke andre plager, inkludert utvikling av kognitiv svikt og demens. Vi har i dag behandling som effektivt lindrer noen av symptomene ved Parkinsons sykdom, men ingen terapi som er rettet mot sykdommens årsak og kan forhindre at tilstanden progredierer og forårsaker alvorlig tap av livskvalitet hos pasienter og pårørende. En bedre forståelse av sykdomsmekanismene på molekylært nivå er avgjørende for å kunne utvikle nye behandlingsstrategier i fremtiden. Kartlegging av sykdommens genetiske årsaker bidrar med viktig nøkkelkunnskap i dette arbeidet. I senere år har en rekke genområder som påvirker risikoen for Parkinsons sykdom, blitt identifisert gjennom storskala genetiske undersøkelser, kjent som helgenoms- assosiasjonsstudier. Dette har vært et vesentlig gjennombrudd, men mange åpne spørsmål må besvares før resultatene fra disse studiene kan få positive ringvirkninger for klinisk medisin. Arbeidet som presenteres i denne avhandlingen, dreier seg om å bygge videre på funn fra helgenoms- assosiasjonsstudier for å kaste nytt lys over genetiske årsaker til Parkinsons sykdom. Først har vi sett på forekomsten av genetiske risikovarianter hos pasienter og friske kontroller i et materiale med over 2500 deltakere fra Norge og Sverige. Vi fant støtte for en assosiasjon med sykdomsrisiko for 11 ulike genområder. Videre har vi presentert et effektivt studiedesign for å undersøke slike assosierte genområder grundigere i store prøvematerialer med moderne sekvenseringsteknologi. Denne metoden har vi deretter anvendt som ledd i en studie der vi har kartlagt nærmere et risiko-område på kromosom 1, kalt PARK16. I den siste studien har vi sett på sammenhengen mellom genetiske risikovarianter i genet *SNCA* og metylering, en epigenetisk modifikasjon av DNA-molekylet som er med på å regulere i hvilken grad genet er uttrykt. Vi fant holdepunkter for at genetikken påvirker sykdomsrisiko gjennom mekanismer som involverer epigenetikk. Studiene som presenteres i denne avhandlingen, involverer samlet sett et bredt spekter av metoder og problemstillinger som forventes å bli sentrale innen genetiske studier av Parkinsons sykdom og andre vanlige sykdommer i årene som kommer.

1. INTRODUCTION

1.1 A brief history of Parkinson's disease

The disorder which came to be known as Parkinson's disease (PD) was first described formally by James Parkinson in his seminal 1817 paper "An essay on the shaking palsy" (Parkinson, 1817). Notably however, a broad range of historical accounts indicate that the syndrome has been recognized in humans from ancient times (Stern, 1989, Raudino, 2012), and in certain cultures even successfully treated by plant medicine (Manyam, 1990). Parkinson's famous monograph is based on clinical descriptions of six patients he had either seen in his own medical practice or observed in the streets close to his home in London. Still, his account captures a remarkable range of clinical features including limb tremor, flexed posture, shuffling gait, asymmetry of symptoms and a progressive disease course. In addition he foresaw the future discovery of an anatomical substrate in the brain explaining the symptoms. The term "Parkinson's disease" was introduced around 1875 by Jean-Martin Charcot, the prominent clinician and professor often referred to as the father of modern neurology (Lees, 2007). Charcot distinguished the motor symptoms of PD from "palsy" or muscle weakness, recognizing all the features still used to define the Parkinsonian syndrome: Bradykinesia, rigidity, tremor and postural instability.

In 1912, Fritz Heinrich Lewy published his first works on the histopathology of PD, describing characteristic eosinophilic, intracellular inclusions in a range of nuclei in the brainstem and basal cerebrum (Lewy, 1912). The lesions were named "Lewy bodies" by Konstantin N. Tretiakoff, who five years later reported the combination of inclusions and neurodegeneration in the substantia nigra (SN) of PD patients (Tretiakoff, 1919). Rolf Hassler later confirmed this finding and refined the anatomical substrate of Parkinsonism to the SN pars compacta (Hassler, 1938). The combination of Lewy pathology and loss of SN pigmented neurons is still considered the pathological hallmark of PD, and both lesions are required for the definitive post-mortem diagnosis (Gelb et al., 1999).

The next crucial milestone in the understanding of Parkinson's disease was the characterization of dopaminergic neuronal transmission in the basal ganglia. The studies of Arvid Carlsson and co-workers in the 1950s provided evidence that the catecholamine neurotransmitter dopamine had a specific role in the corpus striatum (Carlsson et al., 1958). Carlsson further reported that dopa, the molecular precursor of dopamine, was partly able to reverse pharmacologically induced Parkinsonian symptoms in animals (Carlsson et al., 1957). Herbert Ehringer and Oleh Hornykiewicz demonstrated reduced dopamine levels in the striatum in post-mortem tissue from PD patients (Ehringer and Hornykiewicz, 1960). Together with the clinical neurologist Walter Birkmayer, Hornykiewicz went on to test L-dopa injected intravenously in patients with Parkinsonism (Birkmayer and Hornykiewicz, 1961). The initial report described a striking effect, yet subsequent trials were less convincing and limited by pronounced side effects. The clinical breakthrough came in the late 1960s, when George Cotzias developed a therapeutic regime of slowly increasing oral doses of L-dopa, combined with a peripheral dopa decarboxylase inhibitor (Cotzias et al., 1969b, Cotzias et al., 1969a, Yahr et al., 1969). The introduction of L-dopa as symptomatic treatment in PD drastically changed the prognosis for patients with the disorder, and has furthermore become an emblematic example of therapeutic advance following from a fruitful interplay between clinical neurology and basal molecular neuroscience.

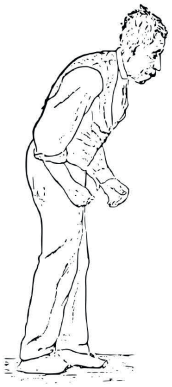
1.2 A modern concept of Parkinson's disease

The clinical diagnosis of PD is currently based on the presence of a Parkinsonian motor syndrome, and exclusion of possible causes of secondary Parkinsonism, such as the use of antipsychotic drugs or cerebrovascular disease affecting the basal ganglia (Figure 1) (Gibb and Lees, 1988). However, the importance of non-motor symptoms, both for the scientific understanding of the PD phenotype and for the disease burden experienced by patients, has been increasingly recognised in recent years (Chaudhuri et al., 2011, Martinez-Martin, 2011). Prevalent non-motor symptoms in PD include obstipation, bladder dysfunction, sleep disturbances, orthostatic hypotension, depression, hallucinations and cognitive decline.

Figure 1. *UK Parkinson's Disease Society Brain Bank clinical diagnostic criteria*

STEP 1. Diagnosis of Parkinsonian syndrome

- BRADYKINESIA
- At least one of the following:
  a. muscular rigidity
  b. 4-6 Hz rest tremor
  c. postural instability (other causes excluded)



STEP 2. Exclusion criteria for Parkinson's disease

- history of repeated strokes with stepwise progression of Parkinsonian features
- history of repeated head injury
- history of definite encephalitis
- oculygyric crisis
- neuroleptic treatment at onset of symptoms
- more than one affected relative*
- sustained remission
- strictly unilateral features after three years
- supranuclear gaze palsy
- cerebellar signs
- early severe autonomic involvement
- early severe dementia with disturbances of memory, language and praxis
- Babinski sign
- Presence of a cerebral tumor or communicating hydrocephalus on CT scan
- negative response to large doses of levodopa (if malabsorption excluded)
- MPTP exposure

STEP 3. Supportive prospective criteria for Parkinson's disease.

Three or more required for diagnosis of definite Parkinson's disease

- unilateral onset
- rest tremor present
- progressive disorder
- persistent asymmetry affecting the side of onset most
- excellent response (70-100% to levodopa)
- severe levodopa-induced chorea
- levodopa response for 5 years or more
- clinical course of 10 years or more

* Many studies apply these criteria, yet allow for more than one affected relative, frequently referred to as "revised UKPDS Brain Bank criteria"

According to currently accepted criteria, at least bradykinesia remains a "sine qua non"; a clinical feature common to all patients diagnosed with PD (Gibb and Lees, 1988). Yet apart from this, there is a remarkable heterogeneity in the distribution, severity and progression rate of both motor and non-motor symptoms across individual patients. This raises the fundamental question of whether it is really justified to consider PD as a distinct, single entity. With respect to the motor syndrome, PD subtypes such as tremor-dominant versus bradykinetic-rigid PD and postural instability/gait disorder (PIGD) have been proposed (Marras and Lang, 2013, Thenganatt and Jankovic, 2014). These motor profiles are to some extent also correlated with long-term prognosis, where tremor-dominant PD is associated with a milder disease course. Studies indicate that as many as 75% of PD patients develop dementia over the course of the disease, yet the temporal relation of cognitive decline to other symptoms varies considerably (Aarsland and Kurz, 2010). According to current criteria, patients with Parkinsonism and dementia are classified as having either PD with dementia (PD-D) or dementia with Lewy bodies (DLB) based on the timing of symptom onset (Dubois et al., 2007, Lippa et al., 2007). DLB is diagnosed when dementia is manifest before an arbitrary cut-off of one year after the onset of a Parkinsonian motor syndrome.

The treatment of PD is essentially symptomatic. Evidence seems to support some degree of neuroprotective properties for the monoamine oxidase inhibitors (Olanow et al., 2009), but there is no available therapy to effectively alter the disease course. In the early phases of the disease, motor symptoms are usually well controlled by the administration of dopaminergic drugs. In later stages, a subset of patients will experience a narrowing of the therapeutic window with problematic fluctuations in motor function. Due to the development in advanced treatment modalities such as deep brain stimulation (DBS) and pump systems for continuous drug administration, effective symptomatic treatment can now be offered even to patients with significant motor complications (Volkmann et al., 2013). However, axial involvement, including disturbances of gait and balance, may still respond poorly to dopaminergic medication and DBS. These more treatment-resistant motor symptoms, in combination with the range of severe non-motor features, are currently emerging as the most debilitating and therapeutically challenging aspects of PD.

It is well established that the neurodegenerative process starts long before the onset of motor symptoms in PD (Kordower et al., 2013). Furthermore, there is also evidence that certain non-motor features, may precede the onset of a Parkinsonian motor syndrome, supporting the notion of a symptomatic, yet pre-motor phase of the disease (Lang, 2011). Pre-motor symptoms include hyposmia, obstipation and rapid eye movement sleep behaviour disorder (RBD), and represent early clinical manifestations of the same neurodegenerative process. Considering future efforts to develop disease-modifying therapies, patients in the pre-motor stage would be interesting candidates for inclusion in clinical trials. Taken together, several lines of evidence weigh in favour of a definition of PD that is less dependent on the motor syndrome alone, taking a broader understanding of the disease process into account. Initiatives are being taken to update the criteria for research purposes (Berg et al., 2014). As pre-motor symptoms are largely unspecific, an earlier diagnosis may have to be supplemented by biomarkers. The development of reliable biomarkers, based on for example imaging techniques or biochemical assays, is therefore an important scientific aim for current PD research.

1.3 The epidemiology of Parkinson's disease

After Alzheimer's disease, PD is the second most common of the neurodegenerative disorders. The incidence and prevalence of PD have been investigated in a number of epidemiological studies, yet published figures vary considerably and direct comparisons are often hampered by differences in methodology and reporting (Wirdefeldt et al., 2011). In a 2011 review, Wirdefeldt et al. found that reported incidence rates ranged between 1.5 and 22 per 100 000 person years, and prevalence rates mostly between 100 and 300 per 100 000. Several reports have indicated a higher prevalence in men than in women, but other studies found no such gender difference. PD is rare before the age of 50 years, and a sharp increase is seen for both incidence and prevalence after the age of 60 (de Lau and Breteler, 2006). A frequently cited figure is an estimated PD prevalence of 1% for people over 60 years of age in industrialized countries (Nussbaum and Ellis, 2003). With a larger proportion of the

population surviving into old age, the total number of PD patients in the world has been projected to double in 25 years, reaching a total of nine million in 2030 (Dorsey et al., 2007).

Sporadic PD is generally considered a complex disease, caused by an interplay between a range of genetic, environmental, and possibly stochastic factors. However, the genetic contribution to PD risk was historically not always recognized. In 1983, Langston et al. reported the cases of four patients who in the context of intravenous drug abuse injected the compound 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP) and developed a severe, chronic Parkinsonian syndrome with response to L-dopa (Langston et al., 1983). The selective toxicity of MPTP against dopaminergic cells provided opportunities for animal models and drew scientific attention towards the possibility of exogenous toxic agents causing PD. An extensive list of environmental and lifestyle exposures have been investigated in epidemiological studies of PD. In large systematic reviews, the evidence for specific environmental risk factors is generally scarce (Wirdefeldt et al., 2011, de Lau and Breteler, 2006). An inverse association with both smoking and coffee consumption has been consistently demonstrated across large studies. There is also suggestive evidence that pesticide exposure may increase PD risk, but the role of specific compounds is unclear. Regarding medication use, there are reports indicating that non-steroid inflammatory drugs (NSAIDs) and calcium channel blockers may be associated with a decreased risk of PD.

The genetic contribution to complex disease is classically addressed in twin studies. Monozygotic twins are genetically identical, whereas dizygotic twins on average share half their genome. Environmental factors are assumed to be equally shared between all twin pairs. Consequently, a genetic contribution to disease risk should result in higher concordance rates for monozygotic than for dizygotic twins. Several early twin studies failed to show differences in concordance rates, but these were generally limited by small sample size and cross-sectional design (Johnson et al., 1990). A significant genetic effect has later been shown in longitudinal twin studies, especially when asymptomatic dopaminergic dysfunction by [^{18}F]dopa positron emission tomography (PET) has been introduced as a more sensitive endpoint to ascertain concordance rate (Burn et al., 1992). On repeated follow up, this approach

has increased the concordance gap between monozygotic and dizygotic twins to 75% versus 22% (Piccini et al., 1999). Studies of familial aggregation represent another approach to investigate the importance of genetics for disease risk. A 2008 meta-analysis based on the highest-quality studies found that individuals with a first-degree relative with PD carried a relative disease risk of 2.9 (Thacker and Ascherio, 2008). Studies from Norway (Kurz et al., 2003), Sweden (Sundquist et al., 2006) and Iceland (Sveinbjornsdottir et al., 2000) have reported slightly higher estimations, but confidence intervals were overlapping with the meta-analysis.

In addition to the available epidemiological evidence, the importance of genetics factors in the aetiology of PD has been firmly established by the identification of genetic loci implicated both in monogenic and sporadic forms of the disease from the 1990s and onwards, as detailed further in the following sections.

1.4 Mendelian forms of Parkinson's disease and implications for pathogenesis

At the time of Charcot, several neurologists believed PD to be inheritable, based on observations of familial cases. In the first half of the twentieth century, PD families were reported from a range of European countries. The first systematic family study of Parkinsonism was published in 1949 by Henry Mjönes, who based on clinical material from central Sweden argued for a dominant inheritance pattern in PD (Mjönes, 1949). Subsequently, however, inheritable forms of PD received little scientific attention until the 1990s when advances in genetic methodology made it feasible to identify disease loci in large kindreds. Over the last 20 years, a number of genes responsible for Mendelian PD have been identified by linkage analysis, or in later years by high throughput sequencing (HTS) methods (Bonifati, 2014, Klein and Westenberger, 2012). Both autosomal recessive, autosomal dominant and X-linked inheritance patterns have been described in monogenic forms of Parkinsonism. While some of these genes give rise to a disease that resembles sporadic PD, others are characterized by a broader phenotype where Parkinsonism is only one of several features constituting a genetic syndrome. Genes causing Mendelian disease where Parkinsonism is the prominent symptom are summarized in Table 1.

Table 1 *Monogenic forms of Parkinsonism*

Gene	Inheritance	Phenotype	Reference
SNCA	Autosomal dominant	Earlier onset PD, agressive course	Polymeropoulos et al., 1997, Singleton et al., 2003
LRRK2	Autosomal dominant	Typical, late-onset PD	Paisan-Ruiz et al., 2004, Zimprich et al., 2004
VPS35	Autosomal dominant	Typical, late-onset PD	Zimprich et al., 2011, Vilarino-Guell et al., 2011
EIF4G1*	Autosomal dominant	Typical, late-onset PD	Chartier-Harlin et al., 2011
DNAJC13*	Autosomal dominant	Typical, late-onset PD	Vilarino-Guell et al., 2014
CHCHD2*	Autosomal dominant	Typical, late-onset PD	Funayama et al., 2015
PARK2 **(Parkin)**	Autosomal recessive	Early-onset PD	Kitada et al., 1998
PINK1	Autosomal recessive	Early-onset PD	Valente et al., 2004
***PARK7* (*DJ-1*)**	Autosomal recessive	Early-onset PD	Bonifati et al., 2003
ATP13A2	Autosomal recessive	Juvenile onset, atypical features	Ramirez et al., 2006
PLA2G6	Autosomal recessive	Juvenile onset, atypical features	Paisan-Ruiz et al., 2009
FBXO7	Autosomal recessive	Juvenile onset, atypical features	Shojaee et al., 2008, Di Fonzo et al., 2009
DNAJC6	Autosomal recessive	Juvenile onset, atypical features	Edvardson et al., 2012, Koroglu et al., 2013
ATP6AP2	X-linked	Juvenile onset, atypical features	Korvatska et al., 2013
RAB39B	X-linked	Juvenile onset, atypical features	Wilson et al., 2014

Parkinsonism may also occur as a clinical feature across a wide range of genetic neurodegenerative disorders not listed here, including DNA repeat expansions, frontotemporal lobe degenerations, mitochondrial disorders, neurodegenerations with brain iron accumulation and several others.

** The pathogenic relevance of some of the recently proposed dominant genes remains to be determined in large independent studies.*

The identification of Mendelian PD genes has been crucially important for the understanding of PD pathogenesis. In the following, this development is illustrated by considering the examples of *SNCA, PARK2* and *LRRK2*.

1.4.1 SNCA - α-synuclein

Two large kindreds originating from the village Contursi in southern Italy, with autosomal dominant PD affecting 41 individuals across four generations, were reported in 1990. The clinical phenotype resembled sporadic PD, but with earlier onset, and typical PD pathology with Lewy bodies was observed on autopsy (Golbe et al., 1990). The PD phenotype was mapped to chromosome 4q21-23 (Polymeropoulos et al., 1996), and a causative missense mutation was identified in *SNCA* (c.157G>A, p.Ala54Thr), which also segregated with disease in three Greek families with a similar PD phenotype and inheritance pattern (Polymeropoulos et al., 1997). This was the first genetic cause of PD to be recognized, and the starting point for extensive subsequent research into α-synuclein as a key protein in the pathogenesis of PD.

Later studies have identified another five rare, pathogenic missense mutations in *SNCA* (Kruger et al., 1998, Zarranz et al., 2004, Lesage et al., 2013, Proukakis et al., 2013, Pasanen et al., 2014). Furthermore, autosomal PD has also been found to be caused by *SNCA* multiplications. A severe phenotype with early onset, rapid progression and dementia was associated with *SNCA* triplication in the large "Iowa kindred" (Spellman, 1962, Muenter et al., 1998, Singleton et al., 2003), as well as a Swedish-American family with triplication of independent genetic origin (Farrer et al., 2004). Screening for *SNCA* multiplications in dominant PD also revealed families with duplications (Chartier-Harlin et al., 2004, Ibanez et al., 2004). Duplications tend to give rise to a milder form of the disease as compared to triplication, pointing towards a dose dependent relationship between copy number and phenotype (Farrer, 2006).

Following the discovery of *SNCA* mutations in PD, studies using immunohistochemical staining revealed it's corresponding protein, α-synuclein, as the main component of Lewy bodies (Spillantini et al., 1997). The application of α-synuclein antibodies to visualise protein deposition in pathological tissue sections has been of great importance in PD research, leading to an increased understanding of the

widespread pathologies in different parts of the nervous system as the disease progresses. An influential hypothesis presented by Braak et al. is based on pathological staging of PD lesions (Braak et al., 2003). It proposes that the pathology spreads in a stereotypic pattern, possibly originating in the gut or nose, explaining the successive onset of different symptoms. Depositions of α-synuclein have further been identified in a range of other neurodegenerative disorders, both in the form of Lewy pathology and different histopathological patterns. These observations have had important implications for our understanding of the interconnections between degenerative brain disorders and have placed PD in a special relation to the broader group of synucleinopaties (Goedert, 2001, Jellinger, 2003).

The normal function of α-synuclein is not fully understood, but evidence supports a role in the recycling of synaptic vesicles (Bendor et al., 2013). The protein exists naturally in an unfolded state, but has the potential to misfold, oligomerize and aggregate (Vekrellis et al., 2011). Understanding the mechanisms that determine the toxicity and spread of α-synuclein is currently a scientific aim receiving high priority across the PD field. Compelling evidence indicate that there may be mechanisms by which α-synuclein aggregation may self-propagate across cell membranes causing the pathology to spread from an initial "seeding" event to neighbouring neurons. This is also consistent with the observation that fetal neurons that have been transplanted into the striatum of PD patients in clinical trials, develop Lewy pathology despite the remarkably short lifespan of the grafted cells themselves (Li et al., 2008, Kordower et al., 2008).

1.4.2 PARK2 - parkin

Homozygous deletions the parkin gene (*PARK2*) were originally identified in a consanguineous Japanese family with juvenile-onset Parkinsonism (Kitada et al., 1998), representing the first autosomal recessive form of PD to be genetically characterized. Parkin was functionally found to have protein-ubiquitin ligase activity (Shimura et al., 2000). Collectively, the discoveries of the recessive genes parkin, *PINK1* and *PARK7* (*DJ-1*) have highlighted the role of mitochondrial function,

oxidative stress and protein degradation pathways in PD pathogenesis (Dodson and Guo, 2007, Pickrell and Youle, 2015).

Parkin mutations are rare overall, yet represent a major fraction of early-onset cases with a recessive inheritance pattern (Lucking et al., 2000). The clinical presentation is typically characterized by marked dystonia and steady progression of motor symptoms with early fluctuations, yet relatively intact cognitive function (Khan et al., 2003). Neuropathologically, most cases of parkin disease differ from *SNCA* mutations and sporadic PD by exhibiting isolated loss of dopaminergic neurons without accompanying Lewy pathology (Poulopoulos et al., 2012). The distinguishing features unique to recessive forms of Parkinsonism are important reminders of how the range of syndromes commonly classified as PD probably have markedly heterogeneous molecular underpinnings.

1.4.3 LRRK2 - leucine-rich repeat kinase 2

A second gene for autosomal dominant PD was identified in 2004 (Zimprich et al., 2004, Paisan-Ruiz et al., 2004). Patients with *LRRK2* mutations have been described with variable clinical and neuropathological characteristics, often indistinguishable from sporadic PD. Large screening efforts have demonstrated *LRRK2* as the most frequent form of monogenic parkinsonism worldwide (Healy et al., 2008), with the G2019S mutation standing out as the most common (Kachergus et al., 2005). The fraction of PD cases caused by *LRRK2* G2019S mutations is particularly high in Ashkenazi Jewish and North-African Arab populations, in the latter explaining over a third of sporadic cases and an even larger proportion of familial disease (Correia Guedes et al., 2010). Among other important insights, studies of *LRRK2* in PD have demonstrated that monogenic disease can be common in particular populations, appear sporadically and have incomplete penetrance, probably modulated by other genetic and environmental factors (Trinh et al., 2014).

1.5 The genetic study of sporadic Parkinson's disease

1.5.1 General aspects of complex genetics

Although valuable insights have resulted from the investigation of Mendelian PD genes, the vast majority of patients have sporadic disease, caused by a complex interplay of both genetic and non-genetic factors. Progress in the study of complex genetics has followed a similar course for many common, non-communicable diseases over the last decades, led on by advances in technology and collaborative efforts to develop crucial database resources, such as the Human Genome Project (Lander, 2011, Lander et al., 2001), the International HapMap Project, the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2010) and the Encyclopedia of DNA Elements (Encode Project Consortium, 2004). Consequently, the opportunities and challenges highlighted in the context of sporadic PD in this thesis, are to a large degree paralleled across a broad range of disorders.

The genetic study of complex disease has several aims. Firstly, the identification of genetic risk loci may pave way for an improved understanding of pathogenic molecular mechanisms. Genes implicated in disease susceptibility are likely to encode proteins that participate in relevant cellular pathways. Mapping of disease-related genes, pathways and networks generates hypotheses for further functional research, and may eventually open new possibilities for therapeutic developments. A second type of aim concerns the utility of genetic information in the assessment of individual patients. Comprehensive knowledge about the genetic architecture of common diseases could conceivably allow for some degree of prediction or identification of individuals at risk. In combination with clinical outcomes, an individual genetic profile could also be useful in distinguishing subgroups of patients with similar characteristics. This strategy could have implications for the selection of patients to future clinical trials and would also be in line with the proposed ideal of tailored, personalized medicine.

1.5.2 Early candidate studies in sporadic Parkinson's disease

The paradigmatic study design for identification of risk-loci in complex disease is the case-control association study. The earliest phase of association studies followed a candidate approach, and the available technology favoured microsatellites as the most convenient class of polymorphism for genotyping. From the late 1990s and onwards, a large body of publications established the Rep1 (D4S3481) mixed dinucleotide repeat upstream of *SNCA* as a susceptibility variant in sporadic PD, importantly demonstrating the principle that one locus can be implicated both in Mendelian and complex forms of the same disease (Kruger et al., 1999, Maraganore et al., 2006).

The microtubule-associated protein tau is a main component of neurofibrillary tangles, a histopathological lesion seen in several neurodegenerative disorders commonly referred to as tauopathies, and distinct from the synuclein-dominant Lewy pathology of classical PD. However, missense and splicing mutations in the tau gene *MAPT* were shown to cause a combined phenotype of frontotemporal dementia and parkinsonism (Hutton, 2001). This and other findings, prompted the investigation of *MAPT* as a candidate gene for sporadic PD. *MAPT* is located on a region of chromosome 17 where a large inversion event has given rise to two distinct haplotype groups termed H1 and H2, which do not recombine. A number of studies have corroborated that the H1 haplotype carries an increased PD risk, although the responsible functional mechanism remains unclear (Martin et al., 2001, Skipper et al., 2004). Importantly, the identification of *MAPT* as a susceptibility gene in PD has contributed to our current understanding of the interrelations and overlaps between different neurodegenerative disorders.

1.5.3 GBA mutations

Pathogenic mutations in *GBA*, encoding the enzyme glucocerebrosidase, cause Gaucher disease, an autosomal recessive lysosomal storage disorder. The clinical phenotypic spectrum of Gaucher disease includes Parkinsonism, and attentive clinicians involved in the follow-up of Gaucher patients came to suspect an overrepresentation of PD amongst the patients' relatives (Goker-Alpan et al., 2004).

Following the original publication, a number of studies have established heterozygous *GBA* mutations as strong risk factors for PD, highlighting the importance of lysosomal pathways in the pathogenesis (Sidransky et al., 2009). Furthermore, the gene has become a textbook example of rare variants with a large effect, representing an intermediate between rare Mendelian mutations with high penetrance and common alleles typically carrying a very small effect on risk.

1.5.4 Rationale for genome-wide association studies

The potential of candidate gene association studies depends on our ability to identify good candidates. Consequently, genetic discovery will rarely be able to uncover completely new pathways based on this approach. Furthermore, if a large number of candidate genes are investigated in small single studies, some will turn out positive by chance. These limitations posed major challenges in the early phases of complex disease genetics, and published associations were frequently not reproduced when independent follow-up studies were performed (Ioannidis et al., 2001). In contrast, the genome-wide association study (GWAS) represents a hypothesis-free approach to complex genetics, which has dominated the field from 2005 and onwards (Hirschhorn and Daly, 2005, McCarthy et al., 2008).

The methodology of GWAS is based on a case-control design where a vast number of single-nucleotide polymorphism (SNP) markers spread across the genome are genotyped in each subject. Advances in array technology in the 2000s enabled large-scale parallelization of genotyping, typically investigating from 500 000 up to several million SNPs on the same chip. These markers are selected to maximize the genetic information captured based on knowledge about common haplotype structures. Genetic variants that tend to be inherited together are in *linkage disequilibrium* (LD). Stretches of high LD typically occur in regions of low recombination, whereas recombination hotspots divide chromosomes into haplotype blocks. The pattern of LD across the genome in different populations has been systematically characterized through the human HapMap-project, providing a crucial foundation for marker selection to GWAS genotyping arrays (International HapMap Consortium, 2003, International HapMap Consortium, 2005). Consequently, although the functionally

relevant genetic variants may not be directly investigated, the association signal will often still be captured by virtue of high LD with a genotyped SNP, *tagging* the functional variant.

When allele frequency for a given SNP differs significantly between patients and healthy controls, it suggests that genetic variation affecting disease risk is present within the relevant LD block. However, given the multitude of SNPs genotyped in GWAS, association analysis may involve several million independent statistical tests, rendering the method vulnerable to false positives by random effects. Statistical significance levels in GWAS are therefore adjusted for multiple testing. A widely accepted threshold for genome-wide significance is $5x10^{-8}$. To achieve statistically robust results below this level, sample size is crucial. Thus, the establishment of large collaborative international consortia has been essential for successful GWAS.

Combining several GWAS in meta-analysis has emerged as an important strategy to boost discovery. Where the included SNPs may vary across original studies due to differing genotyping platforms, compatible data sets may be obtained by *imputation* (Marchini et al., 2007). Based on comparison with comprehensive phased sequence data from a relevant database source such as the 1000 genomes project, imputation employs bioinformatic algorithms to infer the most likely alleles for SNPs that are not directly genotyped. This process greatly increases the density of SNP data, improving the sensitivity to association signals even from low-frequency variants. However, imputation also introduces a possible source of bias, especially if the reference population is not representative for the imputed dataset.

1.5.6 Genome-wide association studies in Parkinson's disease

Initial GWAS in PD were underpowered to detect signals passing a stringent significance threshold (Maraganore et al., 2005, Fung et al., 2006, Evangelou et al., 2010). In 2009, the first genome-wide significant hits in PD were reported, from Caucasian and Japanese populations respectively (Simon-Sanchez et al., 2009, Satake et al., 2009). The findings confirmed *SNCA* as a risk-locus in both populations, and *MAPT* in Caucasians. In addition, three further loci were identified in the Japanese study, including *BST1*, PARK16 and *LRRK2*. Several more GWAS studies followed,

partly replicating these findings, and also adding novel loci to the list such as *GAK* and *HLA* (Hamza et al., 2010, Edwards et al., 2010, Saad et al., 2011, Spencer et al., 2011).

The first PD GWAS based on meta-analysis and imputation was published by the International Parkinson's Disease Genomics Consortium in 2011 (Nalls et al., 2011). It was followed by several similar studies (International Parkinson's Disease Genomics Consortium and Wellcome Trust Case Control Consortium, 2011, Pankratz et al., 2012), and also a GWAS using recruitment of subjects over the internet to obtain a large sample size (Do et al., 2011). Collectively, these efforts greatly expanded the list of nominated risk loci. The cumulative increase in genome-wide significant association signals for PD between 2009 and 2012 is summarized graphically in Paper 1, Figure 1. At the outset of the work described in this thesis, 20 GWAS loci were identified in PD. Seven of these had not been investigated further in independent follow-up studies.

1.5.7 Challenges and unresolved issues for the genetics of sporadic Parkinson's disease

Large-scale GWAS have represented a breakthrough in PD genetics by linking a range of previously unknown loci across the genome to disease risk by means of robust statistical association. However, considerable knowledge gaps need to be filled before translational research and clinical medicine can benefit significantly from such a list of genetic loci. Dedicated follow-up studies, building upon the findings from GWAS, are warranted in order to address a number of unresolved issues.

The variability in disease risk that is accounted for by significant GWAS loci constitutes only a small fraction of the total estimated heritability of PD. This phenomenon, commonly referred to as "missing heritability", has been a general feature of the GWAS era in complex genetics (Maher, 2008, Manolio et al., 2009). A range of hypotheses have been proposed to suggest possible sources of genetic variability not readily detected in GWAS, such as rare variants, structural variation, allelic heterogeneity, gene-environment interaction and gene-gene interaction, also termed epistasis. Dedicated methodologies and study designs must target these

hypotheses in order to gain a comprehensive understanding of the genetic architecture of PD.

A major aim of GWAS is to identify disease-relevant genes. However, associated SNPs are typically located in non-coding genomic regions, within haploblocks of high LD that often encompass several genes. Although one might sometimes speculate based on available knowledge about protein function, it frequently remains an open question precisely which gene is responsible for an association signal. This uncertainty represents a barrier towards modelling of disease pathways and functional studies linking genetic findings to pathogenesis (Ioannidis et al., 2009). Refining our understanding of implicated genes and causal mechanisms behind GWAS signals will be an important step before the findings from large-scale genetic studies can be of greater benefit to neighbouring fields of research and ultimately inspire translational studies aiming for improved therapy.

2. AIMS OF THE STUDY

The principal aim of genetic studies in sporadic PD is to uncover important insights about pathogenic mechanisms, with improvement of therapeutic strategies as the ultimate goal. The aim of the present work was to refine and expand our understanding of the association signals reported in GWAS of PD. This was sought, firstly, by replication of GWAS signals in a Scandinavian population and development of an effective pooled sequencing study design suitable for variant screening in GWAS loci. Subsequently, we concentrated on two specific loci, performing fine mapping and targeted resequencing of the PARK16 locus and assessing GWAS SNPs in relation to DNA methylation and mRNA expression in *SNCA*.

Aim of paper 1: In this paper, we sought to contribute further evidence regarding genetic loci reported in GWAS of sporadic PD and examine the distribution of association signals in a relatively homogeneous Scandinavian population.

Aim of paper 2: In this paper we aimed to develop and report an effective study design for targeted resequencing in large patient cohorts. We investigated monogenic PD genes as an example target to demonstrate the effectiveness and reliability of the method, yet proposing that the same approach is equally suitable for examination of GWAS loci.

Aim of paper 3: In this paper, we aimed to shed further light on the PARK16 association signal. We combined the sequencing approach described in Paper 2 and genotyping of a fine-mapping SNP panel to identify both common and low-frequency variation potentially affecting disease risk. Based on our data we explored previously proposed hypotheses about the PARK16 locus.

Aim of paper 4: In the last paper, we hypothesized that susceptibility SNPs in *SNCA* might correlate with DNA methylation in the promoter region, previously shown to be decreased in PD. We tested this hypothesis by investigating GWAS SNP genotypes, methylation levels and mRNA expression levels in blood and brain tissue.

3. SUMMARY OF RESULTS

Paper 1. Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease

The list of PD risk-loci reported at genome-wide significance levels has expanded steadily from 2009 and onwards. In particular, the larger datasets achieved by imputation and meta-analysis of multiple smaller GWAS have led to higher numbers of positive findings. To further increase the evidence related to various GWAS-linked loci in PD and examine the distribution of association signals in a relatively homogenous Scandinavian sample set, we performed a replication study of 1345 PD patients and 1225 control subjects from Norway and Sweden. We genotyped a panel of SNPs representing 18 loci previously reported at genome-wide significance levels, as well as four near-significant, suggestive loci. Testing for association in a logistic regression model, we replicated 11 loci at nominal significance level ($p<0.05$) (*SNCA*, *STK39, MAPT*, *GPNMB*, *CCDC62/HIP1R*, *SYT11*, *GAK*, *STX1B*, *MCC1/LAMP3*, *ACMSD* and *FGF20*). Stratifying patients according to a cumulative genetic risk profile across all these loci, we found a threefold increase in PD risk for the highest versus the lowest score quintile.

Paper 2. Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease

While the GWAS design is appropriate for the common disease-common variant scenario, resequencing studies will be necessary to clarify the role of rare variants in sporadic PD and other complex disorders. In this article, we present a study design for cost-effective variant detection in large cohorts by targeted deep sequencing of DNA pools. We investigated samples from 387 PD patients, sequenced in pools representing 10 individuals each. A 200kb custom HaloPlex target enrichment kit was used to capture DNA from coding exons of 71 genes relevant to PD and other neurodegenerative disorders, including all PD GWAS loci. In this paper we focused on the analysis of six genes implicated in Mendelian PD, emphasizing quality metrics and evaluation of the method. We detected a total of 17 rare, nonsynonymous variants in

these genes and observed a sensitivity of 97% and specificity of 94% in our experiment when pooled sequencing results were validated against individual genotyping and Sanger sequencing.

Paper 3. Fine mapping and resequencing of the PARK16 locus in Parkinson's disease

The PARK16 locus, entailing five genes on chromosome 1, was among the first regions to be implicated in sporadic PD through GWAS, originally in a Japanese study. The implicated gene(s) and location of functionally relevant variants remain currently unknown, although a few hypotheses have been presented, including a proposed gene-gene interaction between risk alleles of the PARK16 gene *RAB7L1* and *LRRK2*. We genotyped a fine-mapping panel of 17 SNPs spanning the PARK16 locus in the case-control sample set available from the Paper 1 study and found supportive evidence for an association signal near the 5' region of *RAB7L1*, both in single-SNP and sliding window haplotype analysis. As this signal is in low LD with the top-hit from the largest meta-analysis in Caucasians, we argue in favour of allelic heterogeneity. In explorative epistasis-analysis we found suggestive support for an interaction with *LRRK2*. In pooled sequence data from the experiment described in Paper 2, we found no evidence of rare coding variants contributing to the PARK16 association signal.

Paper 4. Parkinson's disease correlates with promoter methylation in the α-synuclein gene

Given the pivotal role of *SNCA* in PD pathogenesis, we aimed to increase the understanding of common association signals at this locus. Previous studies have found evidence that DNA methylation levels of a putative promoter in *SNCA* intron 1 correlate with PD. We hypothesized that methylation is affected by genetic variation, possibly representing a functional mechanism mediating the effects detected in GWAS. We genotyped GWAS SNPs and assessed methylation levels by methylation-sensitive restriction enzyme (MSRE) digestion and quantitative PCR (qPCR) in a case-control

blood sample set (N=72). Intron 1 methylation levels were decreased in PD compared to controls and significantly associated with genotype of a GWAS SNP near the 5' end of *SNCA*. The methylation quantitative-trait locus (mQTL) finding was replicated in a smaller sample set from cerebral cortex (N=24). We detected no significant associations with *SNCA* mRNA levels.

4. METHODOLOGICAL CONSIDERATIONS

4.1 Subjects

Recruitment of study subjects and collection of blood samples to a registered biobank for Parkinsonism at Oslo University Hospital have been performed from 2007 and onwards. Patients were asked to participate in the study during clinical visits at two different hospitals within the larger Oslo region. The majority of patients were recruited from the Department of Neurology at Oslo University Hospital (mainly M. Toft), a tertiary care centre for movement disorders, where patients are referred mostly for diagnostic second opinion or evaluation for advanced treatment options, including DBS. Consequently, this patient group is carachterized by an overrepresentation of early onset and severe motor fluctuations, yet better preserved cognitive function as compared to the average PD cohort. The remaining fraction of PD subjects was included at Drammen Hospital (K. A. Bjørnarå). This is a secondary centre where the age distribution and symptomatology of PD patients reflect a general neurological outpatient practice. Control subjects have been recruited among spouses of patients, volunteers from Rotary clubs in the Oslo area and outpatients in primary care without neurological disease (mainly L. Pihlstrøm). A total of 405 patients and 464 control subjects from Oslo (including Drammen) were included in the final analyses of Paper 1 and Paper 3. The blood samples used in Paper 2 and Paper 4 both represent subsets of this same case-control cohort.

For Paper 1 and Paper 3, four other centres in Norway and Sweden were invited to contribute samples to a collaborative Scandinavian association study. Samples from Oslo, the ParkWest study in Western Norway, Gothenburg, Linköping and Umeå combined into a total of 1380 PD patients and 1295 controls. Details of recruitment and demographic characteristics across study sites are summarized in Paper 1 (Table 1 and Table 2). The variability in age at onset of PD reflects how patients were recruited in different clinical settings. The Oslo series, dominated by a highly specialized clinic, had the earliest mean age at onset, 54 years. At the other extreme was the ParkWest cohort, with average onset at 67 years. This is an incidence-

based study, designed to include a representative cohort of novel cases from four counties in southwestern Norway (Alves et al., 2009).

Strict definition of phenotypes and rigorous patient ascertainment are important ideals in clinical genetic research. Several sets of diagnostic criteria for PD are currently in use. The ParkWest study used the criteria by Gelb et al. (Gelb et al., 1999), whereas the other four sites based diagnosis upon a variant of the United Kingdom Parkinson's Disease Society Brain Bank Criteria (Gibb and Lees, 1988). Regardless of the criteria of choice, even in the hands of expert neurologists the concordance between a clinical diagnosis of PD and classic Lewy pathology on post mortem examination is far from perfect (Litvan et al., 1998). This diagnostic inaccuracy represents a limitation that is likely to reduce the statistical power in genetic association studies. Conversely, there may also be subjects included in the control group which eventually go on to develop PD, or have incidental Lewy body disease without symptoms. However, given that PD patients represent a minor fraction within a random sample of elderly, such misclassification of subjects can be expected to be tolerably low (Moskvina et al., 2005). The control subjects in the present study were all above 40 years of age and did not have any known neurological disease or first-degree relative with Parkinsonism. Ideally, clinical neurological examination of all controls might have excluded some subjects based on subtle clinical signs, yielding a "supernormal" control cohort and theoretically slightly improved statistical power.

In addition to blood samples from Norway and Sweden used in Papers 1-4, Paper 4 also included 24 samples of human frontal cortex, provided by the Banner Sun Health Research Institute.

4.2 Ethical considerations

All participants gave written, informed consent. The study was approved by the Regional Committee for Medical Research Ethics (Oslo, Norway), and sample and data collection at each study site was approved by local ethics committees. The consent form for the Oslo study site covers research into genetic causes of Parkinsonism. Patients were actively informed that genetic analyses would be performed for research purposes only, and that no results would be given back. Only

personnel directly involved in subject recruitment have had access to patient identity. Anonymized IDs were used in all sample handling, experiments and analyses, and data are stored on secure servers dedicated to research databases at Oslo University Hospital.

4.3 DNA and RNA extraction and quality measures

A range of different methods are currently available to extract quality DNA from blood and body tissues, and methodological details are often omitted from scientific papers. However, degraded DNA or highly variable input concentrations may adversely affect the performance of genetic analyses. In the Scandinavian dataset of Paper 1 and 3, 105 samples failed our individual call rate threshold of 0.80 and were excluded from analyses, likely reflecting diluted or low-quality DNA. Ideally, rigorous reassessment of DNA quality could have prompted us too seek out fresh samples from these individuals. The sample set from the ParkWest study and the majority of samples from Umeå were extracted in-house by L. Pihlstrøm by QIAamp DNA Blood Mini kit, yielding good quality DNA and high genotype call rates.

Quality measures are yet more crucial for quantitative RNA analyses. For the gene expression experiments described in Paper 4, L. Pihlstrøm assessed the purity of RNA samples measured as A260/A280-ratios on a Nanodrop spectrophotometer (Thermo Scientific, Waltam, MA, USA). RNA integrity number (RIN) was measured on a Bioanalyzer instrument (Agilent, Santa Clara, CA, USA) by A. Rengmark. Samples extracted from frozen post-mortem tissue expectedly showed lower RIN values than blood RNA collected in PaxGene tubes. Brain sample RIN values in the range of 4.1 to 8.2 indicate RNA integrity below the ideal, yet comparable or even superior to reported figures from large studies in similar tissue (Trabzuni et al., 2011). RIN values showed no correlation with disease status or gene expression. Consequently, variability in RNA quality would be more likely to introduce random noise and type 2 error than systematic bias and type 1 error (see section 4.7 Statistical methods).

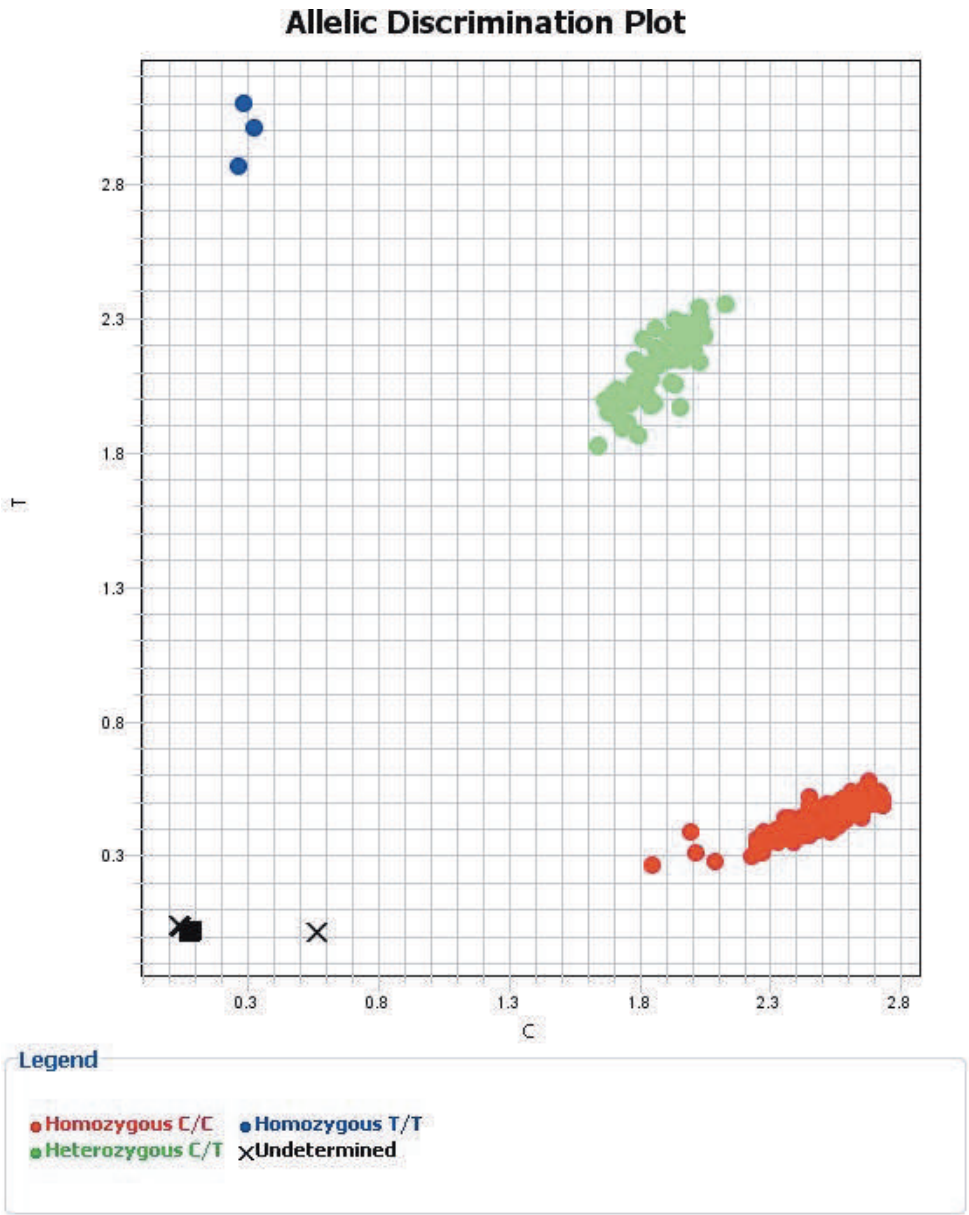
4.4 Genotyping

SNP genotyping was performed locally by TaqMan (Paper 1) or KASPar (Paper 3 and 4) technology on a ViiA7 instrument (Life Technologies, Foster City, CA, USA) or by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry using the Sequenom platform (Sequenom, San Diego, USA) at the Cigene core facility, Norwegian University of Life Sciences. Sequenom technology is based on the principle of allele specific primer extension reactions, where bases incorporated are separated by mass. The platform allows for multiplex reactions suitable for medium-throughput genotyping projects. SNPs of interest for Paper 1 and 3 were submitted to Cigene, and multiplex design and genotype calling were performed at the facility according to default software parameters. For SNPs that came to our attention at later stages, or where Sequenom reactions were not designable, genotyping was performed by L. Pihlstrøm using TaqMan or KASPar assays. These are both based on PCR-reactions where allele-specific amplification is detected by fluorescence. In TaqMan assays, a fluorescent signal is released when DNA polymerase cleaves off the dye from the quencher of an allele-specific Minor Groove Binding (MGB) probe. KASPar technology uses a similar principle, except the fluorescent dyes are incorporated directly into a set of allele-specific PCR primers and released from the quencher when amplification generates double-stranded DNA. The results from genotyping experiments can be visualized through allelic discrimination plots (Figure 2).

In all studies involving genotyping, we set a genotyping rate of 95% as a quality benchmark. As described in Paper 1, two SNPs failed to pass this criteria. One (rs356220) could be replaced in analysis by a different *SNCA* SNP, whereas the other (rs3129882) was regenotyped to achieve an acceptable call rate. All genotyping experiments included negative controls. We also analysed 87 samples as duplicates on the Sequenom platform without observing any discrepant genotype calls. The Hardy-Weinberg equilibrium (HWE) is an equation that describes the relative frequencies of heterozygotes and homozygotes, assuming an outbred population. Departure from HWE may indicate genotyping error and is commonly used as a quality parameter in genetic association studies. In patients however, a true effect of genotype on disease

risk might contribute to HWE deviations. We assessed HWE in controls for all genotyped SNPs, observing no significant departure at a threshold of $p<0.01$.

Figure 2. *Allelic discrimination plot from TaqMan genotyping (rs11248051 in GAK)*



4.5 Methylation and expression analysis by quantitative PCR

In Paper 4, we performed qPCR on a ViiA7 instrument (Life Technologies, Foster City, CA, USA) to assess *SNCA* mRNA expression and intron 1 CpG methylation. Both experiments were designed and performed by L. Pihlstrøm. A range of methodological caveats is relevant in the context of qPCR gene expression experiments. Comprehensive guidelines have been published proposing a set of Minimum Information for publication of Quantitative real-time PCR Experiments (MIQE Guidelines) (Bustin et al., 2009). The common practice in scientific articles however, varies considerably. Firstly, it is appropriate to consider whether the sampled tissue is relevant for the research question of interest. We studied *SNCA* expression in a small, underpowered sample series (12 PD patients, 12 controls) from cerebral cortex. Tissue blocks used for RNA extraction will necessarily contain a mixture of cell types, introducing a possible source of noise to the expression data. The blood sample set was larger (36 PD patients, 36 controls), but only relevant under the assumption that aspects of gene regulation are equal across blood and nervous tissue. Levels of both α-synuclein protein and mRNA in blood have been studied in PD previously with variable results (Pihlstrom and Toft, 2011). *SNCA* mRNA is present in high levels in reticulocytes, but also expressed in white blood cells (Scherzer et al., 2008). In Paper 4, we collected full blood in PAXgene tubes, while blood cell counts were not recorded. Consequently the potential of noisy data arising from sampling a mixed cell population equally applies to the blood series.

RNA extraction is described in the previous section. RNA concentrations were normalized before reverse transcription. We used Applied Biosystems High Capacity Reverse Transcription kit according to the manufacturers instruction for cDNA synthesis. All gene expression assays used had primers crossing exon borders. A set of reverse transcriptase negative controls was run to ensure absence of amplification from genomic DNA. For the gene expression experiment, reactions were performed in technical triplicates, and any outliers in replicate groups were excluded from analysis.

In qPCR, gene expression is measured relative to a set of reference genes. Reference genes, also termed housekeeping or normalization genes, are ideally transcribed at a stable rate across cells. Normalizing the gene of interest mRNA levels

to reference genes will compensate for the variability in input RNA. Which genes are suitable as normalizers will vary across tissues. It is currently widely accepted that normalization in qPCR experiments should be based on the combination of several housekeeping genes. Ideally, the optimal set of reference genes should be empirically determined for each individual study prior to the experiment itself (Vandesompele et al., 2002). In practice, this may often prove too resource demanding, in which case reference genes need to be carefully selected based on published literature. In Paper 4, *GAPDH* and *HPRT* were chosen for normalization in whole blood, minding the high levels of *SNCA* in reticulocytes (Silver et al., 2006). For brain normalization we used *GAPDH*, *SDHA* and *HMBS* as calibrators (Coulson et al., 2008). The normalization genes in qPCR expression experiments all showed pairwise coefficient of variance < 0.5 and standard deviations of normalization factors were < 2-fold from the mean in both data sets, indicating stable expression (Vandesompele et al., 2009). There was no difference in housekeeping gene expression between groups, and RIN values did not correlate with expression of any gene. Results were analysed by the ΔCT-method using the geometric mean of all housekeeping genes as normalization factor. One out of several approaches to the analysis of qPCR data, the ΔCT-method relies on the assumption that amplification efficiency is near 100%, corresponding to a doubling of the PCR product with every cycle (Schmittgen and Livak, 2008). Commercially available gene expression probes as used here, have been validated for this method. For the MRSE qPCR experiment, a standard curve was performed to demonstrate adequate amplification efficiency.

DNA methylation is an epigenetic modification of the cytosine base, occurring in relation to CpG dinucleotides. Methylation-sensitive restriction enzymes will recognize a short nucleotide sequence involving a CpG site and cleave the DNA strand *only if* the cytosine is unmethylated. If a qPCR targeting a differentially methylated region is subsequently performed, in parallel with a reference reaction of the same sample without restriction digestion, the fraction of cleaved versus uncleaved gDNA molecules can be calculated. This is the principle behind the method used to quantitate *SNCA* intron 1 methylation level in Paper 4. Issues related to target selection and primer design are detailed in the article and supplement. The reactions were performed in technical duplicates and analysed according to the protocol for the OneStep

qMethyl kit (Zymo Research, Irvine, CA, USA). It should be noted that the SYTO9 fluorescent dye used to detect amplification in the MRSE qPCR is not target specific, but emits a signal in the presence of any double-stranded DNA in the reaction well. Primer design must be optimized to uniquely target the genomic region of interest. To further exclude the possibility of non-specific amplification, a melt-curve was performed at the end of the experiment, reflecting the composition of double-stranded DNA amplification products in the sample. Our initial test reactions showed an additional smaller peak indicating non-specific amplification. With optimization of the PCR cycling temperatures we obtained single peaks in all reactions (Figure 3).

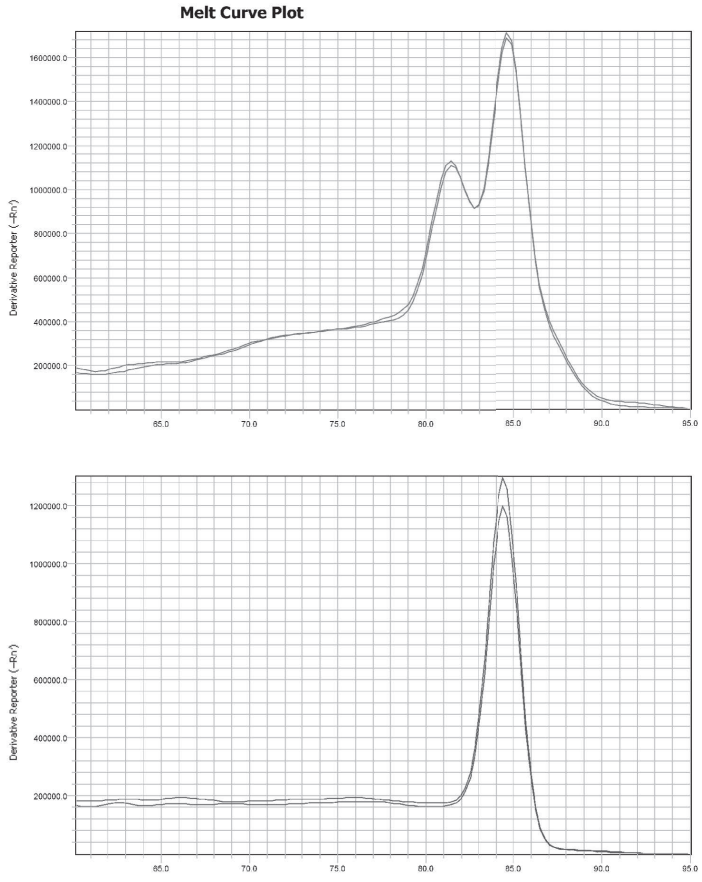


Figure 3. *Melt curve analysis of the SNCA intron 1 PCR product from the same sample duplicate before and after optimization of cycling temperature*

MRSE represents a rapid and affordable approach to quantitation of DNA methylation levels in a defined candidate region, but the method also has clear limitations. The methylation-sensitive restriction enzymes should preferably target several potential cut sites within the amplified region, in our case four motifs. The measured results however, will not give any information about each specific CpG site. The effect of anywhere between one and four cuts to a given molecule will all equally result in non-amplification of that particular fragment. In addition there will typically be many more CpG sites of potential interest in the region, which are not recognized by any of the enzymes used. Consequently, MSRE could be regarded as a low-resolution screening tool for methylation assessment. Ideally, we might have included traditional bisulfite conversion and sequencing as a gold standard reference for at least a subset of samples in Paper 4. While the necessary resources for this were not available to our project, we note that the method has proved reliable in previous studies when methylation results have been validated by bisulfite sequencing in a subset of samples (Holemon et al., 2007, Hua et al., 2011).

4.6 High throughput sequencing

Paper 2 reports an experiment where we performed targeted resequencing in DNA pools, with emphasis on methodology and study design. In Paper 3, this same approach is applied to investigate the coding exons of five genes at the PARK16 locus. Study design for the pooled sequencing experiment was done mainly by M. Toft with contributions from L. Pihlstrøm. DNA pooling, target enrichment and library preparation were performed by A.Rengmark. The sequencing was performed by the core facility at the Norwegian Sequencing Centre, Ullevål. L. Pihlstrøm did all bioinformatic analyses and was responsible for the interpretation of the results.

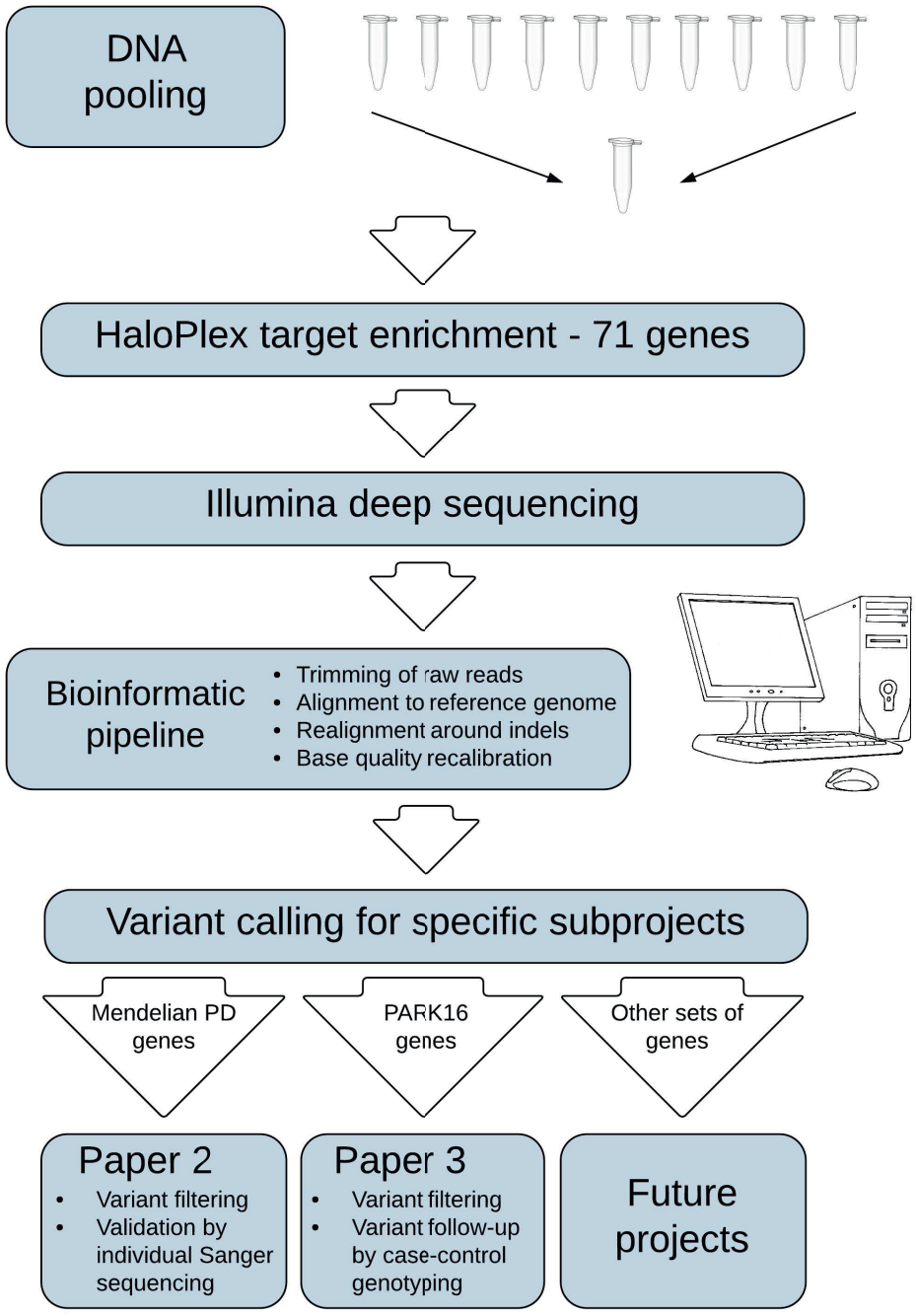
High-throughput or "next-generation" sequencing (HTS or NGS) has revolutionized genetic research (Mardis, 2011). The crucial driving principle behind these technical developments has been the large-scale *parallelization* of reactions, taking place simultaneously on tiny glass plates or arrays while data is recorded. The major breakthrough of HTS technologies has allowed for generation of data on a

previously unimaginable scale. However, the limited availability of standardized, convenient tools for bioinformatic processing and interpretation of sequence data has arguably created a new bottleneck for genetic research.

Traditional Sanger sequencing (Sanger et al., 1977) is still regarded the gold standard for both sensitive and specific sequence analysis. The error rates of different HTS technologies are generally somewhat higher, the major advantage being the high throughput. The bioinformatic processing of HTS data is to a large extent concerned with a range of quality metrics, all carrying important information about the completeness and reliability of the data. For any project using HTS, care must be taken to set up an appropriate bioinformatic pipeline, tailored to the relevant data and research questions. A range of methodological issues regarding DNA pooling, the choice of DNA capture technology, bioinformatic processing and quality metrics are discussed in detail in Paper 2. An overview of the study design is provided in Figure 4.

While we argue that our pooled sequencing design represents a cost-effective and rational approach to targeted resequencing of larger sample cohorts, the work also has clear limitations. We reported a sensitivity estimate of 97% based on the assessment of eight SNPs present as single-alleles in a given pool in a total of 59 instances. Failure to call the variant from pooled sequencing was observed only twice, both times in pools where the relevant position was read less than 80X, corresponding to < 4X per allele on average. Specificity was assessed by Sanger sequencing of pools positive for possibly pathogenic non-synonymous, low-frequency variants, where 17 out of 18 SNPs were confirmed. While these results indicate good sensitivity and specificity, the number of investigated variants is too low for the findings to be confidently generalizable. Ideally, we should have obtained more comprehensive genotype data by an independent method, such as individual sequencing, to allow for statistically robust calculations of sensitivity. We might also have validated a larger number of detected variants within the whole target region of 71 genes, not only those that were possibly pathogenic within Mendelian PD genes.

Figure 4. *Schematic overview of the study design described in Paper 2*



DNA pooling is a strategy motivated by the budget constraints faced by almost any scientist concerned with genetic research. A range of considerations apply regarding the overall study design, the number of individual samples combined in each pool, the tools used for analysis and complementary experiments to validate the results (Schlotterer et al., 2014). For accurate frequency estimation of common variants, large pools are preferred, as the inaccuracy in input DNA from each sample will tend to even out. In the context of rare variant discovery, it is of crucial importance that the sequencing error rate remains well below the expected proportion of reads represented by a single heterozygous variant in the pool (Figure 5). Consequently, we chose relatively small pools in our experiment. To confirm the individual carrier status of interesting variants detected in pooled sequencing, we performed Sanger sequencing of all 10 individual samples in the pool (Paper 2). The workload of this validation stage also weighed in favour of smaller pools. Primer design, PCR amplification, visualization of the PCR product by gel electrophoresis and data analysis in Sequencher 5.1 software (Gene Codes Corporation, Ann Arbor, MI, USA) were performed by L. Pihlstrøm, while the sequencing itself was done as a service from a commercial provider.

4.7 Statistical methods

All statistical analyses in Paper 1-4 were performed by L.Pihlstrøm.

4.7.1 Statistical power

In statistical hypothesis testing, two different scenarios may give rise to false conclusions. The incorrect rejection of a true null hypothesis is traditionally termed Type 1 error, while Type 2 error designates the failure to reject a false null hypothesis (Neyman and Pearson, 1928). When designing a genetic association study, it is of crucial importance to consider whether the expected sample size will give sufficient statistical power to answer the research question of interest (Watanabe, 2011). An

arbitrary, yet widely used threshold for acceptable power is 80%, implying that in the presence of a true disease association, the probability of observing significant results and rejecting the null hypothesis should be at least 80%. When the statistical power is lower, a negative result has very little value and may easily appear misleading, given the high probability of Type 2 error (false negative). In genetic association studies, the factors that determine statistical power are sample size, significance threshold, allele effect size, allele frequency, disease prevalence and genetic model.

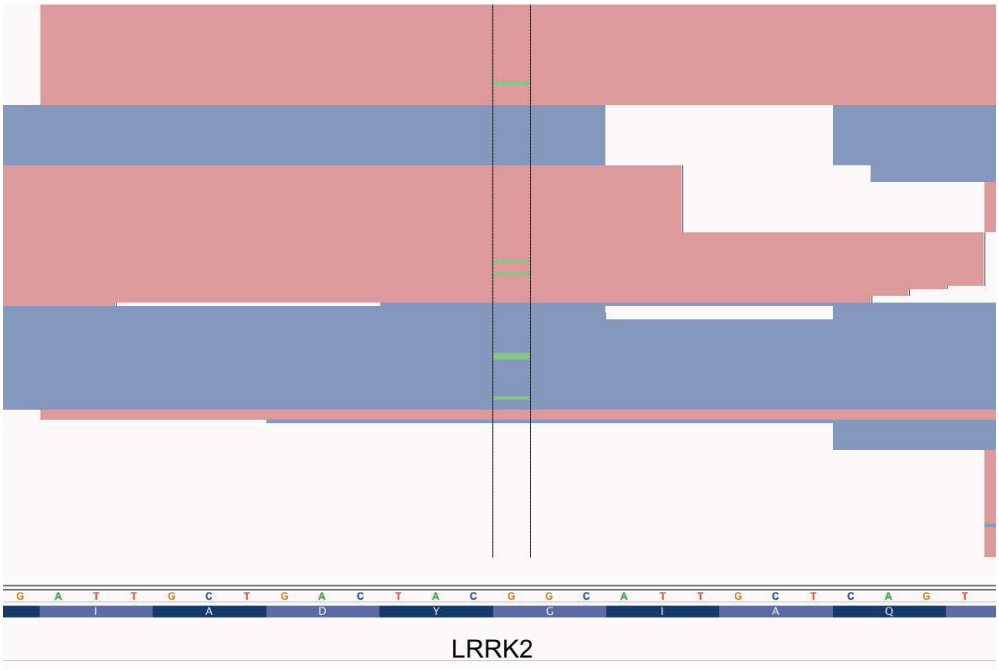


Figure 5. *Visualization of variant reads from pooled sequencing in the Integrative Genome Viewer. The alignment to the reference genome of paired end reads in both directions are shown in pink and blue respectively. A non-reference base G>A appears in green on approximately 5% of reads, representing a single individual in the pool heterozygous for the LRRK2 G2019S mutation. The displayed reads are downsampled from the actual data.*

In Paper 1, power calculations were performed using the Genetic Power Calculator (Purcell et al., 2003). We investigated a total of 22 loci. Based on the odds ratios (ORs) published in relevant GWAS, combined with allele frequencies from the present study and an assumed PD prevalence of 1%, our study had > 80% statistical power for only 10 loci. This represents a major weakness for our replication experiment, particularly limiting the ability to draw any firm conclusions regarding the majority of non-significant SNPs. We have sought to openly highlight these power issues both in Table 3 and the discussion section of Paper 1. Although the limitations in statistical power raise important concerns, there are also a few relevant considerations that may serve to justify the study design. While the statistical power was fully adequate for only 10 loci, another nine had estimated power between 50% and 80%. Of the 18 significant GWAS hits we aimed to replicate, power was <50% for only one SNP. Consequently, our study could be expected to demonstrate a number of replications and arguably provide a valid general picture of GWAS associations in a Scandinavian population, even though a number of non-significant loci might be false negatives. Furthermore, combining association results from a range of different studies in meta-analyses represents an important approach in genetic association research. In this context, a single study like ours is only one small step towards a robust international corroboration of a final list of susceptibility loci for PD. We provided a table with raw allele frequencies for all SNPs in patients and controls, suitable for meta-analysis, in the supplement to Paper 1.

It should be noted that the loci which did not replicate in Paper 1 also included several SNPs were the estimated power to detect association was excellent. In these instances, there must exist a different explanation for the negative result. This observation motivated our extended investigation of the PARK16 locus, as reported in Paper 3. In Paper 4, no formal power analysis was performed. DNA methylation and gene expression are continuous variables, which generally tend to give better statistical power in small samples than binary outcomes.

4.7.2 Association testing

Statistical hypothesis testing reported in Papers 1 and 3-4 were performed partly with the open software package PLINK (Purcell et al., 2007), and partly with IBM SPSS Statistics v.21 (Armonk, NY, USA). As each individual carries two alleles of each variant, association testing in genetic case-control studies traditionally involves choosing a genetic *model* for comparing the observed data across groups. The classical models are genotypic (AA vs. Aa vs. aa), dominant (AA vs. Aa+aa), recessive (AA+Aa vs. aa) and allelic (A vs. a). For all work presented in this thesis, we assumed a simple allelic model, with multiplicative (or log-additive) effects of a given allele in homozygous individuals.

The most straightforward approach to association testing is the chi-square statistic, assessing crude differences in allele frequency between patients and controls. We used this test for SNP association in Paper 3. An alternative for statistical hypothesis testing with binary outcomes is logistic regression. A great advantage of logistic regression is that it allows covariates to be included and accounted for in the statistical model. In Paper 1, we performed logistic regression analysis with age and gender as covariates. This may be considered a somewhat more conservative approach than the chi-square test. Variability in age and gender across patient and control groups reflects the study inclusion practice of the collaborating centres. Unless migration patterns have created genetic differences between age strata in the relevant population, one should not expect the observed allele frequencies to be dependent on age or gender. Nevertheless, including these basic demographic covariables is common practice in genetic association studies, probably due to the large impact they may have in epidemiological studies concerned with environmental risk factors.

The *confounding* phenomenon represents an important factor affecting the validity of association results in epidemiological research. Variable X may appear to affect a given disease of interest, when in fact, the association is driven by variable X being dependent on a second variable Y. The confounding of X by Y may be adjusted for in a regression model, yet if Y is not recorded, the study runs the risk of reporting a spurious association (Type 1 error). Genetic association studies represent a special case in this respect. As an individual's genome generally remains constant throughout

life, the potential of confusing cause and effect is limited in genetic studies. However, a problem may arise if patients and controls are recruited from different population strata with distinct genetic backgrounds. This is the genetic association study's version of confounding, commonly termed *population stratification* (Freedman et al., 2004). In GWAS, the genome-wide SNP dataset does in itself provide information about an individual's genetic background. This opens up a range of opportunities to correct for possible population stratification in GWAS (Zheng et al., 2006, Tian et al., 2008, Yang et al., 2011b), such as genomic control, principal component analysis or multi-dimensional scaling. For studies including a limited number of markers, such as reported in Paper 1 and 3, this option is not available. To minimize the risk of population stratification, we excluded individuals of known non-Scandinavian ancestry. Our study included an uneven proportion of PD patients and controls from each collaborating site. Previous evidence has pointed to local genetic heterogeneity within Sweden (Humphreys et al., 2011), indicating that caution must be paid in the design and interpretation of genetic studies. However, in the context of larger association studies, the Scandinavian peninsula may reasonably be considered relatively homogeneous compared to other sample cohorts broadly recruited from Caucasian populations (Novembre et al., 2008). In Paper 1, we reran the analysis including study site in the logistic regression model and obtained similar results. We also performed Breslow-Day test for heterogeneity of association signals across study sites without significant SNPs. Nevertheless, it is not possible to fully exclude the possibility that population stratification may have affected the statistics somewhat.

Where the case-control design is concerned with binary outcomes, Paper 4 also examined association with numerical outcome variables, namely DNA methylation and mRNA expression levels. For the blood sample analysis (N=72) we performed linear regression with gender and age as covariates, whereas Spearman's rank correlations were calculated for the smaller brain data set (N=24).

4.7.3 Multiple testing and significance threshold

The issue of multiple testing is a major concern for the interpretation of quantitative research data (Ioannidis, 2005). As *p*-values represent the probability of the observed

data occurring by chance under the null hypothesis, seemingly "positive" results will expectedly follow from merely stochastic processes whenever a large number of hypotheses are tested. As a consequence, reliable interpretation of research data relies on appropriate strategies to correct for the multiple testing involved in a given project (Johnson et al., 2010, Rice et al., 2008). A significance level should be defined prior to data analysis, based on the nature of the research question and the statistical hypothesis testing context.

An intuitively appealing and widely applied approach to significance level adjustment is Bonferroni correction, whereby the nominal significance threshold (classically p=0.05) is simply divided by the number of independent tests performed. This strategy has dominated the GWAS field, until a general consensus has more or less established $p=5x10^{-8}$ as defining the "genome-wide significance threshold", regardless of the number of variants investigated (Panagiotou et al., 2012). Another relevant approach is permutation procedures, where the data is rearranged a large number of times, simulating the test distribution under the null hypothesis to generate an empirical p-value (Browning, 2008, Pahl and Schafer, 2010). In line with the principles of Bayesian statistics, the interpretation of test statistics is largely dependent on assumptions about the prior probabilities of a research hypothesis being true. For a random genomic locus in a hypothesis-free GWAS context, this probability is extremely small. However, in a replication study, where strong evidence has already nominated the candidate loci, the prior probability is much larger. For this reason, most GWAS follow a two-stage design, with an initial genome-wide discovery phase and a focused candidate replication phase (Kraft and Cox, 2008).

In the work presented in this thesis, we aimed to rationally balance the strict and liberal arguments when significance thresholds were set. In Paper 1 we investigated 22 independent loci. We decided to highlight and report results both at the Bonferroni corrected (p<0.0023) and the nominal (p<0.05) significance levels. These differed by four versus 11 positive loci. For the discussion and interpretation of the findings, we found it justifiable to regard nominally significant results as supportive of a true association, given the context of a replication study, where previous evidence indicate a higher prior probability of positive findings.

In Paper 3, we investigated 17 SNPs, yet a considerable degree of LD was observed across these markers. Consequently, the SNPs are not independent, and a Bonferroni correction strategy is not straightforward. Given that the PARK16 region is robustly corroborated as a risk-locus for PD, we considered the prior possibility of true association to be sufficiently high to justify a nominal significance threshold in the SNP analysis. This liberal interpretation represents a clear limitation, however, and the data provide only weak support for the PARK16 signal in the Scandinavian population. While single SNP analyses concern only two alleles, association testing on the haplotype level involves comparison across a larger number of haplotypes, introducing more degrees of freedom to the statistics. For the more complicated test scenario of the sliding-window haplotype analysis reported in Paper 3, we generated empirical *p*-values from 100,000 permutations, both across each haplotype window and for the overall experiment. The haplotype showing an experiment-wide significant empirical *p*-value (rs1572931-rs947211-rs1775143, p=0.046), also entailed a SNP reported as a genome-wide top-hit in previous work (Lill et al., 2012), strengthening the posterior probability of a true association. The statistical assessment of gene-gene interaction between PARK16 and *LRRK2* in Paper 3 should be regarded as exploratory and interpreted with caution as the set of tests to be performed was not clearly defined prior to the study, and results were marginal even without any correction for multiple testing.

In Paper 4, the statistical testing followed a two-stage design. Bonferroni correction was applied in the hypothesis testing for the larger blood data set, and the detected association between rs3756063 genotype and methylation level was subsequently replicated in brain at a nomimal significance level. Although sample size was generally small in this study, the concordant finding observed across independent sample sets strengthens the overall reliability of the results.

5. GENERAL DISCUSSION

5.1 Corroborating and refining a list of genetic risk-loci in Parkinson's disease

Over the recent years, results from GWAS have firmly established that multiple risk-loci across the genome affect an individual's risk of developing sporadic PD. This insight provides a strong fundament for further efforts to uncover the full genetic architecture of PD and explore functional hypotheses about disease mechanisms and pathways. In contrast to early genetic association studies of candidate genes, the signals reported at strict genome-wide significance levels have demonstrated high rates of confirmatory replication in independent follow-up studies. Paper 1 represented a contribution to this robust corroboration of GWAS findings in PD.

In parallel to the work described in this thesis, efforts to combine ever larger GWAS data sets for meta-analysis have resulted in the identification of further novel PD loci. The largest meta-GWAS in PD to date was published in 2014 and included more than 13,000 cases and 95,000 controls in the primary analysis (Nalls et al., 2014). Significant SNPs from the discovery phase as well as previously reported loci were genotyped in an independent sample set, where a total of 24 loci were replicated. In addition, four of these loci also showed evidence of a second, independent association signal when the effect of the SNP with minimum *p*-value was adjusted for in conditional analysis.

5.2 Missing heritability

Despite the progress achieved by large meta-GWAS, the current list of identified risk-loci is far from a full account of the heritable variability in PD risk. Two recent studies have applied genome-wide complex trait analysis (GCTA) to estimate PD heritability from GWAS data (Do et al., 2011, Keller et al., 2012). This statistical framework identifies the phenotypic variance explained by genome-wide SNPs, including those that are not significantly associated in a traditional case-control design (Yang et al.,

2011a). Both studies indicate a heritability of about 27%, where GWAS top-hits alone account for only 3-5%. Although six further loci have later been added to the list, these estimates still suggest that the major fraction of genetic risk in sporadic PD remains to be identified. Furthermore, GCTA will not take into account genetic variation that is not captured by GWAS genotyping arrays combined with imputation. Consequently, the true heritability may well be considerably higher than 27%, and the "missing" fraction accordingly larger. Epidemiological studies vary with respect to both methodology and heritability estimates, but at least some seem to support a higher contribution from genetic factors than the GCTA-analyses (Wirdefeldt et al., 2011).

5.2.1 Rare variants

A number of hypotheses have been proposed to explain the missing heritability phenomenon in complex disease genetics (Manolio et al., 2009). A subset of these relate to the effect size and allele frequency of genetic risk variants. The GWAS paradigm is based on the common disease-common variant hypothesis. Basically, the tested markers must be fairly common in order to reach statistical significance when alleles are compared across cases and controls. A larger sample size will improve power to detect low-frequency variants or variants carrying a small effect on risk. The meta-GWAS by Nalls et al. detected significant SNPs of minor allele frequencies (MAFs) down to 0.01 when ORs were strong (1.7-1.8). By contrast, common variants (MAF >0.1) were significant even with ORs as low as 1.1 (Nalls et al., 2014). Figure 6 summarizes the continuum of genetic contributions to disease, spanning from highly penetrant Mendelian mutations to common low-risk variants, highlighting possible sources of missing heritability.

The case of heterozygous *GBA* mutations demonstrates that rare coding variants may contribute to PD without causing a Mendelian inheritance pattern. Association signals near *SNCA* and *LRRK2* also indicate that the same gene may be implicated at several points along the effect-frequency axis with both coding and non-coding variants affecting risk. Consequently, genes located within GWAS regions

provide a natural starting point for further studies of non-synonymous variation contributing to PD risk.

Sequencing studies are expected to provide novel insights about the role of rare variants in complex diseases, including PD. HTS technologies have revolutionized genetic research and provided the means to produce vast amounts of data. Sequencing services are becoming more available and affordable, but high costs still represent a substantial barrier to large-scale sequencing studies. In Paper 2, we reported an effective study design based on high throughput sequencing and DNA pooling to investigate a panel of target genes in large cohorts. Subsequently, the same approach was used to screen for disease-relevant coding variation within the PARK16 locus in Paper 3. Our experiment did not identify low-frequency nonsynonymous variants of clear relevance to PD in PARK16 genes, indicating that the most important contribution to genetic risk at this locus depends on common, non-coding variation. These results need to be interpreted with caution, however, due to several limitations in study design. Although target coverage was relatively good, some regions were not captured, especially in *NUCKS1*, and relevant variants may have been missed. The quality measures reported in Paper 2 indicated that our study design had good sensitivity and specificity, yet we cannot fully exclude that PARK16 variants in regions of relatively low coverage may have passed undetected due to the pooling approach.

Furthermore, the number of patients sequenced was moderate, limiting our capacity to observe variant enrichment in our samples relative to database frequencies below a MAF of about 0.005. Ideally, a larger number of both patients and controls should have been sequenced to allow for a more direct and well-powered comparison between groups. A more comprehensive data set would also have allowed for more advanced statistical approaches, in particular aggregated analyses of rare variant "burden" on a gene level rather than single variant association. Such statistical approaches are challenging, as they involve several non-trivial choices and assumptions, yet rare variant burden testing is becoming an increasingly important tool along with the widespread application of HTS in genetic research (Sham and Purcell, 2014, Bansal et al., 2010).

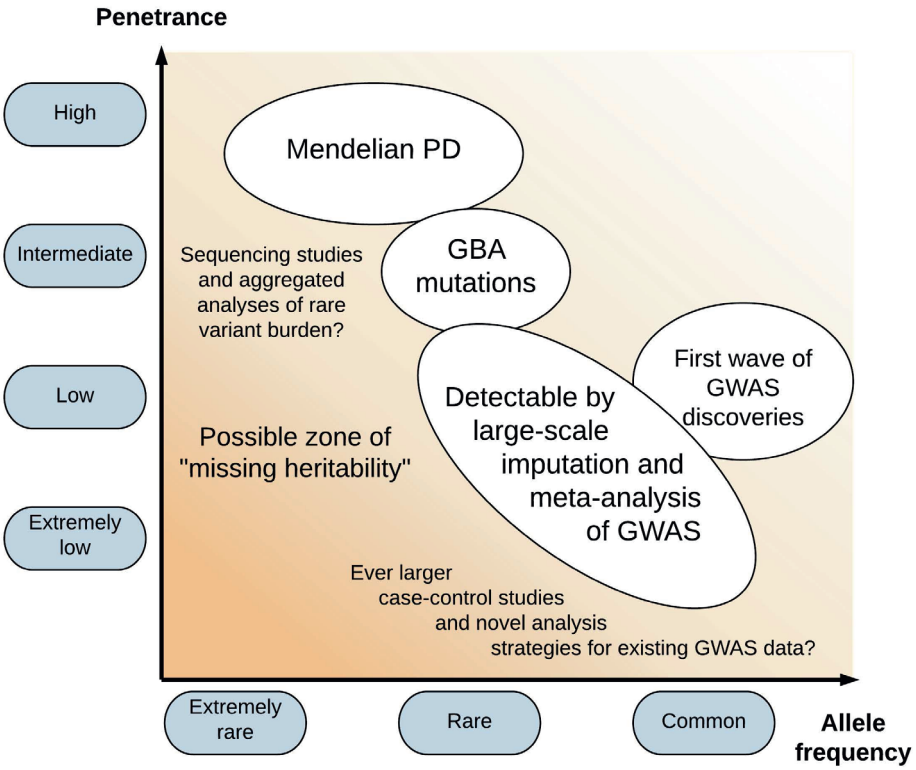


Figure 6. *The spectra of effect sizes and allele frequencies for genetic causes of PD. Inspired by McCarthy et al., 2008.*

Recently, several published studies have investigated rare variants in sporadic PD by targeted resequencing (Foo et al., 2014, Schulte et al., 2015). One study reported enrichment of rare functional variants in Mendelian PD genes, but not in genes at GWAS loci (Spataro et al., 2015). This finding is in line with our negative PARK16 result. Future collaborative studies involving sequencing and rare variant genotyping of much larger sample sets could be expected to further clarify the contribution of rare genetic variants to PD risk.

5.2.2 Allelic heterogeneity

The SNPs genotyped in GWAS each tag a set of variants in high LD at a given genomic locus. Any significant difference in tagSNP allele frequency is commonly ascribed to an unknown functional variant, driving the association signal. Allelic heterogeneity is present when more than one casual variant contribute to disease risk at the same locus. This phenomenon poses challenges to the traditional association study design and represents a possible source of missing heritability. The term "synthetic association" has been proposed for the situation where several rare causal variants are in LD with a single tagSNP (Dickson et al., 2010). In such cases, the tagSNP association will be determined by the combined contribution from all functional variants, weighted by effect sizes and strength of LD. If functional alleles with equal direction of effect tend to occur on the same haplotype, it will boost the observed association for a common tagSNP. Conversely, alleles with opposite effects on risk may neutralize each other and decimate the statistical power to detect a true disease association.

The most widely appreciated approach to assess allelic heterogeneity in a GWAS context is stepwise conditional analysis, where the lowest *p*-value SNP at a given locus is included as a covariate in statistical testing for independent secondary signals. In the recent meta-analysis by Nalls et al., four out of 24 loci showed evidence of secondary associations at the genome-wide significance level using this approach. However, statistical power is reduced for every step in conditional analysis. Alternative bioinformatic strategies to capture the full disease-relevant variation at a given locus have been proposed, yet these have not been applied to large data sets in PD (Yang et al., 2012, Ehret et al., 2012). Genes such as *SNCA* and *LRRK2* are already implicated in both Mendelian and sporadic PD, and recent studies have indicated that regulatory consequences of non-coding variation are widespread accross the genome (Thurman et al., 2012, Gaffney, 2013). It is therefore quite conceivable that important PD loci may harbour larger numbers of independent causal variants of different frequencies and effect sizes.

Paper 3 and Paper 4 of the present thesis are both concerned with allelic heterogeneity of PD GWAS loci. In Paper 3, we fine-mapped the PARK16 locus and

found evidence of association with PD for markers located upstream of *RAB7L1*. Comparing our results to other studies, we note that SNPs highlighted as significant in previous reports have frequently been in low LD with this signal, including the top-hit from the largest GWAS meta-analysis performed in Caucasians (Nalls et al., 2014). We interpret this as indicative of probable allelic heterogeneity. However, robust significance of secondary signals in stepwise conditional analysis has thus far not been demonstrated for PARK16. The study reported in Paper 3 was in this respect unfortunately limited by sample size. Ideally, such a fine mapping study should have been large enough to maintain strong statistical power to detect even small effects when the top-hit SNP is included as a covariate.

In Paper 4 we investigated the association between promoter methylation and GWAS signals of the *SNCA* locus, taking current evidence of allelic heterogeneity into account. In all published PD GWAS, the top *SNCA* signal has been located near the 3' end of the gene, represented by rs356165 in our report. A few studies have also detected significant secondary signals from SNPs near the 5' end of the gene. Interestingly, we found that promoter methylation was associated exclusively with a 5' risk-SNP. This observation illustrates that different functional alleles at the same locus probably operate through different molecular mechanisms, which may have distinct downstream consequences for pathogenic processes.

5.2.3 Gene-gene and gene-environment interaction

The simplest model of polygenic disease risk assumes a multiplicative (or log-additive) relationship between risk alleles. As an example, carrying two risk-alleles each with an OR of 1.1 equals a combined OR of 1.21. In reality, this may not always be the case. If variants at different loci have a synergistic effect on risk, resulting in a larger combined effect than expected under the multiplicative model, it is commonly termed gene-gene interaction, or epistasis. Epistasis has been proposed as a possible source of missing heritability. While the notion of interaction is intuitively appealing, the vast number of possible hypotheses complicates statistical testing for epistasis on a genome-wide scale (Cordell, 2009, Wei et al., 2014). A few PD studies have investigated gene-gene interaction between candidate pairs of SNPs based on plausible

functional hypotheses (Elbaz et al., 2011). In Paper 3, we attempted to replicate a previously reported interaction between *RAB7L1* and *LRRK2* (MacLeod et al., 2013). Our results provided indicative evidence supporting the hypothesis that the effect of common variation in *LRRK2* depends on the individuals' genetic risk-background at the PARK16 locus. These results should be interpreted with caution, as the statistical significance is marginal.

Genetic risk variants may interact not only with each other, but also with environmental factors (Caspi and Moffitt, 2006). Gene-environment interaction may be exemplified by situations where a gene contributes to PD risk only under specific environmental circumstances, or if exposures and genetic risk variants potentiate each other (Ahmed et al., 2014, Hamza et al., 2014). Again, the phenomenon may represent a mechanism explaining missing heritability. None of the papers included in this thesis address gene-environment interaction directly. It might be noted, however, that the interpretation of Paper 4 points towards epigenetics as a possible converging mechanism for genetic and environmental influences on pathogenesis. The main finding of the article is an association between rs3756063 genotype and promoter methylation. Yet within each genotype group, there is still a consistent difference in methylation level between cases and controls, which could conceivably be driven by other, including environmental, risk factors (Paper 4, Figure 2).

5.3 From association to function

5.3.1 Loci versus genes and the limitations of fine-mapping

For a few GWAS loci in PD, the disease-relevant genes are fairly certain. *SNCA* and *LRRK2* are already implicated in monogenic PD, and coding mutations in *GBA* are established as strong risk factors. The pivotal role of the tau protein in neurodegenerative disease also makes it highly likely that *MAPT* is the responsible gene behind the chromosome 17 GWAS signal linked to the H1/H2 inversion haplotype. Various hypotheses have been put forward regarding genes in the HLA region on chromosome 6 (Hamza et al., 2010, Ahmed et al., 2012, Wissemann et al., 2013). While the functionally relevant gene(s) may be debatable, at least the

association signal points to a pathogenic role for the immune system in PD. A common intronic SNP in *GCH1* reached genome-wide significance in the latest PD GWAS meta-analysis (Nalls et al., 2014), shortly after rare, coding variants in the same gene were shown to be enriched in familial and early onset PD cases (Mencacci et al., 2014, Guella et al., 2014). This gene encodes GTP-cyclohydrolase I, a key enzyme in dopamine synthesis, and is furthermore known to be the cause of autosomal dominantly inherited dopa-responsive dystonia, a rare movement disorder which may exhibit Parkinsonian features (Kurian et al., 2011). Consequently, *GCH1* illustrates the minority of cases where a novel GWAS locus immediately implicates a specific gene with a known function, readily suggesting further hypotheses about disease mechanisms.

The above mentioned loci, however, represent the exceptions. For the majority of GWAS signals, the implicated genes remain unclear (Table 2). As a primary challenge, many associated regions contain several genes, with significant SNPs spread across longer stretches of high LD. Many genes encode proteins whose functions are unknown, making it difficult to prioritize biologically plausible candidates within each locus. For these reasons, it has thus far been challenging to map GWAS findings in PD onto protein networks, as has been done to relate genetic findings to relevant disease pathways in other diseases, such as multiple sclerosis (International Multiple Sclerosis Genetics Consortium et al., 2011). A fine-mapping approach is regularly employed as a first step towards a more precise localization of genetic association signals. Genotyping a dense panel of tagging SNPs across an associated locus could narrow down the observed LD block where the most associated SNP is located. In Paper 3, we applied this strategy to the PARK16 locus. Our data provide evidence in favour of an association signal located near the 5' end of *RAB7L1*. This region entails the gene promoter where non-coding genetic variation is likely to have functional consequences, supporting that *RAB7L1* may be relevant for PD pathogenesis.

Table 2. *GWAS loci in PD with examples of tentative functional hypotheses*

Locus name	Tentative implicated protein	Functional hypotheses	References
GBA-SYT11	Beta-glucocerebrosidase	Role in protein homeostasis. Rare coding variants implicated in PD	
PARK16	Ras-related protein Rab-7L1	Role in protein trafficking/degradation. Interacts with *LRRK2*	Beilina et al., 2014, MacLeod et al., 2013
SIPA1L2			
ACMSD			
STK39			
MCCC1-LAMP3			
GAK-DGKQ	Cyclin-G-associated kinase	Role in protein trafficking/degradation, interaction with *LRRK2*	Beilina et al., 2014
BST1			
SCARB2	Lysosymal integral membrane protein -2	Interacts with glucocerebrosidase in α-synuclein clearance	Rothaug et al., 2014
SNCA	α-Synuclein	Implicated in Mendelian PD. Main component of Lewy bodies	
HLA	Human leukocyte antigen - DRB1?	Immunological mechanisms modulated by HLA-alleles	Ahmed et al., 2012
GPNMB			
FGF20	Fibroblast growth factor 20	Neuroprotective properties	Itoh and Ohta, 2013
INPP5F			
MIR4697			
LRRK2	Leucine-rich repeat kinase 2	Implicated in Mendelian PD	
CCDC62-HIP1R			
GCH1	GTP-cyclohydrolase 1	Role in dopamin synthesis. Rare, coding variants implicated in PD	Mencacci et al., 2014
VPS13C			
STX1B			
MAPT	Microtubule-associated protein tau	Implicated in neurodegenerative diseases by genetics and pathology	
SREBF1-RAI1	Sterol regulatory element binding transcription factor 1	Regulates mitophagy. Interacts with parkin	Ivatt et al., 2014
RIT2			
DDRGK1			

The listed loci correspond to the 24 signals reported as significant from the replication phase of the meta-analysis by Nalls et al., 2014. Loci are unsystematically named in the literature, typically after the closest gene(s), the most biologically plausible gene or in relation to chronological discovery (such as PARK16).

Nevertheless, fine-mapping without further functional evidence has clear limitations. It has been shown that a partitioning of chromosomes into neat, neighbouring blocks may be an oversimplification of the LD structure of the human genome (Takeuchi et al., 2005). The most associated SNPs may not necessarily be tagging only variation in the physical proximity, yet these more complex LD patterns may go undetected by established methods used to estimate haplotype blocks (Gabriel et al., 2002). Furthermore, even if the location of an association signal is narrowed down to a smaller region, the relevant gene may not be the closest one. Cis-acting regulatory mechanisms involve DNA-protein interactions over areas of several hundred kb (Sanyal et al., 2012). In addition, some functional mechanisms may be trans-acting. Fine-mapping may be valuable for identifying allelic heterogeneity. Still, as mentioned earlier, complex LD-association patterns may also confuse the interpretation of "synthetic associations" (Dickson et al., 2010). These limitations indicate that fine-mapping data often need to be integrated with other observations in order to truly advance the understanding of GWAS signals in complex disease.

5.3.2 Quantitative-trait loci

Causal susceptibility variants in non-coding regions frequently affect gene regulation (Nicolae et al., 2010). Typically, genetic variants may disrupt the binding motifs of promoters or enhancers, thus altering the affinity of DNA-protein interactions, which in turn determine the rate of mRNA transcription. Genetic variants that correlate with a measurable quantitative trait, such as gene expression, are called quantitative-trait loci (QTLs). Mapping of QTLs may serve an important function in the interpretation of GWAS findings by linking significant loci to the regulation of specific genes (Cookson et al., 2009). Several large QTL datasets are now publicly available in open databases. It should be noted however, that where studies based on genotype may use any DNA, gene regulation is tissue specific. Studies indicate that regulatory QTL-patterns that are relevant to brain disorders may be reflected in blood, yet definitely not always (Hernandez et al., 2012).

The most widely studied quantitative trait in QTL mapping has been mRNA expression. To better understand and interpret our association findings in Paper 3, we queried different PARK16 top-SNPs in publicly available datasets from monocytes (Raj et al., 2014) and brain (Ramasamy et al., 2014). Still, these results point to significant QTL-effects across several of the different PARK16 genes, further complicating rather than clarifying the picture.

Regarding *SNCA*, evidence from families with genomic multiplications have indicated a dose-dependent relationship between gene transcription and disease severity (Farrer, 2006). Consequently, it is natural to assume that common susceptibility variants in the *SNCA* locus may lead to a more subtle increase in expression. Some previous studies seem to support this, yet unequivocal evidence is still lacking for this intuitively appealing hypothesis (Fuchs et al., 2008, Mata et al., 2010, Nalls et al., 2011, Nalls et al., 2014). In Paper 4, we investigated *SNCA* mRNA, but were not able to demonstrate any significant association with GWAS SNPs. Our main hypothesis, however, was that risk variants may affect DNA methylation of a putative promoter in *SNCA* intron 1. Our results indicated that the GWAS signal from the 5' region of *SNCA* is indeed a methylation quantitative-trait locus (mQTL). Although our study was small, it points towards some potentially interesting novel insights. The findings support the hypothesis that multiple functional variants in the *SNCA* locus may influence disease risk through distinct mechanisms. Furthermore, promoter methylation at *SNCA* and other susceptibility loci represent possible points of interaction between genetic risk-variants, ageing and environmental exposures in PD. It is likely that the precise role of epigenetics in the pathogenesis of PD will be the focus of several large-scale studies in the future. Smaller, early reports like ours may have a role by inspiring study design and research hypotheses for later collaborative projects.

5.3.3 Functional variation and causal mechanisms

Even if the disease-relevant gene at a given GWAS locus could be determined with a high degree of certainty based on biological plausibility, fine-mapping and QTL analysis, a comprehensive understanding of the association signal would still require

the identification of the precise functional variation and casual mechanisms (Cooper and Shendure, 2011). Pinpointing the responsible molecular phenomena, such as alterations in transcription factor binding, splicing or interactions with microRNA would open up for even deeper insights into the pathogenic process and also provide indications about systems which might be modifiable by therapy. Functional studies explaining the underlying mechanisms linking GWAS signals to disease are still few in the literature, yet some examples have demonstrated that the task is clearly feasible (Musunuru et al., 2010, Kulzer et al., 2014). The increasing catalogue of genomic annotations provided by the ENCODE project and similar experiments represents an important resource for generating functional hypotheses about non-coding variation (Kellis et al., 2014, Thurman et al., 2012, Neph et al., 2012, Encode Project Consortium, 2012).

At the onset of the present doctoral work, it was a main intention to nominate putative causal variants of the *SNCA* locus and further investigate proposed mechanisms in functional studies. Based on an initial fine-mapping experiment, we selected 48 individuals for deep sequencing of the total *SNCA* region, aiming to identify low-frequency candidate variants. Pinpointing functionally relevant variation from sequence data has proven complicated, and access to novel tools as well as findings published by others have prompted us to modify the project. While this work is still on-going, it is not included in the present thesis.

5.4 Prospects for individual-level genetic assessment in sporadic Parkinson's disease

5.4.1 Genetic score profiling

The genetic study of complex disease is primarily justified as basic research, aimed at elucidating disease pathogenesis. However, with a growing number of established disease loci, many authors also anticipate that genetic data may become practically useful on an individual level, in contexts such as risk prediction, diagnosis, prognosis or stratification of patients in clinical research. One by one, the effect sizes of GWAS variants are typically far too small to have any practical value alone. In PD, no risk

allele identified in GWAS will give an OR>2, and most loci even well below 1.3 (Nalls et al., 2014). Consequently, certain analyses may fruitfully be focused not on individual variants, but on the cumulative burden of genetic risk across a range of susceptibility loci.

In Paper 1 we generated individual cumulative genetic risk scores across all 11 loci that were nominally significant in the replication study. Our analysis showed that individuals within the highest risk score quintile have about a threefold risk of developing PD, compared to the lowest quintile. This result was in accordance with the corresponding figures published by others (Nalls et al., 2011), even though we included a smaller number of SNPs. While the method serves to illustrate the concept of cumulative genetic risk, the effect size would still not be useful for any individual prediction or diagnostics. In the future however, it is conceivable that genetic profile may be successfully integrated as one of many variables in a more comprehensive scoring system for diagnostic or prognostic prediction purposes.

5.4.2 Biomarkers

As mentioned in section 1.2, the identification of informative biomarkers is an important aim for current PD research. Genetic data in itself could serve to classify individuals according to risk strata. Yet as genotypes remain constant, they will tell us little about on-going *processes*. Nevertheless, studies that largely build on genetics may also involve the measurement of dynamic variables, such as gene expression or DNA methylation, which might be conceivably be interesting as candidates for longitudinal biomarkers for PD pathogenesis (Karlsson et al., 2013). In Paper 4, we highlighted this aspect in relation to *SNCA* intron 1 methylation. In particular, the similar findings observed in blood and brain tissue was interpreted as promising, indicating that disease-relevant epigenetic changes may to some extent be mirrored in blood, which can conveniently be sampled and analysed in large numbers of patients at several time-points.

5.5 Future directions

5.5.1 Large-scale international collaborations

The work presented in this thesis largely builds on findings from large GWAS, organized as coordinated efforts within international research consortia. This illustrates the importance of widespread collaboration as a crucial success criterion in the genetic study of complex disease. Our Norwegian research group is currently involved as a partner in a large European project named COmprehensive Unbiased Risk factor Assessment of Genetics and Environment in Parkinson's Disease (COURAGE-PD). This joint effort will involve identification of putative variants through HTS followed by large-scale validation by a customly designed genotyping chip in tens of thousands of samples, providing rare variant data and statistical power for a far more comprehensive characterization of the genetic architecture of PD than traditional GWAS performed to date. As part of the first subproject, all currently known GWAS loci will be investigated by targeted resequencing in more than 3000 samples. This sequencing is now being performed in Oslo, with L. Pihlstrøm as the main responsible researcher, using the method described in Paper 2 of this thesis.

We also have an intention to build on the experience from the work described in Paper 4 and design larger collaborative studies to investigate the epigenetics of PD. A comprehensive epigenetics project should take into account both environmental factors, tissue specificity and longitudinal variability in epigenetic profiles and must therefore involve partners collectively possessing a fruitful combination of scientific expertise and access to relevant biological material and data. A funding application for a collaborative project on the epigenetics of PD is currently in process.

5.5.2 Functional studies

It has been a long-term ambition for our research group to contribute to the characterization of causally relevant non-coding variation at the *SNCA* locus. Through this work, we have gained experience with methods for functional studies of gene

regulation, including DNA-protein interaction assays, cell-based assays involving transfection of reporter constructs and chromatin immunoprecipitation experiments. We will continue this work, building upon robust genetic data, aiming to pinpoint the functional mechanisms affecting disease risk at *SNCA* and/or other GWAS loci in PD.

5.5.3 Refining genotype-phenotype correlations

The main focus of genetic research in sporadic PD thus far has been the identification of variants conferring disease risk through case-control studies. However, the striking heterogeneity in the clinical presentation of the disorder is also likely to have, at least partly, genetic underpinnings. Consequently, an emerging strategy in complex PD genetics is to design studies where the outcomes of interest are specific subphenotypes or clinical variables within a patient cohort (Kasten and Klein, 2015). This approach may represent a first step towards disentangling the relative role of different pathogenic mechanisms and pathways across different subgroups of patients. Such stratification could have important implications for recruitment to future clinical trials, as treatment response to hypothetically disease-modifying therapies may depend on individual molecular states. Current examples of phenotype-oriented genetic studies in PD include investigations of cognitive function and risk variants in relevant candidate genes (Nombela et al., 2014), a study on motor progression and *SNCA* variability (Ritz et al., 2012) and three recent reports showing that cumulative genetic risk scores are associated with age at onset of PD (Nalls et al., 2015, Escott-Price et al., 2015, Lill et al., 2015).

The correlations between genetic makeup and clinical phenotype in PD is a major area of interest for our research group, and we have several on-going projects aiming to study genetics in relation to novel clinical outcome variables.

6. CONCLUSIONS

PD is a common neurodegenerative disorder, severely affecting the quality of life of patients and their families. There is currently no causal treatment available to effectively oppose the progressive course of the disease. Our current understanding of PD aetiology is insufficient, and novel insights into pathogenic mechanisms and pathways are likely to pave way for new therapeutic strategies in the future. Genetic research has a major role to play in this effort to elucidate the pathogenesis of PD on a molecular level. With respect to the sporadic, complex form of the disease, a major advance in recent years has been the identification of susceptibility loci through large-scale GWAS. However, considerable scientific efforts are necessary to refine and expand the understanding of GWAS association signals before the full potential value of these findings can be harvested for further research. The work presented in this thesis involves a range of different approaches, all aiming to shed further light on the genetic architecture of PD by building upon findings from GWAS. We provided further evidence for 11 GWAS loci in a Scandinavian population and presented an effective study design for rare variant detection within GWAS loci by targeted deep sequencing. Furthermore, we explored in depth the genetic variability of the complicated PARK16 locus and reported an association between a GWAS variant and promoter methylation in *SNCA*. Collectively, these studies touch upon many topics expected to become important in the next phase of discoveries in complex genetics, such as rare coding variants, epistasis, epigenetics and cumulative genetic risk score profiling. Further research following a broad range of strategies is needed to map out the complexities of PD and other neurodegenerative disorders and lay the groundwork for improved patient care in the future.

REFERENCES

1000 GENOMES PROJECT CONSORTIUM, ABECASIS, G. R., ALTSHULER, D., AUTON, A., BROOKS, L. D., DURBIN, R. M., GIBBS, R. A., HURLES, M. E. & MCVEAN, G. A. 2010. A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061-73.

AARSLAND, D. & KURZ, M. W. 2010. The epidemiology of dementia associated with Parkinson disease. *J Neurol Sci,* 289**,** 18-22.

AHMED, I., LEE, P. C., LILL, C. M., SEARLES NIELSEN, S., ARTAUD, F., GALLAGHER, L. G., LORIOT, M. A., MULOT, C., NACFER, M., LIU, T., BIERNACKA, J. M., ARMASU, S., ANDERSON, K., FARIN, F. M., LASSEN, C. F., HANSEN, J., OLSEN, J. H., BERTRAM, L., MARAGANORE, D. M., CHECKOWAY, H., RITZ, B. & ELBAZ, A. 2014. Lack of replication of the GRIN2A-by-coffee interaction in Parkinson disease. *PLoS Genet,* 10**,** e1004788.

AHMED, I., TAMOUZA, R., DELORD, M., KRISHNAMOORTHY, R., TZOURIO, C., MULOT, C., NACFER, M., LAMBERT, J. C., BEAUNE, P., LAURENT-PUIG, P., LORIOT, M. A., CHARRON, D. & ELBAZ, A. 2012. Association between Parkinson's disease and the HLA-DRB1 locus. *Mov Disord,* 27**,** 1104-10.

ALVES, G., MULLER, B., HERLOFSON, K., HOGENESCH, I., TELSTAD, W., AARSLAND, D., TYSNES, O. B., LARSEN, J. P. & NORWEGIAN PARKWEST STUDY, G. 2009. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *J Neurol Neurosurg Psychiatry,* 80**,** 851-7.

BANSAL, V., LIBIGER, O., TORKAMANI, A. & SCHORK, N. J. 2010. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet,* 11**,** 773-85.

BEILINA, A., RUDENKO, I. N., KAGANOVICH, A., CIVIERO, L., CHAU, H., KALIA, S. K., KALIA, L. V., LOBBESTAEL, E., CHIA, R., NDUKWE, K., DING, J., NALLS, M. A., INTERNATIONAL PARKINSON'S DISEASE GENOMICS, C., NORTH AMERICAN BRAIN EXPRESSION, C., OLSZEWSKI, M., HAUSER, D. N., KUMARAN, R., LOZANO, A. M., BAEKELANDT, V., GREENE, L. E., TAYMANS, J. M., GREGGIO, E. & COOKSON, M. R. 2014. Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. *Proc Natl Acad Sci U S A,* 111**,** 2626-31.

BENDOR, J. T., LOGAN, T. P. & EDWARDS, R. H. 2013. The function of alpha-synuclein. *Neuron,* 79**,** 1044-66.

BERG, D., POSTUMA, R. B., BLOEM, B., CHAN, P., DUBOIS, B., GASSER, T., GOETZ, C. G., HALLIDAY, G. M., HARDY, J., LANG, A. E., LITVAN, I., MAREK, K., OBESO, J., OERTEL, W., OLANOW, C. W., POEWE, W., STERN, M. & DEUSCHL, G. 2014. Time to redefine PD? Introductory statement of the MDS Task Force on the definition of Parkinson's disease. *Mov Disord,* 29**,** 454-462.

BIRKMAYER, W. & HORNYKIEWICZ, O. 1961. [The L-3,4-dioxyphenylalanine (DOPA)-effect in Parkinson-akinesia]. *Wien Klin Wochenschr,* 73**,** 787-8.

BONIFATI, V. 2014. Genetics of Parkinson's disease--state of the art, 2013. *Parkinsonism Relat Disord,* 20 Suppl 1**,** S23-8.

BONIFATI, V., RIZZU, P., VAN BAREN, M. J., SCHAAP, O., BREEDVELD, G. J., KRIEGER, E., DEKKER, M. C., SQUITIERI, F., IBANEZ, P., JOOSSE, M., VAN DONGEN, J. W., VANACORE, N., VAN SWIETEN, J. C., BRICE, A., MECO, G.,

VAN DUIJN, C. M., OOSTRA, B. A. & HEUTINK, P. 2003. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science,* 299**,** 256-9.

BRAAK, H., DEL TREDICI, K., RUB, U., DE VOS, R. A., JANSEN STEUR, E. N. & BRAAK, E. 2003. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging,* 24**,** 197-211.

BROWNING, B. L. 2008. PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies. *BMC Bioinformatics,* 9**,** 309.

BURN, D. J., MARK, M. H., PLAYFORD, E. D., MARAGANORE, D. M., ZIMMERMAN, T. R., JR., DUVOISIN, R. C., HARDING, A. E., MARSDEN, C. D. & BROOKS, D. J. 1992. Parkinson's disease in twins studied with 18F-dopa and positron emission tomography. *Neurology,* 42**,** 1894-900.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

CARLSSON, A., LINDQVIST, M. & MAGNUSSON, T. 1957. 3,4-Dihydroxyphenylalanine and 5-hydroxytryptophan as reserpine antagonists. *Nature,* 180**,** 1200.

CARLSSON, A., LINDQVIST, M., MAGNUSSON, T. & WALDECK, B. 1958. On the presence of 3-hydroxytyramine in brain. *Science,* 127**,** 471.

CASPI, A. & MOFFITT, T. E. 2006. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nat Rev Neurosci,* 7**,** 583-90.

CHARTIER-HARLIN, M. C., DACHSEL, J. C., VILARINO-GUELL, C., LINCOLN, S. J., LEPRETRE, F., HULIHAN, M. M., KACHERGUS, J., MILNERWOOD, A. J., TAPIA, L., SONG, M. S., LE RHUN, E., MUTEZ, E., LARVOR, L., DUFLOT, A., VANBESIEN-MAILLIOT, C., KREISLER, A., ROSS, O. A., NISHIOKA, K., SOTO-ORTOLAZA, A. I., COBB, S. A., MELROSE, H. L., BEHROUZ, B., KEELING, B. H., BACON, J. A., HENTATI, E., WILLIAMS, L., YANAGIYA, A., SONENBERG, N., LOCKHART, P. J., ZUBAIR, A. C., UITTI, R. J., AASLY, J. O., KRYGOWSKA-WAJS, A., OPALA, G., WSZOLEK, Z. K., FRIGERIO, R., MARAGANORE, D. M., GOSAL, D., LYNCH, T., HUTCHINSON, M., BENTIVOGLIO, A. R., VALENTE, E. M., NICHOLS, W. C., PANKRATZ, N., FOROUD, T., GIBSON, R. A., HENTATI, F., DICKSON, D. W., DESTEE, A. & FARRER, M. J. 2011. Translation initiator EIF4G1 mutations in familial Parkinson disease. *Am J Hum Genet,* 89**,** 398-406.

CHARTIER-HARLIN, M. C., KACHERGUS, J., ROUMIER, C., MOUROUX, V., DOUAY, X., LINCOLN, S., LEVECQUE, C., LARVOR, L., ANDRIEUX, J., HULIHAN, M., WAUCQUIER, N., DEFEBVRE, L., AMOUYEL, P., FARRER, M. & DESTEE, A. 2004. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet,* 364**,** 1167-9.

CHAUDHURI, K. R., ODIN, P., ANTONINI, A. & MARTINEZ-MARTIN, P. 2011. Parkinson's disease: the non-motor issues. *Parkinsonism Relat Disord,* 17**,** 717-23.

COOKSON, W., LIANG, L., ABECASIS, G., MOFFATT, M. & LATHROP, M. 2009. Mapping complex disease traits with global gene expression. *Nat Rev Genet,* 10**,** 184-94.

COOPER, G. M. & SHENDURE, J. 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet,* 12**,** 628-40.

CORDELL, H. J. 2009. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet,* 10**,** 392-404.

CORREIA GUEDES, L., FERREIRA, J. J., ROSA, M. M., COELHO, M., BONIFATI, V. & SAMPAIO, C. 2010. Worldwide frequency of G2019S LRRK2 mutation in Parkinson's disease: a systematic review. *Parkinsonism Relat Disord,* 16**,** 237-42.

COTZIAS, G. C., PAPAVASILIOU, P. S. & GELLENE, R. 1969a. L-dopa in parkinson's syndrome. *N Engl J Med,* 281**,** 272.

COTZIAS, G. C., PAPAVASILIOU, P. S. & GELLENE, R. 1969b. Modification of Parkinsonism--chronic treatment with L-dopa. *N Engl J Med,* 280**,** 337-45.

COULSON, D. T., BROCKBANK, S., QUINN, J. G., MURPHY, S., RAVID, R., IRVINE, G. B. & JOHNSTON, J. A. 2008. Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. *BMC Mol Biol,* 9**,** 46.

DE LAU, L. M. & BRETELER, M. M. 2006. Epidemiology of Parkinson's disease. *Lancet Neurol,* 5**,** 525-35.

DI FONZO, A., DEKKER, M. C., MONTAGNA, P., BARUZZI, A., YONOVA, E. H., CORREIA GUEDES, L., SZCZERBINSKA, A., ZHAO, T., DUBBEL-HULSMAN, L. O., WOUTERS, C. H., DE GRAAFF, E., OYEN, W. J., SIMONS, E. J., BREEDVELD, G. J., OOSTRA, B. A., HORSTINK, M. W. & BONIFATI, V. 2009. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology,* 72**,** 240-5.

DICKSON, S. P., WANG, K., KRANTZ, I., HAKONARSON, H. & GOLDSTEIN, D. B. 2010. Rare variants create synthetic genome-wide associations. *PLoS Biol,* 8**,** e1000294.

DO, C. B., TUNG, J. Y., DORFMAN, E., KIEFER, A. K., DRABANT, E. M., FRANCKE, U., MOUNTAIN, J. L., GOLDMAN, S. M., TANNER, C. M., LANGSTON, J. W., WOJCICKI, A. & ERIKSSON, N. 2011. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet,* 7**,** e1002141.

DODSON, M. W. & GUO, M. 2007. Pink1, Parkin, DJ-1 and mitochondrial dysfunction in Parkinson's disease. *Curr Opin Neurobiol,* 17**,** 331-7.

DORSEY, E. R., CONSTANTINESCU, R., THOMPSON, J. P., BIGLAN, K. M., HOLLOWAY, R. G., KIEBURTZ, K., MARSHALL, F. J., RAVINA, B. M., SCHIFITTO, G., SIDEROWF, A. & TANNER, C. M. 2007. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology,* 68**,** 384-6.

DUBOIS, B., BURN, D., GOETZ, C., AARSLAND, D., BROWN, R. G., BROE, G. A., DICKSON, D., DUYCKAERTS, C., CUMMINGS, J., GAUTHIER, S., KORCZYN, A., LEES, A., LEVY, R., LITVAN, I., MIZUNO, Y., MCKEITH, I. G., OLANOW, C. W., POEWE, W., SAMPAIO, C., TOLOSA, E. & EMRE, M. 2007. Diagnostic

procedures for Parkinson's disease dementia: recommendations from the movement disorder society task force. *Mov Disord,* 22**,** 2314-24.

EDVARDSON, S., CINNAMON, Y., TA-SHMA, A., SHAAG, A., YIM, Y. I., ZENVIRT, S., JALAS, C., LESAGE, S., BRICE, A., TARABOULOS, A., KAESTNER, K. H., GREENE, L. E. & ELPELEG, O. 2012. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. *PLoS One,* 7**,** e36458.

EDWARDS, T. L., SCOTT, W. K., ALMONTE, C., BURT, A., POWELL, E. H., BEECHAM, G. W., WANG, L., ZUCHNER, S., KONIDARI, I., WANG, G., SINGER, C., NAHAB, F., SCOTT, B., STAJICH, J. M., PERICAK-VANCE, M., HAINES, J., VANCE, J. M. & MARTIN, E. R. 2010. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Ann Hum Genet,* 74**,** 97-109.

EHRET, G. B., LAMPARTER, D., HOGGART, C. J., GENETIC INVESTIGATION OF ANTHROPOMETRIC TRAITS, C., WHITTAKER, J. C., BECKMANN, J. S. & KUTALIK, Z. 2012. A multi-SNP locus-association method reveals a substantial fraction of the missing heritability. *Am J Hum Genet,* 91**,** 863-71.

EHRINGER, H. & HORNYKIEWICZ, O. 1960. [Distribution of noradrenaline and dopamine (3-hydroxytyramine) in the human brain and their behavior in diseases of the extrapyramidal system]. *Klin Wochenschr,* 38**,** 1236-9.

ELBAZ, A., ROSS, O. A., IOANNIDIS, J. P., SOTO-ORTOLAZA, A. I., MOISAN, F., AASLY, J., ANNESI, G., BOZI, M., BRIGHINA, L., CHARTIER-HARLIN, M. C., DESTEE, A., FERRARESE, C., FERRARIS, A., GIBSON, J. M., GISPERT, S., HADJIGEORGIOU, G. M., JASINSKA-MYGA, B., KLEIN, C., KRUGER, R., LAMBERT, J. C., LOHMANN, K., VAN DE LOO, S., LORIOT, M. A., LYNCH, T., MELLICK, G. D., MUTEZ, E., NILSSON, C., OPALA, G., PUSCHMANN, A., QUATTRONE, A., SHARMA, M., SILBURN, P. A., STEFANIS, L., UITTI, R. J., VALENTE, E. M., VILARINO-GUELL, C., WIRDEFELDT, K., WSZOLEK, Z. K., XIROMERISIOU, G., MARAGANORE, D. M., FARRER, M. J. & GENETIC EPIDEMIOLOGY OF PARKINSON'S DISEASE, C. 2011. Independent and joint effects of the MAPT and SNCA genes in Parkinson disease. *Ann Neurol,* 69**,** 778-92.

ENCODE PROJECT CONSORTIUM 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science,* 306**,** 636-40.

ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

ESCOTT-PRICE, V., INTERNATIONAL PARKINSON'S DISEASE GENOMICS, C., NALLS, M. A., MORRIS, H. R., LUBBE, S., BRICE, A., GASSER, T., HEUTINK, P., WOOD, N. W., HARDY, J., SINGLETON, A. B., WILLIAMS, N. M. & MEMBERS, I. C. 2015. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol,* 77**,** 582-91.

EVANGELOU, E., MARAGANORE, D. M., ANNESI, G., BRIGHINA, L., BRICE, A., ELBAZ, A., FERRARESE, C., HADJIGEORGIOU, G. M., KRUEGER, R., LAMBERT, J. C., LESAGE, S., MARKOPOULOU, K., MELLICK, G. D., MEEUS, B., PEDERSEN, N. L., QUATTRONE, A., VAN BROECKHOVEN, C., SHARMA, M., SILBURN, P. A., TAN, E. K., WIRDEFELDT, K., IOANNIDIS, J. P. & GENETIC EPIDEMIOLOGY OF PARKINSON'S DISEASE, C. 2010. Non-replication of

association for six polymorphisms from meta-analysis of genome-wide association studies of Parkinson's disease: large-scale collaborative study. *Am J Med Genet B Neuropsychiatr Genet,* 153B**,** 220-8.

FARRER, M., KACHERGUS, J., FORNO, L., LINCOLN, S., WANG, D. S., HULIHAN, M., MARAGANORE, D., GWINN-HARDY, K., WSZOLEK, Z., DICKSON, D. & LANGSTON, J. W. 2004. Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. *Ann Neurol,* 55**,** 174-9.

FARRER, M. J. 2006. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nat Rev Genet,* 7**,** 306-18.

FOO, J. N., TAN, L. C., LIANY, H., KOH, T. H., IRWAN, I. D., NG, Y. Y., AHMAD-ANNUAR, A., AU, W. L., AUNG, T., CHAN, A. Y., CHONG, S. A., CHUNG, S. J., JUNG, Y., KHOR, C. C., KIM, J., LEE, J., LIM, S. Y., MOK, V., PRAKASH, K. M., SONG, K., TAI, E. S., VITHANA, E. N., WONG, T. Y., TAN, E. K. & LIU, J. 2014. Analysis of non-synonymous-coding variants of Parkinson's disease-related pathogenic and susceptibility genes in East Asian populations. *Hum Mol Genet,* 23**,** 3891-7.

FREEDMAN, M. L., REICH, D., PENNEY, K. L., MCDONALD, G. J., MIGNAULT, A. A., PATTERSON, N., GABRIEL, S. B., TOPOL, E. J., SMOLLER, J. W., PATO, C. N., PATO, M. T., PETRYSHEN, T. L., KOLONEL, L. N., LANDER, E. S., SKLAR, P., HENDERSON, B., HIRSCHHORN, J. N. & ALTSHULER, D. 2004. Assessing the impact of population stratification on genetic association studies. *Nat Genet,* 36**,** 388-93.

FUCHS, J., TICHOPAD, A., GOLUB, Y., MUNZ, M., SCHWEITZER, K. J., WOLF, B., BERG, D., MUELLER, J. C. & GASSER, T. 2008. Genetic variability in the SNCA gene influences alpha-synuclein levels in the blood and brain. *FASEB J,* 22**,** 1327-34.

FUNAYAMA, M., OHE, K., AMO, T., FURUYA, N., YAMAGUCHI, J., SAIKI, S., LI, Y., OGAKI, K., ANDO, M., YOSHINO, H., TOMIYAMA, H., NISHIOKA, K., HASEGAWA, K., SAIKI, H., SATAKE, W., MOGUSHI, K., SASAKI, R., KOKUBO, Y., KUZUHARA, S., TODA, T., MIZUNO, Y., UCHIYAMA, Y., OHNO, K. & HATTORI, N. 2015. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet Neurol,* 14**,** 274-82.

FUNG, H. C., SCHOLZ, S., MATARIN, M., SIMON-SANCHEZ, J., HERNANDEZ, D., BRITTON, A., GIBBS, J. R., LANGEFELD, C., STIEGERT, M. L., SCHYMICK, J., OKUN, M. S., MANDEL, R. J., FERNANDEZ, H. H., FOOTE, K. D., RODRIGUEZ, R. L., PECKHAM, E., DE VRIEZE, F. W., GWINN-HARDY, K., HARDY, J. A. & SINGLETON, A. 2006. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol,* 5**,** 911-6.

GABRIEL, S. B., SCHAFFNER, S. F., NGUYEN, H., MOORE, J. M., ROY, J., BLUMENSTIEL, B., HIGGINS, J., DEFELICE, M., LOCHNER, A., FAGGART, M., LIU-CORDERO, S. N., ROTIMI, C., ADEYEMO, A., COOPER, R., WARD, R., LANDER, E. S., DALY, M. J. & ALTSHULER, D. 2002. The structure of haplotype blocks in the human genome. *Science,* 296**,** 2225-9.

GAFFNEY, D. J. 2013. Global properties and functional complexity of human gene regulatory variation. *PLoS Genet,* 9**,** e1003501.

GELB, D. J., OLIVER, E. & GILMAN, S. 1999. Diagnostic criteria for Parkinson disease. *Arch Neurol,* 56**,** 33-9.

GIBB, W. R. & LEES, A. J. 1988. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry,* 51**,** 745-52.

GOEDERT, M. 2001. Alpha-synuclein and neurodegenerative diseases. *Nat Rev Neurosci,* 2**,** 492-501.

GOKER-ALPAN, O., SCHIFFMANN, R., LAMARCA, M. E., NUSSBAUM, R. L., MCINERNEY-LEO, A. & SIDRANSKY, E. 2004. Parkinsonism among Gaucher disease carriers. *J Med Genet,* 41**,** 937-40.

GOLBE, L. I., DI IORIO, G., BONAVITA, V., MILLER, D. C. & DUVOISIN, R. C. 1990. A large kindred with autosomal dominant Parkinson's disease. *Ann Neurol,* 27**,** 276-82.

GUELLA, I., SHERMAN, H. E., APPEL-CRESSWELL, S., RAJPUT, A., RAJPUT, A. H. & FARRER, M. J. 2014. Parkinsonism in GTP cyclohydrolase 1 mutation carriers. *Brain*, 138, e349.

HAMZA, T. H., HILL-BURNS, E. M., SCOTT, W. K., VANCE, J. M., FACTOR, S. A., ZABETIAN, C. P. & PAYAMI, H. 2014. Glutamate receptor gene GRIN2A, coffee, and Parkinson disease. *PLoS Genet,* 10**,** e1004774.

HAMZA, T. H., ZABETIAN, C. P., TENESA, A., LAEDERACH, A., MONTIMURRO, J., YEAROUT, D., KAY, D. M., DOHENY, K. F., PASCHALL, J., PUGH, E., KUSEL, V. I., COLLURA, R., ROBERTS, J., GRIFFITH, A., SAMII, A., SCOTT, W. K., NUTT, J., FACTOR, S. A. & PAYAMI, H. 2010. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat Genet,* 42**,** 781-5.

HASSLER, R. 1938. Zur Pathologie der Paralysis agitans und des postenzephalitischen Parkinsonismus. *Journal für Psychologie und Neurologie,* 48**,** 387-455.

HEALY, D. G., FALCHI, M., O'SULLIVAN, S. S., BONIFATI, V., DURR, A., BRESSMAN, S., BRICE, A., AASLY, J., ZABETIAN, C. P., GOLDWURM, S., FERREIRA, J. J., TOLOSA, E., KAY, D. M., KLEIN, C., WILLIAMS, D. R., MARRAS, C., LANG, A. E., WSZOLEK, Z. K., BERCIANO, J., SCHAPIRA, A. H., LYNCH, T., BHATIA, K. P., GASSER, T., LEES, A. J., WOOD, N. W. & INTERNATIONAL, L. C. 2008. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol,* 7**,** 583-90.

HERNANDEZ, D. G., NALLS, M. A., MOORE, M., CHONG, S., DILLMAN, A., TRABZUNI, D., GIBBS, J. R., RYTEN, M., AREPALLI, S., WEALE, M. E., ZONDERMAN, A. B., TRONCOSO, J., O'BRIEN, R., WALKER, R., SMITH, C., BANDINELLI, S., TRAYNOR, B. J., HARDY, J., SINGLETON, A. B. & COOKSON, M. R. 2012. Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol Dis,* 47**,** 20-8.

HIRSCHHORN, J. N. & DALY, M. J. 2005. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet,* 6**,** 95-108.

HOLEMON, H., KORSHUNOVA, Y., ORDWAY, J. M., BEDELL, J. A., CITEK, R. W., LAKEY, N., LEON, J., FINNEY, M., MCPHERSON, J. D. & JEDDELOH, J. A. 2007.

MethylScreen: DNA methylation density monitoring using quantitative PCR. *Biotechniques,* 43**,** 683-93.

HUA, D., HU, Y., WU, Y. Y., CHENG, Z. H., YU, J., DU, X. & HUANG, Z. H. 2011. Quantitative methylation analysis of multiple genes using methylation-sensitive restriction enzyme-based quantitative PCR for the detection of hepatocellular carcinoma. *Exp Mol Pathol,* 91**,** 455-60.

HUMPHREYS, K., GRANKVIST, A., LEU, M., HALL, P., LIU, J., RIPATTI, S., REHNSTROM, K., GROOP, L., KLARESKOG, L., DING, B., GRONBERG, H., XU, J., PEDERSEN, N. L., LICHTENSTEIN, P., MATTINGSDAL, M., ANDREASSEN, O. A., O'DUSHLAINE, C., PURCELL, S. M., SKLAR, P., SULLIVAN, P. F., HULTMAN, C. M., PALMGREN, J. & MAGNUSSON, P. K. 2011. The genetic structure of the Swedish population. *PLoS One,* 6**,** e22547.

HUTTON, M. 2001. Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms. *Neurology,* 56**,** S21-5.

IBANEZ, P., BONNET, A. M., DEBARGES, B., LOHMANN, E., TISON, F., POLLAK, P., AGID, Y., DURR, A. & BRICE, A. 2004. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet,* 364**,** 1169-71.

INTERNATIONAL HAPMAP CONSORTIUM 2003. The International HapMap Project. *Nature,* 426**,** 789-96.

INTERNATIONAL HAPMAP CONSORTIUM 2005. A haplotype map of the human genome. *Nature,* 437**,** 1299-320.

INTERNATIONAL MULTIPLE SCLEROSIS GENETICS CONSORTIUM, WELLCOME TRUST CASE CONTROL CONSORTIUM, SAWCER, S., HELLENTHAL, G., PIRINEN, M., SPENCER, C. C., PATSOPOULOS, N. A., MOUTSIANAS, L., DILTHEY, A., SU, Z., FREEMAN, C., HUNT, S. E., EDKINS, S., GRAY, E., BOOTH, D. R., POTTER, S. C., GORIS, A., BAND, G., OTURAI, A. B., STRANGE, A., SAARELA, J., BELLENGUEZ, C., FONTAINE, B., GILLMAN, M., HEMMER, B., GWILLIAM, R., ZIPP, F., JAYAKUMAR, A., MARTIN, R., LESLIE, S., HAWKINS, S., GIANNOULATOU, E., D'ALFONSO, S., BLACKBURN, H., MARTINELLI BONESCHI, F., LIDDLE, J., HARBO, H. F., PEREZ, M. L., SPURKLAND, A., WALLER, M. J., MYCKO, M. P., RICKETTS, M., COMABELLA, M., HAMMOND, N., KOCKUM, I., MCCANN, O. T., BAN, M., WHITTAKER, P., KEMPPINEN, A., WESTON, P., HAWKINS, C., WIDAA, S., ZAJICEK, J., DRONOV, S., ROBERTSON, N., BUMPSTEAD, S. J., BARCELLOS, L. F., RAVINDRARAJAH, R., ABRAHAM, R., ALFREDSSON, L., ARDLIE, K., AUBIN, C., BAKER, A., BAKER, K., BARANZINI, S. E., BERGAMASCHI, L., BERGAMASCHI, R., BERNSTEIN, A., BERTHELE, A., BOGGILD, M., BRADFIELD, J. P., BRASSAT, D., BROADLEY, S. A., BUCK, D., BUTZKUEVEN, H., CAPRA, R., CARROLL, W. M., CAVALLA, P., CELIUS, E. G., CEPOK, S., CHIAVACCI, R., CLERGET-DARPOUX, F., CLYSTERS, K., COMI, G., COSSBURN, M., COURNU-REBEIX, I., COX, M. B., COZEN, W., CREE, B. A., CROSS, A. H., CUSI, D., DALY, M. J., DAVIS, E., DE BAKKER, P. I., DEBOUVERIE, M., D'HOOGHE M, B., DIXON, K., DOBOSI, R., DUBOIS, B., ELLINGHAUS, D., et al. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature,* 476**,** 214-9.

INTERNATIONAL PARKINSON'S DISEASE GENOMICS CONSORTIUM & WELLCOME TRUST CASE CONTROL CONSORTIUM 2011. A two-stage

meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet,* 7**,** e1002142.

IOANNIDIS, J. P. 2005. Why most published research findings are false. *PLoS Med,* 2**,** e124.

IOANNIDIS, J. P., NTZANI, E. E., TRIKALINOS, T. A. & CONTOPOULOS-IOANNIDIS, D. G. 2001. Replication validity of genetic association studies. *Nat Genet,* 29**,** 306-9.

IOANNIDIS, J. P., THOMAS, G. & DALY, M. J. 2009. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet,* 10**,** 318-29.

ITOH, N. & OHTA, H. 2013. Roles of FGF20 in dopaminergic neurons and Parkinson's disease. *Front Mol Neurosci,* 6**,** 15.

IVATT, R. M., SANCHEZ-MARTINEZ, A., GODENA, V. K., BROWN, S., ZIVIANI, E. & WHITWORTH, A. J. 2014. Genome-wide RNAi screen identifies the Parkinson disease GWAS risk locus SREBF1 as a regulator of mitophagy. *Proc Natl Acad Sci U S A,* 111**,** 8494-9.

JELLINGER, K. A. 2003. Neuropathological spectrum of synucleinopathies. *Mov Disord,* 18 Suppl 6**,** S2-12.

JOHNSON, R. C., NELSON, G. W., TROYER, J. L., LAUTENBERGER, J. A., KESSING, B. D., WINKLER, C. A. & O'BRIEN, S. J. 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics,* 11**,** 724.

JOHNSON, W. G., HODGE, S. E. & DUVOISIN, R. 1990. Twin studies and the genetics of Parkinson's disease--a reappraisal. *Mov Disord,* 5**,** 187-94.

KACHERGUS, J., MATA, I. F., HULIHAN, M., TAYLOR, J. P., LINCOLN, S., AASLY, J., GIBSON, J. M., ROSS, O. A., LYNCH, T., WILEY, J., PAYAMI, H., NUTT, J., MARAGANORE, D. M., CZYZEWSKI, K., STYCZYNSKA, M., WSZOLEK, Z. K., FARRER, M. J. & TOFT, M. 2005. Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. *Am J Hum Genet,* 76**,** 672-80.

KARLSSON, M. K., SHARMA, P., AASLY, J., TOFT, M., SKOGAR, O., SAEBO, S. & LONNEBORG, A. 2013. Found in transcription: accurate Parkinson's disease classification in peripheral blood. *J Parkinsons Dis,* 3**,** 19-29.

KASTEN, M. & KLEIN, C. 2015. Genetic risk loci for Parkinson's disease: Moving from state to trait? *Mov Disord*. Epub ahead of print. DOI: 10.1002/mds.26246.

KELLER, M. F., SAAD, M., BRAS, J., BETTELLA, F., NICOLAOU, N., SIMON-SANCHEZ, J., MITTAG, F., BUCHEL, F., SHARMA, M., GIBBS, J. R., SCHULTE, C., MOSKVINA, V., DURR, A., HOLMANS, P., KILARSKI, L. L., GUERREIRO, R., HERNANDEZ, D. G., BRICE, A., YLIKOTILA, P., STEFANSSON, H., MAJAMAA, K., MORRIS, H. R., WILLIAMS, N., GASSER, T., HEUTINK, P., WOOD, N. W., HARDY, J., MARTINEZ, M., SINGLETON, A. B., NALLS, M. A., INTERNATIONAL PARKINSON'S DISEASE GENOMICS, C. & WELLCOME TRUST CASE CONTROL, C. 2012. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum Mol Genet,* 21**,** 4996-5009.

KELLIS, M., WOLD, B., SNYDER, M. P., BERNSTEIN, B. E., KUNDAJE, A., MARINOV, G. K., WARD, L. D., BIRNEY, E., CRAWFORD, G. E., DEKKER, J., DUNHAM, I.,

ELNITSKI, L. L., FARNHAM, P. J., FEINGOLD, E. A., GERSTEIN, M., GIDDINGS, M. C., GILBERT, D. M., GINGERAS, T. R., GREEN, E. D., GUIGO, R., HUBBARD, T., KENT, J., LIEB, J. D., MYERS, R. M., PAZIN, M. J., REN, B., STAMATOYANNOPOULOS, J. A., WENG, Z., WHITE, K. P. & HARDISON, R. C. 2014. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A,* 111**,** 6131-8.

KHAN, N. L., GRAHAM, E., CRITCHLEY, P., SCHRAG, A. E., WOOD, N. W., LEES, A. J., BHATIA, K. P. & QUINN, N. 2003. Parkin disease: a phenotypic study of a large case series. *Brain,* 126**,** 1279-92.

KITADA, T., ASAKAWA, S., HATTORI, N., MATSUMINE, H., YAMAMURA, Y., MINOSHIMA, S., YOKOCHI, M., MIZUNO, Y. & SHIMIZU, N. 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature,* 392**,** 605-8.

KLEIN, C. & WESTENBERGER, A. 2012. Genetics of Parkinson's disease. *Cold Spring Harb Perspect Med,* 2**,** a008888.

KORDOWER, J. H., CHU, Y., HAUSER, R. A., FREEMAN, T. B. & OLANOW, C. W. 2008. Lewy body-like pathology in long-term embryonic nigral transplants in Parkinson's disease. *Nat Med,* 14**,** 504-6.

KORDOWER, J. H., OLANOW, C. W., DODIYA, H. B., CHU, Y., BEACH, T. G., ADLER, C. H., HALLIDAY, G. M. & BARTUS, R. T. 2013. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. *Brain,* 136**,** 2419-31.

KOROGLU, C., BAYSAL, L., CETINKAYA, M., KARASOY, H. & TOLUN, A. 2013. DNAJC6 is responsible for juvenile parkinsonism with phenotypic variability. *Parkinsonism Relat Disord,* 19**,** 320-4.

KORVATSKA, O., STRAND, N. S., BERNDT, J. D., STROVAS, T., CHEN, D. H., LEVERENZ, J. B., KIIANITSA, K., MATA, I. F., KARAKOC, E., GREENUP, J. L., BONKOWSKI, E., CHUANG, J., MOON, R. T., EICHLER, E. E., NICKERSON, D. A., ZABETIAN, C. P., KRAEMER, B. C., BIRD, T. D. & RASKIND, W. H. 2013. Altered splicing of ATP6AP2 causes X-linked parkinsonism with spasticity (XPDS). *Hum Mol Genet,* 22**,** 3259-68.

KRAFT, P. & COX, D. G. 2008. Study designs for genome-wide association studies. *Adv Genet,* 60**,** 465-504.

KRUGER, R., KUHN, W., MULLER, T., WOITALLA, D., GRAEBER, M., KOSEL, S., PRZUNTEK, H., EPPLEN, J. T., SCHOLS, L. & RIESS, O. 1998. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat Genet,* 18**,** 106-8.

KRUGER, R., VIEIRA-SAECKER, A. M., KUHN, W., BERG, D., MULLER, T., KUHNL, N., FUCHS, G. A., STORCH, A., HUNGS, M., WOITALLA, D., PRZUNTEK, H., EPPLEN, J. T., SCHOLS, L. & RIESS, O. 1999. Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype. *Ann Neurol,* 45**,** 611-7.

KULZER, J. R., STITZEL, M. L., MORKEN, M. A., HUYGHE, J. R., FUCHSBERGER, C., KUUSISTO, J., LAAKSO, M., BOEHNKE, M., COLLINS, F. S. & MOHLKE, K. L. 2014. A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *Am J Hum Genet,* 94**,** 186-97.

KURIAN, M. A., GISSEN, P., SMITH, M., HEALES, S., JR. & CLAYTON, P. T. 2011. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. *Lancet Neurol,* 10**,** 721-33.

KURZ, M., ALVES, G., AARSLAND, D. & LARSEN, J. P. 2003. Familial Parkinson's disease: a community-based study. *Eur J Neurol,* 10**,** 159-63.

LANDER, E. S. 2011. Initial impact of the sequencing of the human genome. *Nature,* 470**,** 187-97.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LANG, A. E. 2011. A critical appraisal of the premotor symptoms of Parkinson's disease: potential usefulness in early diagnosis and design of neuroprotective trials. *Mov Disord,* 26**,** 775-83.

LANGSTON, J. W., BALLARD, P., TETRUD, J. W. & IRWIN, I. 1983. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science,* 219**,** 979-80.

LEES, A. J. 2007. Unresolved issues relating to the shaking palsy on the celebration of James Parkinson's 250th birthday. *Mov Disord,* 22 Suppl 17**,** S327-34.

LESAGE, S., ANHEIM, M., LETOURNEL, F., BOUSSET, L., HONORE, A., ROZAS, N., PIERI, L., MADIONA, K., DURR, A., MELKI, R., VERNY, C., BRICE, A. & FRENCH PARKINSON'S DISEASE GENETICS STUDY, G. 2013. G51D alpha-synuclein mutation causes a novel parkinsonian-pyramidal syndrome. *Ann Neurol,* 73**,** 459-71.

LEWY, F. J. H. 1912. Pathologische Anatomie. *In:* LEWANDOWSKY, M. A., G. (ed.) *Handbuch der Neurologie Vol. 3.* Berlin: Springer-Verlag.

LI, J. Y., ENGLUND, E., HOLTON, J. L., SOULET, D., HAGELL, P., LEES, A. J., LASHLEY, T., QUINN, N. P., REHNCRONA, S., BJORKLUND, A., WIDNER, H., REVESZ, T., LINDVALL, O. & BRUNDIN, P. 2008. Lewy bodies in grafted neurons in

subjects with Parkinson's disease suggest host-to-graft disease propagation. *Nat Med,* 14**,** 501-3.

LILL, C. M., HANSEN, J., OLSEN, J. H., BINDER, H., RITZ, B. & BERTRAM, L. 2015. Impact of Parkinson's disease risk loci on age at onset. *Mov Disord*, 30, 847-50.

LILL, C. M., ROEHR, J. T., MCQUEEN, M. B., KAVVOURA, F. K., BAGADE, S., SCHJEIDE, B. M., SCHJEIDE, L. M., MEISSNER, E., ZAUFT, U., ALLEN, N. C., LIU, T., SCHILLING, M., ANDERSON, K. J., BEECHAM, G., BERG, D., BIERNACKA, J. M., BRICE, A., DESTEFANO, A. L., DO, C. B., ERIKSSON, N., FACTOR, S. A., FARRER, M. J., FOROUD, T., GASSER, T., HAMZA, T., HARDY, J. A., HEUTINK, P., HILL-BURNS, E. M., KLEIN, C., LATOURELLE, J. C., MARAGANORE, D. M., MARTIN, E. R., MARTINEZ, M., MYERS, R. H., NALLS, M. A., PANKRATZ, N., PAYAMI, H., SATAKE, W., SCOTT, W. K., SHARMA, M., SINGLETON, A. B., STEFANSSON, K., TODA, T., TUNG, J. Y., VANCE, J., WOOD, N. W., ZABETIAN, C. P., YOUNG, P., TANZI, R. E., KHOURY, M. J., ZIPP, F., LEHRACH, H., IOANNIDIS, J. P. & BERTRAM, L. 2012. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet,* 8**,** e1002548.

LIPPA, C. F., DUDA, J. E., GROSSMAN, M., HURTIG, H. I., AARSLAND, D., BOEVE, B. F., BROOKS, D. J., DICKSON, D. W., DUBOIS, B., EMRE, M., FAHN, S., FARMER, J. M., GALASKO, D., GALVIN, J. E., GOETZ, C. G., GROWDON, J. H., GWINN-HARDY, K. A., HARDY, J., HEUTINK, P., IWATSUBO, T., KOSAKA, K., LEE, V. M., LEVERENZ, J. B., MASLIAH, E., MCKEITH, I. G., NUSSBAUM, R. L., OLANOW, C. W., RAVINA, B. M., SINGLETON, A. B., TANNER, C. M., TROJANOWSKI, J. Q., WSZOLEK, Z. K. & GROUP, D. P. W. 2007. DLB and PDD boundary issues: diagnosis, treatment, molecular pathology, and biomarkers. *Neurology,* 68**,** 812-9.

LITVAN, I., MACINTYRE, A., GOETZ, C. G., WENNING, G. K., JELLINGER, K., VERNY, M., BARTKO, J. J., JANKOVIC, J., MCKEE, A., BRANDEL, J. P., CHAUDHURI, K. R., LAI, E. C., D'OLHABERRIAGUE, L., PEARCE, R. K. & AGID, Y. 1998. Accuracy of the clinical diagnoses of Lewy body disease, Parkinson disease, and dementia with Lewy bodies: a clinicopathologic study. *Arch Neurol,* 55**,** 969-78.

LUCKING, C. B., DURR, A., BONIFATI, V., VAUGHAN, J., DE MICHELE, G., GASSER, T., HARHANGI, B. S., MECO, G., DENEFLE, P., WOOD, N. W., AGID, Y., BRICE, A., FRENCH PARKINSON'S DISEASE GENETICS STUDY, G. & EUROPEAN CONSORTIUM ON GENETIC SUSCEPTIBILITY IN PARKINSON'S, D. 2000. Association between early-onset Parkinson's disease and mutations in the parkin gene. *N Engl J Med,* 342**,** 1560-7.

MACLEOD, D. A., RHINN, H., KUWAHARA, T., ZOLIN, A., DI PAOLO, G., MCCABE, B. D., MARDER, K. S., HONIG, L. S., CLARK, L. N., SMALL, S. A. & ABELIOVICH, A. 2013. RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron,* 77**,** 425-39.

MAHER, B. 2008. Personal genomes: The case of the missing heritability. *Nature,* 456**,** 18-21.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R.,

CHAKRAVARTI, A., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G., HAINES, J. L., MACKAY, T. F., MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MANYAM, B. V. 1990. Paralysis agitans and levodopa in "Ayurveda": ancient Indian medical treatise. *Mov Disord,* 5**,** 47-8.

MARAGANORE, D. M., DE ANDRADE, M., ELBAZ, A., FARRER, M. J., IOANNIDIS, J. P., KRUGER, R., ROCCA, W. A., SCHNEIDER, N. K., LESNICK, T. G., LINCOLN, S. J., HULIHAN, M. M., AASLY, J. O., ASHIZAWA, T., CHARTIER-HARLIN, M. C., CHECKOWAY, H., FERRARESE, C., HADJIGEORGIOU, G., HATTORI, N., KAWAKAMI, H., LAMBERT, J. C., LYNCH, T., MELLICK, G. D., PAPAPETROPOULOS, S., PARSIAN, A., QUATTRONE, A., RIESS, O., TAN, E. K., VAN BROECKHOVEN, C. & GENETIC EPIDEMIOLOGY OF PARKINSON'S DISEASE, C. 2006. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *JAMA,* 296**,** 661-70.

MARAGANORE, D. M., DE ANDRADE, M., LESNICK, T. G., STRAIN, K. J., FARRER, M. J., ROCCA, W. A., PANT, P. V., FRAZER, K. A., COX, D. R. & BALLINGER, D. G. 2005. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet,* 77**,** 685-93.

MARCHINI, J., HOWIE, B., MYERS, S., MCVEAN, G. & DONNELLY, P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet,* 39**,** 906-13.

MARDIS, E. R. 2011. A decade's perspective on DNA sequencing technology. *Nature,* 470**,** 198-203.

MARRAS, C. & LANG, A. 2013. Parkinson's disease subtypes: lost in translation? *J Neurol Neurosurg Psychiatry,* 84**,** 409-15.

MARTIN, E. R., SCOTT, W. K., NANCE, M. A., WATTS, R. L., HUBBLE, J. P., KOLLER, W. C., LYONS, K., PAHWA, R., STERN, M. B., COLCHER, A., HINER, B. C., JANKOVIC, J., ONDO, W. G., ALLEN, F. H., JR., GOETZ, C. G., SMALL, G. W., MASTERMAN, D., MASTAGLIA, F., LAING, N. G., STAJICH, J. M., RIBBLE, R. C., BOOZE, M. W., ROGALA, A., HAUSER, M. A., ZHANG, F., GIBSON, R. A., MIDDLETON, L. T., ROSES, A. D., HAINES, J. L., SCOTT, B. L., PERICAK-VANCE, M. A. & VANCE, J. M. 2001. Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. *JAMA,* 286**,** 2245-50.

MARTINEZ-MARTIN, P. 2011. The importance of non-motor disturbances to quality of life in Parkinson's disease. *J Neurol Sci,* 310**,** 12-6.

MATA, I. F., SHI, M., AGARWAL, P., CHUNG, K. A., EDWARDS, K. L., FACTOR, S. A., GALASKO, D. R., GINGHINA, C., GRIFFITH, A., HIGGINS, D. S., KAY, D. M., KIM, H., LEVERENZ, J. B., QUINN, J. F., ROBERTS, J. W., SAMII, A., SNAPINN, K. W., TSUANG, D. W., YEAROUT, D., ZHANG, J., PAYAMI, H. & ZABETIAN, C. P. 2010. SNCA variant associated with Parkinson disease and plasma alpha-synuclein level. *Arch Neurol,* 67**,** 1350-6.

MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R., GOLDSTEIN, D. B., LITTLE, J., IOANNIDIS, J. P. & HIRSCHHORN, J. N. 2008. Genome-wide association

studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet,* 9**,** 356-69.

MENCACCI, N. E., ISAIAS, I. U., REICH, M. M., GANOS, C., PLAGNOL, V., POLKE, J. M., BRAS, J., HERSHESON, J., STAMELOU, M., PITTMAN, A. M., NOYCE, A. J., MOK, K. Y., OPLADEN, T., KUNSTMANN, E., HODECKER, S., MUNCHAU, A., VOLKMANN, J., SAMNICK, S., SIDLE, K., NANJI, T., SWEENEY, M. G., HOULDEN, H., BATLA, A., ZECCHINELLI, A. L., PEZZOLI, G., MAROTTA, G., LEES, A., ALEGRIA, P., KRACK, P., CORMIER-DEQUAIRE, F., LESAGE, S., BRICE, A., HEUTINK, P., GASSER, T., LUBBE, S. J., MORRIS, H. R., TABA, P., KOKS, S., MAJOUNIE, E., RAPHAEL GIBBS, J., SINGLETON, A., HARDY, J., KLEBE, S., BHATIA, K. P., WOOD, N. W., INTERNATIONAL PARKINSON'S DISEASE GENOMICS, C. & CONSORTIUM, U. C.-E. 2014. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain,* 137**,** 2480-92.

MJÖNES, H. 1949. *Paralysis Agitans: A Clinical and Genetic Study*, E. Munksgaard.

MOSKVINA, V., HOLMANS, P., SCHMIDT, K. M. & CRADDOCK, N. 2005. Design of case-controls studies with unscreened controls. *Ann Hum Genet,* 69**,** 566-76.

MUENTER, M. D., FORNO, L. S., HORNYKIEWICZ, O., KISH, S. J., MARAGANORE, D. M., CASELLI, R. J., OKAZAKI, H., HOWARD, F. M., JR., SNOW, B. J. & CALNE, D. B. 1998. Hereditary form of parkinsonism--dementia. *Ann Neurol,* 43**,** 768-81.

MUSUNURU, K., STRONG, A., FRANK-KAMENETSKY, M., LEE, N. E., AHFELDT, T., SACHS, K. V., LI, X., LI, H., KUPERWASSER, N., RUDA, V. M., PIRRUCCELLO, J. P., MUCHMORE, B., PROKUNINA-OLSSON, L., HALL, J. L., SCHADT, E. E., MORALES, C. R., LUND-KATZ, S., PHILLIPS, M. C., WONG, J., CANTLEY, W., RACIE, T., EJEBE, K. G., ORHO-MELANDER, M., MELANDER, O., KOTELIANSKY, V., FITZGERALD, K., KRAUSS, R. M., COWAN, C. A., KATHIRESAN, S. & RADER, D. J. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature,* 466**,** 714-9.

NALLS, M. A., ESCOTT-PRICE, V., WILLIAMS, N. M., LUBBE, S., KELLER, M. F., MORRIS, H. R., SINGLETON, A. B. & INTERNATIONAL PARKINSON'S DISEASE GENOMICS, C. 2015. Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov Disord,* 30**,** 850-4.

NALLS, M. A., PANKRATZ, N., LILL, C. M., DO, C. B., HERNANDEZ, D. G., SAAD, M., DESTEFANO, A. L., KARA, E., BRAS, J., SHARMA, M., SCHULTE, C., KELLER, M. F., AREPALLI, S., LETSON, C., EDSALL, C., STEFANSSON, H., LIU, X., PLINER, H., LEE, J. H., CHENG, R., IKRAM, M. A., IOANNIDIS, J. P., HADJIGEORGIOU, G. M., BIS, J. C., MARTINEZ, M., PERLMUTTER, J. S., GOATE, A., MARDER, K., FISKE, B., SUTHERLAND, M., XIROMERISIOU, G., MYERS, R. H., CLARK, L. N., STEFANSSON, K., HARDY, J. A., HEUTINK, P., CHEN, H., WOOD, N. W., HOULDEN, H., PAYAMI, H., BRICE, A., SCOTT, W. K., GASSER, T., BERTRAM, L., ERIKSSON, N., FOROUD, T. & SINGLETON, A. B. 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet,* 46, 989-93.

NALLS, M. A., PLAGNOL, V., HERNANDEZ, D. G., SHARMA, M., SHEERIN, U. M., SAAD, M., SIMON-SANCHEZ, J., SCHULTE, C., LESAGE, S., SVEINBJORNSDOTTIR, S., STEFANSSON, K., MARTINEZ, M., HARDY, J.,

HEUTINK, P., BRICE, A., GASSER, T., SINGLETON, A. B. & WOOD, N. W. 2011. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet,* 377**,** 641-9.

NEPH, S., VIERSTRA, J., STERGACHIS, A. B., REYNOLDS, A. P., HAUGEN, E., VERNOT, B., THURMAN, R. E., JOHN, S., SANDSTROM, R., JOHNSON, A. K., MAURANO, M. T., HUMBERT, R., RYNES, E., WANG, H., VONG, S., LEE, K., BATES, D., DIEGEL, M., ROACH, V., DUNN, D., NERI, J., SCHAFER, A., HANSEN, R. S., KUTYAVIN, T., GISTE, E., WEAVER, M., CANFIELD, T., SABO, P., ZHANG, M., BALASUNDARAM, G., BYRON, R., MACCOSS, M. J., AKEY, J. M., BENDER, M. A., GROUDINE, M., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature,* 489**,** 83-90.

NEYMAN, J. & PEARSON, E. S. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference Part I. *Biometrika,* 20A**,** 175-240.

NICOLAE, D. L., GAMAZON, E., ZHANG, W., DUAN, S., DOLAN, M. E. & COX, N. J. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet,* 6**,** e1000888.

NOMBELA, C., ROWE, J. B., WINDER-RHODES, S. E., HAMPSHIRE, A., OWEN, A. M., BREEN, D. P., DUNCAN, G. W., KHOO, T. K., YARNALL, A. J., FIRBANK, M. J., CHINNERY, P. F., ROBBINS, T. W., O'BRIEN, J. T., BROOKS, D. J., BURN, D. J., GROUP, I.-P. S. & BARKER, R. A. 2014. Genetic impact on cognition and brain function in newly diagnosed Parkinson's disease: ICICLE-PD study. *Brain,* 137**,** 2743-58.

NOVEMBRE, J., JOHNSON, T., BRYC, K., KUTALIK, Z., BOYKO, A. R., AUTON, A., INDAP, A., KING, K. S., BERGMANN, S., NELSON, M. R., STEPHENS, M. & BUSTAMANTE, C. D. 2008. Genes mirror geography within Europe. *Nature,* 456**,** 98-101.

NUSSBAUM, R. L. & ELLIS, C. E. 2003. Alzheimer's disease and Parkinson's disease. *N Engl J Med,* 348**,** 1356-64.

OLANOW, C. W., RASCOL, O., HAUSER, R., FEIGIN, P. D., JANKOVIC, J., LANG, A., LANGSTON, W., MELAMED, E., POEWE, W., STOCCHI, F., TOLOSA, E. & INVESTIGATORS, A. S. 2009. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. *N Engl J Med,* 361**,** 1268-78.

PAHL, R. & SCHAFER, H. 2010. PERMORY: an LD-exploiting permutation test algorithm for powerful genome-wide association testing. *Bioinformatics,* 26**,** 2093-100.

PAISAN-RUIZ, C., BHATIA, K. P., LI, A., HERNANDEZ, D., DAVIS, M., WOOD, N. W., HARDY, J., HOULDEN, H., SINGLETON, A. & SCHNEIDER, S. A. 2009. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. *Ann Neurol,* 65**,** 19-23.

PAISAN-RUIZ, C., JAIN, S., EVANS, E. W., GILKS, W. P., SIMON, J., VAN DER BRUG, M., LOPEZ DE MUNAIN, A., APARICIO, S., GIL, A. M., KHAN, N., JOHNSON, J., MARTINEZ, J. R., NICHOLL, D., CARRERA, I. M., PENA, A. S., DE SILVA, R., LEES, A., MARTI-MASSO, J. F., PEREZ-TUR, J., WOOD, N. W. & SINGLETON,

A. B. 2004. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron,* 44**,** 595-600.

PANAGIOTOU, O. A., IOANNIDIS, J. P. & GENOME-WIDE SIGNIFICANCE, P. 2012. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol,* 41**,** 273-86.

PANKRATZ, N., BEECHAM, G. W., DESTEFANO, A. L., DAWSON, T. M., DOHENY, K. F., FACTOR, S. A., HAMZA, T. H., HUNG, A. Y., HYMAN, B. T., IVINSON, A. J., KRAINC, D., LATOURELLE, J. C., CLARK, L. N., MARDER, K., MARTIN, E. R., MAYEUX, R., ROSS, O. A., SCHERZER, C. R., SIMON, D. K., TANNER, C., VANCE, J. M., WSZOLEK, Z. K., ZABETIAN, C. P., MYERS, R. H., PAYAMI, H., SCOTT, W. K. & FOROUD, T. 2012. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann Neurol,* 71**,** 370-84.

PARKINSON, J. 1817. *An essay on the shaking palsy,* London, Sherwood, Neely and Jones.

PASANEN, P., MYLLYKANGAS, L., SIITONEN, M., RAUNIO, A., KAAKKOLA, S., LYYTINEN, J., TIENARI, P. J., POYHONEN, M. & PAETAU, A. 2014. Novel alpha-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol Aging,* 35**,** 2180 e1-5.

PICCINI, P., BURN, D. J., CERAVOLO, R., MARAGANORE, D. & BROOKS, D. J. 1999. The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. *Ann Neurol,* 45**,** 577-82.

PICKRELL, A. M. & YOULE, R. J. 2015. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron,* 85**,** 257-73.

PIHLSTROM, L. & TOFT, M. 2011. Genetic variability in SNCA and Parkinson's disease. *Neurogenetics,* 12**,** 283-93.

POLYMEROPOULOS, M. H., HIGGINS, J. J., GOLBE, L. I., JOHNSON, W. G., IDE, S. E., DI IORIO, G., SANGES, G., STENROOS, E. S., PHO, L. T., SCHAFFER, A. A., LAZZARINI, A. M., NUSSBAUM, R. L. & DUVOISIN, R. C. 1996. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science,* 274**,** 1197-9.

POLYMEROPOULOS, M. H., LAVEDAN, C., LEROY, E., IDE, S. E., DEHEJIA, A., DUTRA, A., PIKE, B., ROOT, H., RUBENSTEIN, J., BOYER, R., STENROOS, E. S., CHANDRASEKHARAPPA, S., ATHANASSIADOU, A., PAPAPETROPOULOS, T., JOHNSON, W. G., LAZZARINI, A. M., DUVOISIN, R. C., DI IORIO, G., GOLBE, L. I. & NUSSBAUM, R. L. 1997. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science,* 276**,** 2045-7.

POULOPOULOS, M., LEVY, O. A. & ALCALAY, R. N. 2012. The neuropathology of genetic Parkinson's disease. *Mov Disord,* 27**,** 831-42.

PROUKAKIS, C., DUDZIK, C. G., BRIER, T., MACKAY, D. S., COOPER, J. M., MILLHAUSER, G. L., HOULDEN, H. & SCHAPIRA, A. H. 2013. A novel alpha-synuclein missense mutation in Parkinson disease. *Neurology,* 80**,** 1062-4.

PURCELL, S., CHERNY, S. S. & SHAM, P. C. 2003. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics,* 19**,** 149-50.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I., DALY, M. J. & SHAM, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**,** 559-75.

RAJ, T., ROTHAMEL, K., MOSTAFAVI, S., YE, C., LEE, M. N., REPLOGLE, J. M., FENG, T., LEE, M., ASINOVSKI, N., FROHLICH, I., IMBOYWA, S., VON KORFF, A., OKADA, Y., PATSOPOULOS, N. A., DAVIS, S., MCCABE, C., PAIK, H. I., SRIVASTAVA, G. P., RAYCHAUDHURI, S., HAFLER, D. A., KOLLER, D., REGEV, A., HACOHEN, N., MATHIS, D., BENOIST, C., STRANGER, B. E. & DE JAGER, P. L. 2014. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science,* 344**,** 519-23.

RAMASAMY, A., TRABZUNI, D., GUELFI, S., VARGHESE, V., SMITH, C., WALKER, R., DE, T., COIN, L., DE SILVA, R., COOKSON, M. R., SINGLETON, A. B., HARDY, J., RYTEN, M. & WEALE, M. E. 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci,* 17**,** 1418-28.

RAMIREZ, A., HEIMBACH, A., GRUNDEMANN, J., STILLER, B., HAMPSHIRE, D., CID, L. P., GOEBEL, I., MUBAIDIN, A. F., WRIEKAT, A. L., ROEPER, J., AL-DIN, A., HILLMER, A. M., KARSAK, M., LISS, B., WOODS, C. G., BEHRENS, M. I. & KUBISCH, C. 2006. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nat Genet,* 38**,** 1184-91.

RAUDINO, F. 2012. The Parkinson disease before James Parkinson. *Neurological Sciences,* 33**,** 945-948.

RICE, T. K., SCHORK, N. J. & RAO, D. C. 2008. Methods for handling multiple testing. *Adv Genet,* 60**,** 293-308.

RITZ, B., RHODES, S. L., BORDELON, Y. & BRONSTEIN, J. 2012. alpha-Synuclein genetic variants predict faster motor symptom progression in idiopathic Parkinson disease. *PLoS One,* 7**,** e36199.

ROTHAUG, M., ZUNKE, F., MAZZULLI, J. R., SCHWEIZER, M., ALTMEPPEN, H., LULLMANN-RAUCH, R., KALLEMEIJN, W. W., GASPAR, P., AERTS, J. M., GLATZEL, M., SAFTIG, P., KRAINC, D., SCHWAKE, M. & BLANZ, J. 2014. LIMP-2 expression is critical for beta-glucocerebrosidase activity and alpha-synuclein clearance. *Proc Natl Acad Sci U S A,* 111**,** 15573-8.

SAAD, M., LESAGE, S., SAINT-PIERRE, A., CORVOL, J. C., ZELENIKA, D., LAMBERT, J. C., VIDAILHET, M., MELLICK, G. D., LOHMANN, E., DURIF, F., POLLAK, P., DAMIER, P., TISON, F., SILBURN, P. A., TZOURIO, C., FORLANI, S., LORIOT, M. A., GIROUD, M., HELMER, C., PORTET, F., AMOUYEL, P., LATHROP, M., ELBAZ, A., DURR, A., MARTINEZ, M. & BRICE, A. 2011. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum Mol Genet,* 20**,** 615-27.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SANYAL, A., LAJOIE, B. R., JAIN, G. & DEKKER, J. 2012. The long-range interaction landscape of gene promoters. *Nature,* 489**,** 109-13.

SATAKE, W., NAKABAYASHI, Y., MIZUTA, I., HIROTA, Y., ITO, C., KUBO, M., KAWAGUCHI, T., TSUNODA, T., WATANABE, M., TAKEDA, A., TOMIYAMA, H., NAKASHIMA, K., HASEGAWA, K., OBATA, F., YOSHIKAWA, T., KAWAKAMI, H., SAKODA, S., YAMAMOTO, M., HATTORI, N., MURATA, M., NAKAMURA, Y. & TODA, T. 2009. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet,* 41**,** 1303-7.

SCHERZER, C. R., GRASS, J. A., LIAO, Z., PEPIVANI, I., ZHENG, B., EKLUND, A. C., NEY, P. A., NG, J., MCGOLDRICK, M., MOLLENHAUER, B., BRESNICK, E. H. & SCHLOSSMACHER, M. G. 2008. GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein. *Proc Natl Acad Sci U S A,* 105**,** 10907-12.

SCHLOTTERER, C., TOBLER, R., KOFLER, R. & NOLTE, V. 2014. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nat Rev Genet,* 15**,** 749-63.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc,* 3**,** 1101-8.

SCHULTE, E. C., FUKUMORI, A., MOLLENHAUER, B., HOR, H., ARZBERGER, T., PERNECZKY, R., KURZ, A., DIEHL-SCHMID, J., HULL, M., LICHTNER, P., ECKSTEIN, G., ZIMPRICH, A., HAUBENBERGER, D., PIRKER, W., BRUCKE, T., BEREZNAI, B., MOLNAR, M. J., LORENZO-BETANCOR, O., PASTOR, P., PETERS, A., GIEGER, C., ESTIVILL, X., MEITINGER, T., KRETZSCHMAR, H. A., TRENKWALDER, C., HAASS, C. & WINKELMANN, J. 2015. Rare variants in beta-Amyloid precursor protein (APP) and Parkinson's disease. *Eur J Hum Genet*. Epub ahead of print. DOI: 10.1038/ejhg.2014.300.

SHAM, P. C. & PURCELL, S. M. 2014. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet,* 15**,** 335-46.

SHIMURA, H., HATTORI, N., KUBO, S., MIZUNO, Y., ASAKAWA, S., MINOSHIMA, S., SHIMIZU, N., IWAI, K., CHIBA, T., TANAKA, K. & SUZUKI, T. 2000. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat Genet,* 25**,** 302-5.

SHOJAEE, S., SINA, F., BANIHOSSEINI, S. S., KAZEMI, M. H., KALHOR, R., SHAHIDI, G. A., FAKHRAI-RAD, H., RONAGHI, M. & ELAHI, E. 2008. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. *Am J Hum Genet,* 82**,** 1375-84.

SIDRANSKY, E., NALLS, M. A., AASLY, J. O., AHARON-PERETZ, J., ANNESI, G., BARBOSA, E. R., BAR-SHIRA, A., BERG, D., BRAS, J., BRICE, A., CHEN, C. M., CLARK, L. N., CONDROYER, C., DE MARCO, E. V., DURR, A., EBLAN, M. J., FAHN, S., FARRER, M. J., FUNG, H. C., GAN-OR, Z., GASSER, T., GERSHONI-BARUCH, R., GILADI, N., GRIFFITH, A., GUREVICH, T., JANUARIO, C., KROPP, P., LANG, A. E., LEE-CHEN, G. J., LESAGE, S., MARDER, K., MATA, I. F., MIRELMAN, A., MITSUI, J., MIZUTA, I., NICOLETTI, G., OLIVEIRA, C., OTTMAN, R., ORR-URTREGER, A., PEREIRA, L. V., QUATTRONE, A., ROGAEVA, E., ROLFS, A., ROSENBAUM, H., ROZENBERG, R., SAMII, A., SAMADDAR, T., SCHULTE, C., SHARMA, M., SINGLETON, A., SPITZ, M., TAN, E. K., TAYEBI, N., TODA, T., TROIANO, A. R., TSUJI, S., WITTSTOCK, M., WOLFSBERG, T. G., WU, Y. R., ZABETIAN, C. P., ZHAO, Y. & ZIEGLER, S. G.

2009. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med,* 361**,** 1651-61.

SILVER, N., BEST, S., JIANG, J. & THEIN, S. L. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol Biol,* 7**,** 33.

SIMON-SANCHEZ, J., SCHULTE, C., BRAS, J. M., SHARMA, M., GIBBS, J. R., BERG, D., PAISAN-RUIZ, C., LICHTNER, P., SCHOLZ, S. W., HERNANDEZ, D. G., KRUGER, R., FEDEROFF, M., KLEIN, C., GOATE, A., PERLMUTTER, J., BONIN, M., NALLS, M. A., ILLIG, T., GIEGER, C., HOULDEN, H., STEFFENS, M., OKUN, M. S., RACETTE, B. A., COOKSON, M. R., FOOTE, K. D., FERNANDEZ, H. H., TRAYNOR, B. J., SCHREIBER, S., AREPALLI, S., ZONOZI, R., GWINN, K., VAN DER BRUG, M., LOPEZ, G., CHANOCK, S. J., SCHATZKIN, A., PARK, Y., HOLLENBECK, A., GAO, J., HUANG, X., WOOD, N. W., LORENZ, D., DEUSCHL, G., CHEN, H., RIESS, O., HARDY, J. A., SINGLETON, A. B. & GASSER, T. 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet,* 41**,** 1308-12.

SINGLETON, A. B., FARRER, M., JOHNSON, J., SINGLETON, A., HAGUE, S., KACHERGUS, J., HULIHAN, M., PEURALINNA, T., DUTRA, A., NUSSBAUM, R., LINCOLN, S., CRAWLEY, A., HANSON, M., MARAGANORE, D., ADLER, C., COOKSON, M. R., MUENTER, M., BAPTISTA, M., MILLER, D., BLANCATO, J., HARDY, J. & GWINN-HARDY, K. 2003. alpha-Synuclein locus triplication causes Parkinson's disease. *Science,* 302**,** 841.

SKIPPER, L., WILKES, K., TOFT, M., BAKER, M., LINCOLN, S., HULIHAN, M., ROSS, O. A., HUTTON, M., AASLY, J. & FARRER, M. 2004. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet,* 75**,** 669-77.

SPATARO, N., CALAFELL, F., CERVERA-CARLES, L., CASALS, F., PAGONABARRAGA, J., PASCUAL-SEDANO, B., CAMPOLONGO, A., KULISEVSKY, J., LLEO, A., NAVARRO, A., CLARIMON, J. & BOSCH, E. 2015. Mendelian genes for Parkinson's disease contribute to the sporadic forms of the diseasedagger. *Hum Mol Genet,* 24**,** 2023-34.

SPELLMAN, G. G. 1962. Report of familial cases of parkinsonism. Evidence of a dominant trait in a patient's family. *JAMA,* 179**,** 372-4.

SPENCER, C. C., PLAGNOL, V., STRANGE, A., GARDNER, M., PAISAN-RUIZ, C., BAND, G., BARKER, R. A., BELLENGUEZ, C., BHATIA, K., BLACKBURN, H., BLACKWELL, J. M., BRAMON, E., BROWN, M. A., BROWN, M. A., BURN, D., CASAS, J. P., CHINNERY, P. F., CLARKE, C. E., CORVIN, A., CRADDOCK, N., DELOUKAS, P., EDKINS, S., EVANS, J., FREEMAN, C., GRAY, E., HARDY, J., HUDSON, G., HUNT, S., JANKOWSKI, J., LANGFORD, C., LEES, A. J., MARKUS, H. S., MATHEW, C. G., MCCARTHY, M. I., MORRISON, K. E., PALMER, C. N., PEARSON, J. P., PELTONEN, L., PIRINEN, M., PLOMIN, R., POTTER, S., RAUTANEN, A., SAWCER, S. J., SU, Z., TREMBATH, R. C., VISWANATHAN, A. C., WILLIAMS, N. W., MORRIS, H. R., DONNELLY, P. & WOOD, N. W. 2011. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum Mol Genet,* 20**,** 345-53.

SPILLANTINI, M. G., SCHMIDT, M. L., LEE, V. M., TROJANOWSKI, J. Q., JAKES, R. & GOEDERT, M. 1997. Alpha-synuclein in Lewy bodies. *Nature,* 388**,** 839-40.

STERN, G. 1989. Did parkinsonism occur before 1817? *J Neurol Neurosurg Psychiatry,* Suppl**,** 11-2.

SUNDQUIST, K., LI, X. & HEMMINKI, K. 2006. Familial risks of hospitalization for Parkinson's disease in first-degree relatives: a nationwide follow-up study from Sweden. *Neurogenetics,* 7**,** 231-7.

SVEINBJORNSDOTTIR, S., HICKS, A. A., JONSSON, T., PETURSSON, H., GUGMUNDSSON, G., FRIGGE, M. L., KONG, A., GULCHER, J. R. & STEFANSSON, K. 2000. Familial aggregation of Parkinson's disease in Iceland. *N Engl J Med,* 343**,** 1765-70.

TAKEUCHI, F., YANAI, K., MORII, T., ISHINAGA, Y., TANIGUCHI-YANAI, K., NAGANO, S. & KATO, N. 2005. Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. *Genetics,* 170**,** 291-304.

THACKER, E. L. & ASCHERIO, A. 2008. Familial aggregation of Parkinson's disease: a meta-analysis. *Mov Disord,* 23**,** 1174-83.

THENGANATT, M. A. & JANKOVIC, J. 2014. Parkinson disease subtypes. *JAMA Neurol,* 71**,** 499-504.

THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., EBERSOL, A. K., FRUM, T., GISTE, E., JOHNSON, A. K., JOHNSON, E. M., KUTYAVIN, T., LAJOIE, B., LEE, B. K., LEE, K., LONDON, D., LOTAKIS, D., NEPH, S., NERI, F., NGUYEN, E. D., QU, H., REYNOLDS, A. P., ROACH, V., SAFI, A., SANCHEZ, M. E., SANYAL, A., SHAFER, A., SIMON, J. M., SONG, L., VONG, S., WEAVER, M., YAN, Y., ZHANG, Z., ZHANG, Z., LENHARD, B., TEWARI, M., DORSCHNER, M. O., HANSEN, R. S., NAVAS, P. A., STAMATOYANNOPOULOS, G., IYER, V. R., LIEB, J. D., SUNYAEV, S. R., AKEY, J. M., SABO, P. J., KAUL, R., FUREY, T. S., DEKKER, J., CRAWFORD, G. E. & STAMATOYANNOPOULOS, J. A. 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

TIAN, C., GREGERSEN, P. K. & SELDIN, M. F. 2008. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet,* 17**,** R143-50.

TRABZUNI, D., RYTEN, M., WALKER, R., SMITH, C., IMRAN, S., RAMASAMY, A., WEALE, M. E. & HARDY, J. 2011. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem,* 119**,** 275-82.

TRETIAKOFF, C. 1919. *Contribution á l'ètude de l'anatomie pathologique du locus niger de Soemmering avec quelques déductions relatives à la pathogénie des troubles du tonus musculaire et de la maladie de Parkinson.* University of Paris.

TRINH, J., GUELLA, I. & FARRER, M. J. 2014. Disease penetrance of late-onset parkinsonism: a meta-analysis. *JAMA Neurol,* 71**,** 1535-9.

VALENTE, E. M., ABOU-SLEIMAN, P. M., CAPUTO, V., MUQIT, M. M., HARVEY, K., GISPERT, S., ALI, Z., DEL TURCO, D., BENTIVOGLIO, A. R., HEALY, D. G., ALBANESE, A., NUSSBAUM, R., GONZALEZ-MALDONADO, R., DELLER, T., SALVI, S., CORTELLI, P., GILKS, W. P., LATCHMAN, D. S., HARVEY, R. J.,

DALLAPICCOLA, B., AUBURGER, G. & WOOD, N. W. 2004. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science,* 304**,** 1158-60.

VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol,* 3**,** RESEARCH0034.

VANDESOMPELE, J., KUBISTA, M. & PFAFFL, M. W. 2009. Reference Gene Validation Software for Improved Normalization. *In:* LOGAN, J., EDWARDS, K. & SAUNDERS, N. (eds.) *Real-Time PCR: Current Technology and Applications.* London: Caister Academic Press.

VEKRELLIS, K., XILOURI, M., EMMANOUILIDOU, E., RIDEOUT, H. J. & STEFANIS, L. 2011. Pathological roles of alpha-synuclein in neurological disorders. *Lancet Neurol,* 10**,** 1015-25.

VILARINO-GUELL, C., RAJPUT, A., MILNERWOOD, A. J., SHAH, B., SZU-TU, C., TRINH, J., YU, I., ENCARNACION, M., MUNSIE, L. N., TAPIA, L., GUSTAVSSON, E. K., CHOU, P., TATARNIKOV, I., EVANS, D. M., PISHOTTA, F. T., VOLTA, M., BECCANO-KELLY, D., THOMPSON, C., LIN, M. K., SHERMAN, H. E., HAN, H. J., GUENTHER, B. L., WASSERMAN, W. W., BERNARD, V., ROSS, C. J., APPEL-CRESSWELL, S., STOESSL, A. J., ROBINSON, C. A., DICKSON, D. W., ROSS, O. A., WSZOLEK, Z. K., AASLY, J. O., WU, R. M., HENTATI, F., GIBSON, R. A., MCPHERSON, P. S., GIRARD, M., RAJPUT, M., RAJPUT, A. H. & FARRER, M. J. 2014. DNAJC13 mutations in Parkinson disease. *Hum Mol Genet,* 23**,** 1794-801.

VILARINO-GUELL, C., WIDER, C., ROSS, O. A., DACHSEL, J. C., KACHERGUS, J. M., LINCOLN, S. J., SOTO-ORTOLAZA, A. I., COBB, S. A., WILHOITE, G. J., BACON, J. A., BEHROUZ, B., MELROSE, H. L., HENTATI, E., PUSCHMANN, A., EVANS, D. M., CONIBEAR, E., WASSERMAN, W. W., AASLY, J. O., BURKHARD, P. R., DJALDETTI, R., GHIKA, J., HENTATI, F., KRYGOWSKA-WAJS, A., LYNCH, T., MELAMED, E., RAJPUT, A., RAJPUT, A. H., SOLIDA, A., WU, R. M., UITTI, R. J., WSZOLEK, Z. K., VINGERHOETS, F. & FARRER, M. J. 2011. VPS35 mutations in Parkinson disease. *Am J Hum Genet,* 89**,** 162-7.

VOLKMANN, J., ALBANESE, A., ANTONINI, A., CHAUDHURI, K. R., CLARKE, C. E., DE BIE, R. M., DEUSCHL, G., EGGERT, K., HOUETO, J. L., KULISEVSKY, J., NYHOLM, D., ODIN, P., OSTERGAARD, K., POEWE, W., POLLAK, P., RABEY, J. M., RASCOL, O., RUZICKA, E., SAMUEL, M., SPEELMAN, H., SYDOW, O., VALLDEORIOLA, F., VAN DER LINDEN, C. & OERTEL, W. 2013. Selecting deep brain stimulation or infusion therapies in advanced Parkinson's disease: an evidence-based review. *J Neurol,* 260**,** 2701-14.

WATANABE, R. M. 2011. Statistical issues in gene association studies. *Methods Mol Biol,* 700**,** 17-36.

WEI, W. H., HEMANI, G. & HALEY, C. S. 2014. Detecting epistasis in human complex traits. *Nat Rev Genet,* 15**,** 722-33.

WILSON, G. R., SIM, J. C., MCLEAN, C., GIANNANDREA, M., GALEA, C. A., RISELEY, J. R., STEPHENSON, S. E., FITZPATRICK, E., HAAS, S. A., POPE, K., HOGAN, K. J., GREGG, R. G., BROMHEAD, C. J., WARGOWSKI, D. S., LAWRENCE, C. H., JAMES, P. A., CHURCHYARD, A., GAO, Y., PHELAN, D. G., GILLIES, G., SALCE,

N., STANFORD, L., MARSH, A. P., MIGNOGNA, M. L., HAYFLICK, S. J., LEVENTER, R. J., DELATYCKI, M. B., MELLICK, G. D., KALSCHEUER, V. M., D'ADAMO, P., BAHLO, M., AMOR, D. J. & LOCKHART, P. J. 2014. Mutations in RAB39B cause X-linked intellectual disability and early-onset Parkinson disease with alpha-synuclein pathology. *Am J Hum Genet,* 95**,** 729-35.

WIRDEFELDT, K., ADAMI, H. O., COLE, P., TRICHOPOULOS, D. & MANDEL, J. 2011. Epidemiology and etiology of Parkinson's disease: a review of the evidence. *Eur J Epidemiol,* 26 Suppl 1**,** S1-58.

WISSEMANN, W. T., HILL-BURNS, E. M., ZABETIAN, C. P., FACTOR, S. A., PATSOPOULOS, N., HOGLUND, B., HOLCOMB, C., DONAHUE, R. J., THOMSON, G., ERLICH, H. & PAYAMI, H. 2013. Association of Parkinson disease with structural and regulatory variants in the HLA region. *Am J Hum Genet,* 93**,** 984-93.

YAHR, M. D., DUVOISIN, R. C., SCHEAR, M. J., BARRETT, R. E. & HOEHN, M. M. 1969. Treatment of parkinsonism with levodopa. *Arch Neurol,* 21**,** 343-54.

YANG, J., FERREIRA, T., MORRIS, A. P., MEDLAND, S. E., GENETIC INVESTIGATION OF, A. T. C., REPLICATION, D. I. G., META-ANALYSIS, C., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., WEEDON, M. N., LOOS, R. J., FRAYLING, T. M., MCCARTHY, M. I., HIRSCHHORN, J. N., GODDARD, M. E. & VISSCHER, P. M. 2012. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet,* 44**,** 369-75, S1-3.

YANG, J., LEE, S. H., GODDARD, M. E. & VISSCHER, P. M. 2011a. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet,* 88**,** 76-82.

YANG, J., WEEDON, M. N., PURCELL, S., LETTRE, G., ESTRADA, K., WILLER, C. J., SMITH, A. V., INGELSSON, E., O'CONNELL, J. R., MANGINO, M., MAGI, R., MADDEN, P. A., HEATH, A. C., NYHOLT, D. R., MARTIN, N. G., MONTGOMERY, G. W., FRAYLING, T. M., HIRSCHHORN, J. N., MCCARTHY, M. I., GODDARD, M. E., VISSCHER, P. M. & CONSORTIUM, G. 2011b. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet,* 19**,** 807-12.

ZARRANZ, J. J., ALEGRE, J., GOMEZ-ESTEBAN, J. C., LEZCANO, E., ROS, R., AMPUERO, I., VIDAL, L., HOENICKA, J., RODRIGUEZ, O., ATARES, B., LLORENS, V., GOMEZ TORTOSA, E., DEL SER, T., MUNOZ, D. G. & DE YEBENES, J. G. 2004. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol,* 55**,** 164-73.

ZHENG, G., FREIDLIN, B. & GASTWIRTH, J. L. 2006. Robust genomic control for association studies. *Am J Hum Genet,* 78**,** 350-6.

ZIMPRICH, A., BENET-PAGES, A., STRUHAL, W., GRAF, E., ECK, S. H., OFFMAN, M. N., HAUBENBERGER, D., SPIELBERGER, S., SCHULTE, E. C., LICHTNER, P., ROSSLE, S. C., KLOPP, N., WOLF, E., SEPPI, K., PIRKER, W., PRESSLAUER, S., MOLLENHAUER, B., KATZENSCHLAGER, R., FOKI, T., HOTZY, C., REINTHALER, E., HARUTYUNYAN, A., KRALOVICS, R., PETERS, A., ZIMPRICH, F., BRUCKE, T., POEWE, W., AUFF, E., TRENKWALDER, C., ROST, B., RANSMAYR, G., WINKELMANN, J., MEITINGER, T. & STROM, T. M. 2011. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet,* 89**,** 168-75.

ZIMPRICH, A., BISKUP, S., LEITNER, P., LICHTNER, P., FARRER, M., LINCOLN, S., KACHERGUS, J., HULIHAN, M., UITTI, R. J., CALNE, D. B., STOESSL, A. J., PFEIFFER, R. F., PATENGE, N., CARBAJAL, I. C., VIEREGGE, P., ASMUS, F., MULLER-MYHSOK, B., DICKSON, D. W., MEITINGER, T., STROM, T. M., WSZOLEK, Z. K. & GASSER, T. 2004. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron,* 44**,** 601-7.