The Genetic Architecture of Parkinson's Disease:

Emphasis on Genetic Susceptibility

José Miguel Tomás Brás

Universidade de Coimbra

2010

Dissertação apresentada à

Faculdade de Medicina da Universidade de Coimbra

Para prestação de Provas de Doutoramento em Ciências Biomédicas

“Sometimes you see beautiful people with no brains.
Sometimes you have ugly people who are intelligent,
like scientists.”

José Mourinho, 2005

À Aninhas, ao Pedrinho, ao Miguelinho,

à Isabelinha, à Susaninha e,

em especial, ao Carlinhos

Acknowledgments

This study was carried out in the Laboratory of Neurogenetics, National Institutes on Aging, National Institutes of Health and the Center for Neuroscience and Cell Biology in Coimbra during the years 2006-2010.

I am grateful to all the people who contributed to this work and helped me during this process.

I wish to express my deepest gratitude to my co-supervisor Dr. Andrew Singleton for offering me the opportunity to join his group, introducing me to the world of genomics and for the warm and encouraging attitude towards this research, combined with his extensive genetics knowledge.

I also want to thank my co-supervisor, Professor Catarina Oliveira for her teachings and support and for always being present when help was needed.

I would also want to thank Dr. John Hardy, who was at the very beginning of this work. You have a unique way of lifting a student's spirits, particularly when things don't go as planned.

I am grateful to former and present students and postdoctoral researchers at the LNG for creating a friendly and productive environment. I owe my special gratitude to Mar and Erinn who were always able to lighten up a bad day, and to Javi, for what I've learned from you and for some of the funnest times in the lab. I would also like to thank Maria Helena, who has helped me immensely since the very early days of my undergraduate studies.

To Rita, who has been by my side through the best and the worse that life has thrown our way.

My deepest gratitude to my family, specially my mom and dad, for always being able to cheer me up and keep my mind off work. To my little sister, Martinha, a true case study in dedication and hard work.

This work was funded, in part, by Fundação para a Ciência e Tecnologia and Fundo Social Europeu, grant SFRH / BD / 29647 / 2006.

List of original publications

This thesis is based on the following original publications that are referred to in the text by the Roman numerals I-XI.

I. **Bras JM**, Singleton A. Genetic susceptibility in Parkinson's disease. Biochim Biophys Acta. 2009 Jul;1792(7):597-603. Epub 2008 Nov 20.

II. **Bras J**, Singleton A, Cookson MR, Hardy J. Emerging pathways in genetic Parkinson's disease: Potential role of ceramide metabolism in Lewy body disease. FEBS J. 2008 Dec;275(23):5767-73.

III. **Bras JM**, Guerreiro RJ, Ribeiro MH, Januario C, Morgadinho A, Oliveira CR, Cunha L, Hardy J, Singleton A. G2019S dardarin substitution is a common cause of Parkinson's disease in a Portuguese cohort. Mov Disord. 2005 Dec;20(12):1653-5.

IV. **Bras J**, Guerreiro R, Ribeiro M, Morgadinho A, Januario C, Dias M, Calado A, Semedo C, Oliveira C, Hardy J, Singleton A. Analysis of Parkinson disease patients from Portugal for mutations in SNCA, PRKN, PINK1 and LRRK2. BMC Neurol. 2008 Jan 22;8:1.

V. Okubadejo N, Britton A, Crews C, Akinyemi R, Hardy J, Singleton A, **Bras J**. Analysis of Nigerians with apparently sporadic Parkinson disease for mutations in LRRK2, PRKN and ATXN3. PLoS One. 2008;3(10):e3421. Epub 2008 Oct 17.

VI. Guerreiro RJ#, **Bras JM#**, Santana I, Januario C, Santiago B, Morgadinho AS, Ribeiro MH, Hardy J, Singleton A, Oliveira C. Association of HFE common mutations with Parkinson's disease, Alzheimer's disease and mild cognitive impairment in a Portuguese cohort. BMC Neurol. 2006 Jul 6;6:24.

VII. **Bras J#**, Simón-Sánchez J#, Federoff M, Morgadinho A, Januario C, Ribeiro M, Cunha L, Oliveira C, Singleton AB.Lack of replication of association between GIGYF2 variants and Parkinson disease. Hum Mol Genet. 2009 Jan 15;18(2):341-6. Epub 2008 Oct 15.

VIII. **Bras J**, Paisan-Ruiz C, Guerreiro R, Ribeiro MH, Morgadinho A, Januario C, Sidransky E, Oliveira C, Singleton A. Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal. Neurobiol Aging. 2009 Sep;30(9):1515-7. Epub 2007 Dec 21.

IX. Neumann J, **Bras J**, Deas E, O'Sullivan SS, Parkkinen L, Lachmann RH, Li A, Holton J, Guerreiro R, Paudel R, Segarane B, Singleton A, Lees A, Hardy J, Houlden H, Revesz T, Wood NW. Glucocerebrosidase mutations in clinical and pathologically proven

Parkinson's disease. Brain. 2009 Jul;132(Pt 7):1783-94. Epub 2009 Mar 13.

X. Sidransky E, Aasly J, Aharon-Peretz J, Annesi G, Barbosa ER, Bar-Shira A, Berg D, **Bras J**, Brice A, Chen C-M, Clark LN, Condroyer C, Marco EV, Dürr A, Eblan MJ, Fahn S, Farrer M, Fung H-C, Gan-Or Z, Gasser T, Gershoni-Baruch R, Giladi N, Griffith A, Gurevich T, Januario C, Kropp P, Lang A, Lee-Chen G-J, Lesage S, Marder K, Mata I, Mirelman A, Mitsui J, Mizuta I, Nalls MA, Nicoletti G, Oliveira C, Ottman R, Orr-Urtreger A, Pereira L, Quattrone A, Rogaeva E, Rolfs E, Rosenbaum H, Rozenberg R, Samii A, Samaddar T, Schulte C, Sharma M, Singleton A, Spitz M, Tan EK, Tayebi N, Toda T, Troiano A, Tsuji T, Wittstock M, Wolfsberg T, Wu Y-R, Zabetian C, Zhao Y, Ziegler S*. International multi-center analysis of glucocerebrosidase mutations in Parkinson disease. N Engl J Med. 2009 Oct 22;361(17):1651-61

XI. Simon-Sanchez S#, Schulte C#, **Bras JM#**, Sharma M#, Gibbs J, Berg D, Paisan-Ruiz C, Lichtner P, Scholz S, Hernandez D, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls M, Illig T, Gieger C, Houlden H, Steffens M, Okun M, Cookson M, Foote K, Fernandez H, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock S, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood N, Lorenz D, Deuschl G, Chen H, Riess O, Hardy J, Singleton A, Gasser T. Genome-Wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet. 2009 Dec;41(12):1308-12. Epub 2009 Nov 15*

# Equally contributing authors
* Authors listed in alphabetical order

Summary

Parkinson's disease (PD) is the second most frequent neurodegenerative disease, following Alzheimer's disease, and has been commonly designated a sporadic disorder with few environmental triggers. To date, the single most important risk factor for the disease is ageing. Since the proportion of the elderly is growing steadily as the longevity of the population increases, this leads to greater numbers of patients suffering age-associated neurodegenerative diseases, including PD. Parkinson's disease not only has a devastating effect on the individual patients and their families but it also imposes an enormous socioeconomic burden on society.

Nevertheless, about 10% of cases present early-onset of disease, commonly defined at below 40 years of age, and familial clustering, suggesting that genetic factors play a pivotal role in these cases.

This led, in the late 1990s, to the identification of mutations that presented clear segregation in families and were, thus, considered pathogenic. Several *loci* and genes have since been identified where mutations are disease causing, shedding light on the genetic background of mendelian forms of the disease.

The work presented herein follows 4 lines of research with the ultimate goal of clarifying the role of genetics in PD.

The first chapter was aimed at characterizing mendelian PD in populations where it was mostly unknown. To this end, we have conducted a screening of mendelian genes in a Portuguese cohort, representative of the population of the center region of Portugal. In addition to the Portuguese population, we have also screened a small cohort of samples from sub-Saharian Africa. This

was particularly important since it was the first study to address the genetics of PD in a population from this region, where genetic diversity is known to be far greater than in Caucasian Europeans.

An approach that has yielded promising results in the past is candidate gene association studies. Broadly, these studies look at common variability in genes that fit the pathogenesis of the disease, and determine if alleles are more frequent in cases when compared to controls. Chapter two deals with two association studies: the first aimed at determining if variants in the gene *HFE* were associated with PD; while the second was an attempt to replicate recent results implicating the gene *GIGYF2* in PD. While the first study was performed only in a Portuguese cohort, the second also looked at an extended cohort from North America.

The third chapter builds upon chapter two, still dealing with candidate gene association studies, however, here we looked at a particular gene, where recent results have been very promising. This chapter includes three separate studies: the first performs a standard association study of variants in *GBA* in the Portuguese cohort. The results from this study prompted us to perform a similar study in different and extended cohorts, which we have done in Study IX using a British cohort. Still in chapter three, we have performed a meta-analysis of association studies of *GBA* in PD that includes 16 international centers and nearly 11,000 samples.

Chapter four deals with a genome-wide association study in PD, where we have tested a very large number of PD cases and controls for markers spread throughout the genome. This approach has the benefit of not making *a priori*

assumptions of genes' relevance. This is particularly significant when studying a disease like PD where the etiology still remains largely elusive.

Resumo

A Doença de Parkinson (DP) é a segunda doença neurodegenerativa mais frequente, seguindo a Doença de Alzheimer, tendo sido frequentemente designada como uma doença esporádica com potenciais causas ambientais. Até à data, o factor de risco mais relevante para o desenvolvimento da DP é a idade. Uma vez que a proporção da população idosa tem vindo a aumentar, devido ao aumento da longevidade, o número de indivíduos afectados por doenças neurodegenerativas, incluindo DP é, também, cada vez maior. A DP tem, não só um efeito devastador para os indivíduos afectados e suas famílias, mas também um impacto socioeconómico enorme para a sociedade. Apesar da denominação comum de doença esporádica, cerca de 10% de indivíduos com DP apresentam um inicio precoce da doença – definido como início antes dos 40 anos de idade – e indícios de história familiar, sugerindo que factores genéticos desempenham um papel de relevo nestas formas da doença. Este facto esteve na origem, no final da década de 1990, da identificação de mutações que apresentam clara segregação com a doença em algumas famílias, tendo sido, por isso, consideradas mutações patogénicas. Vários genes e localizações cromossómicas foram desde então identificados, nos quais mutações levam invariavelmente ao início da doença, clarificando, em parte, os mecanismos genéticos que estão na base destas formas mendelianas da DP.

O trabalho que se segue foi delineado com base em 4 linhas de investigação, com o objectivo final de clarificar o papel dos factores genéticos na DP.

O primeiro capitulo deste trabalho teve como objectivo caracterizar formas mendelianas da DP em populações onde este campo não tinha, até a data,

sido extensivamente estudado. Com esse fim, levamos a cabo um *screening* de mutações nos genes mendelianos conhecidos num amostragem da população portuguesa, representativa da população da região Centro. Para além da população Portuguesa, também estudámos uma amostra de origem Subsariana. O estudo deste grupo de indivíduos de origem Africana é de particular importância, uma vez que foi o primeiro trabalho levado a cabo com o objectivo de elucidar a genética da DP numa população daquela região, onde a variabilidade genética é reconhecidamente maior do que na Europa Caucasiana.

Um tipo de trabalhos que tem originado resultados promissores até a data são os estudos de associação. Aqui, de uma forma genérica, é estudada a forma como a variabilidade em genes, conhecidos como estando envolvidos na doença, influencia o desenvolvimento da mesma, através da frequência dessas variações em casos e controlos. O segundo capitulo deste trabalho faz uso deste tipo de estudos com dois objectivos: o primeiro prende-se com o gene *HFE*, verificando se mutações neste gene são factores de risco para a DP; enquanto o segundo objectivo é o de tentar replicar resultados recentes que implicam o gene *GIGYF2* na patogénese da DP. Enquanto o primeiro estudo foi realizado apenas numa amostragem da população Portuguesa, o segundo fez uso de uma amostragem adicional de indivíduos Norte Americanos.

O terceiro capítulo complementa o capítulo dois, utilizando ainda estudos de associação. No entanto, aqui apenas nos debruçámos sobre um gene em particular (*GBA*), o qual tem originado resultados muito promissores. Este capítulo inclui três estudos: no primeiro estudo realizamos uma associação

entre variantes na *GBA* na amostragem de indivíduos de origem Portuguesa. Estes resultados levaram-nos a prosseguir o trabalho, desta feita num grupo significativamente maior, de amostras de origem britânica. Ainda no terceiro capitulo realizámos uma meta-análise de estudos de associação da *GBA* com DP que inclui dados de 16 centros internacionais e cerca de 11,000 amostras.

No quarto capitulo realizámos um estudo de associação do genoma completo com a DP. Neste trabalho testámos um numero significativo de marcadores, posicionados ao longo de todo o genoma, em casos e controlos. Este tipo de estudo tem o benefício de não presumir a priori quanto a potencial relevância de genes para a doença. Este facto é de particular importância para uma doença como a DP, onde a etiologia permanece ainda grandemente por explicar.

Contents

CHAPTER 1

INTRODUCTION

Based on the following studies:

I) Bras JM, Singleton A. **Genetic susceptibility in Parkinson's disease**. Biochim Biophys Acta. 2009 Jul;1792(7):597-603. Epub 2008 Nov 20.

II) Bras J, Singleton A, Cookson MR, Hardy J. **Emerging pathways in genetic Parkinson's disease: Potential role of ceramide metabolism in Lewy body disease**. FEBS J. 2008 Dec;275(23):5767-73.

Introduction

Parkinson's Disease

Parkinson's disease is a common progressive bradikynetic disorder, characterized by the presence of pars compacta nigral-cell loss, and accumulation of aggregated alpha-synuclein in brain stem, spinal cord and cortical regions [1]. Symptoms usually appear when a significant proportion of nigrostriatal dopaminergic neurons have been lost (~50-70%).
Although PD is, by and large, considered a sporadic disorder, few environmental triggers have been identified [2, 3]. As with other neurodegenerative diseases, ageing is the major risk factor, however incidence appears to decrease in the ninth decade of life [4]. A small proportion of cases (~10%) present an onset earlier than 45 years of age.
PD commonly presents with impairment of dexterity, however since the onset is gradual, the earlier symptoms may be unnoticed or misinterpreted for a long time. Diagnosis of PD remains largely a clinical one and is given by the clinician from the cardinal features of bradykinesia with at least one or more of the following: resting tremor, gait difficulties, postural instability, and/or rigidity. Responsiveness to dopamine replacement treatments is taken as supportive evidence for the diagnosis. These criteria are indicative of dysfunction in the substantia nigra and have been formalized into the London Brain Bank criteria for the diagnosis of PD [5].

Mendelian Forms of PD

Although PD was long considered a non-genetic disorder of sporadic origin, 5–10% of patients are now known to have monogenic forms of the disease. At least, 13 loci and 9 genes have been linked with both autosomal dominant and recessive forms of the disease (Table 1).

Table 1: Loci and genes known to be involved in PD

LOCUS	*MOI*	*ONSET*	*DESIGNATION*	*CHR.*	*GENE*
PARK-1	AD	~45	-	4q	SNCA
PARK-2	AR	7-60	ARJP	6p	PRKN
PARK-3	AD	59	-	2p13	-
PARK-4	AD	30s	-	4q	SNCA
PARK-5?	AD	30-60	-	4p14	UCH-L1?
PARK-6	AR	36-60	-	1p36	PINK-1
PARK-7	AR	27-40	-	1p36	DJ-1
PARK-8	AD	45-57	-	12p-q	LRRK2
PARK-9	AR	Teens	Kufor-Rakeb	1p36	ATP13A2
PARK-10	-	Late	Icelandic	1p36	-
PARK-11	AD	58±12	-	2q36-37	-
PARK-12	-	Late	-	Xq21-25	-
PARK-13	-	Late	-	2p12	-
PARK-14	AR	Teens	-	22q13.1	PLA2G6
PARK-15	AR	Teens	-	22q12-13	FBXO7

Dominant mutations can exert their effect in several ways: they can act as gain-of-function, where toxicity is achieved by the amplification of the normal function of the protein or the gaining of a new toxic function; they can cause simple loss-of-function, usually associated with nonsense mutations; or they can act through a dominant-negative mechanism, whereby the mutant allele interferes with the function of another wild-type allele and leads to loss-of-function.

Mutations in alpha-synuclein were the first genetic cause of PD to be identified. A point mutation in the alpha-synuclein gene was initially discovered in a large Greek/Italian kindred with autosomal dominant Parkinson's disease with a mean onset age in the 50s [6]. Subsequently, two other point mutations have been described, and alpha-synuclein was shown to be the major component of Lewy bodies [7]. Additionally, several families have been described with gene duplications and triplications [8, 9]. Interestingly, the families with the gene triplications get affected in their thirties and those with gene duplications, in their fifties. These data show that an increase in alpha-synuclein dose of 50% leads inevitably to disease at age 50, in a clear dose specific manner. Given that mutations in *SNCA* cause dominant disease that can be due to gene duplications shows that the mechanism of alpha-synuclein's toxicity likely relates to an exaggeration of its normal propensity to aggregate.

LRRK2 mutations were initially found in families originally from the Basque Country and England [10]. In these and following studies, a very large number of mutations was found, with varying degrees of confirmed pathogenicity. Some of these mutations are very common: R1441G causes a large number

of Basque cases [11], G2019S underlies a considerable percentage of cases in Europeans [12], and G2385R and R1628P both explain a large proportion of disease among eastern Asian people [13, 14]. The mechanism by which these mutations cause disease is not fully understood. However, it has been shown that the common Caucasian mutation, G2019S, located in the kinase domain of the protein, leads to disease by a gain of function effect [15].

The vast majority of the recessive alleles causing PD, act in a simple loss-of-function manner. A significant proportion of these are predicted to lead to a non-functional protein, because they cause either deletions or frameshifts. However, some of the mutations are missense, and attributing pathogenicity in these cases is slightly more complicated, particularly if complete segregation within the family is not clear. Another confounding factor in these cases is the genetic background of the individuals – not identifying a variant in a cohort of healthy controls does not necessarily mean that it causes disease, it may be that it is simply a rare variant, and thus other populations should be analyzed. In addition, recessive alleles may also be pathogenic when heterozygous if they are in *trans* with yet another heterozygous pathogenic variant. One must bear this in mind when finding heterozygous variants in these recessive *loci* and the complete gene has not been thoroughly screened.

Parkin mutations were initially discovered in Japanese families with a juvenile form of PD [16]. Parkin is an E3 ligase whose functions in the cell may include preparing defined proteins for proteasomal degradation. A large number of mutations has since been identified in this gene, ranging from point mutations

to large copy number changes (deletion/duplications). Onset in these cases is generally below 40 years [17].

DJ-1 mutations were first identified in a Dutch population isolate [18]. The phenotype may be similar to the one caused by mutations in *PRKN*, however, mutations are exceedingly rare. The protein structure has been identified and it is known that some mutations, particularly the L166P, prevents dimerization and leads to protein degradation [19]. Nevertheless, the precise function of the protein is not known.

In a similar manner to *DJ-1*, mutations in *PINK1* are also quite rare. The phenotype, however, closely resembles that in *PRKN* cases. Mutations in *PINK1* were initially identified in Spanish and Italian families [20]. Also similarly to *PRKN*, compound heterozygous and single mutations have been reported [21, 22]. It is known that *PINK1* is a mitochondrial kinase, however, neither its direct activators or repressors, nor downstream targets have been identified.

Mutations in *ATP13A2* were first described in a family with very early-onset parkinsonism from Jordania [23]. It is known that ATP13A2 is a lysosomal pump, but its substrates are yet to be identified. The phenotype also resembles those from lysosomal storage disorders, thus providing a link between parkinsonism and these disorders, similarly to *GBA*.

FBXO7 was the most recent to be identified, in which recessive mutations lead to parkinsonism [24, 25]. Like *PRKN*, FBXO7 is part of an E3 ubiquitin ligase, but its precise function is yet to be identified [26].

Genetic Susceptibility in PD

In parallel to work on monogenic PD a large amount of research has focused on identifying genetic variability that confers risk for, rather than causes, PD. This work aims not only to add insight into the molecular pathogenesis of PD, but also to create a risk profile for disease in the general population. For the most part risk variant identification is based on the tenet of the common-disease common-variant hypothesis. This theory operates on the premise that common genetic variants underlie susceptibility for common diseases such as PD. The common-disease common-variant hypothesis is an idea that is the basis for the vast majority of genetic case control association studies and the impetus for initiatives such as the International Human Haplotype Map Project (www.hapmap.org). There has been significant contention over the common disease common variant hypothesis with substantive support for the idea that rare mutations underlie the etiology of complex disorders. While the common disease common variant and rare variant hypotheses are often proposed as opposing theories they are not mutually exclusive; however, the current accessible technologies do not readily allow the investigation of rare mutations as a cause of common disease in a complete way.

In general, within the field of disease genetics, the search for common genetic risk variants has been difficult and has had a low hit rate. There are of course several reasons for this failure, particularly in a disease such as PD. Most prominent is the size of effect and thus the number of samples required to

detect an effect. Many studies were predicated on the idea that risk variants with effect sizes comparable to APO E ε4 in Alzheimer's disease may exist for PD; thus these generally only included sample sizes in the low hundreds. Early genome wide association data shows, quite convincingly, that in terms of common genetic risk factors, there are no risk loci with an odds ratio greater than two in the North American White population. It is likely that sample sizes of more than a thousand are required to detect effects with such a low odds ratio. The second limitation of such studies is that because they are largely low-throughput in nature, typing usually only one gene and one or a few variants, the prior odds of selecting the correct gene and the correct variant to type were very small (i.e. if there are 10 genes that exert a measurable risk in PD and 20 SNPs, the odds are against the probability of choosing the correct gene, out of 25,000 in the genome or the correct SNPs, out of 2 million in the genome).

Thus the majority of previous studies were not only unlikely to have selected a genuine risk locus or variant for interrogation but further were likely not powered to detect risk effects, should they exist. The exception to this work is that which has centered on exhaustive analysis of genes already implicated in familial forms of related disorders, emphasizing that candidate analysis in the absence of prior genetic evidence implicating the locus in disease, is likely to fail. Such analyses provide two of the most convincing sets of genetic association data for PD, implicating the genes *SNCA* and *MAPT* as risk loci.

SNCA (encoding alpha-synuclein)

Genetic variability at *SNCA* is arguably the most reliable association of a common genetic risk locus with PD identified to date. The impetus for examination of this locus resulted from the cloning of *SNCA* mutations as a cause of a rare familial form of PD. Closely after the initial identification of the first disease cause mutation in *SNCA,* the protein product of this gene, alpha-synuclein, was shown to be a major constitutive part of the pathologic hallmark inclusion of PD, the Lewy body. The relevance of studying rare familial forms of PD to understanding the common non-familial form of this disease had remained in question, however these two findings elegantly linked these disease entities; this work simply shows that the mutant form of alpha-synuclein causes a familial form of PD and that the alpha-synuclein protein is a component of the pathology of all cases of PD. Shortly after this work Krüger and colleagues [27] reported an association between common genetic variability in *SNCA* and risk for PD, specifically an association with an imperfect dinucleotide repeat approximately 10kb 5' to the translation start site of *SNCA*. Variability at and proximal to REP has been examined in a large number of studies [2-9] and most recently a meta-analysis of published studies and combination with novel data revealed a consistent association between risk for disease and the longer REP allele [10]. Experimental *in vitro* evidence suggests that this allele is associated with increased expression of *SNCA*. From an etiologic perspective this observation fits well with the discovery that multiplication of the *SNCA* locus, which leads to increased levels of the wild type alpha-synuclein, causes familial PD [11, 12]. Further the multiplication mutations, which are to date both triplication and duplication

events, appear to have a dose dependent effect on disease severity; thus triplications that double the genomic copy number of *SNCA* result in disease with an onset in the 4th decade of life and duplications, which increase *SNCA* load by 50%, result in disease onset in the 5th or 6th decade of life. Given this finding it is quite reasonable to suppose that common genetic variability at *SNCA*, which may increase expression by a small amount, is a risk factor for the late onset sporadic form of the disease. In addition to the work characterizing the role of the REP alleles in risk for PD, more complete analyses have shown an association between genetic variability at other parts of *SNCA* with risk for disease; in particular variability in the 3' half of the gene [13-15]. Although these works do not resolve the issue of which are the pathobiologically relevant variants, they do suggest that investigation of REP alleles alone may miss, or underestimate, the contribution of genetic variability at *SNCA* to sporadic PD.

MAPT (encoding Microtubule associated protein tau)

There are six major brain isoforms of the microtubule associated protein tau (hereafter called tau) generated by alternate splicing of exons 2, 3 and 10 of the gene *MAPT*. Alternate splicing of exon 10 results in tau with 3 or 4 microtubule binding respeats (3 repeat tau or 4 repeat tau). Tau is a major protein component of neurofibrillary tangles, a hallmark lesion of Alzheimer's disease; in these lesions tau appears to be deposited as hyperphosphorylated insoluble filaments. Tau deposition is a hallmark of several other neurodegenerative disorders, including Picks disease (OMIM #257220), argyrophilic grain disease (OMIM #172700), corticobasal degeneration,

progressive supranuclear palsy (PSP; OMIM #601104) and frontotemporal dementia (FTD; OMIM #600274).

An initial link between the *MAPT* locus and disease was first reported in 1997 by Conrad and colleagues [28] who showed an association between a repeat polymorphism close to the gene and PSP, indicating that variability at this locus is a risk factor for this disorder. In 1998 definitive evidence linking *MAPT* to neurodegenerative disease was provided when Hutton and colleagues identified mutations in this gene as a cause of frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP17) [29]. To date more than 35 mutations have been identified at the *MAPT* locus associated with this disease (for a good review see [30]). Many of these mutations are predicted to alter the alternative splicing of *MAPT*, altering the ratio of 3 repeat tau to 4 repeat tau [29, 31].

In addition to rare causal mutations, common variability in *MAPT* has been linked to disease; notably robust association between *MAPT* and risk for PSP, AD and most recently PD has been reported. From a genomic perspective the architecture of the *MAPT* locus is unusual; the gene sits within a large block approximately 1.6 million base pairs long that shows reduced recombination and thus high levels of linkage disequilibrium. This appears to be a result of a common genomic inversion in the Caucasian population; this inversion inhibits recombination between genomic fragments that are in the opposite orientation. This phenomenon results in two common Caucasian haplotype groups across this locus; often termed H1 and H2. Association between

MAPT H1 sub-haplotypes and risk for PD has been tested by many groups [32-36], and the results in general show a consistent association with disease, the H1 haplotype conferring a risk with an odds ratio of approximately 1.3 (for summary statistics see the PD Gene database http://www.pdgene.org/meta.asp?geneID=14). Evidence is also mounting that the *MAPT* risk alleles for these disorders are associated with increased *MAPT* expression; either in total or specific to four-repeat tau splice variants (i.e. those containing exon 10) [37, 38]. Most recently Tobin and colleagues [39] have shown association between PD risk and a sub-haplotype of H1; these authors then extended this work to show over-expression of 4 repeat tau in the brains of PD patients.

From a pathological standpoint the relationship between tau and PD remains enigmatic: in general the brains of PD patients do not show abundant tau positive neuropathology; however the data supporting genetic association between *MAPT* and risk for PD continues to grow and is certainly one of the more robust findings in the field of risk variant in neurogenetics.

GBA; Glucocerebrosidase

Glucocerebrosidase is a lysosomal enzyme that hydrolyses the beta-glycosidic linkage of glucosylceramide, a ubiquitous sphingolipid present in the plasma membrane of mammalian cells, originating ceramide and glucose. The human *GBA* gene is located on chromosome 1q21 and comprises 11 exons and 10 introns spanning over 7kb.

A 5.5kb pseudogene, which shares over 96% homology with GBA, is located just 16kb downstream of the functional gene. The difference in size between the two is due to several *Alu* insertions in intronic regions of GBA. On the other hand, the lack of functionality of the pseudogene is attributed to two exonic deletions: a 4bp deletion in exon 4 and a 55bp deletion in exon 9 [40]. The pseudogene is absent in non-primate species and it has been suggested that the duplication event that originated the two genes, occurred about 40 million years ago. Interestingly, it has been shown that the orangutan does not present a pseudogene, but instead two functional genes, hence potentially four copies of GBA [41].

Mutations in GBA are the cause of a recessive lysosomal storage disorder – Gaucher disease. Patients with Gaucher present macrophages enlarged with deposits of glucosylceramide, suggesting that mutations in GBA act in a loss-of-function fashion [42]. Over 200 mutations have been described in GBA, including point mutations, deletions and recombination alleles derived from the pseudogene sequence. It has been estimated that approximately 20% of the pathogenic mutations in GBA are caused by recombination or gene conversion between the two genes. Although mutations are distributed over the entire GBA coding region, pathogenic mutations seem to cluster in the carboxyl-terminal region, which encodes the catalytic domain [43].

Phenotypes of Gaucher and Parkinson's diseases do not overlap significantly, but the first indication for a relationship between the two, actually came from

clinical descriptions. These reported patients with Gaucher disease who developed early-onset, treatment-refractory parkinsonism [44].

The first report of an increased frequency of mutations in GBA in Parkinson disease patients was published online in 2003 [45]. Here, the authors screened 57 brain samples from subjects with PD and 44 brain samples from adult subjects without a diagnosis of PD. Mutations in GBA were identified in 14% of the PD samples, and no mutations were found in the control samples. The percentage found in PD patients was of particular relevance, given that the carrier frequency for Gaucher disease-causing alleles is estimated at 0.006.

In 2004 Aharon-Peretz J et al. [46] reported a screening of 99 Ashkenazi PD patients, 74 Ashkenazi Alzheimer's disease patients and 1543 healthy Ashkenazi Jews for six GBA mutations, considered to be the most common cause of Gaucher among Ashkenazi Jews. A surprising percentage of 31% of PD patients had one or two mutant alleles, when compared to only 6% of controls with mutations in GBA. Also, among the PD patients, those who were carriers of GBA mutations had significantly earlier age-of-onset than those who presented no mutations.

The following year, Clark LN et al. [47] presented a report on 160 Ashkenazi Jewish probands with Parkinson's disease and 92 clinically evaluated, age-matched controls of Jewish ancestry. Subjects were screened only for the N370S mutation, which was the most frequent variant in the previous Ashkenazi Jewish study. Seventeen probands (10.7%) were identified when

compared to 4.3% controls; however these results did not reach statistical significance. While this variant was described as the most frequent among this population, one cannot help but think that a complete gene screening may have yielded positive results.

Sato C. et al. performed a screening for seven of the most common variants in GBA in a series of 88 unrelated Caucasian subjects of Canadian origin, selected for early age of onset and/or positive family history; additionally a group of 122 healthy controls was also screened. Mutations were enriched in the PD group when compared with the controls (5.6% vs 0.8%); these results just about reached statistical significance (p=0.048) [48].

In a smaller series of cases and controls collected in Venezuela (33 PD samples, 31 controls), Eblan M. et al. screened the entire coding region of the GBA gene and described an increase in mutation frequency among the early-onset PD samples when compared to the controls (12% vs 3.2%) [49].

Toft M. et al published a report on the screening of two variants in GBA in individuals from Norway [50]. This was the first report on northern European subjects. The authors screened 311 PD patients and 474 healthy controls for the two common mutations N370S and L444P. They did not find an increased frequency of mutations in PD samples when compared to controls, however, the frequency of the mutations was surprisingly high when compared to estimates in white individuals (1.7% vs 0.6%). Though no statistical significance was obtained, the fact that only two variants were screened in a population not previously studied, may account for the lack of association.

Several other studies reported positive associations of GBA mutations with PD, and one with Lewy bodies disorders. The most recent studies, which performed complete screenings of the gene, all found an association with PD in their study populations. Clark, L. et al. performed a study with two subsets of PD patients: one with Jewish ancestry and another without Jewish background. Controls were also selected to match each of these groups. The frequency of GBA mutations is always greater in PD samples when compared to controls, particularly if only early-onset PD samples are considered [51].

Our group recently published a report on a cohort of Portuguese samples, where the enrichment of mutations among PD samples is also clear. This is the first report to obtain a clear statistically significant association in a population other than Ashkenazi Jewish. However, this same population has been shown to present an increased frequency of the G2019S mutation in the gene LRRK2; which is known to be a Jewish mutation [52].

An interesting result regarding GBA, and lysosomal enzymes in general, comes from the report of Balducci C. et al. who tested PD patients and controls for activity of lysosomal hydrolases in the CSF. GBA activity was significantly reduced in the CSF of PD patients, as would be expected by a loss-of-function model of mutations in these samples [53].

Given all these results it seems clear that mutations in GBA are a risk factor for the development of PD, particularly early-onset PD. The mechanism by which mutations exert their effect and act as a risk factor is not yet

understood, but several tentative explanations have been provided in the literature, relating to decrease in lysosomal function or involvement of the ubiquitin proteasome system.

Lrrk2

Mutations in *LRRK2* were identified as a cause of PD in 2004 [10, 54]. A single mutation, G2019S, is a relatively common cause of PD in Caucasian populations, underlying approximately 2% of sporadic PD cases in North America and Northern Europe and 5% of cases with a positive family history for disease [55-57]. This mutation is more common in populations such as those from Portugal, those of Ashkenazi Jewish origin and from North African Arab populations; underlying 8%, 21% and 41% of disease in these populations respectively [58-60]. The G2019S variant does not however occur at appreciable frequency in control cohorts from these populations, so it cannot be designated this as a susceptibility variant in these populations. Two variants reported from Asian populations, however, do appear to be true risk variants for PD. The first G2385R was initially described in a Taiwanese family [61]. Assessment of this variant in large Asian populations showed association with risk for disease in Taiwanese [13, 62, 63], Japanese [64], Hong Kong Chinese [65] and mainland Chinese [66, 67] populations. In general this work showed that the risk variant, 2385R is present in PD populations at a frequency of ~10%, whereas it is only found in 0.5%-5% of controls. Taking a fairly conservative view of these results they would suggest that carrying the risk allele imparts a two-fold risk increase in the chance of Parkinson's disease. Given that this association appears robust across Asian

populations, this risk allele is an underlying factor in a very large number of PD cases worldwide. More recently a second *LRRK2* risk allele, also identified within Asian PD populations, has been described [68, 69].

OMI/HTRA2

Variants in the gene *OMI/HTRA2* (OMIM #606441) were recently associated with an increased risk for PD [70]. This gene encodes a serine-protease with proapoptotic activity containing a mitochondrial targeting sequence at its N-terminal region. Several lines of evidence in the literature support a role for *OMI/HTRA2* in neurodegeneration, the first of which was produced by Gray and colleagues [71] when they showed that Omi/HtrA2 interacts with presenilin-1, which is encoded by a gene known to be involved in Alzheimer's disease. Moreover beta-amyloid, which plays a pivotal role in the pathogenesis of Alzheimer's, was shown to be cleaved by Omi/HtrA2 [72]. Mouse models also provided support for the involvement of this protein in neurodegeneration; a mutation in the protease domain of Omi/HtrA2 was found as the genetic cause underlying the disease in the mnd2 mutant mouse [73] and the knockout mouse showed loss of neurons in the striatum concomitant with parkinsonian features [74].

The description of mutations associated with PD came from the work of Strauss and colleagues in 2005 [70]. Here, they screened a large cohort of 518 German PD patients and 370 healthy control individuals for mutations in *OMI/HTRA2*. One variant (p.G399S) was found only in PD patients (n=4) suggesting that it would be a pathogenic mutation. The other variant (p.A141S) was found significantly overrepresented in the PD group (p=0.039)

suggesting that it would act as a risk factor for PD. In *vitro* studies of both variants provided evidence of a functional role. Additionally, Omi/HtrA2 was detected in Lewy bodies in brain tissue from PD patients.

More recently, Simon-Sanchez and Singleton [75] presented a thorough analysis of the coding region of *OMI/HTRA2* in a case-control study, which comprised a large cohort of PD patients (n=644) and neurologically normal controls (n=828). The mutation initially thought to be pathogenic was found at the same frequency in PD samples and controls (0.77% and 0.72% respectively), indicating that it is not disease causing, but probably a rare variant in the German population. Similarly for the p.A141S variant, no association with PD was found.

Evidences for the involvement of Omi/HtrA2 in neurodegenerative diseases are quite compelling at this point, but the genetic basis for this involvement is still very much debatable.

Heterozygous Mutations in Genes that are Recessive loci for Monogenic PD as Risk Factors for Sporadic PD

Mutation of the gene *PARK2*, was the second genetic cause of parkinsonism identified. Mutation of this gene was found to cause an autosomal recessive juvenile form of parkinsonism. As with many autosomal recessive diseases, most proven pathogenic mutations are loss of function variants, involving large structural genomic disruption of the coding region of the gene or premature termination of the transcript; in addition several missense mutations have been identified with varying levels of proof *vis a vis* association with disease. More controversial still, is the role of single

heterozygous mutations as risk factors for late onset typical PD. Much of this work has been driven by observations in family based studies where later onset typical PD was seen in relatives of patients with young onset parkinsonism – analysis of these cases revealed homozygous or compound heterozygous *PARK2* mutations in the young-onset cases and possession of single *PARK2* mutations in the later onset affected relatives. These studies have been criticized because of the significant confound of ascertainment bias, i.e. the families that tend to be collected and analyzed are those with many affected family members. Further support for the role of single *PARK2* mutations as a risk factor for disease is the observation that heterozygous mutation carriers display dopamine reuptake deficiency in f-dopa PET analysis [76]. The strength of the case for a role of heterozygous *PARK2* mutations as risk loci based on these observations, and on data arising from case control analyses examining this issue, has been hindered by the lack of studies that have performed full sequence and gene dosage analysis of *PARK2* in large groups of cases and controls. Only a small number of studies have taken this approach and so far that data does not strongly support a large role for these mutations in typical PD, identifying pathogenic mutations in the heterozygous state in both cases and controls. This issue warrants a large scale sequencing effort, however, it is likely that several thousand cases and controls will need to be fully sequenced to finally prove or disprove this putative association.

Genome Wide Association Studies

Genome wide association studies (GWAS) were a much-anticipated technology, and the application of this approach is expected to make major inroads into our understanding of the genetic basis of disease [77]. The basic tenet underlying GWAS is the common disease common variant hypothesis. A growing number of GWAS are being published and these are proving a valuable approach in understanding the genetics of complex disease. Two studies have been published thus far in PD; the success of these studies was limited somewhat by the sample size used, a point that is illustrated by their failure to identify *SNCA* or *MAPT* as potential risk loci. However, given the increasing investment in this technology [78] it is probable that several laboratories around the world are investing in large-scale GWAS in PD and that in the next 2 years we will see the identification of novel risk loci for this disease.

Because GWAS require large sample series they necessitate inter-laboratory collaborations and large consortia, individual investigators are by and large unable to accumulate large enough series. Clearly the formation of such collaborations is a good thing for research; they facilitate communication between scientists, maximize the chances of finding positive signals and engender trust between research groups that allows collaboration outside the immediate aims of the collaborative framework. This latter point is particularly critical; as technologies improve, GWAS will be expanded to include better genomic coverage (currently even the most dense SNP platforms probably only capture ~70% of common variation) in more samples.

The most immediate challenge following GWAS will be in understanding the pathobiological consequences of identified risk variants. These will not easily

be amenable to traditional disease model approaches used now in cell biological and transgenic research, the primary limitation being that the biological effects of risk variants are likely extremely subtle. Parsimony would suggest that the majority of identified risk variants will be non-coding, in all likelihood exerting an effect through expression, either modulating constitutive levels, expression level in response to a stimulus, sub-cellular expression and/or splicing. The easiest of these to catalog are alterations in expression and indeed some effort has gone toward characterizing in a genome wide manner, the effects of individual genetic variants on expression of proximal (*cis*) and distal (*trans*) transcripts [79]. The creation of standard genotype-expression transcript maps will be a critical step in understanding the effects of disease associated genetic variants, and there have already been moves to create such a resource http://nihroadmap.nih.gov/GTEx/.

While GWAS are providing a unique set of insights into complex diseases, it is only the first of many burgeoning technologies that will impact our understanding of biology and disease. Most anticipated of these is cost-effective genome wide resequencing; the launch of this type of work is an implicit goal of the 1000 genomes project (http://www.1000genomes.org/page.php), which explicitly aims to catalog rare and common human variation by sequencing the genome of at least 1000 individuals from around the World. Currently this is a huge endeavor and genome resequencing for case-control analysis is cost- and time-prohibitive; however, this is likely to change over the next 2-5 years. The type of next generation sequencing employed for this research will not only facilitate genome resequencing but also allows us to analyze other features previously

impractical; these includes genome wide assays of DNA methylation, histone modifcation, analysis and identification of transcription binding sites, full transcriptome sequencing and identification of allelic imbalance in expression. Each of these approaches provides revolutionizing data in their own right; however, the true power of such data will become evident as we integrate these datasets to garner a systems based understanding of biological and disease processes.

In summary, there are several common risk loci unequivocally associated with risk for PD; in each instance these genes were originally implicated in the disease process by studying families with disease. The advent and application of novel technologies promises to define other common genetic variants that exert risk for disease, help in the identification of rare risk variants and facilitate in the understanding of the pathobiological consequences of genetic variants linked with disease.

Emerging Pathways in Genetic PD: A role for Ceramide?

Genetic research in the past decade has changed the view of PD from an archetypical non-genetic disease to one having a clear genetic basis in a percentage of patients [2].

Classically, the approach taken to the study of genetic forms of PD has relied on a clinical definition of disease and PARK loci have been assigned on this clinical basis. It is known what clinical features are primarily associated with each locus and a great deal of attention has been focused on this association [80]. However, if one wants to identify pathways of pathogenicity for a given disorder, arguably, one should start by analyzing the genetics of disease based on pathology. In this review, we start from the position that it is more likely to find a common pathway if there is a common pathology rather than common clinical characteristics. We and others have suggested that, for the early onset recessive diseases (encoded at the parkin, PINK1 and DJ-1 loci), in which Lewy bodies are either usually absent (parkin) or where no neuropathological data is available (PINK1 and DJ-1), the evidence for a mitochondrial pathway to cell death is overwhelming [2].

The inspiration for our attempt to re-evaluate a Lewy body pathway to cell death has come from the recent observation that mutations in glucosecerebrosidase (GBA) which when homozygous, lead to Gaucher's disease but when heterozygous, predispose to PD [45]. GBA catalyses the breakdown of glucosecerebroside to ceramide and glucose. Gaucher's disease is caused by a lysosomal build up of glucosecerebroside, but this occurs only when GBA activity is almost completely lost. In the heterozygous

state this is unlikely to be a problem. We therefore began to consider that ceramide metabolism, more generally, may be an initiating problem in PD. The genes associated with Lewy bodies that will be dealt with within this review are presented in Table 2.

Table 2: Genes associated with Lewy body inclusions and their potential role in ceramide metabolism

	Gene	*Chr*	*Function*	*Disease*
Ceramide metabolism and Lewy body inclusions	GBA	1q21	Lysosomal Hydrolase	Gaucher disease/Parkinson's disease in heterozygotes
	PANK2	20p13-p12.3	Pantothenate kinase	Neurodegeneration with brain iron accumulation type 1 (NBIA-1)
	PLA2G6	22q13.1	A2 phospholipase	Neurodegeneration with brain iron accumulation 2 (NBIA2)
Probably Ceramide metabolism ; Possibly Lewy body inclusions	NPC1	18q11-q12	Regulation of intracellular cholesterol trafficking	Niemann-Pick disease type C1
	SPTLC1	9q22.1-22.3	Transferase activity	Hereditary sensory neuropathy type I (HSN1)
	ATP13A2	1p36	ATPase	Kufor-Rakeb Syndrome
Possibly Ceramide metabolism; Definite Lewy body inclusions	SNCA	4q21	Dopamine transmission and synaptic vesicle dynamics	Parkinson's disease
Unknown Ceramide; Usually Lewy body inclusions	LRRK2	12q12	Protein Kinase	Parkinson's disease

These are divided in three categories: the first are genes clearly involved in ceramide metabolism and that cause diseases where Lewy bodies are known to be abundant; the second category groups genes that may be involved in

ceramide metabolism and cause diseases where Lewy bodies have been described; the third category presents genes for which, while they do give rise to Lewy body disease, there is currently little or no evidence suggesting a role in ceramide metabolism. Levels of cellular ceramide are regulated by the *de novo* pathway and the recycling pathway. The former relates to the synthesis of ceramide through the condensation of palmitate and serine in a series of reactions that are ultimately dependent on Co-Enzyme A. The latter pathway is slightly more intricate, since several outcomes are possible depending on the enzymes involved. The simplified metabolism is shown in Figure 1.

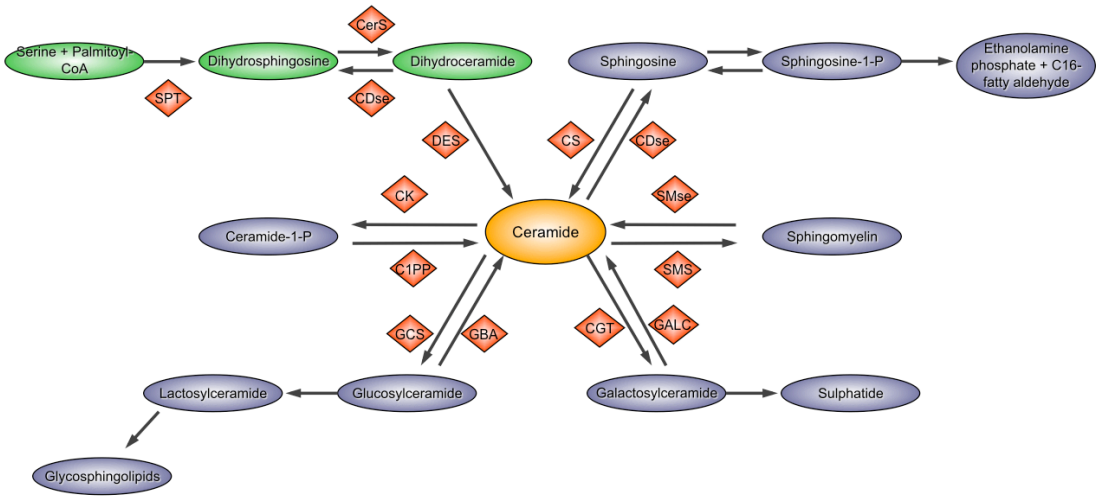


Figure 1: Simplified representation of ceramide metabolism. SPT, serine palmitoyl transferase; CerS, ceramide synthase; CDse, Ceramidase; DES, Desaturase; CS, Ceramide synthase; CK, Ceramide Kinase; SMse, sphingomyelinase; C1PP, Phosphatase; GCS, glucosylceramide synthase; GBA, glucosylceramidase; CGT, UDP glycosyltransferase; GALC, Galactosylceramidase; SMS, sphingomyelin synthase. Red represents enzymes directly involved in ceramide metabolism, in which mutations are associated with Lewy body inclusions.

The gene GBA encodes a lysosomal enzyme, glucocerebrosidase, that catalyses the breakdown of the glycolipid glucosylceramide to ceramide and glucose [81]. Over 200 mutations have been described in GBA, most of which are known to cause Gaucher disease, in the homozygous or compound heterozygous condition (for a review see [43]). Gaucher patients typically

present enlarged macrophages resulting from the intracellular accumulation of glucosylceramide. The fact that these patients show increased levels of the enzyme's substrate indicates that pathogenic variants act as loss-of-function mutations. GBA mutations, in addition to causing Gaucher disease when homozygous, have recently been established to act as a risk factor for PD [13, 14] and for Lewy body disorders [82].

Neurodegeneration with brain iron accumulation-1 (NBIA-1), formerly known as Hallervorden-Spatz disease is a form of neurodegeneration caused by mutations in the pantothenate kinase gene, PANK2. Clinically the condition is characterized by progressive rigidity, first in the lower and later in the upper extremities. Both involuntary movements and rigidity may involve muscles supplied by cranial nerves, resulting in difficulties in articulation and swallowing. Mental deterioration and epilepsy occur in some. Onset is in the first or second decade and death usually occurs before the age of 30 years [83]. Neuropathological studies have shown that patients with NBIA-1 present extensive Lewy bodies [84-86]. Pantothenate kinase is an essential regulatory enzyme in CoA biosynthesis, catalyzing the cytosolic phosphorylation of pantothenate (vitamin B5), N-pantothenoylcysteine, and pantetheine [87]. PANK2 is also involved in ceramide metabolism as the de novo pathway for ceramide formation relies on the presence of CoA [88]. Hence, there is a direct, though not specific, connection to ceramide metabolism.

Neurodegeneration with brain iron accumulation-2 (NBIA-2) is characterized by the disruption of cellular mechanisms leading to the accumulation of iron in the basal ganglia. Mutations in the gene PLA2G6 were recently described as

the cause of NBIA-2 [89]. Phenotypically similar to NBIA-1, Lewy bodies were also described in patients with NBIA-2, particularly in the brainstem nuclei and cerebral cortex [90]. PLA2G6 belongs to the family of A2 phospholipases, which catalyze the release of fatty acids from phospholipids and play a role in a wide range of physiologic functions [91]. Interestingly, it has been recently demonstrated that PLA2G6 plays a role in the ceramide pathway; activation of this enzyme promotes ceramide generation via neutral sphingomyelinase-catalyzed hydrolysis of sphingomyelins [92]. Similarly to what happens with GBA or PANK2, mutations in PLA2G6 that diminish its activity are expected to reduce the levels of ceramide formed through the breakdown of sphingomyelin.

Niemann-Pick type C (NPC) disease is an autosomal recessive lipid storage disorder characterized by progressive neurodegeneration with a highly variable clinical phenotype. Patients with the 'classic' childhood onset type C usually appear normal for 1 or 2 years with symptoms appearing between 2 and 4 years. They gradually develop neurologic abnormalities, which are initially manifested by ataxia, grand mal seizures, and loss of previously learned speech. Spasticity is striking and seizures are common [93]. Approximately 95% of cases are caused by mutations in the NPC1 gene, referred to as type C1. This gene encodes a putative integral membrane protein containing motifs consistent with a role in intracellular transport of cholesterol to post-lysosomal destinations. Cells from NPC subjects show a decrease in acid sphingomyelinase activity, leading to the accumulation of sphingomyelin [94]. Since one of the pathways for ceramide recycling is the sphingomyelin pathway, it is conceivable that associated to the accumulation

of sphingomyelin, a decrease of ceramide may also be present. Some cases of NPC1 were described as presenting Lewy bodies [95].

Mutations in SPTLC1 are the cause of hereditary sensory neuropathy type I (HSAN I) [96], a dominantly inherited sensorimotor axonal neuropathy with onset in the first or second decades of life. SPTLC1 is a key enzyme in sphingolipids biosynthesis, catalyzing the pyridoxal-5-prime-phosphate-dependent condensation of L-serine and palmitoyl-CoA to 3-oxosphinganine [97]. Patients usually present neuropathic arthropathy, recurrent ulceration of the lower extremities, signs of radicular sensory deficiency in both the upper and the lower extremities without any motor dysfunction [98]; restless legs and lancinating pain are other presentations of the disorder, which often results in severe distal sensory loss, and mutilating acropathy [99]. Although mutations in SPTLC1 cause neurological disease, there is, as yet, no description of the pathology of the disorder. We would hypothesize that this disease will have Lewy body pathology.

Kufor-Rakeb syndrome (KRS) is a form of autosomal recessive hereditary parkinsonism with dementia. It was recently described that loss-of-function mutations in the predominantly neuronal P-type ATPase gene ATP13A2 are the cause of Kufor-Rakeb syndrome [23]. The clinical features of KRS are similar to those of idiopathic Parkinson disease and pallidopyramidal syndrome, including mask-like face, rigidity, and bradykinesia [100]. Although ATP13A2 does not play an obvious role in the ceramide pathway it is a lysosomal transport protein thought to be responsible for the maintenance of the ideal pH in the lysosome. This function, albeit potentially implying a much broader effect of mutations, might also mean that ATP13A2 may be related to

the recycling pathways of ceramide metabolism. Interestingly, it has been suggested that alpha-synuclein turnover may occur via chaperone-mediated autophagy (CMA), a specialized form of lysosomal turnover [101-105]. It has also been shown that Alpha-synuclein turnover is slowed in mouse models of lysosomal storage disorders [106].

Alpha-synuclein (SNCA) is the major component of Lewy bodies and mutations in this gene are a rare cause of PD. Only three point mutations have been described to date, but duplication and triplication of the entire SNCA locus has also been discovered [6, 8, 107-110]. PD cases with underlying SNCA mutations have extensive Lewy bodies, since these mutations are known to increase aggregation of the protein [111]. SNCA may also be involved, albeit in a more indirect manner, in the ceramide pathway. It has been shown that deletion of the gene decreases brain palmitate uptake [112] and that the presence of palmitic acid increases the de novo synthesis of ceramide significantly [113]. However, known pathogenic mutations in SNCA are likely gain-of-function mutations, suggesting that, in these cases, the mutations drive the aggregation of alpha-synuclein, while in cases where ceramide metabolism is affected, Lewy Body inclusions may be a cellular response to this altered ceramide metabolism. Also connecting the ceramide pathway to alpha-synuclein deposition is the recent description of an increase in alpha-synuclein inclusions in C. elegans when LASS2, a ceramide synthase, is knocked-down [114]. This result should obviously be taken with some caution, since it was obtained in a non-mammalian organism, but nevertheless it further connects ceramide to synuclein deposition.

Mutations in the gene encoding the leucine-rich repeat kinase 2 (LRRK2) are a common cause of PD [57, 60, 115]. The function of LRRK2 is not clear, but it has been shown to possess two enzymatic domains as well as several potential protein-protein interaction motifs [116]. The phenotype attributed to LRRK2 PD is usually not different from the idiopathic form of the disease [117]. However, discrepant results have been presented by neuropathological studies; while some cases have no Lewy bodies [118], most have typical Lewy body disease [119]. The mechanism of this variability is not clear. Similarly, it is not obvious that LRRK2 plays a role in the ceramide pathway as no studies of this question have been published to date.

With our work, we have brought together data suggesting that some of the genes involved in the genetics of Lewy body disease have in common the fact that they impinge on ceramide metabolism. One shortfall of the present theory is the lack of neuropathological data regarding cases with PINK1 or DJ-1 mutations. However, we may see studies addressing this same issue in the near future.

A major premise of this theory is the fact that Lewy body inclusions should have a key role in our understanding of the mechanisms of the disease. We propose that pathology data will, in most cases, be more insightful than clinical data in defining the disease. This is based on what we have learned by other neurodegenerative diseases with inclusion pathology. For Alzheimer's disease, when pathology was used as a basis to understand the disease, pathways involved became evident. This would be most unlikely to happen if, instead, clinical data was used.

These data are incomplete and there have been few relevant studies directly addressing neuronal ceramide metabolism in this context. However, the hypothesis we present has the benefit of making several predictions amongst which are:

1) Mutations in other genes, which alter neuronal ceramide metabolism, should lead to Lewy body diseases, and plausibly ATP13A2 and HSAN mutation carriers should have Lewy bodies.

2) alpha-Synuclein and LRRK2 should have roles in ceramide metabolism.

This notion also suggests that it may be profitable to consider other genes in these pathways as risk factors for Lewy body disease, and in particular, to consider whether they influence the penetrance of the GBA mutations.

CHAPTER 2

OBJECTIVES

Objectives

The work detailed herein intends to shed light on the genetics underlying both familial and common sporadic forms of Parkinson's disease.

We plan to use three individual approaches that, together, have the potential to clarify a proportion of the genetics that is involved in this disease. The first approach focuses on genes where mutations are known to cause the disease. We propose to study these genes in two separate cohorts: the first is comprised by PD patients and controls originating from the center of Portugal, while the second cohort is comprised by samples originating from Nigeria. We chose a Portuguese cohort because studies addressing the genetic variability in PD in this population were scarce. Since some of the genes known to be involved in PD do so at very small frequency, we decided to focus on the most frequent genes. To this end, we studied *LRRK2, SNCA, PINK1,* and *PRKN*.

The second cohort derived from Nigeria. We decided to study this cohort, because the genetics of PD in Sub-Saharian populations was, by and large, unknown. Since the genetic background of the disease is likely to be distinct from the European population, abd thus, the Poertuguese population as well, we decided to focus on the two genes more frequent in worldwide populations (*LRRK2* and *PRKN*) and *ATX3*, a gene previously identified in an African-American kindred presenting with parkinsonsim.

The second approach is based on association studies. Here we compare frequencies of variants in a group of cases with a group of matched controls. If they differ significantly it implies that they may be associated with the disease. The difference to the first approach relies on the fact that these

variants only exert risk to the onset of disease instead of being the underlying cause. We decided to look at genes where associations were established and attempted to replicate those. We selected *HFE,* a gene involved in iron metabolism; *GIGYF2*, a gene where a compelling association was recently proposed based on family studies; and *GBA*, a gene involved in lysosomal degradation where homozygous mutations cause Gaucher's disease and heterozygous mutations have been associated with PD. We plan to use a variety of cohorts for these association studies: *HFE* will tested in the aforementioned Portuguese cohort, *GIGIF2* will additionally be tested in an extended North American cohort comprising close to 1,000 samples, *GBA* will be tested in the Portuguese and a large british cohorts, including pathologically proven PD samples. In addition, the study of *GBA* will also comprise a large meta-analysis with participants from 16 international centers. This approach will clarify the role of these genes in PD, either by confirming or excluding their involvement in the pathogenesis of the disease.

The third approach relies on genome-wide genotyping to uncover genes where common variants act as susceptibility factors for PD. We will perform genome-wide genotyping (in excess of 550,000 markers will be assayed) in a large number of cases and controls, following-up the most suggestive markers in a additional cohort of cases and controls. Samples will originate from North America and Northern Europe. This approach has the added benefit of not making prior assumptions regarding the potential involvement of genes. All genes are assayed in an identical manner, regardless of the biological plausibility of their involvement.

Specific aims are:

1. Identify genetic variability in genes known to cause disease in cohorts of Portuguese and Nigerian ancestry;
2. Replicate previous associations with PD in distinct populations;
3. Identify common genetic variability that plays a role in the common forms of sporadic PD.

CHAPTER 3

MENDELIAN GENES IN PD

Based on the following studies:

III) Bras JM, Guerreiro RJ, Ribeiro MH, Januario C, Morgadinho A, Oliveira CR, Cunha L, Hardy J, Singleton A. **G2019S dardarin substitution is a common cause of Parkinson's disease in a Portuguese cohort**. Mov Disord. 2005 Dec;20(12):1653-5.

IV) Bras J, Guerreiro R, Ribeiro M, Morgadinho A, Januario C, Dias M, Calado A, Semedo C, Oliveira C, Hardy J, Singleton A. **Analysis of Parkinson disease patients from Portugal for mutations in SNCA, PRKN, PINK1 and LRRK2**. BMC Neurol. 2008 Jan 22;8:1.

V) Okubadejo N, Britton A, Crews C, Akinyemi R, Hardy J, Singleton A, Bras J. **Analysis of Nigerians with apparently sporadic Parkinson disease for mutations in LRRK2, PRKN and ATXN3**. PLoS One. 2008;3(10):e3421. Epub 2008 Oct 17.

Mendelian genes in PD

In this chapter we aimed to screen genes known to cause PD in a clear Mendelian fashion in two separate populations. The first population comprises a cohort of PD cases and controls, where all individuals are Caucasian and of apparent Portuguese ancestry. Two studies (III and IV) were performed in the Portuguese cohort. Although the number of samples is not the same, given they were carried out two years apart, and sample collection had continued, they overlap significantly. Nevertheless, the materials and methods are detailed for each individual study.

The usage of this cohort enabled us to have a picture of the genetics underlying this disease in Portugal. We have screened the most common genes known to harbor mutations, these included *LRRK2, SNCA, PRKN*, and *PINK1*. While Study III only tested for mutations in *LRRK2*, Study IV expanded upon this work screening the remaining genes in an extended cohort.

Study V included the screening of the most common Mendelian genes (*LRRK2 and PRKN*) in a small cohort of cases and controls from Nigeria. Additionally, in the African cohort we also tested for the CAG expansion variant in *ATX3*. The rationale for this screening was based upon data generated in 2001 by Gwinn-Hardy and colleagues, who reported a family of sub-Saharan African descent with several individuals displaying parkinsonism suggestive of PD [120]. In this family the ATXN3 mutation segregated completely with the suggestive PD phenotype. Several cases of PD caused by repeat expansion mutation have been described in African-Americans, and

it has been suggested that genetic background may modulate the expressivity of this mutation [121]. It was quickly clear to us that in spite of the upsurge in research and publications relating to PD genetics in the past decade or so, much less is known about the genetics of PD in the African subcontinent, with the majority of publications to date focusing on the North African population [122]. Understanding the genetic associations of PD in Africans will improve our understanding of disease pathogenesis, and improve decision making relating to the usefulness of commercially available predictive genetic tests and preventive and therapeutic interventions that may become available in the future.

Materials and Methods

Methods for Study III

One hundred twenty-eight cases of clinically typical Parkinson's disease were collected at the Movement Disorders Clinic at the Coimbra University Hospital. This is a consecutive clinic case series comprised of patients who gave permission for sampling. The patients were all Caucasian and of apparent Portuguese ancestry, although a detailed genealogical history outside of the nuclear family was not taken. In this clinic, more than 90% of cases consent for blood sampling. The control series, which was from the same region, largely consisted of spouses of affected individuals and comprised 126 individuals. The criteria for Parkinson's disease diagnosis

were the United Kingdom Brain Bank Criteria[5]. The clinical PD evaluation was done using Unified Parkinson's Disease Rating Scale (UPDRS) and Hoehn and Yahr scale. All the cases were diagnosed with levodopa-responsive Parkinson's disease by a neurologist with experience in PD. All control individuals were examined by a neurologist and were found to be free of any movement disorder or neurodegenerative disease. Mini-Mental State Examination (MMSE) was used as a screening test and no other cognitive test was performed. MMSE scores used for the diagnosis are less than 15 for individuals who never went to school, less than 22 for individuals with 1 to 11 years of school, and less than 27 for individuals with more than 11 years of school.

After obtaining informed consent, approved by the Coimbra University Hospital Ethical Committee, a 10 ml blood sample was taken and DNA was extracted by standard procedures. For sequencing of *LRRK2*, exons 31 and 41 of all the case and control DNA were PCR-amplified from the genomic sample using appropriate primers. In addition, we sequenced all 51 exons of the gene in 16 of the familial cases using primers previously described for exons 1–5 and 7–51 and forward primer 5-GGAAGGGCTGCTTCACAGAAAT-3 and reverse primer 5-GAATGGGTTGAGCATCCACAAG-3 for exon 6[10]. In all cases, the products were sequenced using the same forward and reverse primers with Applied Biosystems BigDye terminator and run on an ABI3100 genetic analyzer as per the manufacturer's instructions (Applied Biosystems, Foster City, CA). The

sequences were analyzed with *Sequencher* software (Genecodes, Ann Arbor, MI).

Methods for Study IV

After obtaining informed consent, 132 PD patients underwent a standardized neurological examination by a movement disorder specialist. The diagnosis of PD was based on the UK Brain Bank diagnostic criteria (family history was not used as an exclusion criterion) and those published by Gelb et. al.[5, 123]. Family history was considered positive if parkinsonism was reported in at least a first- or second-degree relative. Collection of these 132 patients was performed at the Movement Disorder Clinics of both the University of Coimbra Hospital and the Lisbon Hospital Center – Center Region EPE in Lisbon, in a consecutive manner, all patients consent to participate. This cohort is identical to that previously described by us except for the inclusion of 4 additional PD patients (Study III). From this series of 132 subjects we have selected 66 unrelated patients to include only those with a positive family history for parkinsonism, or early-onset disease (age at onset <50 years of age). The remaining 66 patients failed to meet either of these criteria, were related to a proband already included or had previously been found to carry the *LRRK2* c.6055G>A; p.G2019S mutation (n = 11). This selection led to the inclusion of 39 patients with positive family history and 46 patients with early-onset PD; 19 patients presented with both an early-onset phenotype and a positive family history, thus the net number of patients from both inclusion groups is 66.

Table 3: Features of patients studied

Characteristic	*Subjects (n=66)*
Age at collection (mean ± SD)	60.1 ± 11.1
Age at onset (mean ± SD)	44.5 ± 9.3
Range of age at onset	20 – 60
Family history	
Positive	39
Negative	27

Additionally we have included a control group comprised of 126 healthy subjects as previously described (Study III). Briefly, this control group consisted primarily of spouses accompanying patients to the clinic (~80%); the remaining controls were recruited from non-neurology outpatient clinics, after observation by the movement disorders specialist. This series presented a mean age of 60.5 ± 23.1 years. Apart from the spouses of the patients, no other familiarity with movement disorders patients was found. All individuals are Caucasian and of apparent Portuguese ancestry.

Genomic DNA was extracted from peripheral blood using standard methods. We screened the genes *SNCA*, *PRKN*, *PINK1* and *LRRK2* for sequence variants and, with the exception of *LRRK2*, for genomic copy number variants. The reference sequence used for the *PRKN* gene throughout this study is based on the accession number NM_004562 and codon counting starts from the first ATG.

For *SNCA*, *PRKN* and *PINK1*, all exons were polymerase chain reaction-amplified and sequenced in both directions using BigDye chemistry (Applied Biosystems, Foster City, CA) on an ABI 3100 Genetic Analyzer as previously

described[16, 124, 125]. While for the *LRRK2* gene, only exon 41 was screened for mutations, using conditions previously described (Study III). Gene dosage analysis was performed using the ABI 7900 Sequence Detection System. Exons 1,2, 4–9 and 11–12 of *PRKN* and exons 1 and 2 of *SNCA*, as well as the complete coding region of *PINK1* were individually co-amplified with β-globin, which served as an endogenous reference gene. Each plate contained six replicates of every genomic DNA sample, control DNA, and a no-template water control. The cycle in the log phase of PCR amplification at which a significant fluorescence threshold was reached (*Ct*) was used to quantify each exon relative to β-globin. The dosage of each exon relative to β-globin and normalized to control DNA was determined using the 2-ΔΔ*Ct* method (Applied Biosystems, Foster City, CA).

Methods for Study V

The study protocol was approved by the Research and Ethics Committee of the Lagos University Teaching Hospital, Lagos, Nigeria. Written informed consent was obtained from all patients and controls. Using a case-control design, 57 unrelated black African PD patients (43 males and 14 females) aged 43 to 80 years and 51 age-matched healthy individuals without a family history of PD or tremor (35 males and 16 females; age range 42 to 87) were recruited from sequentially attending patients at the Neurology Out-patients clinic of the Lagos University Teaching Hospital, Lagos, Nigeria. All patients were evaluated by a neurologist specializing in movement disorders, with keen attention to excluding patients with a possible secondary etiology. The

PD cases recruited were those with clinically definite PD only. The inclusion criteria were the presence of all five of: a) at least two of three cardinal signs of tremor, rigidity, bradykinesia (with or without postural or gait abnormality); b) an asymmetric onset; c) no identifiable secondary cause (e.g. repeated stroke, exposure to medications capable of causing PD within 6 months before onset); d) responsiveness to levodopa therapy (applicable to treated patients only); e) absence of signs of more extensive nervous system involvement (e.g. early autonomic features or cognitive impairment within 2 years of onset, otherwise unexplained corticospinal tract dysfunction, and cerebellar signs). All PD cases were evaluated using a standard protocol that included a historical account, neurological examination, Unified Parkinson's Disease Rating Scale (UPDRS) assessment[126], Hoehn and Yahr staging[127], and Folstein's Mini Mental State examination[128]. Control subjects had an abridged neurologic examination to exclude parkinsonism, cognitive impairment, corticospinal tract dysfunction, or any overt neurologic illness.

The mean age at onset (based on patient's or caregiver's recollection of age at onset of first cardinal symptom of PD) for this group is 58.2 years (range 40–75). The majority of patients presented no apparent family history for parkinsonism, while nine presented at least one first-degree relative with a history of tremor. This fact may suggest that, for these nine individuals, an autosomal dominant mode of inheritance is possible. Hence, this study will fail to rule out the possibility of mutations in one of the known autosomal dominant PD genes, the SNCA gene. Although all patients are from Nigeria, their specific ethnic origins were as follows: Yoruba – 35 (61.4%), Igbo – 11

(19.3%), and Edo (Ijaw/ Itsekiri/ Urhobo/ other south-south ethnicity) – 11 (19.3%). The ethnic origins of the controls were as follows: Yoruba – 36 (70.6%), Igbo – 9 (17.6%), Edo – 5 (9.8%) and Hausa – 1 (2%).

Genomic DNA was extracted from saliva using the Oragene kit (DNA Genotek). PRKN, LRRK2 and ATXN3 were screened for mutations. For PRKN, all exons and intron/exon boundaries were polymerase chain reaction-amplified and sequenced in both directions using BigDye chemistry (Applied Biosystems, Foster City, CA) on an ABI3730xl as previously described[16]. PRKN mutations are numbered according to GenBank accession number NP_004553 for the protein (p.) and NM_004562 for the cDNA (c.). LRRK2 was screened by sequence analysis for the most common mutations occurring in exons 31 and 41 as previously described (Study III). ATXN3 pathogenic repeat expansion size was assessed in all samples using methods described previously by us [121]. In addition to sequencing PRKN, we screened PRKN for copy number mutations in two samples carrying heterozygous mutations not found in controls (p.P153R; c.C458G and p.R334H; c.G1001A), using the Illumina Infinium HumanHap550 BeadChips (version 3; Illumina Inc, San Diego, CA, USA) as previously described[129]. Copy number analysis by genotyping was done according to the manufacturer's protocol (Illumina Inc.) using 750 ng of genomic DNA. Data was analysed with BeadStudio v3 (Illumina Inc.) using the Human Genome build 35. Two metrics were visualized using this tool, B allele frequency and log R ratio. The former is the theta value for an individual SNP, which gives an estimate of the proportion of times an individual allele at each polymorphism

is called A or B. The log R ratio is the log2 ratio of the observed normalized R value for the SNP divided by the expected normalized R value for the SNPs theta value. An R above 1 is indicative of an increase in copy number, and values below 1 suggest a deletion. We have shown previously that this is a reliable method for detecting large genomic copy number mutation in PRKN[129].

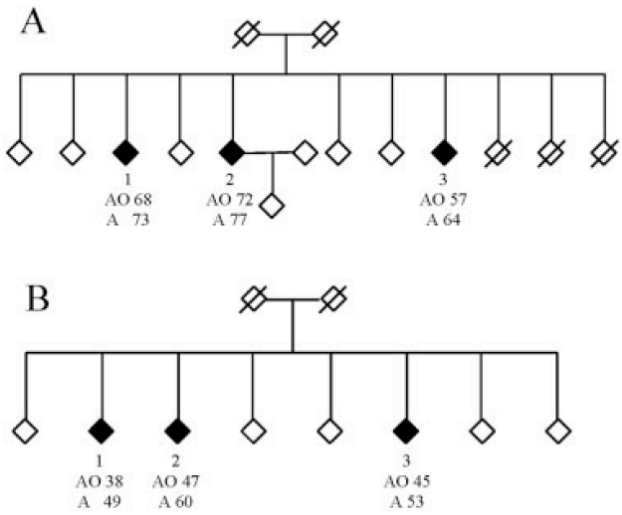
Results

Results for Study III

In our series of 128 consecutive patients, 11 (9%) had the G2019S mutation; none had the R1441G mutation. The mutations occurred in five sporadic cases of disease and in two families. Discounting the secondary cases in these families, the mutation prevalence in probands is 7/124 (6%). In addition, our sequencing of the rest of the gene in a subset of familial disease failed to identify other mutation carriers. The mutation-positive patients presented with an akinetic-rigid syndrome (n = 5) or a tremor-predominant disease (n = 6). In both families with the G2019S mutation, the mutation occurred in all three affected family members who had attended the clinic together. In Family A, both parents in this kindred were apparently healthy and by history died of nonneurological disease in their 80s. The three individuals are siblings and appear to be the only affected family members. Two of the patients show an akinetic-rigid syndrome while the other presented with resting tremor. Two

individuals also had cognitive impairment beginning after their parkinsonian syndrome. Family members A2 and A3 showed cognitive decline (MMSE scores 22 and 13, respectively). In the second family, two of the mutation carriers (B2 and B3) presented with painful cervical dystonia, which was responsive to levodopa, and neither showed tremor or dementia. In this family, the parents died at ages 69 and 80 without neurological illness according to the family report. For the five sporadic cases, parental mortality information was available for three of them. In two, the father died at less than 55 years and the mother died at 70 years; in the third, both parents died at an age greater than 70. It is worth noting that the case with the youngest age of onset (38 years) is an identical twin, whose twin remains clinically unaffected. Neither of the families and none of the sporadic cases are known to be related to each other, nor did they come from the same villages. Thus, there is no suggestion that these individuals share a recent founder. In addition, though the ethnic background of the two families from the United States is not known, in 12 families there is no hint of Portuguese ancestry; rather, German and Irish are discussed as the likely ethnic backgrounds.

Figure 2: Families A and B, positive for LRRK2 G2019S mutation. All affected family members are carriers of the mutation. AO: Age at onset; A: current age.



Results for Study IV

Analysis of both sequence and copy number yielded several parkin mutations in our subset of patients. The positive results found are represented in Table 4, and the electropherograms corresponding to point mutations are shown in Figure 3. We found four patients in whom both alleles were mutated; three of these patients had the same homozygous mutation (c.154delA; p.N52fsX80), a single base pair deletion that inserts a premature stop codon downstream; one patient showed deletion of exon 2 and duplication of exon 5; analysis of a sixth sample (S4) showed data consistent with a homozygous c.1183G > T; p.E395X mutation and a duplication of exon 9; because the co-occurrence of 3 mutations in the same gene is unlikely we designed an additional forward primer that flanked the E395X mutation as close as possible on the 5' side; sequencing of the PCR product generated by amplification with this new primer and primer PRKNexon11R showed the E395X in a heterozygote state, suggesting that the duplication of exon 9 (and presumably exon 10, which we were unable to assay) interfered with the original PCR and sequencing reaction. Thus in Table 4 this mutation is denoted as a compound heterozygous E395X/exon 9 duplication mutation.

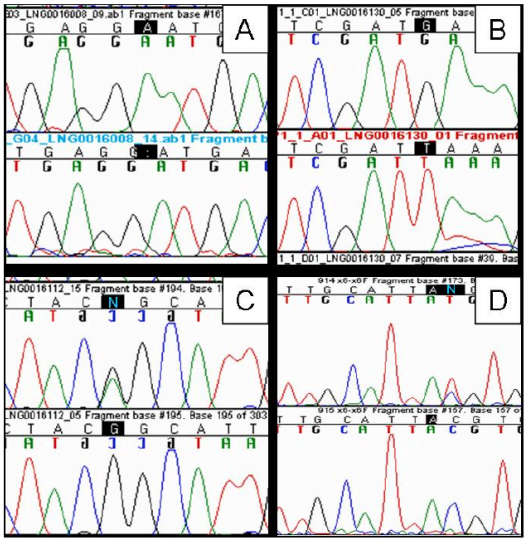


Figure 3: Chromatograms of the mutations found.

Two samples tested positive for the *LRRK2* c.6055G > A; p.G2019S mutation, sample S12 and sample S7. Notably, analysis of sample S7 also showed a

heterozygous duplication of *PRKN* exon 9. Screening of variants E395X, G2019S and 154delA in 252 control chromosomes failed to reveal any control subjects harboring these mutations. Additionally, we have identified three patients with heterozygous mutations in *PRKN*: one harboring the T240M variant; another with a deletion of exons 8 through 11 and one with an exon 8 duplication. Furthermore, we have found one patient with two heterozygous dosage variants: an exon 2 deletion and an exon 5 duplication.

Table 4: Variants found. Homo: Homozygous; Het: Heterozygous; * previously undescribed mutation; # variants of unknown pathogenicity; AAO: age at onset.

Sample	*Gene*	*Nucleotide change*	*Amino acid change*	*Copy variation*	*Exon*	*Zygosity*	*AAO*
S1	PRKN	154delA	N52fsX80	N/A	2	Homo	30
S2	PRKN	154delA	N52fsX80	N/A	2	Homo	35
S3	PRKN	154delA	N52fsX80	N/A	2	Homo	21
S4	PRKN	G1183T	E395X*	N/A	11	Het	53
	PRKN	N/A	N/A	Duplication	9	Het	
S5	PRKN	C719T	T240M	N/A	6	Het	55
S6	PRKN	N/A#	N/A	Duplication	8	Het	33
S7	PRKN	N/A	N/A	Duplication	9	Het	38
	LRRK2	G6055A	G2019S	N/A	41	Het	
S8	PRKN	N/A#	N/A	Deletion	8-11	Het	32
S10	PRKN	N/A#	N/A	Deletion	2	Het	35
	PRKN	N/A#	N/A	Duplication	5	Het	
S12	LRRK2	G6055A	G2019S	N/A	41	Het	41

No mutations were found in *SNCA* or *PINK1*.

Results for Study V

We did not find any variants in exons 31 and 41 of LRRK2. Likewise, the screening for the ATXN3 repeats revealed that no samples contained pathogenic expansions. The PRKN gene yielded several variants.

Table 5: Variants found

Gene	Heterozygous variants	*Cases* N (%)	*Controls* N (%)
PARK2	p.E16E; c.G48T	3 (5.3)	1 (2.0)
	p.P37P; c.G111A	8 (14.0)	6 (11.8)
	p.A46T; c.G136A	1 (1.8)	0
	p.P153R; c.C458G	0	1 (2.0)
	p.S167N; c.G500A	4 (7.0)	4 (7.8)
	p.M192L; c.A574G	7 (12.3)	6 (11.8)
	p.L261L; c.A783G	13 (22.8)	14 (27.4)
	p.G319G; c.T957C	5 (8.8)	4 (7.8)
	p.R334H; c.G1001A	1 (1.8)	0
	p.V380L; c.G1138A	2 (3.5)	3 (5.9)
	Homozygous variants		
	p.P37P; c.G111A	1 (1.8)	3 (5.9)
	p.S167N; c.G500A	1 (1.8)	0
	p.L261L; c.A783G	1 (1.8)	3 (5.9)
	p.V380L; c.G1138A	1 (1.8)	1 (2.0)

With the exception of the V380L and S167N polymorphisms, no other missense homozygous variants were found. All the PRKN variants are shown in Table 5.

Additionally, samples that showed only one heterozygous mutation in PRKN that was not present in controls were screened for copy number variants in this gene, in order to assess if these were in fact compound heterozygous for one point and one genomic copy number mutation. Genomic copy number analysis did not detect any mutations in these samples.

Discussion

The pathogenicity of the G2019S is clear for three reasons. First, our previous data showed segregation of this mutation in two families from North America. Second, the limited segregation in the two families we have examined also shows segregation of mutation with disease. Third, we did not find this mutation in controls in our control series here, nor have we found this variant in more than 1,500 control subjects from North America. These data are remarkable and of clinical importance for the simple reason that they show that a high proportion of clinically typical Parkinson's disease in this population, as in the Basque population, carries a pathogenic mutation. Many of these cases do not have familial disease, and, given the information from the two familial cases, the likely reason for this non-penetrance even at high age in the parents is unclear. These findings will have an impact on clinical practice. Neurologists treating Parkinson's disease have usually assured their patients that the disease was not genetic in etiology and certainly have not routinely suggested genetic testing. These data suggest that this widespread advice and practice will have to change. A large proportion of cases, certainly from Portugal and from the Basque country, carry mutations that put their family members at very high risk for disease, although the likely ages of onset are extremely difficult to predict.

In addition to the common mutation G2019S in *LRRK2* a detailed mutation analysis of *PRKN*, *PINK1 and SNCA* was performed. We have included PD patients with a positive family history (n = 19 age at onset <50 years, n = 20 age at onset ≥ 50 years), or early-onset sporadic disease (n = 27) in order to maximize our chances of identifying mutations. This approach has led us to

find 6 subjects (9.1%) with pathogenic mutations in *LRRK2* or *PRKN*, in addition to 4 variants of unknown significance in 4 patients.

We showed that the c.6055G > A; p.G2019S LRRK2 mutation underlies about 6% of late-onset PD in the Portuguese population (Study IV). While we did not find any c.6055G > A; p.G2019S carriers in the 20 late-onset patients, we did identify c.6055G > A; p.G2019S in 2 of 46 early-onset cases (Study III). One of these individuals also carried a heterozygous duplication of *PRKN* exon 9 consistent with the notion of digenic parkinsonism, as previously described [130]. This patient presented no family history consistent with PD, while the other *LRRK2* patient had positive family history. Taking into account the removal of samples previously found to carry the c.6055G > A; p.G2019S mutation we calculate that this mutation is present in 9 probands out of the entire series of 132 patients; this represents 2 of 46 early-onset patients, counting only sporadic cases and a single proband from each family (4.3%) and 7 of 76 late-onset patients counting only sporadic cases and a single proband from each family (9.2%). We found several *PRKN* mutations as either homozygous or compound heterozygous loss of function changes. The N52fsX80 variant was the most frequent mutation identified in *PRKN*. It was present as a homozygous mutation in three unrelated young onset patients (of 46, 6.5%). Analysis of relatives of these patients failed to show any heterozygous carriers of this mutation with parkinsonism. We identified a heterozygous deletion of exon 2 and a duplication of exon 5 in *PRKN* in one early-onset patient with positive family history. The only affected family member that was available for testing was the sibling of S10 who presented with the same two variants, albeit with a remarkably different age at onset (50

years vs 35 years of patient S10) (Figure 2). While parsimony suggests that these mutations are *in trans* we were unable to unequivocally establish phase as DNA was unavailable from other family members. We also identified a heterozygous deletion of exons 8 through 11 in a female patient with an age at onset of 32 years. Additional family members were unavailable, so we were unable to determine whether this mutation represented a single contiguous mutation or two mutations existing *in trans* and thus the pathogenicity of the observed changes remains unknown. We identified a novel mutation in *PRKN* exon 11 (E395X) as a heterozygous alteration. This patient also presented a heterozygous duplication of exon 9. The pathogenicity of the new E395X mutation is clear since it is a nonsense mutation that occurs upstream of a functional domain of the protein. Of these patients presenting either homozygous or clear compound heterozygous mutations in *PRKN*, four (80%) have an age at onset below 40 years. Only one (patient S4) presents late-onset disease (53 years). In addition we identified several *PRKN* variants of unknown significance. The T240M alteration, an exon 8 deletion and an exon 8 duplication were each identified as heterozygous mutations in single patients. In the absence of additional mutations in *PRKN* in these subjects, we have not considered these as disease causing variants in these patients. We make this statement with caution, since we cannot rule out copy number mutations in exons 3 or 10, which we were unable to assay effectively.
With Study V we have performed the first study screening a sub-Saharan African cohort of apparently sporadic PD cases for mutations in genes commonly associated with PD. The results from this study are of clear importance not only for Nigerian PD patients, but also because they shed light

on the genetic background associated with PD in the African population. It should be noted that the number of individuals included in this preliminary report is clearly small, and thus, definite conclusions about frequencies of variants are difficult to achieve.

We have performed a screening for the most common autosomal recessive variants in three genes associated with PD (LRRK2, PRKN and ATXN3). We decided not to screen for mutations in the genes PINK1, DJ-1 and ATXN2 given the low frequency of mutations in these genes in worldwide populations. Moreover, we did not screen SNCA as mutations in this gene are not only rare but they are also associated with an autosomal dominant mode of inheritance. Given that the majority of our patients have no affected family members with any form of parkinsonism, mutations in SNCA would be unlikely to occur in our cohort. Mutations in the LRRK2 are the most common cause of PD in several populations, including populations of Northern African ancestry. In particular, Lesage and colleagues found a high frequency (41%) of the G2019S mutation on exon 41 in a study of North African Arabs that included both familial and apparently sporadic PD cases [60]. Thus, it would be interesting to determine if there is a similarly high frequency of LRRK2 G2019S mutations in other geographically and ethnically distinct parts of Africa. However, we did not find any mutations in either exons 31 and 41 of LRRK2 in our cohort, suggesting that mutations in these domains of LRRK2 are not a common cause of PD in sub-Saharan populations. A recent study performed a comprehensive analysis of the entire coding region of LRRK2 in a large cohort of American PD cases and healthy controls [131]. Of the seven mutations found to be segregating with disease, five were in either exon 31 or

exon 41, indicating these as clear mutational hotspots. The noteworthy difference in mutation frequency among populations from the same continent is, in all probability, due to the occurrence of the founding G2019S mutation event happening after human populations moved out of sub-Saharan Africa as this is most consistent with the dating of this mutational event [12, 132].
Parkinsonism due to ATXN3 repeat expansion mutation has been previously described in one single large African descent family. We failed to find any samples harboring the increased repeat expansion size. This result suggests that ATXN3 repeat expansion mutations are not a frequent cause of parkinsonism in this population.
Even though a considerable number of PRKN mutations are dosage mutations, the majority are sequence variants; hence, we decided to perform the initial screen of our cohort only for these variants. Subsequently, we performed gene dosage analysis in two samples as previously detailed. Again, our study did not identify any pathogenic mutations in PRKN in our subset of PD patients. We found several heterozygous variants both in patients and controls. Two of the variants are novel and present only in PD cases (p.R334H and p.A46T). Given the fact that these are novel variants, we ran the analysis software SIFT (available at http://blocks.fhcrc.org/sift/) [133] in order to have some insight into the potential effects of these variants. The p.A46T was predicted by the software to potentially affect protein function, whereas the p.R334H was predicted to have no functional effect. However, these results are merely based on a similarity score in comparison to other proteins, and hence it must be stressed that this is not true functional data for these variants. Nonetheless, in the absence of a second mutation, these

cannot be described as pathogenic, thus we decided to classify these as variants of unknown pathogenicity. It should be noted that for these two samples, an additional screen for gene dosage mutations was performed using the Illumina BeadChips. In addition, one homozygous variant was found only in the PD group (p.S167N; c.G500A). However, this has previously been classified as a polymorphism [134, 135].

All populations showed polymorphisms with varying frequencies. Three variants p.A46T, p.P153R and p.R334H were found in a single sample each. Variant p.A46T was present in one PD sample from Igbo, p.P153R was found in a Yoruban sample, and p.R334H was present in a sample with Edo background. Two individuals, both from the same ethnic region (Edo), presented with two missense variants each (p.M129L and p.S167N; p.M129L and p.V380L). Although the present study cannot completely rule out that these compound hetererozygous events could potentially be pathogenic, two facts suggest otherwise: 1) in each case the second variant is a well known and described SNP; 2) one control sample presented the same combination of two of the variants (p.M129L and p.V380L). We report the first genetic screening for PD genes in a sub-Saharan population. We found no pathogenic mutations in the genes most commonly known to cause PD in European North American, or North African populations. Although the cohort studied is clearly small and definite conclusions regarding frequencies are unachievable, a trend for different genetic basis of PD in this sub-Saharan population is, in our opinion, noteworthy. Two main caveats are present in this work: gene dosage mutations in PRKN were only screened for in two samples and only exons 31 and 41 of LRRK2 were sequenced. Nevertheless, the aim

of this study was to ascertain a preliminary frequency of the most common variants known to cause PD in a sub-Saharan population. We report here that the most common variant associated with PD in several world-wide populations, the p.G2019S mutation in LRRK2, is not overrepresented in this Nigerian population of PD patients; similarly, sequence variants in PRKN, which represent a significant proportion of PRKN mutations underlying PD, are also not significantly present in the studied cohort. It is thus likely that the differences in the genetic background of these populations mean that other genes or different variants are underlying the disease. Therefore, a search for these is clearly warranted, since they will, in all probability, shed more light on different pathways leading to PD.

It is now clear that genetics plays an important role in the pathogenesis of PD. Specifically, in the Portuguese population, we have found a reasonable number of mutations: the frequency of the c.6055G > A; p.G2019S is one of the highest in Europe, and we have found that 8.7% (4 out of 46 cases) of early-onset cases are attributable to *PRKN* mutations. Similar to other reports we found *PINK1* and *SNCA* mutations to be a rare cause of disease in our families [136]. Taken as a whole these results have implications mainly for clinicians in Portugal; in particular showing that genetic screening may aid the diagnosis of PD in this population. However, even with the combination of gene dosage and sequencing, a significant proportion of mutations might remain undetected, probably due to the size and the complexity of the *PRKN* gene. In this way, negative results should be interpreted with caution, as well as heterozygous mutations in this gene.

CHAPTER 4

CANDIDATE GENE ASSOCIATION STUDIES IN PD

Based on the following studies:

VI) Guerreiro RJ#, Bras JM#, Santana I, Januario C, Santiago B, Morgadinho AS, Ribeiro MH, Hardy J, Singleton A, Oliveira C. **Association of HFE common mutations with Parkinson's disease, Alzheimer's disease and mild cognitive impairment in a Portuguese cohort**. BMC Neurol. 2006 Jul 6;6:24.

VII) Bras J#, Simón-Sánchez J#, Federoff M, Morgadinho A, Januario C, Ribeiro M, Cunha L, Oliveira C, Singleton AB.**Lack of replication of association between GIGYF2 variants and Parkinson disease**. Hum Mol Genet. 2009 Jan 15;18(2):341-6. Epub 2008 Oct 15.

Candidate gene association studies in PD

Candidate gene studies have had considerable success in identifying loci associated with PD. Thus far, 14 chromosomal *loci* have been described (PARK1-PARK14), in which mutations in seven genes are unequivocally linked to rare forms of PD (SNCA, PARK1 OMIM #168601 and PARK4, PRKN, PARK2 OMIM #602544; PINK1, PARK6 OMIM #605909; DJ-1, PARK7 OMIM #606324; LRRK2, ATP13A2, PARK9 OMIM #610513 and PLA2G6, PARK14 OMIM #603604) [6, 10, 16, 18, 20, 23, 137]. Although the mechanism through which mutations in these genes exert their pathogenicity is not fully understood, SNCA and LRRK2 are known to cause autosomal-dominant disease, while the remaining cause autosomal-recessive PD.

More controversial results have been obtained for loci such as PARK5 [138] and PARK13 [70, 139]. Candidate gene studies are usually prompted by insights into the molecular mechanisms underlying the disease. In this sense, post-mortem examinations of PD brains and magnetic resonance imaging of PD patients that have revealed increased iron contents in the substantia nigra led to the analysis of iron metabolism related genes. Classic Hemochromatosis is an autosomal recessive disorder that is associated with a deregulation of the iron metabolism [140]. Clinical features often include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Hemochromatosis is most often caused by mutations in the gene *HFE* on chromosome 6p21.3. The most common mutation, C282Y, was

initially found in a subset of patients with hereditary hemochromatosis, in a total of 83% of all individuals. A second mutation, H63D, was also described, although the clinical effects of this modification are clearly more limited. However, about 1 to 2 percent of individuals with compound heterozygous *HFE* mutations appear to be at risk for hemochromatosis [141].
Previous studies assessing the effect of *HFE* variants on the onset of PD have been contradictory [142, 143]. Thus, to ascertain if *HFE* mutations are a risk factor for the development of this disease, we conducted a genetic screening for the most common *HFE* mutations in a series of patients and healthy controls.

PARK11 is located in chromosome 2q36–37 and was initially described by a whole-genome linkage analysis in a population of familial PD patients [144-146]. Although conflicting results followed shortly [147], it was also detected in an earlier association analysis [148]. The PARK11 locus spans 18cM encompassing 73 candidate genes where the highest LOD score was obtained for the marker D2S206, located within intron 21 of GIGYF2, a gene encoding a 1320 amino acid protein (Grb10-Interacting GYF Protein 2, gigyf2). Because of this and because gigyf2 has been shown to interact with grb10 and consequently have a potential role in insulin and insulin-like growth factor signaling [149], Lautier et al. [150] recently performed a screening of pathogenic mutations in GIGYF2 in a series of 249 familial PD Cases and 237 Controls from two different populations in Europe. The authors reported 10 different mutations spread in an even manner throughout the gene in PD patients but not in Controls, suggesting that these variants would be the

cause of the disease in these patients. These results prompted us to undertake a complete screening of GIGYF2 mutations in a large series of 724 Cases and 911 neurologically normal Controls from two different populations.

Materials and Methods

Methods for Study VI

A total of 132 PD patients were selected according to the United Kingdom Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria (UK PDS Brain Bank)[5]. Patients comprised a consecutive clinic based cohort (over 90% of cases consent for blood sampling), diagnosed by a movement disorder specialist at the movement disorder clinic of the University of Coimbra Hospital. This series included 62 males and 70 females, with mean of ages of 66,7 ± 10,7 years, and mean age at onset of 57,2 ± 12,0 years. From these, 28 patients presented with a positive family history for PD, while the remaining 104 showed no evidence of family history for PD or any form of parkinsonism. The control group included 115 healthy controls with a mean age of 70,7 ± 10,3 years, 38 males and 77 females. All subjects were examined by a neurologist and were free of any clinical signs or symptoms of neurodegeneration. This group comprised mainly spouses of patients and caregivers that were accompanying patients to the clinic. All individuals included in this study are Caucasian with an apparent Portuguese ancestry. The study was submitted to the Ethics board of the University Hospital of Coimbra and all the subjects involved gave their informed consent.

Genotyping

Genomic DNA was isolated from whole blood by means of standard procedures and the samples were genotyped for the *HFE* mutations C282Y and H63D using the polymerase chain reaction (PCR) technique with subsequent restriction and gel electrophoresis, as previously described[151]. Similarly, *APOE* genotypes were assessed by a PCR-based methodology, as previously described[152].

Statistical analysis

Observed genotype distributions were compared with those expected by cross-tabulation and analyzed using Chi-square and Fisher Exact-tests. Means of quantitative variables were compared using Student's *t*-test. Kaplan-Meier (KM) survival analysis was used to analyze the effects of the *HFE* mutations on the age of PD onset. The log-rank test was employed to determine whether genotype-specific survival functions were significantly different from one another. All tests were interpreted at the 0,05 level of significance. All statistical analyses were performed with the SPSS package, version 10.0 (SPSS, Chicago, IL, USA).

Methods for Study VII

Portuguese series

The series originating from Portugal comprised 267 PD patients and 451 healthy Controls, their characteristics are presented in Table 1. Patients were selected in a consecutive manner, in the Movement Disorder clinic of the University of Coimbra Hospital. Diagnosis followed the UK Brain Bank criteria[5]. Control samples were collected from healthy unrelated individuals

from the same population and geographical regions. All Controls were subjected to a neurological examination and found free of any symptoms suggestive of parkinsonism.

US series

The US series were taken directly from pre-compiled panels from the National Institute of Neurological Disorders and Stroke (NINDS) funded Neurogenetics repository hosted by the Coriell Institute for research (NJ, USA). All participants provided written informed consent. Neurologically normal Controls were derived from five different panels of DNA: NDPT002, NDPT006, NDPT009, NDPT022 and NDPT024, containing DNA from total of 460 unrelated individuals from North America, including 225 males and 235 females. All individuals were Caucasian and lacked history of Alzheimer's disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, brain aneurism, dementia, dystonia, or PD. None had any first-degree relative with a known primary neurological disorder and the mean age of participants was 68.57 (range 55–95). PD Cases were taken from five panels of DNA: NDPT001, NDPT005, NDPT007, NDPT017 and NDPT018. These panels contain DNA from 460 unrelated Caucasian individuals from North America with PD, including 258 males and 202 females. The mean age at onset is 66.36 years (range 50–87) and they all showed at least one of the main clinical signs of PD such as resting tremor, rigidity, bradykinesia, gait disorder and postural instability at the disease onset. All subjects were questioned regarding family history of parkinsonism, dementia, tremor, gait disorders and other neurological dysfunction. Subjects both with and without a reported

family history of PD were included. None were included who had three or more relatives with parkinsonism, nor with clear Mendelian inheritance of PD. A more detailed description of both Case and Control samples, can be found at http://ccr.coriell.org/Sections/Collections/NINDS/DNAPanels.aspx?PgId=195&coll=ND.

Sequencing analysis

Screening of GIGYF2 was carried out using genomic DNA of a total of 727 Cases and 911 neurologically normal Controls from two different populations. Polymerase chain reaction (PCR) amplification was performed in a final volume of 16 ml containing 10 ng of genomic DNA, 10 pmol of forward and reverse primers and 12 ml of FastStart PCR Master mix (Roche). Primers for all coding exons and intron/exon boundaries were designed using ExonPrimer (http://ihg2.helmholtz-muenchen.de/ihg/ExonPrimer.html) for isoform a (NM_001103147.1) of GIGYF2, which is the longest transcript of the gene, encoding 31 exons. Note that this isoform is different than that sequenced by Lautier et al. in their series (NM_015575.3). Therefore amino acid and cDNA positions are different. Each purified product was sequenced using Applied Biosystems BigDye terminator v3.1 sequencing chemistry as per manufacturer's instructions (Applied Biosystems, Foster City, CA, USA). The resulting reactions were purified and resolved on an ABI3730XL genetic analyzer (Applied Biosystems) and analyzed with Sequencher software v4.1.4 (Gene Codes Corporation, VA, USA).

Statistical analysis

Statistical analyses included Hardy–Weinberg equilibrium, x2 and Bonferroni correction tests and were performed using PLINK v1.03[153].

Results

Results for Study VI

To test the association between the presence of the C282Y and H63D mutations and the development of PD, we screened these series of patients and a series of healthy controls. The genotypes in these cohorts were at or near Hardy-Weinberg equilibrium. Analysis of the genotypes in the PD series revealed a significant overrepresentation of 282Y carriers and of the 282Y allele compared to controls (p = 0.01) (Table 6).

Table 6: Genotype frequencies for HFE mutations

	C282Y			P	H63D			P
	AA	**GA**	**GG**		**GG**	**CG**	**CC**	
Controls (n=115)	0	5 (4.3%)	110 (95.7%)		2 (1.7%)	39 (33.9%)	74 (64.3%)	
PD (n=132)	0	18 (13.6%)	114 (86.4%)	0.01	5 (3.8%)	38 (28.8%)	89 (67.4%)	0.47

The outcome of the genetic mutations studied may also affect the age at onset of the studied disorders. Therefore we used Kaplan-Meier survival curves to determine this outcome (Figure 4).

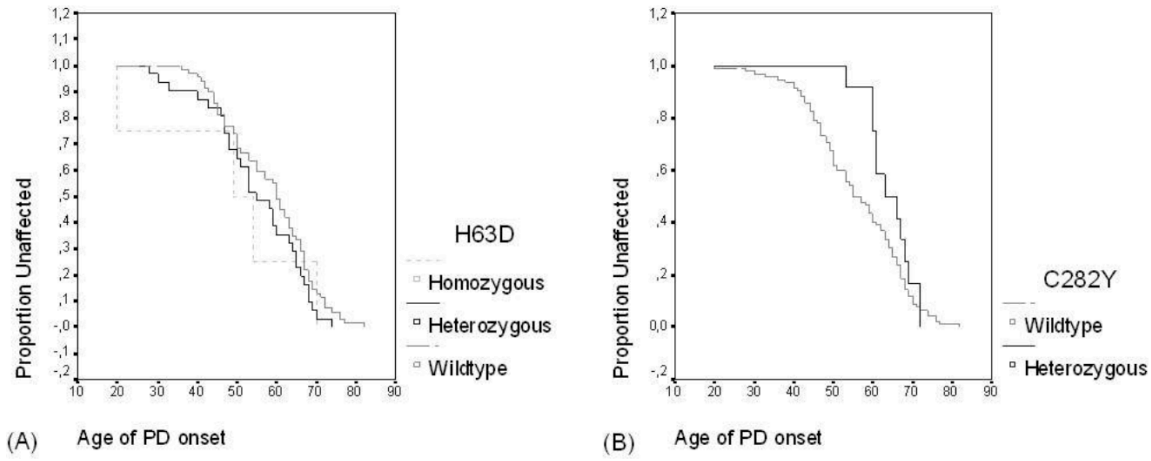


Figure 4: Kaplan-Meier survival curves indicating the effect of H63D and C282Y mutations on age of PD onset. (A) There are no statistically significant differences in the age at onset of PD between wild type, heterozygous and homozygous patients for H63D mutation ($\chi 2(2df) = 2.4$, P = 0.30). (B) There are no statistically significant differences in the age at onset of PD between wild type, heterozygous and homozygous patients for C282Y mutation ($\chi 2(1df) = 1.66$, P = 0.20).

We failed to find any association between the mutations studied and the age at onset of PD.

Results for Study VII

Portuguese series

The Portuguese cohort yielded 31 variants (Supplementary Table 1), of these; seven were known polymorphisms present in dbSNP. Four non-synonymous variants were found only in Cases, while 16 non-synonymous variants were present only in Controls (Supplementary Table 1). Of the remaining 24 variants not present in dbSNP, six were synonymous changes, suggesting no functional change for the protein. None of the variants present in both Cases and Controls showed association with disease. Although the variants rs12328151 and rs2289912 showed a P-value ,0.05 after x2, they failed to maintain the association after Bonferroni correction (P . 0.7). Interestingly, one

of the variants present only in the Control group (p.N478T) had previously been reported as a pathogenic mutation (18). In addition, the exon 29 p.L1230_Q1237del, which had also been suggested to be disease-causing, is present in this population with an increased frequency in the Control group. We failed to find any of the remaining 8 variants previously associated with disease in our cohort.

US series

The US series harbored 40 different variants (Supplementary Table 2) including seven described SNPs in exons 15, 16, 26, 29 and 31; and introns 3 and 28. None of these polymorphisms showed association with PD after x2 test of association. Although rs2305138 showed a P-value of 0.04, this is not considered significant after multiple-test correction. Interestingly, two of the variants Lautier et al. described as mutations in their series were found in this cohort in both Cases and Controls. These variants (c.3689-3712del24 and p.H1192R) were not associated with PD in our cohort after x2 test for association (P 1/4 0.385 and 0.1621, respectively). In addition, we have found 26 novel variants of which 16 were non-synonymous. Of these, three (p.G108R, p.L1230_Q1236del and p.P1231_Q1232insQQ) were found in both Cases and Controls but were not associated with disease (P 1/4 0.58, 0.90 and 0.29, respectively). The remaining variants were found in either a single Control sample (p.S395T, p.R982Q, p.S1269G and p.Q1272E) or a single PD Case (p.D453G, p.M586T, p.Q599K, p.S729G, p.R778G, p.R895G, p.S1056C, p.N1270D and p.I1314V) (Supplementary Table 2). Except for p.A995A and p.K1015K which were found in homozygous state in a single

Control sample, the rest of silent changes were found in heterozygous state in only one sample.

Discussion

The data generated in Study VI showed a significant increase of the prevalence of 282Y carriers in the PD cohort compared to controls. A previous study examining the relationship between *HFE* variants and PD reported an opposite effect to the data presented here: the authors presented data suggesting that individuals with C282Y mutation have a decreased risk of developing PD [143], in contrast an additional study suggests no role of *HFE* variants in risk for PD [154] and recent work describes a positive relationship between the 282Y variant and PD risk, consistent with the data presented in the current study [142].

The discordant results may be explained by several factors: first, the results of the current study and those of Dekker and colleagues represent false positive findings; second, the results of Buchannan and colleagues represent false positive findings; third, 282Y is not a causal variant but is in linkage disequilibrium with another variant that underlies disease risk. The degree and direction of a disease association when genotyping what is in effect a tagging SNP, are both sensitive to the structure and content of a given block of linkage disequilibrium; these factors are both potentially different between populations. While it is tempting to speculate that differences in iron handling may differentiate the molecular underpinnings of these two disorders, the current data is too far removed from this mechanistically and too preliminary

to make this a convincing argument. The infrequency of C282Y mutations obviously limits the statistical power of this analysis, thus, studies in larger samples from diverse populations are needed to clarify the relationship between variability in *HFE* and PD The small number of individuals in this study makes an ultimate assessment of the biological and genetic significance of these data clearly impossible. Thus we have analyzed all previous studies published so far on this subject, in order to perform a meta-analysis of the data, and hopefully shed some light on these mechanisms (Tables 7 and 8).

Table 7: Meta-analysis of the C282Y variation in the three published studies regarding Parkinson's disease

	C282Y						
	Patients			Controls			
Author	wt/wt	wt/mut	mut/mut	wt/wt	wt/mut	mut/mut	p
Dekker, M.	125 (91.2%)	10 (7.3%)	2 (1.5%)	2616 (89.7%)	290 (10%)	8 (0.3%)	
Dekker, M.	54 (90%)	6 (10%)	0	2616 (89.7%)	290 (10%)	8 (0.3%)	
Buchanan, D.	391 (89.3%)	46 (10.5%)	1 (0.2%)	405 (83.5%)	76 (15.7%)	4 (0.8%)	
Borie, C.	66 (93%)	5 (7%)	0	53 (91.4%)	5 (8.6%)	0	
Total	636	67	3	3074	371	20	0.55

Table 8: Meta-analysis of the H63D variation in the two published studies regarding Parkinson's disease

	H63D						
	Patients			Controls			
Author	wt/wt	wt/mut	mut/mut	wt/wt	wt/mut	mut/mut	p
Dekker, M.	104 (76%)	31 (22.6%)	2 (1.4%)	2185 (75%)	661 (22.7%)	68 (2.3%)	
Dekker, M.	44 (73.3%)	16 (26.7%)	0	2185 (75%)	661 (22.7%)	68 (2.3%)	
Borie, C.	42 (63.6%)	23 (34.8%)	1 (1.5%)	39 (66.1%)	20 (33.9%)	0	
Total	190	70	3	2224	681	68	0.21

When considering the results from the meta-analysis, no statistically significant association between any of the variants and PD is detected.

Overall, Study VI suggests that genetic variability in *HFE* may be a risk factor for PD. The rarity of *HFE* 282Y limits the statistical power of this analysis, thus reinforcing that studies in larger samples and in diverse cohorts are needed to make clarify the relation between variability in *HFE* and PD.

Study VII presents a detailed analysis of the genetic variability in the GIGYF2 gene and its association with PD in two large sets of Cases and age-matched Controls, from two geographically distinct populations. The sample originating from Portugal comprised a total of 267 PD samples and 451 healthy age-matched Controls, while the US series comprised 460 Cases and 460 Controls. A significant difference for the previously published study on GIGYF2 variants and PD is the fact that we used a different transcript—isoform ‘a’ (NM_001103147.1)—whereas the transcript studied before was isoform ‘b’ (NM_015575.3), which lacks one exon when compared with the former. Due to this difference, the present mutation numbering differs from the work previously published.

The combined analysis of both cohorts yielded 46 variants of which seven are SNPs already present in dbSNP. Most of the remaining variants were present in both Case and Control groups and, additionally, none was shown to be statistically associated with PD. Lautier et al. presented a list of 10 mutations that their results suggested to be pathogenic, and an additional mutation present in one Control individual. Although we failed to find all these mutations in our combined series, we did find three of them: p.L1230_Q1237del (described as Del LPQQQQQQ 1209–1216 by Lautier et al.), p.N478T (first described as Asn457Thr) and p.H1992R (described as His1171Arg). Although p.L1230_Q1237del was found in a similar number of

Cases and Controls in both populations (P 1/4 0.37 and 0.84 in the Portuguese and US series respectively), p.N487T was only found in one Portuguese Control sample, and p.H1192R, identified by Lautier et al. in one Control individual only; it was present in both Cases and Controls in our US series, suggesting it to be a benign variant.

Interestingly we have found 37 new single nucleotide variants in our cohorts of which 26 are non-synonymous, and five deletion or insertion mutations in both PD and Control groups (Supplementary Tables 1 and 2); this high number is probably reflective of the repeat rich nature of exons 26 and 29. In order to test if there was an enrichment of rare non-synonymous mutations in Cases when compared with Controls, we compared the collective frequency of nonsynonymous alterations that were identified only in Cases, with the collective frequency of non-synonymous changes found only in Controls; a 2x2 Fischer exact test of association showed no statistically significant difference.

In comparison with the report of Lautier et al., some differences should be noted. The first is the ethnicity of the cohorts studied. Although Lautier et al. screened samples originating in Italy and France, our study included samples from Portugal and the United States. It is possible, although we believe unlikely, that differences in the genetic background between these cohorts, result in disparities in the pathogenicity of the variants. A second difference to the previous study is the sample selection criteria. Although Lautier et al. selected samples with positive family history, we included samples with and without family history representative of PD of their respective populations. Although the number of samples with family history may be smaller in our

combined cohort, the effect of pathogenic mutations in GIGYF2 in these samples would be evident.

Another possibility is that the healthy individuals harboring mutations may in future convert to disease; the fact that we used age-matched Controls and that there is no enrichment of these mutations in the Case group support our supposition that this is also unlikely.

The literature is clearly scarce in results relating genetic variability of GIGYF2 with PD. We present the first follow-up study to the results published by Lautier et al. Although we cannot rule out a small genetic contribution of GIGYF2 to PD, our data seem to point in the direction that the pathogenic variants previously published are rare polymorphisms. We support this statement based on the fact that two of such mutations were found in our Control groups and that SNPs across GIGYF2 did not show any association with PD; we cannot of course unequivocally rule out the other previously identified mutations from having a role in disease.

The previous study used a rather small Control group to verify the presence of the variants (n 1/4 227), and thus rare variants may have been missed. We believe this is a critical and important finding; this gene has already been assigned a PARK designation (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=607688) and we feel, given the evidence that we present here, that such a designation may be premature or misleading.

CHAPTER 5

GLUCOCEREBROSIDASE AND PD

Based on the following studies:

VIII) Bras J, Paisan-Ruiz C, Guerreiro R, Ribeiro MH, Morgadinho A, Januario C, Sidransky E, Oliveira C, Singleton A. **Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal**. Neurobiol Aging. 2009 Sep;30(9):1515-7. Epub 2007 Dec 21.

IX) Neumann J, Bras J, Deas E, O'Sullivan SS, Parkkinen L, Lachmann RH, Li A, Holton J, Guerreiro R, Paudel R, Segarane B, Singleton A, Lees A, Hardy J, Houlden H, Revesz T, Wood NW. **Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease**. Brain. 2009 Jul;132(Pt 7):1783-94. Epub 2009 Mar 13.

X) Sidransky E, Aasly J, Aharon-Peretz J, Annesi G, Barbosa ER, Bar-Shira A, Berg D, Bras J, Brice A, Chen C-M, Clark LN, Condroyer C, Marco EV, Dürr A, Eblan MJ, Fahn S, Farrer M, Fung H-C, Gan-Or Z, Gasser T, Gershoni-Baruch R, Giladi N, Griffith A, Gurevich T, Januario C, Kropp P, Lang A, Lee-Chen G-J, Lesage S, Marder K, Mata I, Mirelman A, Mitsui J, Mizuta I, Nalls MA, Nicoletti G, Oliveira C, Ottman R, Orr-Urtreger A, Pereira L, Quattrone A, Rogaeva E, Rolfs E, Rosenbaum H, Rozenberg R, Samii A, Samaddar T, Schulte C, Sharma M, Singleton A, Spitz M, Tan EK, Tayebi N, Toda T, Troiano A, Tsuji T, Wittstock M, Wolfsberg T, Wu Y-R, Zabetian C, Zhao Y, Ziegler S*. **International multi-center analysis of glucocerebrosidase mutations in Parkinson disease.** NEJM. In press

Glucocerebrosidase and PD

This chapter includes Studies VIII, IX and X and deals with the role of variants in the *GBA* gene in PD. We have performed two case-control association studies in two different populations: Portuguese population (Study VIII) and British population (Study IX). Additionally, in Study X we have collaborated with Dr. Ellen Sidransky at NIMH in performing a meta-analysis that included sixteen worldwide centers that have been ascertaining the frequency of *GBA* mutations in PD.

Gaucher's disease (GD) is the most common lysosomal storage disorder and results from the deficiency of the lysosomal enzyme glucocerebrosidase (GBA). It is caused by mutations in the gene coding for GBA and follows an autosomal recessive mode of inheritance. GBA deficiency leads to the accumulation of its substrate, glucosylceramide, within the lysosomes of a variety of cell types, including neurons and macrophages [155]. Clinically, Gaucher's disease is highly variable and the spectrum of disease correlates, at least in part, with residual enzyme activity [155]. In its most severe, infantile-onset form which has traditionally been termed type 2 disease, there is glucosylceramide accumulation in a variety of cell types, including neurons, which leads to rapidly fatal neurodegenerative disease. In contrast, in late-onset, type 1 disease, there is enough residual enzyme activity to prevent storage in all cell types except macrophages which are exposed to an exceptionally high glycosphingolipid load due to their role in phagocytosis of effete blood cells. These lipid-laden macrophages, termed Gaucher cells,

infiltrate the liver, spleen and bone marrow, and patients can present with organomegaly, hypersplenism and, in its most severe form, bone infarction [155]. Recently, it has become clear that these subtypes are part of a spectrum of disease. In particular, there appears to be a greater range of neurological involvement than previously recognized and a variety of neurological deficits have been described in patients who have what has classically been thought of as type 1, non-neuronopathic disease [156, 157]. The human GBA gene (MIM# 606463) is located on chromosome 1q21 in a gene rich area. GBA comprises 11 exons and 10 introns, spanning 7.6 kb of sequence. A non-processed pseudogene (GBAP) which shares 96% exonic sequence homology is located 16 kb downstream of the functional GBA gene [158]. The presence of this highly homologous pseudogene along with another six genes at the locus increases the occurrence of chromosomal rearrangements and misalignments in this region. These processes provide an explanation for the high number of complex recombinant alleles that have been detected in Gaucher's disease [43].

Although the GBA genotype plays a role in determining the type of Gaucher's disease, there is still enormous clinical variation between patients who have the same genotype, including twins, and genotype–phenotype correlations are difficult to make [157, 159]. Surveys in the Ashkenazi Jewish population suggest that up to 60% of patients homozygous for the common N370S mutation may never present clinically [160]. There is consensus, however, that heterozygosity for this relatively 'mild' allele protects the individual from neuronopathic involvement. In contrast, homozygosity for the L444P allele is invariably associated with brain involvement [161].

Parkinsonism is one of the neurological symptoms described in Gaucher's disease and affected individuals exhibit classical symptoms, including tremor, rigidity and bradykinesia [44, 162, 163]. A relatively common finding is the early age of disease onset (Age of onset ≤50 years) of parkinsonian symptoms in Gaucher's disease and the presence of cognitive symptoms, such as dementia [164]. Pathological evaluation of brains from Gaucher patients revealed Parkinsonian like features, including alpha-synuclein immunoreactive cortical and brainstem-type Lewy bodies [165]. An increased frequency of parkinsonism was noted amongst otherwise healthy relatives of Gaucher patients [166]. Further analysis of these Gaucher relatives revealed a possible association between heterozygous changes in GBA and Parkinsonism. These findings led to the hypothesis that even heterozygous mutations in GBA might constitute a genetic risk factor for the development of Parkinsonism.

Subsequent GBA genotyping studies on various cohorts of Parkinson's disease patients showed an increased frequency of GBA mutations. Whilst most studies focused on the more common pathogenic mutations in GBA, such as N370S and L444P [46-48, 82, 167], some smaller studies also performed a complete sequencing screen of the entire GBA gene [45, 51, 168, 169]. Carrier frequencies for GBA mutations differed between 10% and 31% in the Ashkenazi Jewish Parkinson's disease population, and 2.9% and 12% in Parkinson's disease cohorts of non-Ashkenazi-Jewish origin, such as North American (with European background), Taiwanese, and Italian. The lowest carrier frequency was reported to be 2.3% in Norwegian Parkinson's disease patients as compared to 1.7% in controls [50].

Materials and Methods

Methods for Study VIII

In this Study we report the results of complete sequence analysis of GBA in a series of 230 Portuguese patients with PD, collected sequentially at the Coimbra University Hospital in Portugal and in 430 healthy age-matched controls. All subjects were Caucasian and of apparent Portuguese ancestry. Diagnosis was made in accordance with the UK Brain Bank criteria. Family history was considered positive if at least one family member presented with PD. The entire coding region and the exon/intron boundaries of GBA gene were sequenced in all subjects, using previously published conditions[168]. To exclude false positives due to co-amplification of the pseudogene, all mutations were confirmed by sequencing bands excised from an agarose gel, following PCR amplification of a fresh DNA aliquot.

Methods for Study IX

In this study, we explored the association between mutations in GBA and Parkinson's disease by performing a sequencing screen of 790 British patients with the disease and 257 age and ethnicity matched controls. Our study examined all GBA exons and flanking intronic sequences where possible and, to our knowledge, represents the largest study to date on a non-Ashkenazi-Jewish Parkinson's disease patient sample in which an extensive

review of clinical data on all GBA mutation carriers and a pathological evaluation of GBA carriers was performed. The aim of this study was to more accurately define the role of GBA mutations in a non-Ashkenazi-Jewish population and to provide detailed phenotype and neuropathological correlation data to clarify the clinical parameters.

Genomic DNA samples from 790 patients diagnosed with Parkinson's disease were screened for mutations in the GBA gene. A total of 380 cases had been diagnosed with pathologically proven Parkinson's disease and were procured from the Queen Square Brain Bank at the Institute of Neurology, UCL (346 cases) or from the Parkinson's Disease Society Tissue Bank, Imperial College London, UK (34 cases). 410 cases were from a series collected by the Department of Molecular Neuroscience at the Institute of Neurology, UCL. All subjects met the UK Brain Bank Clinical Criteria for Parkinson's disease [5]. The mean age of disease onset was 58.7±12.3 years. The male-to- female ratio in this series was 3.5:1. Among the 790 Parkinson's disease samples, 83 were associated with familial Parkinsonism, whereas 707 samples were diagnosed with sporadic Parkinson's disease and showed no pattern of Mendelian inheritance. Familial Parkinson's disease was defined as showing a positive family history compatible with the diagnosis of Parkinsonism in at least one first or second degree relative. Control DNA was procured from three different sources. The first control DNA set, control group 1, consisted of 115 samples of the European Collection of Cell Cultures (ECACC), the Human Random Control (HRC) panels 1 and 2. DNA was extracted from lymphoblastoid cell lines generated from peripheral blood lymphocytes of healthy donors. The male-to-female-ratio in this set was 1:1 and the mean

age at donation was 38 years. Another subset of 73 DNA controls, control group 2, was extracted from brain tissue that was derived from the Newcastle Brain Tissue Resource. The male-to-female-ratio in this control group was approximately 3:2 and the mean age at death was 57 years. The control DNAs of group 1 and 2 were kindly provided by Dr Rohan de Silva, Reta Lila Weston Institute, UCL. A last group of 69 DNA control samples, control group 3, derived from brain samples or blood samples, was provided by the Department of Molecular Neuroscience at the Institute of Neurology. The mean age at death or at sample collection was 72 years. No information was available about the gender distribution in this last control group. All three control groups were of UK Caucasian origin and no individual reported an Ashkenazi Jewish background. The DNA samples were obtained according to ethical guidelines and all donor individuals gave written consent.

For amplification of the GBA gene, three different PCR reactions were performed as described previously[170]. In order to avoid amplification of the pseudogene, primer sequences were designed to DNA regions exclusively found within the GBA gene. Three distinct fragments were amplified spanning all exonic and most intronic sequences of GBA. As an internal control, the size of the PCR products resulting from amplification of the pseudogene for these three fragments was calculated and confirmed as being of an alternative size to those amplified from GBA. Different PCR conditions were set up to optimize the annealing temperature and extension time for each fragment. The following reagents were used for the PCR in a total reaction volume of 15 ul: 7.5ul fast start PCR master mix (Roche), 1ul of 10uM forward

primer, 4.5ul deionised water and 1ul genomic DNA template (50 ng/ml). All PCR products were run on a 1% agarose gel with ethidium bromide and size checked to rule out amplification of the GBA pseudogene. Cycle sequencing was performed for each exon and the flanking intronic sequences using the Dye Terminator Sequencing Kit (Applied Biosystems) and run on an ABI 3700xl genetic analyzer (Applied Biosystems). Reactions were conducted as described previously[170]. However, for some exons, sequencing with the mentioned primers did not result in a sequencing read over the entire exon. Therefore, an alternative set of sequencing primers was designed. All identified mutations were confirmed by re-amplification of the individual patient DNA and sequenced both in the forward and the reverse direction. Sequence chromatograms were analysed using the Sequencher software (Genecodes) and a cDNA reference sequence for GBA was taken from GenBank (NM_001005749). All exons and the flanking intronic regions were analysed when clean, complete sequence reads were obtained. This approach allowed us to take into account all successful sequencing reads for each exon rather than excluding data when complete GBA gene reads could not be obtained for individual patients. The overall number of mutations found was then used to calculate the carrier frequencies. To evaluate the degree of conservation of amino acids, which were altered due to novel missense mutations, an online version of the ClustalW2 software was used. Protein sequences for glucosylceramidase were obtained from the UniProt (www.uniprot.org) and Ensembl (www.ensembl.org/index.html) database. The amino acid sequences of eight different species were compared: Homo sapiens (human), Pan troglodytes (chimpanzee), Pongo abelii (sumatran

orangutan), Sus scrofa (pig), Bos Taurus (cow), Mus musculus (mouse), Rattus norvegicus (rat), Drosophila melanogaster (fruit fly), Caenorhabditis elegans (worm) and Danio rerio (zebrafish).

Clinical notes of all GBA mutation carriers were reviewed independently by two experienced neurologists. The data were analysed with the main focus on age of disease onset, age of death (in the case of brain derived samples), sex, levodopa (L-Dopa) responsiveness, motor symptoms and non-motor symptoms, especially the presence of cognitive impairment, visual hallucinations and depression. The following criteria were applied for the assessment of L-Dopa responsiveness: a reported improvement of at least 30% after first introduction of L-Dopa was regarded as being a positive response. The degree of improvement was based on the clinical impression documented by the treating clinician, with specific changes in formal rating scales of Parkinsonism, such as the Unified Parkinson's Disease Rating Scale. Visual hallucinations which were considered as side effects of L-Dopa or dopamine agonist therapy and which resolved after changing medications were not counted. Similarly, hallucinations which occurred in the context of febrile illnesses and delirium were not taken into account.

Seventeen Parkinson's disease brains from GBA mutation carriers and from 16 sporadic Parkinson's disease control brains without GBA mutations matched for age at onset, disease duration and gender had been fixed in 10% buffered formalin and dissected according to the standardized protocol used in the Queen Square Brain Bank for Neurological Disorders. Brain samples from selected regions were embedded in paraffin, cut into 8 mm thick tissue sections and were deparaffinized and rehydrated according to established

procedures. For alpha-synuclein (a-syn) immunohistochemistry, the sections were autoclaved in citrate buffer for 10 min and pre-treated with 98% formic acid at room temperature for 15 min. Following epitope unmasking, a monoclonal antibody to human alpha-synuclein1–140 (Novocastra, Newcastle upon Tyne, UK) was applied at a dilution of 1 : 1000 and incubated overnight at + 4C. For detection, the Histostain SP kit (Zymed, San Francisco, CA, USA) was used with Romulin AEC chromogen (Biocare Medical, Walnut Creek, CA, USA). Finally, the expression of a-syn was assessed in ten brain regions: (i) medulla with dorsal motor nucleus of vagus (DMV); (ii) pons with locus ceruleus (LC); (iii) midbrain with substantia nigra (SN); (iv) basal forebrain (BFB) including the nucleus basalis of Meynert (NBM) and amygdaloid complex (AC); (v) posterior hippocampus including the CA2 subregion at the level of the lateral geniculate body; (vi) entorhinal cortex; (vii) medial temporal gyrus; (viii) anterior cingulate gyrus; (ix) anterior frontal cortex; and (x) inferior parietal cortex. The selection of regions was based on the currently used staging and grading systems for Lewy body disorders[171, 172].

Genotype frequencies in Parkinson's disease patients and controls were compared using Fisher's exact test, statistical significance was considered to be P50.05 using a one-tailed test. To determine the odds ratio (OR) and the 95% confidence interval (95% CI) an online calculator was used (DJR Hutchon Calculator; http://www.hutchon.net/ConfidOR.htm). The statistical differences in Braak staging and McKeith grading were estimated by Fisher's exact test. The differences in Lewy body scores between GBA carriers and sporadic Parkinson's disease controls were estimated using the non-

parametric Mann–Whitney U-test. For the statistical analyses, SPSS (version 14.0) for Windows (SPSS Inc., Chicago, IL, USA) was used.

Methods for Study X

Patients and procedures

Researchers known to be genotyping Parkinson disease cohorts for *GBA* mutations were solicited for this collaboration. Sixteen centers participated and data were collected and analyzed at the National Human Genome Research Institute (NGHRI) Bethesda, Maryland (Supplementary Table 3). The centers included subjects from North America (four groups), South America (one group), Asia (three groups), Israel (two groups) and Europe (six groups). Ethnicity was by self-report. Ashkenazi Jews provided the origin of grandparents. Informed consent was obtained under the supervision of each local ethics committee. All subjects fulfilled the UK Parkinson Disease Brain Bank Clinical Diagnostic Criteria for Parkinson disease[173].

As the detection methods and number of mutations that could be identified varied greatly from center to center, 12 standard DNA samples from patients with Gaucher disease were genotyped by each center and the results were analyzed at the NHGRI. All participating groups could reliably detect mutations N370S and L444P, unless the mutant allele included large stretches of *GBA* pseudogene sequence. Four centers could identify 4-9 specific selected point mutations (Supplementary Table 3). In addition, five participating centers sequenced all exons of *GBA*, and in a sixth, a subgroup

was sequenced. Thus, results evaluating two mutations, 4-9 mutations, and the entire coding sequence were analyzed separately. Two frequent *GBA* variants, E326K and T369M, were evaluated, but not included as mutant alleles.

The extent of clinical data collected from each site varied (Supplementary Table 3). Some study centers provided only age, family history, sex, ethnicity and diagnosis, while others reported more complete data, including presenting symptoms, age at disease onset, specific clinical manifestations, response to medications and standardized Hoehn and Yahr and/or Unified Parkinson's Disease Rating Scale scores. Only one proband was collected per family and subjects with diagnoses other than Parkinson disease were excluded.

Controls were screened for signs or symptoms of parkinsonism and centers attempted to match for age, sex and ethnicity. Controls with a family history of Parkinson disease were removed. For regression modeling 266 patients and 261 controls were excluded because age, sex or ethnicity data were incomplete.

Study Design

The a priori aim in this study was to conduct a combined analysis of risk associated with *GBA* mutations from different centers including both published and unpublished data. Analysis of the global risk of Parkinson disease associated with *GBA* mutations was approached employing the fixed-effects

Mantel-Haenszel test to combine data across studies. This methodology is similar to that used in meta-analyses, although since patient level data from both published and unpublished/in-press studies were provided, this is not a true meta-analysis.

The study was designed to assess risk of Parkinson disease associated with *GBA* mutations not only in all available genotyped samples, but also in distinct subgroups of the total population. Due to the variety of recruitment practices, available data, and laboratory capabilities of the multiple study centers involved, we opted to partition our analysis into pre-specified subgroups by study center, as well as to examine the effects of two common mutations, N370S and L444P. In addition, stratified analyses of subgroups based on Ashkenazi status and sequencing-depth were pre-specified in our study design. Post-hoc subgroups were defined as studies reporting recruitment of family-based case and control sets (the study of Norwegian families[50] and the Japanese study[174]) as a means of testing the effect of familial recruitment on the homogeneity of effects in the Mantel-Haenszel model for risk associated with mutations in *GBA*.

Statistical Analyses

Descriptive statistics were calculated and stratified by study center. Mutation frequencies for Ashkenazi and non-Ashkenazi cases and controls and the total number of mutations detected, as well as the mean age at sampling,

male to female ratios, and level of sequencing depth were summarized for each center based on the data provided by study centers.

Available data on nine clinical features of Parkinson disease (asymmetric onset, bradykinesia, dementia, dyskinesia, family history of Parkinson disease, orthostatic hypotension, postural instability, rest tremor and rigidity) were analyzed. Means were compared using two-tailed Student's T-tests and frequency differences between patients with and without *GBA* mutations were assessed using chi-squared tests.

Multivariate logistic regression models were used to ascertain the odds of developing Parkinson disease in varied populations with *GBA* mutations. An initial series of models was constructed comparing the presence or absence of any mutation in *GBA* as the primary predictor of Parkinson disease. These models were adjusted for gender, age at sampling and ethnicity, and stratified by the specificity of sequencing coverage in the *GBA* region. In these models, ethnicity and site were considered collinear. Primary predictors of mutations N370S and L444P were then assessed in subsequent iterations of these models to evaluate risk. Logistic regressions included all samples with complete outcome, predictor and covariate data; missing data was the only exclusion criterion.

A second series of multivariate logistic regression models was used to compare associations of clinical and demographic features with genetic factors in Ashkenazi and non-Ashkenazi cases. Models were stratified by

level of genomic coverage as described previously, using covariates of gender and age at sampling. Regression models in Ashkenazi samples were adjusted for site, and in non-Ashkenazi samples were adjusted for self-reported ethnicity. Identical parameters were used for the construction of logistic regression models utilized to test association with E326K and T369M variants. Chi-square tests of heterogeneity were used to compare effect size differences between Ashkenazi and non-Ashkenazi stratified models assessing risk attributable to all *GBA* mutations, and N370S and L444P separately across all sequencing depths.

The primary aim was to summarize Parkinson disease risk due to mutations in *GBA* across cohorts. To accomplish this we conducted fixed effects Mantel-Haenszel analyses using all available cases and controls from each study center. These models estimate risk attributable to counts of mutations in cases and controls from standard contingency tables. Three separate Mantel-Haenszel analyses were performed, using any mutation, N370S and L444P respectively as the primary predictor of Parkinson disease. These three Mantel-Haenszel analyses are not completely independent of each other, and therefore, do not necessitate correction for multiple testing phenomena.

The heterogeneity of effects in the Mantel-Haenszel analyses were evaluated using Woolf's test for heterogeneity [175]. Possible analyses of interactions contributing to heterogeneity of odds ratios (ORs) were limited by the data available for analysis, so several additional Mantel-Haenszel analyses were carried out. First, the centers that utilized family-based recruitment (Japan

and Norway) were excluded, based on both the lack of additional data sufficient for testing of interactions consistently across all subgroups, and the assumption that general genetic homogeneity among the Norwegian and Japanese samples was compounded by active recruitment of family members of cases that could influence the independence (or in the case of the Norwegian cohort, the non-independence) of mutation frequency differences between cases and controls. Mantel-Haenszel analyses excluding either the data from Japan or from Norway independently were also performed. In addition, to confirm that the most robust OR (with respect to the standard error of the estimate) in the Mantel-Haenszel analysis (the Tel Aviv center) was not inflating the combined OR, an additional Mantel-Haenszel analysis was carried out omitting data from this center.

All data analyses were conducted using R 2.8.0[176]. Source code for plotting of meta-analysis results is available in the package *r.meta* maintained by Thomas Lumley (available from http://cran.cnr.berkeley.edu/).

Results

Results for Study VIII

The PD group yielded 14 carriers of previously described pathogenic GBA mutations (N370S, N396T, D409H and L444P), all heterozygous, while the control group yielded 3 N370S carriers, also heterozygotes (Table 9).

Table 9: As described in the text, all variants were present in heterozygous state.

Mutations		***PD Patients***			***Controls***		
Protein[a]	*Allele name*[b]	*N*	*% cases*	*% carriers*	*N*	*% controls*	*% carriers*
N435T	N396T	5	2.2	35.7	0	-	-
L483P	L444P	3	1.3	21.4	0	-	-
N409S	N370S	5	2.2	35.7	3	0.7	100
D448H	D409H	1	0.4	7.1	0	-	-
Polymorphic variants							
K13R[c]	K(-27)R[c]	1	0.4	20	0	-	-
E365K	E326K	2	0.9	40	3	0.7	27.3
T408M	T369M	2	0.9	40	5	1.2	45.5
R41L[c]	R2L[c]	0	-	-	1	0.2	9.1
E427K[c]	E388K[c]	0	-	-	2	0.5	18.2

[a] Amino acid designations are based on the primary GBA translation product, including the 39-residue signal peptide.
[b] Common nomenclature attributed to mutations; does not include the 39-residue signal peptide.
[c] These represent previously unpublished mutations, therefore pathogenicity or functional effects are unknown.

Two variants, E326K and T369M, previously described as non-pathogenic polymorphisms, were identified in both patients and controls. In addition, 2 novel variants and one previously described variant of unknown significance were identified (p.K13R; p.R41L and p.E427K)[177].

Results for Study IX

In this study, a total number of 33 mutations were found in 790 screened Parkinson's disease patient samples (4.18%) as compared to three sequence changes in 257 controls (1.17%) (Supplementary Table 3).

The frequency of GBA mutations detected in the patients is statistically significantly higher than the frequency observed in age and ethnicity matched

controls (P = 0.01; OR= 3.7; 95% CI = 1.12–12.14). Due to technical difficulties, clear sequencing reads could not be obtained for all exons. We therefore decided to analyse each exon and the flanking intronic region separately in an exon-by-exon approach. This approach allowed us to determine the maximum number of mutations for all successfully sequenced exons of GBA in our 790 patients. The sequencing changes included 30 missense mutations, one deletion and two complex alleles resulting from recombination events with the GBA pseudogene (GBAP). Out of the 33 mutations observed in Parkinson's disease patients, 11 individuals were found to be heterozygous for L444P (Carrier frequency 1.39%) a mutation which in homozygous carriers is unequivocally associated with the neuronopathic type 3 of Gaucher's disease. In addition, eight heterozygous carriers of the N370S allele could be identified (Carrier frequency 1.01%). In patients of non-Ashkenazi-Jewish origin, these two mutations (N370S and L444P) represent the most frequent changes in GBA. Three individuals were carriers of the complex alleles RecNciI (Carrier frequency 0.25%) and RecA456P (Carrier frequency 0.13%), respectively. These alleles include the non-synonymous changes L444P and A495P and are reported to result from a recombination between GBA and GBAP (Latham et al., 1990; Hatton et al., 1997). Therefore, whilst these three individuals carry the L444P mutations, they have not been counted as L444P exclusive carriers. The third most common change in sequence was R463C (carrier frequency 0.38%). All allele names used in this report follow the common nomenclature and refer to the processed protein, not including the 39-residue signal peptide.

One individual carried a 55 bp deletion in exon 9 (c.1263–1317 del55). This 55 bp deletion, along with additional DNA base changes, is present in exon 9 of the pseudogene. Therefore, the presence of this deletion suggests that a gene conversion or another recombination event between the functional gene and the pseudogene must have occurred. In support of this, no other DNA alterations normally present within the pseudogene sequence were identified in exon 9 in this individual confirming that the presence of this deletion was not due to an accidental amplificiation of the pseudogene. The c.1263–317 del55 results in a non-functional gene and has been associated with severe clinical manifestations of Gaucher's disease (Beutler et al., 1993b). Two more previously undescribed point mutations were found in exons 10 and 6 – D443N and G193E (carrier frequencies 0.13%, respectively), both resulting in amino acid changes (p.Asp482Asn and p.Gly232Glu). Whilst most of the mutations identified were clustered in the region spanning exons 9 and 10, we discovered another novel change in exon 3 resulting in the amino acid change p.Lys46Glu (K7E) (carrier frequency 0.13%). An interspecies comparison of the amino acids affected by these novel mutations revealed that K7E and D443N are highly conserved in most mammalian species, but not in rat, zebrafish, C. elegans and D. melanogaster (data not shown). Interestingly, G193E is conserved in all species screened except for zebrafish, indicating that this amino acid is particularly well conserved during evolution. These findings suggest that the three novel GBA mutations not only cause an alteration in the amino acid sequence but also are likely to be pathogenic mutations. However, the precise functional effects of these novel mutations remain to be investigated. In the control groups, three individuals (1.2%) were

heterozygous for the following changes: N370S (control group 1), R257Q (control group 2), and a previously unpublished alteration V458L (control group 1).

Of the 33 GBA mutation carriers within the patient group, four (12%) have been diagnosed with familial Parkinsonism as compared to 29 patients (88%) with sporadic Parkinson's disease (Supplementary Table 4). Therefore, the prevalence of GBA mutations in British patients diagnosed with sporadic Parkinson's disease can be estimated at ~3.7% (29/790). No clinical data was available for patients 27 and 31, and therefore these individuals were not included in the clinical data analyses. The mean ±ST age of onset (AoO) of all GBA mutation carriers in the Parkinson's disease group was 52.7 ± 11.3 years. Twelve patients had an AoO ≤ 50 years (38.71%) which represents the cut off value for early-onset PD. The mean AoO of the 790 PD patients in this study was 58.7 ± 12.3 years, which is statistically significantly higher than in the *GBA* mutant group (t-test for equality of means: t = 2.658; p = 0.008). Comparing these results to previous studies our findings confirm that mutations in *GBA* are associated with an early onset of PD. The male-to-female ratio of *GBA* carriers within the PD group was 5.2 (26 male: 5 female), which is considerably higher than the overall male-to-female ratio of 3.5 in the total study group (Pearson's Chi-Square test: 5.12; p = 0.024). 28 out of 31 (90.32%) PD patients who carried a *GBA* mutation were initially responsive to L-Dopa treatment. Patients 28 and 29 did not respond to L-Dopa therapy and patient 8 showed a minimal response to L-Dopa. Notably, patient 9 was initially responsive but became unresponsive to L-Dopa treatment over the course of 5 years. Fifteen out of the 31 (48.39%) PD patients with *GBA*

mutations developed symptoms of cognitive decline during the course of the disease. Patients 1, 13, 16, 21, 22, 23, 24, 25, 26, and 30 were diagnosed with PD and dementia (PDD) or probable dementia, whereas patients 4, 7, 10, 12, and 28 had not been formally given the diagnosis of dementia but showed clear symptoms of cognitive degeneration (e.g. memory loss, cognitive slowing, confusion). None of these fifteen cases had a reported onset of cognitive symptoms prior to or within the first year after diagnosis of PD, thus no patient fulfilled the formal criteria for Dementia with Lewy Bodies (DLB). Information about the mean disease duration was available for 12 of these cognitively impaired PD patients and was 11.7 ± 5 years. Interestingly, 40% (6/15) of the patients with cognitive symptoms and/or dementia had an AoO ≤ 50 years.

In this study, we also evaluated the presence of visual hallucinations in PD patients with *GBA* mutations. Visual hallucinations (VH) were present in 45.16% (14/31) of patients with PD, of which 44.86% (6/14) had an AoO ≤ 50 years. None of the GBA mutants had an occurrence of VH prior to or concurrent with the onset of PD motor symptoms. The minimum interval to developing VH was 42 months and the average interval was 125 months after motor symptom onset. To conclude, we can summarize that the clinical features of PD patients with *GBA* mutations comprise an early age of disease onset (AoO ≤ 50 years) and a good responsiveness to L-Dopa treatment. Symptoms of cognitive decline and/or dementia were a common finding and non-treatment associated hallucinations were present frequently in almost 45% of the cases.

All pathologically examined cases with *GBA* mutations (*N* = 17) showed morphological changes, which were within the spectrum of classical (idiopathic) PD and were not considered to represent a form of atypical PD. Braak staging, which is used to establish the topographical extent of alpha-syn-immunopositive inclusions (Lewy bodies and neurites), revealed that, in addition to involvement of subcortical structures, cortical areas were also affected by alpha-syn-immunoreactive inclusions corresponding to Braak stages 5-6 in all 17 patients (Table 3). There was no statistically significant difference in Braak stages between the *GBA* carriers and sporadic PD controls ($p = 0.537$, Fisher's exact test). However, 13 of the 17 *GBA* carriers (76%) and 6 of the 16 PD controls (38%) fulfilled the McKeith criteria for diffuse neocortical Lewy body pathology. This shows a positive trend for a higher McKeith grade among the *GBA* mutation carriers, as the difference between the two groups just reached statistical significance ($p = 0.049$, Fisher's exact test). LB scores generated by the McKeith protocol were used to give an indicative of the overall cortical burden and did not differ between the two groups; *GBA* carriers 7.3 ± 3.0 (mean ± ST), PD controls 6.3 ± 2.8 ($p > 0.5$, Mann-Whitney U test).

Results form Study X

A total of 5691 genotyped patients with Parkinson disease were evaluated, including 780 Ashkenazi Jewish subjects and 4911 patients with no known Ashkenazi ancestry. The 4898 controls genotyped included 387 Ashkenazi

Jewish individuals and 4511 with other ethnicities. Table 1 lists the frequency of mutations and demographics for each individual study.

In Figure 5, the odds ratios, standard errors and confidence limits for each independent study are shown graphically, where the precision of the effect estimate is reflected in the size of the squares, then combining odds across study sites. Panel A summarizes the results using any mutation as a predictor. Each center had an over-representation of mutations among patients as compared to controls, with an OR above 1, although confidence intervals varied considerably. Eight centers had an OR greater than 5. Because one center did not provide individual controls (Haifa, Israel) and three centers did not find mutations among their controls (Brazil, Singapore and Tubingen, Germany), they do not appear in the forest plots. The overall combined OR denoted by the diamond symbol demonstrates how greatly the confidence interval is reduced when the individual studies are combined. Panels B and C show the individual ORs for *GBA* mutations L444P and N370S respectively. While the results are overwhelmingly positive for each mutation, the ORs in the individual studies were higher for L444P.

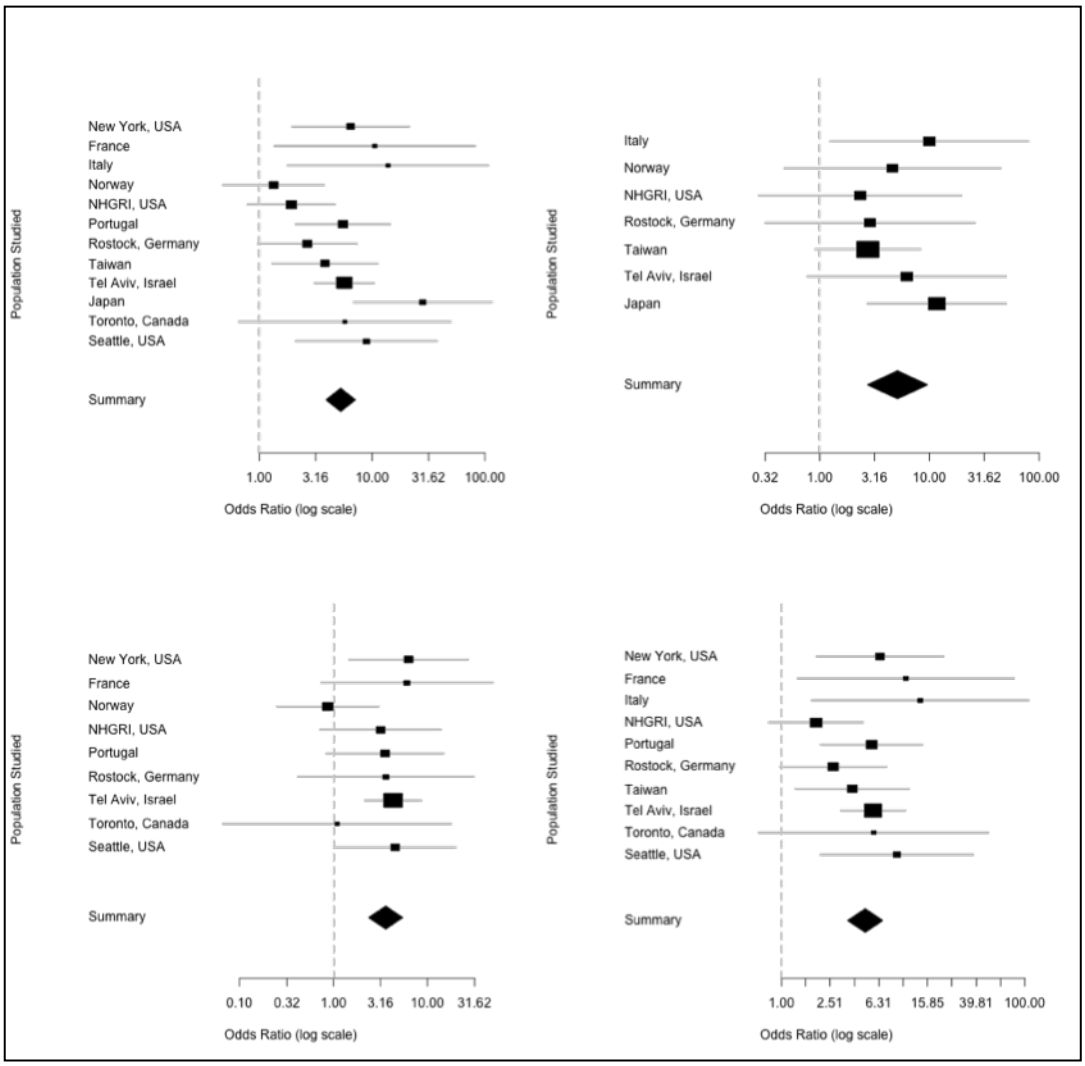


Figure 5: Odds ratios from meta-analyses. Panels correspond to combined risk estimates associated with possessing (A) any *GBA* mutation, (B) mutation L444P and (C) mutation N370S (D) any *GBA* mutation excluding Japan and Norway study centers. Horizontal grey lines indicate 95% confidence intervals of estimates. Point estimates per study population are indicated by squares where height is inversely proportional to the standard error of the estimates. Diamonds represent the summary odds ratio whose width indicates the 95% confidence intervals.

Combined ORs for Mantel-Haenszel analyses were homogenous for the risk attributable to L444P and N370S (Woolf's test for heterogeneity, p-values of 0.347 and 0.4988 respectively). This allows for confidence in reporting the Mantel-Haenszel combined ORs attributable to N370S (OR = 3.96, 95% CI 2.6-6.02) and L444P (OR = 6.73, 95% CI 4.5-15.42). However, the Mantel-Haenszel OR in the model for risk attributable to all *GBA* mutations was significantly heterogeneous (Woolf's test for heterogeneity, p-value = 0.0207).

After excluding the two study centers that actively recruited family members, Tokyo and Norway, the OR was slightly attenuated (OR = 5.43, 95% CI 3.89-7.57), but heterogeneity was not longer significant (Woolf's test for heterogeneity, p-value = 0.414). To ensure that the highest individual OR in this follow-up model was not inflating the Mantel-Haenszel OR, we carried out a subsequent analysis excluding data from Tel Aviv, Japan and Norway. For the remaining nine study centers, the homogenous (Woolf's test for heterogeneity, p-value = 0.3386) OR only decreased slightly (OR = 5.34, 95% CI 3.59-7.94). Independent exclusions of data from either Norway or Japan showed homogenous ORs (Japan excluded: OR = 4.91, 95% CI 3.6-6.7, Woolf's test p-value = 0.1447; Norway excluded: OR = 6.35, 95% CI 4.6-8.75, Woolf's test p-value = 0.0991) (Panel 1D). Thus, we can confidently report a Mantel-Haenszel OR of 5.43 associated with *GBA* mutations even when these centers are excluded.

Overall, when screening solely for N370S and L444P, one of these two mutations were found in 15.3% of Ashkenazi Jewish patients versus 3.4% of controls, and in 3.2% of non-Ashkenazi patients versus 0.6% of controls. The frequency of these mutations differed greatly between studies. As expected, N370S was particularly prevalent among Ashkenazi Jews and was not seen among any of the Asian patients or controls. For N370S, the OR for Ashkenazi subjects was 5.6 (95%CL 3.04-10.39) and 3.3 in non-Ashkenazim (95%CL 1.79-6.10), with p-values below 0.001 for both groups (Table 2). These ORs differed significantly between Ashkenazi and non-Ashkenazi subjects for L444P and N370S risk across all sequencing levels (p-values <

0.01), although interaction analyses in combined multivariate models utilizing an identical covariate set showed the interactions between Ashkenazi-status and either mutation to be non-significant.

All Ashkenazi Jewish subjects were screened for the presence or absence of 6-8 different *GBA* mutations. Including this larger number of screened mutations increased the OR for any mutation to 6.48 (95%CL 3.78-11.09). The distribution of specific *GBA* mutations among Ashkenazi Jewish patients with Parkinson disease, where ~20% of patients carried a *GBA* mutation shows that 20% of the identified mutant alleles were not L444P or N370S.

Among non-Ashkenazi Jewish subjects the entire *GBA* coding region was sequenced in 1682 patients and 609 controls. In patients where full sequencing was performed, the OR for any *GBA* mutation was 6.51 (95%CL 3.62-11.7). The mutations identified indicate that as many as 46% of mutant alleles could be missed when focusing solely on N370S and L444P. Moreover, 22 of 43 subjects with L444P carried other pseudogene sequence and hence had recombinant alleles.

Full sequencing data demonstrated that two *GBA* variants, E326K and T369M which are not pathogenic in subjects with Gaucher disease[178], were common in both white patients and controls. Neither variant demonstrated a significant association with Parkinson disease.

Seventeen patients (15 Ashkenazi) carried two *GBA* mutations. Genotype N370S/N370S was observed in fourteen, N370S/ R496M in two and N370S/V394L in one.

Age at onset was provided for almost all subjects, and was found to be significantly lower among subjects with *GBA* mutations (p-value <0.001), with a mean age of 54.9 in subjects with *GBA* mutations as compared to 58.8 in subjects without. The mean length of disease duration from diagnosis to evaluation did not vary significantly, and was 7.8 years both groups.

Information about family history of parkinsonism was available on 4401 of the patients studied. 17.8% of participants without *GBA* mutations reported a relative with Parkinson disease, as compared to 24.0% of subjects with a *GBA* mutation (p-value =0.0057).

Generally, the symptom profile for the two groups (with and without GBA mutations) was similar, although mutations were associated with a significantly lower frequency of asymmetric onset (p-value<0.0001), bradykinesia (p-value=0.0001), resting tremor (p-value=0.0298), and rigidity (p-value<0.0001). There were no significant differences between orthostatic blood pressure changes (p-value=0.2591) or postural instability (p-value=0.1194) among the subjects with and without *GBA* mutations.
The presence or absence of cognitive changes was recorded for 1948 patients, reported as present in 26.3% of patients with mutations and 19.0%

without (p-value= 0.0071). Furthermore, dyskinesias were observed in 42% of mutation carriers versus 32% of those without (p-value=0.05668).

Although not a primary objective of the study, the vast amount of data collected also provided an opportunity to explore the carrier frequency of *GBA* mutations in a non- Ashkenazi population. Never before has the gene been sequenced in such a large number of individuals without Gaucher disease. Overall, among the 1609 control individuals with full *GBA* sequencing, the carrier frequency for any *GBA* mutation was found to be 0.013.

Discussion

Study VIII was performed in Portuguese samples, a cohort with a different and defined ethnicity, different than those where *GBA* mutations have been studied thus far. Here, we found a frequency of 6.1% (14/230) known pathogenic mutations in the PD series and 0.7% (3/430) in the control group. These results represent a significantly higher frequency of mutations in GBA in PD patients when compared to controls ($p < 0.001$; OR= 9.2; 95% CI 2.6–32.4). Of note, the control group shows no mutations associated with severe GD; while they exist in the PD group – 4/14 patients with L444P or D409H. If we consider the variants of unknown pathogenicity (p.K13R; p.R41L and p.E427K) as potentially causative, this association still remains ($p < 0.001$; OR= 4.9; 95% CI 1.9–12.9).

The most common mutation identified was N370S, the most frequently identified pathogenic mutation in Ashkenazi Jewish as well as Portuguese

patients with GD. Based on screening of 2000 random cord blood samples in Portugal, the carrier frequency of this mutation is estimated to be 0.0043 in this population [179]. This mutation is believed to account for 63% of the mutant alleles in Portuguese patients with GD. Although the N370S mutation was three times more frequent in the PD group when compared to controls, we did not have sufficient power to identify a statistically significant association analyzing this mutation alone (p = 0.079; OR 3.3; 95% CI 0.75–13.4). Mutation N396T, encountered in 5 subjects in this study, was first identified in Portugal and has proven to be a relatively common mutation in this population [180].

This study substantiates the need to sequence GBA in non-Ashkenazi cohorts in order to accurately determine the frequency of mutations in this gene. Had we screened only for common Gaucher mutations, we would have missed 43% of the mutant alleles in this population.

The association of GBA mutations with PD in the Portuguese population is particularly interesting when it is noted that the mutation driving this association is one associated with Jewish ancestry, and that another PD causing mutation, p.G2019S of LRRK2, underlying ~6% of Portuguese PD cases [58], is also associated with Ashkenazi Jewish ancestry; these data clearly illustrate the contribution of Jewish ancestry to the modern Portuguese population.

In summary, Study VIII demonstrates that GBA mutations are significantly more common in patients with PD than in neurologically normal controls. These findings illustrate that the identification of such an association requires large sample series, even when using populations where GBA mutations are

enriched; thus detection of such an association in populations with non-Ashkenazi ancestry is likely to require thousands of samples.

Building upon the work detailed in Study VIII, we have decided to examine a different population for frequency of *GBA* variants. We had access to a cohort of cases and controls of British ancestry, which included a very significant brain bank collection of samples. This enabled us to perform Study IX.

The frequency of GBA mutations found in the British Parkinson's disease population is clearly a striking result as it represents the highest frequency of mutations of a single gene related to the development of the idiopathic disease in this population. Although former studies on the same series of British patients showed that other genes such as PTEN induced putative kinase 1 (PINK1), leucine-rich repeat kinase 2 (LRRK2) and DJ-1 may also play a role in the sporadic form of the disease [57, 181, 182] mutations in GBA have the highest prevalence with ~3.7% of all sporadic Parkinson's disease cases being affected. Several studies have evaluated the frequency of GBA mutations among Parkinson's disease patients and show similar results. However, the majority of previous studies specifically screened the GBA gene for previously reported common mutations and did not attempt sequencing analysis of the complete gene. Whilst we were unable to obtain a complete gene sequencing read for all 790 of our patients, our compiled data from all clear exon and intronic sequence reads represents the largest sequencing study of GBA mutations in Parkinson's disease patients to date. However, one needs to take into consideration the fact that our data is based on a subset of patients who were referred to a specialized university clinic or who have donated their brains to a brain bank for research. We acknowledge

the limitations of the retrospective nature of our data and the selection bias that is expected in this series. Nevertheless, the pathological evaluation of the GBA mutant cases revealed that the observed morphological features were typical for sporadic Parkinson's disease, suggesting that a proportion of classical sporadic Parkinson's disease might indeed be caused by mutations in the GBA gene.

The mutant GBA gene frequency in the general population in the UK has been estimated at 0.0016 in contrast to 0.034 in the Ashkenazi-Jewish population [183]. Studies on control subjects from other non-Ashkenazi-Jewish populations have found very different frequencies ranging from 0.004 in a North American cohort (with a European ethnic background) to 0.017 in a Norwegian control sample [50, 82]. Thus, the observed frequency of 0.012 in our British control group is representative for a European population, and provides the best estimate for the British population to date. Regarding the clinical data on the Parkinson's disease patient group with GBA mutations, it can be summarized that, in general, our findings confirm previously published results, stating that GBA mutation carriers frequently have an earlier age of disease onset (Age of onset<50 years), show a good response to L-Dopa treatment, and have an increased likelihood to present with symptoms of cognitive decline and dementia. In addition to that, we looked at the occurrence of other non-motor features such as visual hallucinations, which have been associated with Parkinson's disease.

Symptoms of cognitive decline are a common feature in parkinsonism. In a systematic review of prevalence studies which looked at dementia in the disease, a proportion of ~24–31% has been suggested [184]. In our

Parkinson's disease patient group of GBA mutation carriers, 48% had been given the diagnosis Parkinson's disease with dementia or showed clear symptoms of cognitive decline. Moreover, 40% of the patients with cognitive symptoms had an age of onset<50 years. Therefore, we hypothesize that mutations in GBA might increase the risk of developing dementia or cognitive impairment in individuals with an early disease onset. This finding might be of importance given that patients rarely show symptoms of cognitive decline at an age younger than 55 years. Hence, in future research it will be interesting to determine whether GBA mutations have an impact on the development of dementia in younger patients with an early disease onset.

The male-to-female ratio in our series was 3.5:1 which is comparable to the published range of 3:2 [185]. In studies on Parkinson's disease patients which carry a GBA mutation, the male-to-female ratio has been reported to be higher, ranging between 2:1 to 5:2 [50, 82, 167]. In the present study, male GBA mutation carriers were by far more frequently affected than women (26 male : 5 female; Pearson's chi-squared test: 5.12; P = 0.024). Thus, the ratio observed in GBA mutation carriers suggests that male individuals who have a mutated GBA gene are more susceptible to develop Parkinson's disease than female mutation carriers.

Overall, the initial response to L-Dopa treatment was good to very good. This finding is in accordance with results from previous research, which described an excellent response to L-Dopa therapy in Parkinson's disease probands heterozygous for GBA mutations [82, 169]. However, one of the characteristics of Gaucher patients with parkinsonism is that their symptoms are mostly refractory to standard Parkinson therapy. Thus, it is possible that

identical mutations in GBA result in different phenotypic traits of Parkinson's disease (e.g. good to no response to L-Dopa treatment) and that other genetic modifiers play a role in the susceptibility to the disease.

In our subset of Parkinson's disease patients with GBA mutations, non-treatment associated visual hallucinations were present in almost half of the cases. Visual hallucinations are a common feature in Parkinson's disease and have been estimated to occur in up to 50% of patients [186]. As in idiopathic Parkinson's disease, the occurrence of visual hallucinations in patients with GBA mutations is likely to be the consequence of the extension of the Lewy body pathology in the temporal lobe [187].

These data implicating GBA mutants in Parkinson's disease pathogenesis strongly motivates an evaluation of potential pathways. There are two broad possibilities. First, haploinsufficiency of GBA leads directly to an accumulation of glucosylceramide and a concomitant impairment of ceramide metabolism and thereby increases the risk of developing the disease. The second possibility is that a novel property of the mutant enzyme is contributing to the risk of developing parkinsonism.

If one considers the neuronopathic form of Gaucher's disease it seems unlikely that the association between mutant GBA alleles and parkinsonism relates solely to the catalytic activity of the mutant enzyme although it is possible that there will be a subtle dysregulation in ceramide metabolism. In heterozygote mutant carriers, the unaffected allele would likely provide adequate GBA activity to degrade most of the glucosylceramide entering the lysosome.

If there is a novel toxic function playing a role it is of note that most of the mutations described here are missense alleles, which would be predicted to produce a protein product. A precise toxic function is unclear but for a number of these alleles, it has been demonstrated that the mutant enzyme produced is unstable and, instead of being targeted to the lysosome, is diverted by the quality control mechanisms of the cell to proteosomal degradation [188]. Lewy bodies are seen in the brains of Gaucher patients who develop Parkinson's disease and a particularly severe involvement of neuronal populations of the CA2-CA4 hippocampal subregions has been documented. Immunohistochemical studies have demonstrated that constitutive levels of GBA expression are high in these hippocampal subregions [165]. Therefore it seems likely that the expression of high levels of the unstable mutant enzyme may play a role in the formation of Lewy bodies in Gaucher patients with parkinsonism.

In the cases presented in this study, neuropathological analysis (including Braak staging and grading using consensus criteria) demonstrated extensive Lewy body pathology in a pattern identical to that seen in sporadic Parkinson's disease controls matched for age at disease onset, disease duration and gender. Furthermore a larger proportion of the cases with GBA mutations tended to have neocortical Lewy body pathology than the sporadic cases, although investigation of larger cohorts is required to confirm this. The autophagy-lysosome pathway, including chaperone-mediated autophagy and macroautophagy is an important mechanism for the degradation of cellular alpha-synuclein [189]. The findings of this study support the hypothesis that mutant glucocerebrosidase may interfere with cellular pathways related to

lysosomal degradation of cellular alpha-synuclein and that these mechanisms might also be fundamental in Lewy body formation in the sporadic form of the disease. Further research will be essential to establish whether neuronal cell death might be a consequence of a misfolded GBA enzyme or how alteration of the glucosylceramide/ceramide metabolism could contribute to the development of Parkinson's disease.

The results from this study and those from other reports in Parkinson's disease patients clearly establish that GBA mutations account for a significant minority of cases. The clinical and pathological data reported here emphasise that these cases are indistinguishable from what is normally considered as idiopathic Parkinson's disease. This has important implications for genetic counselling of such patients and indeed relatives of patients with Gaucher's disease. The classical scenario of autosomal recessive disease is that carriers are both unaffected and the recurrence risk to their offspring is incredibly low in the absence of a consanguineous relationship. These data and findings emerging from studies of proven autosomal recessive Parkinson's disease genes (e.g. parkin, DJ-1 and PINK1) in which there is some, but controversial, evidence to support a role of heterozygous mutations have changed the terrain and suggest that carrying a single heterozygous mutation is associated with increased risk. To provide accurate information to patients and their families one really requires a reliable and accurate estimate of prevalence to be made. This will be difficult but will probably require international collaboration to achieve sufficient numbers of cases. Even then given the allelic variability, which may in part, influence penetrance means that accurate predictive risk counseling will be fraught. However it is obligatory

for the clinicians who are making these genetic diagnoses that a discussion of these difficulties is conducted with the patients and their families. Our data reinforce the proposed association between GBA mutations and Parkinson's disease.

Study IX was clearly the largest single effort to fully characterize *GBA* variation in a well-defined population of PD cases and controls. While conducting this study, we have also embarked on a parallel study, which aimed at performing a meta-analysis of all the data generated regarding *GBA* variants and PD. This project managed to bring together a significant collaboration effort, between sixteen international centers working on this subject.

The results of this analysis overwhelmingly support the association between *GBA* mutations and Parkinson disease. The combined study demonstrates that this finding is not exclusive to a specific ethnicity. Furthermore, it is not associated with any specific *GBA* mutation. In fact the OR for all combined mutations was higher than for the common N370S allele alone, suggesting that alleles other than N370S might confer a greater risk, as previously proposed [167]. As expected from studies in Gaucher disease [190, 191], the distribution of mutations varied among diverse ethnicities, with N370S being prevalent among Ashkenazi Jews yet absent in Asian subjects.

The major limitation of this study was the unavoidable differences in data ascertainment among the different sites. Moreover, some sites were more successful in matching cases and controls with regard to age and gender. We attempted to account for these differences in our logistic regression models using age at sampling; self reported ethnicity and gender as covariates. In the

analysis, multivariate models included only samples with complete covariate data, and the entire data set was utilized in the Mantel-Haenszel analyses only. Data from centers with inadequate controls were only used in the stratified multivariate logistic models. To ensure that the analysis was not driven by a small subset of centers, the Woolf test for heterogeneity was used to evaluate the effect of the variability of ORs across centers. Excluding the center with the most precise estimate (Tel Aviv), and the centers with the most extreme ORs (Norway and Japan) only resulted in a slight attenuation of the combined OR, which in all analyses remained 5.4 or higher.

This study confirms the need to perform full exon sequencing to accurately ascertain the frequency of *GBA* mutations in both patients and controls. Our data demonstrate that among non-Ashkenazi cases as many as 46% of mutant alleles can be missed when screening for only two mutations. Furthermore, analysis of specific mutations may produce a serious bias. The data also demonstrate that *GBA* variants E326K and T369M do not confer a significant risk for Parkinson disease.

Focusing solely on sequenced samples, the frequency of *GBA* mutations was 6.9% among 1642 patients (OR 6.51). However only 36% of the samples included in our entire analysis were fully sequenced. Thus it was not possible to accurately determine the frequency of all mutations in different populations or to ascertain if symptoms in mutation carriers versus controls were estimated accurately in different populations. The most common mutation reported in both Ashkenazi (75%) and non- Ashkenazi (60%) controls was

N370S, reflecting the higher frequency of this allele in Ashkenazi and European populations. Adequate sample size and accurate genotyping in controls are imperative to avoid underestimating rare variants.

Despite the difficulty in determining the phenotypic profile associated with *GBA* mutations from this study, which intentionally only included subjects that met diagnostic criteria for Parkinson disease, some trends are apparent. Subjects carrying mutations presented on average 4 years earlier, were more likely to have a family history of Parkinson disease, and had less bradykinesia and rest tremor and a tendency toward dementia and dyskinesias. The general trend supports other reports in the literature that *GBA* mutations are associated with an earlier age at onset and more prominent cognitive findings.[44, 45, 51, 164, 192] However since the diagnostic criteria for Parkinson disease used in this analysis excluded more severe and progressive forms of parkinsonism such as Lewy body dementia, our findings would not accurately reflect the full spectrum of parkinsonian symptoms associated with *GBA* mutations. An increased frequency of *GBA* mutations has also been described in cohorts with Lewy body disorders [168, 193, 194] although not in multiple system atrophy [195], and a meta-analysis of *GBA* mutations in subjects with other parkinsonian diagnoses is in progress to better elucidate this issue.

Now that *GBA* is a well-validated risk factor for Parkinson disease, the ultimate challenge is to establish the mechanisms resulting in this association. Both a gain of function mechanism due to enhanced protein aggregation or

lysosmal dysfunction [196] or a loss of function related to fluctuations in levels of ceramide [197] have been postulated. Further research is in progress to elucidate the pathophysiology of both Parkinson and Gaucher disease, facilitate more accurate genetic counseling and develop new therapeutic strategies.

CHAPTER 6

GENOME-WIDE ASSOCIATION STUDY IN PD

Based on the following study:

Simon-Sanchez S#, Schulte C#, Bras JM#, Sharma M#, Gibbs J, Berg D, Paisan-Ruiz C, Lichtner P, Scholz S, Hernandez D, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls M, Illig T, Gieger C, Houlden H, Steffens M, Okun M, Cookson M, Foote K, Fernandez H, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock S, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood N, Lorenz D, Deuschl G, Chen H, Riess O, Hardy J, Singleton A, Gasser T. **Genome-Wide association study reveals genetic risk underlying Parkinson's disease**. Nat Genet. *Nat Genet. 2009 Dec;41(12):1308-12. Epub 2009 Nov 15*

Genome-Wide Association Study (GWAS) in PD

Major advances in genotyping technology have allowed rapid genome-wide screening of common variants in large populations and launched a new era in the investigation of the genetic basis of complex diseases. To an extent, GWAS have revolutionized the way genetics deals with disease. This occurred because of two major facts: first, GWAS require no prior knowledge of the disease biology and, thus can be used in a similar manner to study any disease, regardless of how well understood it is; second, because of the nature of GWAS, they require massive amounts of samples to be tested simultaneously, which often means that individual laboratories are not capable of undertaking such projects on their own. The need for large number of samples derives from the fact that, with GWAS, one is usually analyzing several hundred thousand markers simultaneously, which leads to spurious associations when underpowered studies are performed[198]. A parallel concept relates to the fact that the larger the sample size, the more likely one is to confidently detect smaller effect sizes. This becomes obvious when comparing results from initial GWAS with more recent ones, where sample size has been greatly increased.

To date, results from five different GWAS in PD have been reported, which, to some extent, reproducible results [148, 199-202]. These studies are of great importance not only for the results they report, but also because most of them have made their data publicly available, allowing for any researcher to

conduct meta-analysis or *in-silico* replication of their own data in a genome-wide level. The three initial GWAS published were all severely underpowered to detect the small effect sizes that are expected (odds ratios of 1.2-1.4). Additionally, only one of these used a two-stage approach, which has been considered a powerful, cost-effective design for GWAS. None of these studies produced genome-wide significant results.

We have thus decided to perform a GWAS in PD. This study, which was based on an international collaboration, included over 5,000 cases and 8,000 controls, making it the largest GWAS in PD to date.

Materials and Methods

Study design

The approach taken followed the two-stage design common to several published GWAS. Table 10 shows the characteristics of the approach. Stage I samples were genotyped for markers distributed across the genome. The most significant markers were then assessed in a second, independent cohort of samples.

Table 10: Study characteristics of cases and controls in stage I and II

		Cases			Controls		
		N	**aao (s.d.)**	**M/F**	**N**	**aae (s.d.)**	**M/F**
Stage I	USA	988	55.9 (15.1)	1.09	3071	62 (15.6)	0.96
	Germany	757	56 (11.64)	1.49	976	NA	1.08
Stage II	USA	1528	62.5 (8.55)	2.44	2044	63 (15.6)	2.45
	Germany	1100	61 (11.32)	1.37	2168	57 (10.54)	1.4
	UK	824	59 (12.3)	3.5	544	NA	0.57

aao = age at onset; aae = age at examination; s.d. = standard deviation; NA = not available

Stage I Subjects

Stage I of our study comprised a total of 5,820 individuals originally from the US and Germany. Each cohort is described in detail in the following sections.

US cohort

The total number of cases and controls from the United States included in stage I of this project was 4,134, comprising 1,063 cases and 3,071 controls.

PD samples: 988 of the patients were derived from the NINDS-funded Neurogenetics Repository at the Coriell Institute for Medical research (Camden, NJ, USA, www.coriell.org). Samples from the precompiled panels NDPT001, NDPT005, NDPT007, NDPT014, NDPT015, NDPT016, NDPT017

and NDPT018, as well as 250 non-paneled samples, were included in the experiments. In addition, 75 PD cases collected by a movement disorders specialist in the Laboratory of Neurogenetics were also included.

All patients were Caucasian individuals with idiopathic Parkinson's disease from the United States. The mean age at onset of the parkinsonian syndrome was 55.91 years, ranging from 7 to 98 years. Age at onset was defined as the time when symptom(s) of PD were first noted (including at least one: resting tremor, rigidity, bradykinesis, gait disorder, postural instability). Coriell Institute samples required complete NINDS Repository Clinical Data Elements, in order to be included. According to those criteria, all subjects had bradykinesia, and at least one of the following: muscular rigidity, 4-6 Hz resting tremor, postural instability (not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction). None had exclusionary features. All had documentation of sustained, excellent response to anti-parkinsonian therapy. Informed consent was obtained for every participant under locally approved protocols. All subjects were queried regarding family history of parkinsonism, dementia, tremor, gait disorders, and other neurological dysfunction. Subjects with and without family history of Parkinson's disease were included in this panel. However, patients with three or more relatives with parkinsonism or with an apparent Mendelian inheritance of PD were excluded.

Coriell Institute neurologically normal controls: Samples included in precompiled panels NDPT002, NDPT006, NDPT009, NDPT019, NDPT020, NDPT021, NDPT022, NDPT023 and NDPT024 were used for this study, leading to a total of 828 control individuals. All individuals are reported to be unrelated Caucasians free from any neurological disorders. All individuals

were asked specifically regarding the following disorders: Alzheimer's disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, cerebrovascular disease, dementia, dystonia, Parkinson's disease, and schizophrenia. None had any first-degree relative with a known primary neurological disorder. The mean age of participants was 58 years, ranging from 15 to 98 years. For more information about controls and PD cases from the Coriell institute see http://ccr.coriell.org.

CGEMS initiative controls: The Cancer Genetic Markers of Susceptibility initiative (CGEMS, Bethesda, USA, http://cgems.cancer.gov/) is a three-year, $14 million initiative aiming to identify genetic alterations that make individuals susceptible to prostate and breast cancer, funded by the National Cancer Institute. For this purpose they have collected not only cancer patients, but also 1,101 male and 1,142 female controls. Genotyping data from all these 2,243 control samples was generously shared by the National Cancer Institute and included in our study.

German cohort

The German cohort consisted of 757 PD cases and 976 population based controls from the KORA and POPGEN surveys.

PD samples: The PD cases were collected by movement disorders specialists of the Universities of Munich and Tuebingen, who established the diagnosis according to the UK Brain Bank criteria[5]. The mean age at onset was 56 years, ranging from 28 to 86 years. Both patients with and without a reported

family history of Parkinson's disease were included in this panel. However, cases showing clear evidence of dominant inheritance were excluded. In the German cohort 20% reported a family history of PD. All samples and data were collected with informed consent under locally approved protocols.

KORA survey controls: 488 control individuals were selected from the KORA survey (Cooperative Health Research in the Region of Augsburg, www.helmholtz-muenchen.de/kora), a population based study holding more that 18,000 individuals representative of the general population living in or near the region of Augsburg, Germany. All 488 samples were recruited from the KORA F3 survey in which a total of 3,006 subjects were studied in 2005. The age at sampling ranges from 34 to 84 years[203, 204].

POPGEN survey controls: 488 healthy control individuals were collected by the 'Population Based Assessment of Genetic Risk Factors' (POPGEN) - Project (www.popgen.de), an on-going cross-sectional epidemiological survey of the population in the most northern part of Germany with Kiel Canal as the southern border. The region covers 1.1 Mio inhabitants. The control individuals were identified through the official population registry of the state of Schleswig-Holstein and were assessed by trained physicians to exclude neurological and other disorders in particular PD[205].

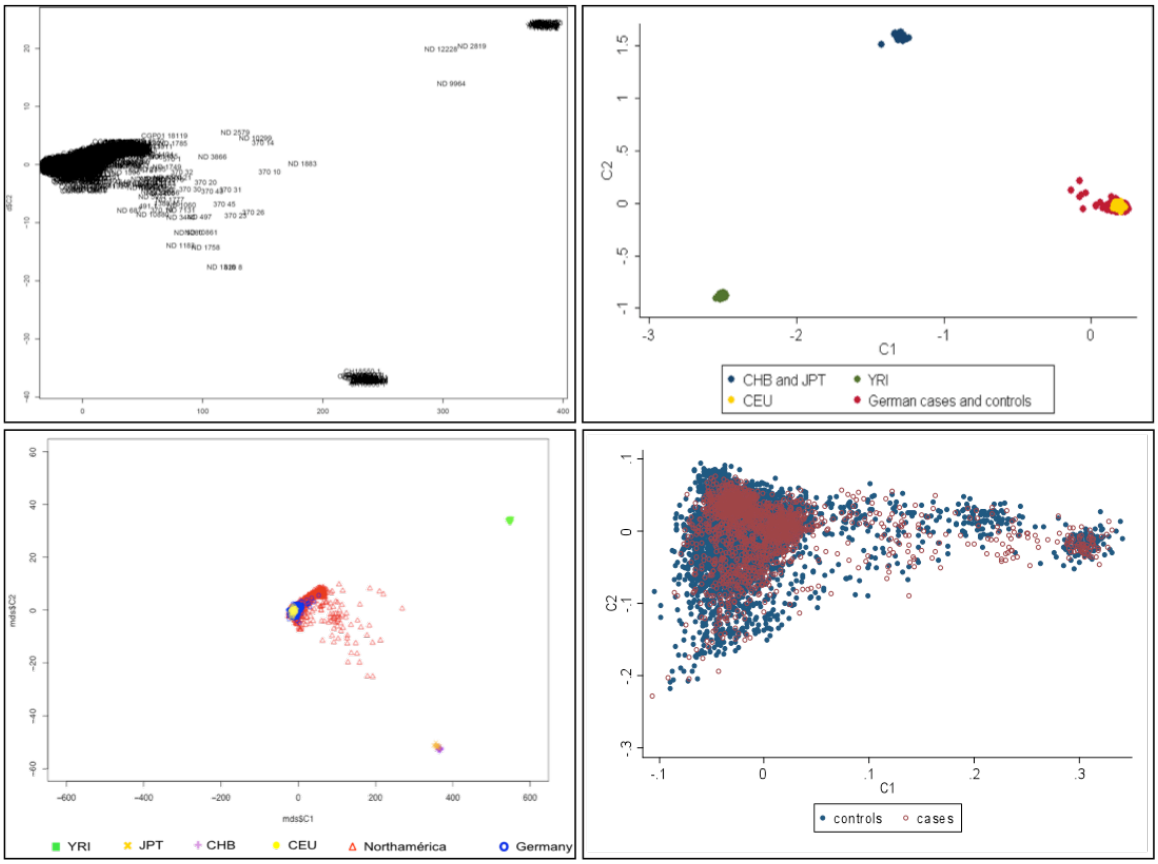


Figure 6: Population stratification plots. A) US samples with HapMap populations; B) German samples with HapMap populations; C) Merged US and German samples with HapMap populations; D) US and German samles merged.

STAGE I GENOTYPING

All samples were genotyped using Infinium Beadchips from Illumina. These genotyping chips contain tagSNPs derived from the recently completed Phase I and Phase II International HapMap Project[206] and display a comprehensive genomic coverage across the Caucasian population. 90% of all Phase I + II HapMap *loci* (MAF ≥ 0.05) are covered by at least one SNP in the CEU population. Additionally, SNPs were added evenly spaced across the genome to ensure a comprehensive coverage. On average, there is 1 common SNP every 5.5 kb across the genome in the CEU population. For more details about these genotyping platforms and Infinium workflow, see

www.illumina.com. Genotypes were called with the BeadStudio software (Illumina, inc.).

US cohort

Genotyping of the DNA panels NDPT014, NDPT015, NDPT016, NDPT017, NDPT018, NDPT019, NDPT020, NDPT021, NDPT022, NDPT023, NDPT024, and those 252 non-paneled from the Coriell Institute was performed using HumanHap550 version 3 beadchips, attempting to genotype 555,363 SNPs. The samples collected by the Laboratory of Neurogenetics in Bethesda were assayed with HumanHap550 version 1 beadchips, attempting to genotype 561,467 SNPs. Samples from the Coriell Institute within DNA panels NDPT001, NDPT002, NDPT005, NDPT006, NDPT07 and NDPT009 had previously been genotyped with HumanHap300 beadchips[207]. For the present study these samples were additionally assayed with HumanHap240S beadchips, to provide (combined) the same genotype information as the HumanHap550 version 1 beadchips. The CGEMS controls were also genotyped with HumanHap300 and HumanHap240S beadchips. Using these genotyping platforms, 545,066 unique SNPs were genotyped for each sample of our cohort.

German cohort

Genotyping of all samples was performed with HumanHap550 version 1 beadchips, attempting to genotype 561,467 SNPs. Samples were assayed at three different sites (GSF, Munich, Germany; Illumina, SanDiego, USA; Dept.

of Medical Genetics, Tuebingen, Germany). To assess the accuracy of genotyping, eleven samples were genotyped in duplicates across all batches. The concordance rate of all duplicates was 99.99%, assuring high genotype accuracy.

STAGE I QUALITY CONTROL PROCEDURES

Although it provides the opportunity to scan the whole genome in a relative short period of time, the microarray based sequencing approach also has a major problem: the high rate of false positive results. Thus, eliminating any systematic bias like population stratification (existing when the case and control groups are not well-matched genetically or if several distinct, but unrecognized, sub-populations exist in a cohort) is required to minimize the rate of false positives. All statistical analyses were performed using PLINK[153].

US cohort

Low quality genotyping: Samples with call rates below 95% were repeated using fresh DNA aliquots and if the call rate persisted below this level, the samples were excluded from the analysis. Low-quality genotyping led us to repeat 57 individual samples, of which 41 were ultimately excluded from the analysis, including 16 cases and 25 controls.

Gender ambiguity: Individuals with gender ambiguity were flagged based on heterozygosity on chromosome X genotypes (inbreeding coefficient [F] in this chromosome). A male call is made if F is more than 0.8 and a female call if F is less than 0.2. Samples with an ambiguous F score or discrepancies between genotyped and reported sex, were considered as problematic. These samples were analyzed by visual examination of log R ratio and B allele frequency metrics with the Illumina Genome Viewer (IGV) tool within BeadStudio to rule out whether this discrepancy was caused because of copy number variation or extended homozygosity in chromosome X. These analyses led to the exclusion of 15 samples, including 11 cases and 4 controls.

Population substructure: In an attempt to detect the presence of population substructure or ethnically mismatched individuals, pairwise Identity By State (IBS) distances were calculated. Consequently, IBS distance to its "nearest neighbor" was calculated for each individual in our cohort along with 30 trios from Yoruba (Nigeria), 45 unrelated individuals from the Tokyo area in Japan, 45 unrelated individuals from Beijing (China) and 30 US-resident trios with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH, Paris, France); data downloaded from the HapMap website (www.hapmap.org). This distribution was standardized (by the sample mean and variance of nearest neighbor) and inspected for outliers. For this last purpose Multidimensional scaling (MDS) was performed. This analysis showed that except for three individuals with genetic background indicative of African ancestry. These samples were removed from

further analysis. The remaining samples clearly shared Caucasian ancestry (supplementary figure 1).

Non-reported relatedness: The pairwise clustering based on IBS distances (see previous section) is useful for making estimations of pairwise Identity by Descent (IBD) to find pairs of individuals who look more similar than expected by chance, in a random sample. By estimating the probability of sharing 0, 1, or 2 alleles IBD for any two individuals, a proportion of IBD can be calculated (PI-HAT = P [IBD = 2] + 0.5 x P [IBD = 1]). Using 0.2 as a threshold for PI-HAT, 17 sample pairs were considered too similar to each other. Thus, one member of each pair was removed from further SNP association tests (11 cases and 6 controls). Additionally, PI-HAT data revealed 50 replicates within our dataset including 49 cases and one control. All these samples were also dropped from further analysis.
After this extensive quality-control phase, the final number of fully genotyped samples from the United States was 4,005 including 971 cases and 3,034 controls.

SNP quality control: Only those SNPs successfully genotyped in at least 95% of our final set of samples (18,579 SNPs removed) as well as those with a minimum allele frequency (MAF) above 5% (50,758 SNPs removed) and with no extreme departure from Hardy-Weinberg equilibrium (HWE) in controls ($p > 0.01$; 9,043 SNPs removed) were included in our Stage I statistical analyses. These procedures gave us a total of 474,995 SNPs in the US cohort.

German cohort

Low quality genotyping: Any sample with a call rate below 95% was excluded from the analysis. This led us to exclude 18 samples from the analysis (4 cases and 14 controls).

Gender ambiguity: Heterozygosity on chromosome X was used to detect gender discrepancies in our sample. Three individuals (2 cases and 1 control) were identified in which ambiguity could not be resolved, thus they were removed from further analysis. Moreover, we assessed the heterozygosity on all autosomes in our population. Excess of heterozygosity reflects genotyping error or contamination of the sample. We excluded 11 individuals (5 cases and 6 controls), which showed more than 4 standard deviations from the sample mean.

Population substructure: Population structure was assessed based upon the genome wide average proportion of alleles shared identical by state between two individuals. IBS distances were calculated between all study subjects and additional individuals, for whom genotype data was downloaded from the HapMap. These individuals originated from Nigeria, China, Japan and the United States with European ancestry. Visualization of sub-structuring in our population was done by the multi-dimensional scaling (MDS) approach, implemented in PLINK[153]. Inspection of the MDS plot (supplementary figure 2) led us to further exclude 6 individuals from our analysis (3 cases and 3 controls).

Non-reported relatedness: We excluded close relatives based on IBD estimates. Nine samples were identified as 1st and 2nd degree relatives and excluded from further analyses (1 case and 8 controls).
After applying the stringent filtering criteria as described above, 1686 samples were included in the statistical analyses (742 cases and 944 controls).

SNP quality control: Only those SNPs genotyped in at least 95% of our final set of samples (5,387 SNPs removed) as well as those with a minimum allele frequency (MAF) above 5% (51,834 SNPs removed) and with no extreme departure from Hardy-Weinberg equilibrium (HWE) in controls ($p>0.01$; 5,685 SNPs removed) were included in our Stage I statistical analyses. These procedures gave us a total of 498,560 SNPs in the German cohort.

Combined cohort

Genotyping results obtained for both the US and the German populations were merged and a further quality control step was taken. This included the removal of SNPs that presented a MAF below 5% (589 SNPs removed), or a genotyping rate below 95% (42,169 SNPs removed) or extreme deviation from HWE ($p<0.01$) (2,463 SNPs removed). These filters were applied to the combined controls and flagged SNPs were removed from the complete combined cohort. This led us to obtain a total of 463,185 unique SNPs genotyped in 5,691 individuals (including 1,713 cases and 3,978 controls). Although the evidence generated so far supports the idea that population effects are unlikely to mask association or produce false positives when

pooling white northern European and North American populations[208], we reassessed the effect of population structuring in our cohort. Therefore, pair wise Identity by State (IBS) distances were re-calculated for all 5,691 individuals in our cohort with the same procedure as mentioned above. This distribution was standardized (by the sample mean and variance of nearest neighbor) and inspected visually for outliers with a multidimensional scale plot (MDS). As expected, our observation showed that our cohort clearly shared common Caucasian ancestry as shown in supplementary figure 3. Hereafter, we describe the US and the German cohort as combined cohort.

STAGE I STATISTICAL ANALYSIS

Power calculation: Using Quanto software for sample size and power calculations, we simulated different scenarios to estimate the power to detect association in our combined cohort dataset. By plotting the power to detect association with different odds ratios, considering a p value of 1.7×10^{-7} (genome wide significance level after Bonferroni correction) and different minor allele frequencies, these simulations showed that we have 80% power to detect a variant exerting a risk with an OR as low as 1.3 in our stage I cohort (Figure 7).

Association analysis: All estimates and tests were performed with the PLINK toolset[153]. For each SNP that successfully passed the genotyping, MAF and HWE filters an additive Cochran-Armitage trend test of association was

applied based on the Fisher theory of additivity. OR, 95% CI and p values were calculated for each test performed (each SNP). Since a large number of tests were performed, conservative Bonferroni corrections were applied to correct for multiple testing.

STAGE II SUBJECTS

For our replication stage, we included a total of 7,672 Caucasian individuals (3,460 cases and 4,212 controls), originally from the United Kingdom (824 cases and 7 controls), North America (1,528 cases and 2,044 controls) and Germany (1,100 cases and 2,168 controls). A brief description of these samples is listed below and in table 1 of the main text:

US cohort

Coriell PD samples: A total of 207 PD samples that where not available at the time of the Stage I genotyping execution where included in the replication stage. These included 140 males and 65 females from the United States. The age of PD onset ranges from 16 to 80 years with a mean of 54.6 years, defined as when symptom(s) of PD were first noted (including at least one: resting tremor, rigidity, bradykinesis, gait disorder, postural instability).

The Parkinson's Genes and Environment Study (*PAGE) samples:* These include 840 PD cases (643 males and 196 females) and 1700 controls (1329

males and 371 females) identified from a large population-based cohort. Cases were initially identified by self-reported with subsequent verification with patients treating neurologist. The age of onset ranges from 42 to 78 (average 65.8±7.4).

Washington University at St. Louis: Patients and spouse controls were recruited consecutively from the Movement Disorders Center at Washington University in St. Louis. PD diagnosis was made using the UK Brain Bank criteria[5]. All controls had normal neurologic examinations. This included 818 samples including 481 (299 males and 182 females) cases and 337 controls (118 males, 219 females).

UK cohort

824 cases were included, comprising 466 neuropathologically diagnosed PD cases and 358 clinically diagnosed PD cases. The male to female ratio was 3.5:1, age at onset ranged from 28:86 years (mean 59 years). Diagnosis was made using the UK Brain Bank criteria (Hughs et al, 1992). Additionally, 544 healthy controls were also collected. Male to female ratio was 0.57.

German cohort

1323 German controls were selected from the population-based KORA cohort described in stage I. Additionally; sample collection of 793 German controls was performed as part of the “Prospective validation of risk markers for the development of idiopathic Parkinson’s disease (PRIPS)” study a longitudinal

cohort study in Tuebingen. For this study participants were recruited using two main sources: advertisement in local newspapers, and employees from local companies. Inclusion criteria of this longitudinally designed study were age older than 50 years and no diagnosis of PD. Moreover, 62 German controls were either recruited as spouse controls or through advertisements in the clinic or local press at the movement disorder outpatient clinics at the Departments of Neurology at the Universities of Luebeck, Germany. All underwent a detailed neurological and movement disorders examination. 607 patients with PD were recruited within the 'POPGEN Parkinson's Disease' - Project (POPGEN-PD). All patients were identified through the databases and charts office-based neurologists or neurological hospitals in the popgen region. Only patients fulfilling the British Brain Bank criteria[1] were included. They were contacted by mail and asked for their participation. The protocol was approved by the local ethical committees. 286 German sporadic and familial PD patients were recruited at two university clinics for neurology in Bochum and Tuebingen. All patients were evaluated by a neurologist experienced in movement disorders and were diagnosed as idiopathic PD, based on the UK Parkinson's disease brain bank criteria. 163 cases were collected by movement disorders specialists of the Universities of Munich and Tuebingen, who established the diagnosis according to the UK Brain Bank criteria[5]. Both patients with and without a reported family history of Parkinson's disease were included in this panel. However, cases showing clear evidence of dominant inheritance were excluded. 52 German patients were recruited at the movement disorder outpatient clinics at the Department of Neurology at the University of Luebeck, Germany. Consecutive patients

willing to participate were included in the study. All patients underwent a detailed neurological and movement disorders examination.

All subjects gave written informed consent. The study, including DNA collection, was approved by the local ethical committees.

STAGE II GENOTYPING

GoldenGate genotyping

For stage II, 384 SNPS were selected purely based on the *p* value observed in stage I, under the additive model of association, and predicted assay creation, following Illumina guidelines (see supplementary table 1). Each of the samples was independently typed for these 384 SNPs using GoldenGate technology for VeraCode platform from Illumina (www.illumina.com). Briefly, 250 ng of DNA from each sample were activated through a chemical reaction with biotin. After purifying from excess biotin, assay oligonucleotides were added and hybridized to the DNA, and the mixture bound to streptavidin-conjugated paramagnetic particles (SA-PMPs). After the oligo hybridization, mis- and non-hybridized oligos were washed away and allele-specific extension (ASE) and ligation of the hybridized oligos was performed. The extended and ligated products formed synthetic templates that were then amplified through a PCR reaction. The strands containing the fluorescent signal in the PCR products were then isolated and hybridized to the VeraCode beads via an address sequence. After the hybridization, the VeraCode beads were washed and scanned in the BeadXpress Reader (Illumina). After

scanning, raw data was imported into BeadStudio v3.1.12 (Illumina) for analysis and genotype calling.

MALDI-TOF genotyping

In addition to the 384 SNPs from the Caucasian stage I analysis, 12 SNPs were genotyped on chromosome 1 and chromosome 4 (see supplementary table 2). These SNPs are located in *loci* detected by our collaborators studying a Japanese cohort[209] and were successfully replicated by them. Some of these SNPs had allele frequencies below the quality control cut-off of 5% in our Caucasian population, and were therefore excluded in our stage I analysis. These SNPs were genotyped by primer extension of multiplex PCR products with the detection of the allele-specific extension products using the matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry method (MassArray system, Sequenom, San Diego, CA).

STAGE II QUALITY CONTROL

In order to check the accuracy of the GoldenGate genotyping method, a set of 96 samples previously genotyped in stage I of this project was re-assayed with GoldenGate. Those SNPs in which genotypes from these two different platforms were not identical across 10 or more samples were considered as consistent failures and removed from further analyses. This approach led to the exclusion of 11 non-concordant SNPs. After removing these 11 SNPs from our dataset, genotyping concordance between HumanHap550 v3

beadchips and GoldenGate technology was greater than 99.2%. In addition, visual examination of SNP clusters using Beadstudio v3 led us to the exclusion of 12 more SNPs, resulting in a total of 23 SNPs to be excluded from further analysis. Moreover, those SNPs genotyped in less than 90% of the samples, with a MAF below 5% and with an extreme departure of HWE (1 x 10^{-7}) were also removed. After the initial removal of all samples with a call rate below 90% and following the quality control procedures described above, our stage II dataset consisted of 345 SNPs successfully genotyped in a total of 3,452 cases and 4,173 controls.

STAGE II STATISTICAL ANALYSIS

Power calculations

Power calculations were performed as described in stage I. Considering a p value of 1.44 x 10^{-4} (genome wide significance level after Bonferroni correction for 345 SNPs) and different minor allele frequencies, these simulations showed that our stage II sample had 80 to 90% power to detect a variant exerting a risk with an OR as low as 1.2 in our stage II cohort (Figure 7).

Association analysis

As described above (stage I), assuming an additive model for association, a test statistic was computed for each SNP. OR, 95% CI and p values were calculated for each test performed.

LD STRUCTURE ANALYSIS

The LD structure of the identified *loci* was analyzed using Haploview 4.1 (http://www.broad.mit.edu/mpg/haploview [210]) and LD blocks delimited using the D'-based confidence interval method developed by Gabriel *et al* [211].

For the *SNCA locus*, haplotype counts were performed with haploview 4.1 and risk values (OR) plotted with R (Figure 12). To check if the association signal detected in *SNCA* was in significant LD with that previously reported in REP1, genotype data at this marker was included in a subset of 1,774 US PD cases and controls from our Stage I analysis (REP1 genotyping performed as previously described[212]). LD calculation performed with Haploview 4.1[210] revealed that both our risk allele at the 3' end of *SNCA* and the risk allele of REP1 tag the same risk haplotype (D' = 0.872 and r^2 = 0.365) (Figure 13).

To further demonstrate that both risk alleles at REP1 and at the 3' end of SNCA tag the same risk haplotype, we performed a logistic regression analysis conditioned on the genotypes in REP1. Results derived from this analysis show a drop in the association detected at rs2736990 from p = 0.003649 to p = 0.03934, suggesting that variation at REP1 and at this SNP are not independent of each other in the genetic aetiology of the disease.

To compare the signals of the *MAPT locus* with previous studies, we included genotype data of the H1/H2 deletion/insertion polymorphism. 154 individuals

of the Stage II sample were genotyped with previously described methods [213]. With this subset of samples the rest of H1/H2 genotypes were imputed in the combined cohort. These genotypes were used to compute LD between the H1/H2 polymorphism and the genotyped SNPs (D' = 0.902 and r^2 = 0.742).

Test of epistatic interaction

A stepwise procedure was used to test for independent effect. In brief, the most significant SNP (rs2736990 for SNCA and rs393152 for MAPT) was modeled to condition on all other alleles. We did observe some marginal effects but these were not significant after correcting for multiple testing, arguing that all significant SNPs refer to one single causal variant.

To test for an interaction between the *SNCA* and *MAPT loci*, we used a likelihood ratio test (LRT). Forward selection procedure was used to develop a final model to test an epistatic interaction. In brief, both unrestricted and restricted models were fitted using the maximum likelihood method. The advantage of using LRT is that distribution of test statistic is approximately chi-square distributed with degrees of freedom equal to the difference of the numbers of unrestricted and restricted parameters.

PAR: We furthermore computed a population attributable risk (PAR) for SNCA and MAPT. PAR% was calculated using the formula:

PAR=(p [OR-1])/(p [OR-1]+1)]* 100,

where p is the prevalence of the risk allele in the population and OR is the odds ratio.

Since, there was no evidence of interaction, we further estimated the combined population attributable (cPAR) risk for these genes using the following formula:

$cPAR=1-(1-PAR_{SNCA})\ (1-PAR_{MAPT})$

Comparison of MSA, PD and control *SNCA* risk genotypes

Genotype and allele frequencies were selected from 92 pathologically proven MSA cases that had previously been genotyped for the selected *SNCA* variants and published by us[214]. The frequency of these variants was compared to that observed in the stage II PD and control groups using Haploview 4.1[210] (Table 11).

EXPRESSION ANALYSIS

Frozen tissue samples of the frontal cortex were obtained from 133 neurologically normal Caucasian subjects. 100-200mg aliquots of frozen tissue were sub-dissected from each of the samples and used for genotyping and expression assays. Genotyping was performed using Infinium HumanHap550 beadchips (Illumina Inc) followed by imputation to ~1.6 million SNPs after data cleaning, profiling of 22,000 mRNA transcripts was performed using HumanRef-8 Expression BeadChips (Illumina Inc) as previously described [215].

A regression analysis was performed on the expression intensities generated for mRNA. Gender, age, post-mortem interval, tissue source and hybridization batch were included as covariates. Residuals from the regression analysis for each probe were then used as the quantitative trait for that probe in genome-wide association analysis looking for quantitative trait loci, performed using the *assoc* function within PLINK, which correlates allele dosage with change in the trait [153]. To correct for the large number of SNPs tested per trait, a genome-wide empirical p-value was computed for the asymptotic p-value for each SNP using 1,000 permutations of sample-label swapping. To correct for the number of traits being tested per tissue region, a false discovery rate (FDR) threshold was determined based on the empirical p-values. Empirical p-values were allowed to exceed this threshold if the linkage disequilibrium r^2 was greater than or equal to 0.7 with a SNP with empirical values within the FDR threshold. The sequences of probes with significant correlation to a trait were examined for the presence of polymorphisms using CEU HapMap data, and, if present, that QTL was removed from the result set. Notably, we failed to detect sufficient expression at the probe for LRRK2, thus precluding analysis of the levels of this transcript as a quantitative trait.

Results

To assess the homogeneity of our cohort, pair-wise Identity by State distances were calculated[216],[153] using HapMap data as a reference[217]. The results of these analyses reveal that our samples share common Caucasian ancestry (Figure 6)[218]. We chose not to use genomic control in subsequent analyses as false positive association, possibly caused by

population substructure (lambda=1.17), would be controlled for by our two-stage design[153, 219]. Power calculations showed that our sample had 80% power to detect variants conferring an odds ratio (OR) of 1.3 with an allele frequency of 10% (Figure 7).

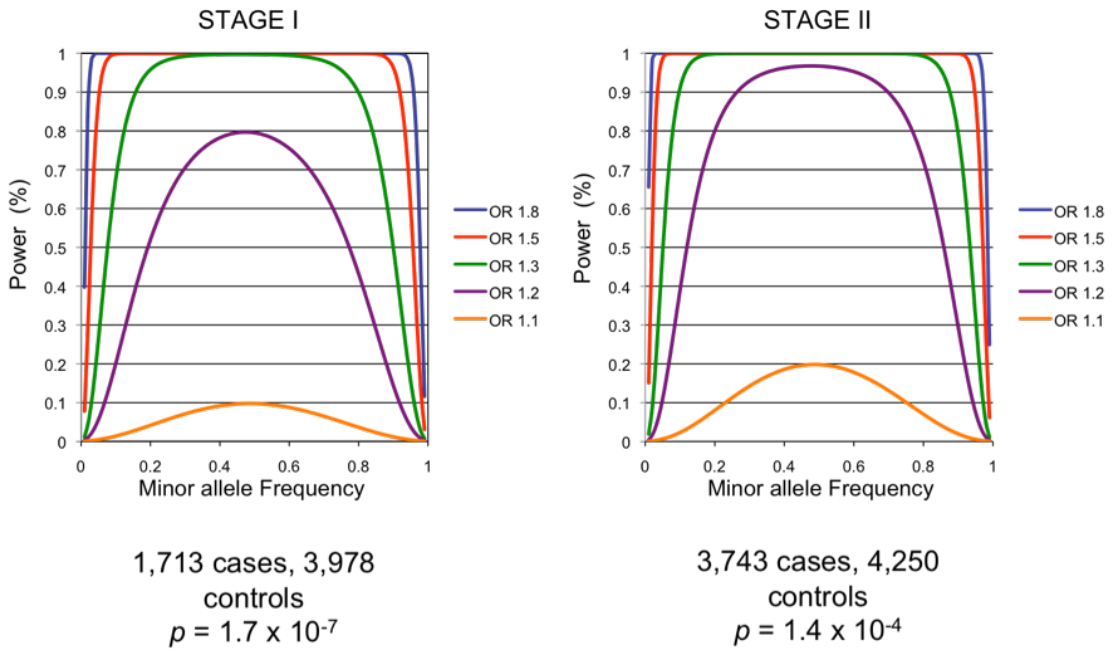


Figure 7: Power estimates for stage I (A) and stage II (B) of the study. Power on the y-axis is plotted against minor allele frequency with different thresholds for Odds ratio.

Each SNP was tested for association using an additive model. Four SNPs on chromosome 4q22 within the *SNCA* locus exceeded Bonferroni corrected genome-wide significance threshold in stage I (most significant p=5.69x10^{-9}, rs2736990; Figure 8, Supplementary Table 5). Three SNPs at the *MAPT* locus on chromosome 17q21 also surpassed genome-wide significance in stage I (most significant p=5.05x10^{-8}, rs199533).

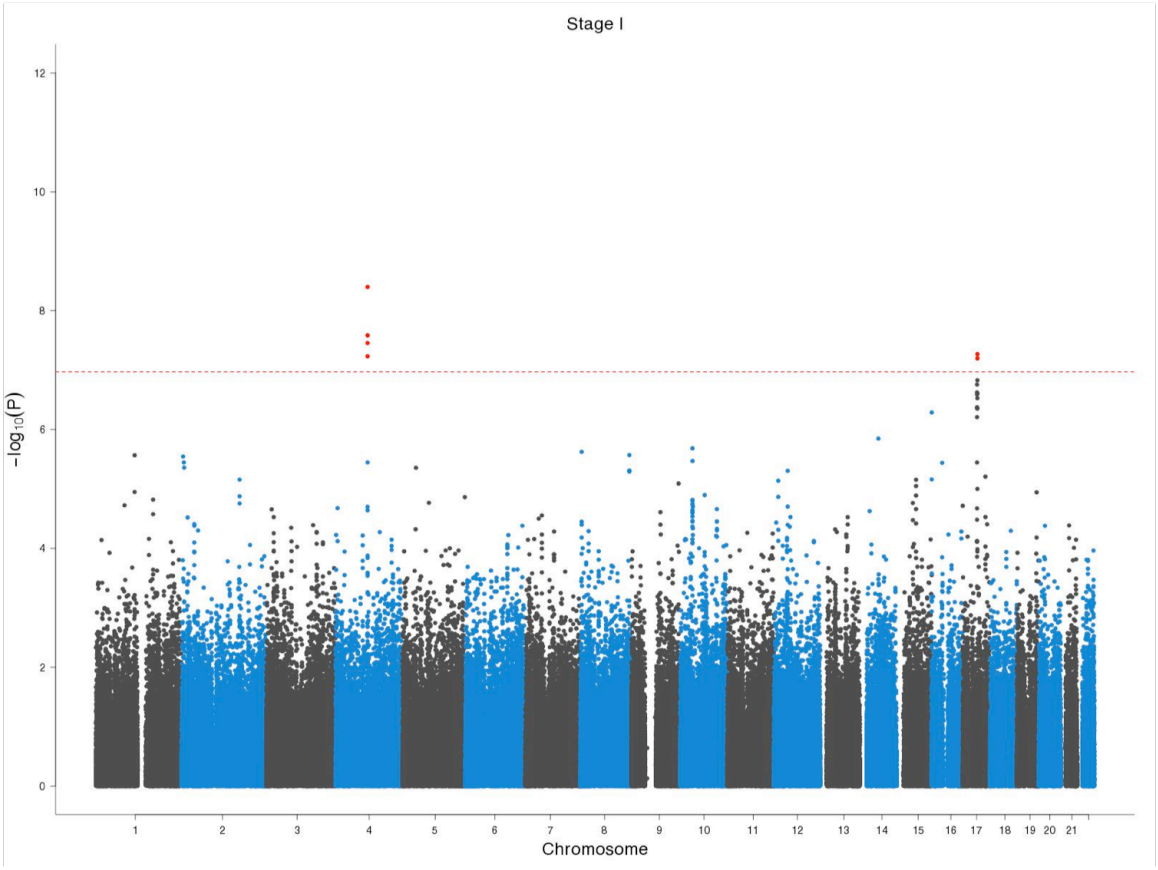


Figure 8: Graphical representation of p values in stage I. p values are log transformed (y-axis) and plotted against chromosomes (x-axis). The red line indicates the Bonferroni threshold. Signals indicated in red are on chromosome 4 and chromosome 17 and surpass Bonferonni threshold for genome wide significance.

Replication comprised genotyping of 384 SNPs selected based on the *p* value observed in stage I under the additive model (least significant $p=2.87 \times 10^{-4}$). Genotyping was performed in an independent cohort of 3,545 cases and 4,855 controls from the US, Germany and Britain. Taking into account the results obtained from our pair-wise Identity by State distances calculations and considering that genetic heterogeneity and allelic heterogeneity are not likely to produce type I and type II errors when pooling white North American and white North European populations, we decided to analyze all Stage II samples together. Following quality control filtering, 345 SNPs were analyzed in 3,452 cases and 4,710 controls.

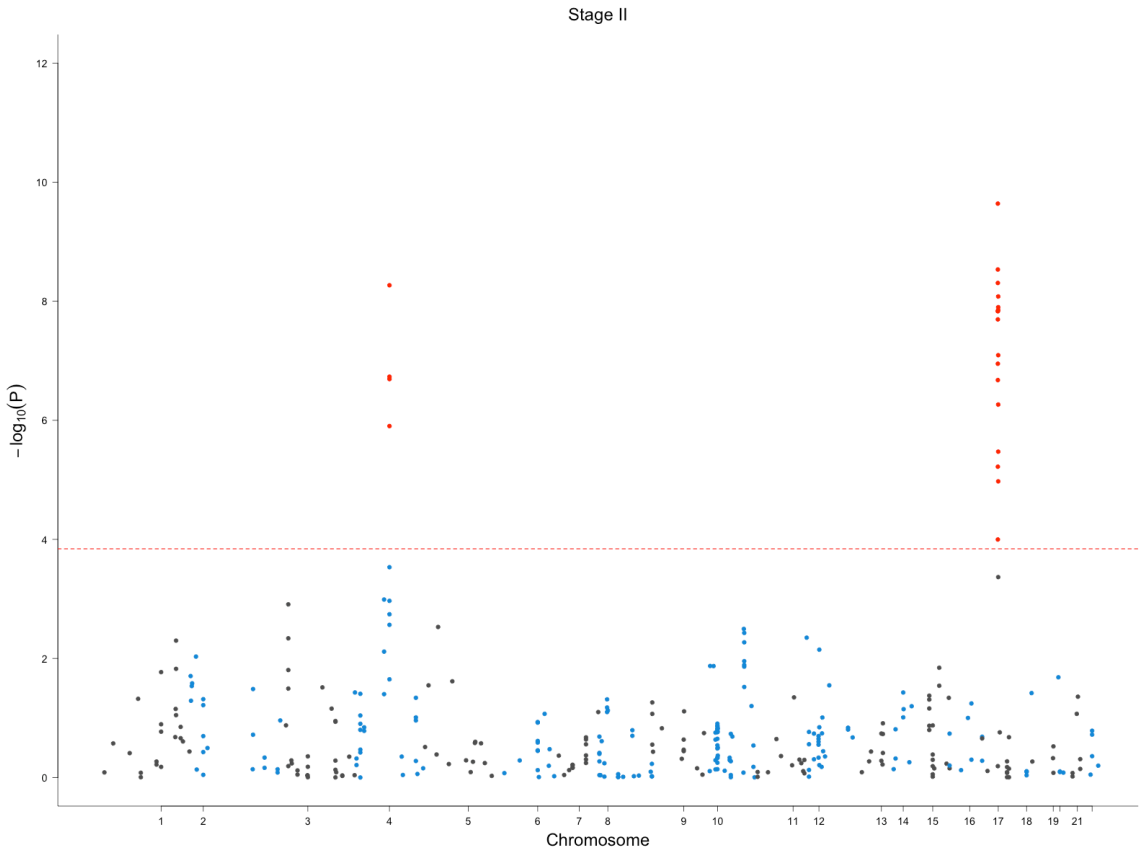


Figure 9: Graphical representation of p values in stage II. log transformed p-values of Stage II SNPs (y-axis) are plotted against chromosomes (x-axis). Signals indicated in red are on chromosome 4 and chromosome 17 and surpass Bonferonni threshold for multiple testing.

Twenty-one SNPs within the *SNCA* and *MAPT* loci surpassed Bonferroni threshold for significance (p<0.000145), highlighting *SNCA* and *MAPT* as top hits and providing unequivocal evidence that these loci are risk factors for PD (Supplementary Table 5).

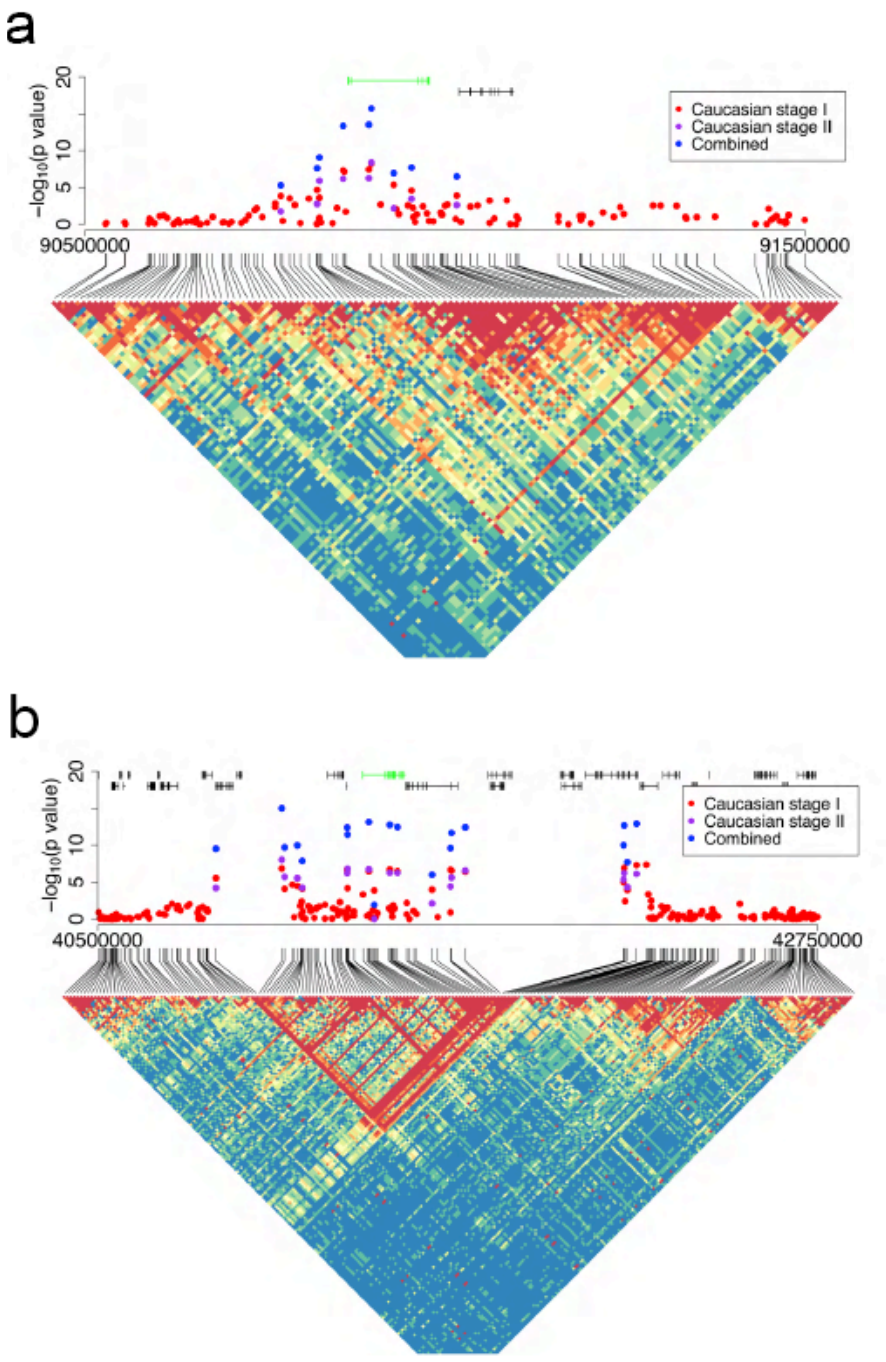


Figure 10: Association and LD across *SNCA*, *MAPT*. -log$_{10}$ p values are shown for stage I and II analyses, annotated transcripts are shown across the top of each plot. A) One million base pairs across the *SNCA* locus, *SNCA* transcript indicated in green. B) 2.25Mb across the *MAPT* locus, *MAPT* transcript indicated in green.

In an effort to further delineate the signals on *SNCA* and *MAPT*, allelic association of significant SNPs was tested, conditioned on alleles of other significant SNPs at the same locus [220]. No independent signals were identified, suggesting that variants at each locus point to a single pathogenic

variant. We did not find evidence for epistasis between *SNCA* and *MAPT* risk alleles.

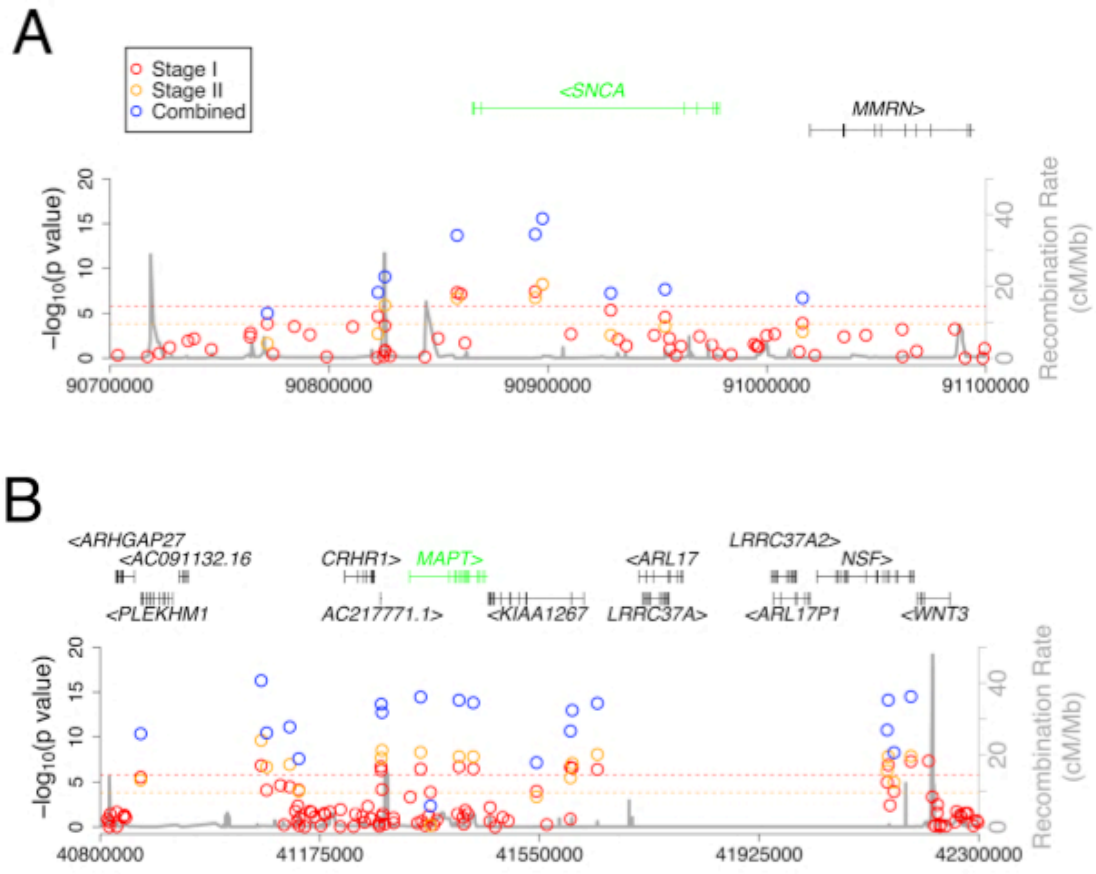


Figure 11: Association and recombination rates across *SNCA, MAPT*. -log$_{10}$ p values are shown for stage I and II analyses, annotated transcripts are shown across the top of each plot. Red dotted line indicates threshold for genome wide significance in stage I and orange line indicates threshold for significance of stage II.

Analysis of the linkage disequilibrium (LD) structure across the *SNCA* locus revealed two blocks of LD (Figure 10A). The 3' block contains three of the four significantly associated SNPs, suggesting that the causal variant is located in the 3' region of the *SNCA* gene. This is strengthened by analysis of the haplotype frequencies at this locus and previous studies [221, 222].

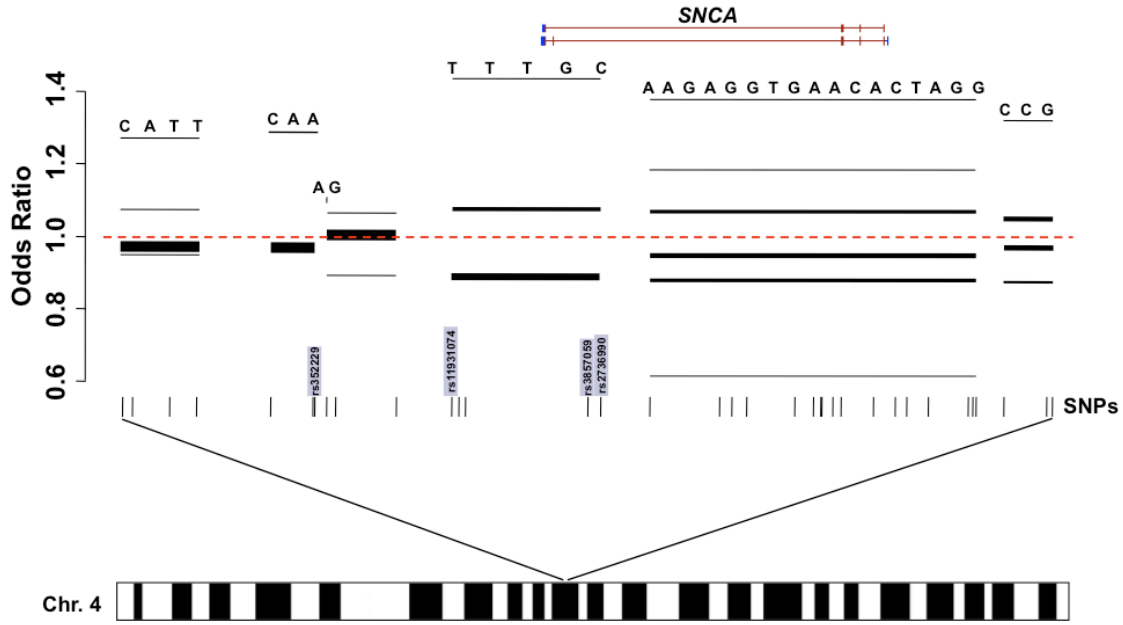


Figure 12: Odds ratio exerted by each of the haplotypes present in the two LD blocks identified across *SNCA locus*. Each haplotype is represented by a single line, which is wider according to the haplotype frequency. Those SNPS significantly associated with PD after stage II of our analyses are shaded in grey.

The REP1 microsatellite in the promoter region of SNCA was previously associated with PD [212] and its pathological effect has been suggested to be mediated by gene expression [223], Analysis of REP1 genotype data in 1,774 samples from the US cohort revealed that the risk allele of REP1 is in LD with the 3' risk alleles identified here (r^2 = 0.365 with rs3857059), suggesting that the association identified at the REP1 locus and the SNPs identified here may be the result of residual LD between these loci.

Figure 13: LD structure across stage I American population in the SNCA locus. REP1 genotypes (D4S3481) have been included for these samples.

This was further supported by a logistic regression analysis conditioned on REP1 genotypes, showing that association at REP1 is not independent from the association identified here. We have recently reported a significant association of *SNCA* SNPs with another synucleinopathy, multiple system atrophy (p=5.5x10^{-12}, MSA) [224]; comparison of these data reveals disparate *SNCA* risk SNPs in MSA and PD, a finding that may shed light on the exact pathogenic substrate and molecular etiology of these disorders (Table 11).

Table 11: SNPs at the SNCA locus in PD, Controls and MSA.

SNP	Allele	Minor Allele Frequency			P values		
		PD	MSA	CON	PD vs Con	MSA vs Con	MSA vs PD
rs1430961	C	0.093	0.11	0.082	0.0191	0.1741	0.4283
rs12644119	A	0.124	0.163	0.108	0.0016	0.0178	0.1193
rs356229	G	0.402	0.359	0.363	1.12E-06	0.8937	0.2371
rs11931074	T	0.098	0.174	0.076	7.87E-07	8.91E-07	7.00E-04
rs3857059	G	0.098	0.152	0.075	6.46E-07	1.00E-04	0.0152
rs2736990	T	0.509	0.473	0.46	2.90E-09	0.7379	0.3365
rs3775439	A	0.145	0.217	0.13	0.007	5.00E-04	0.0062
rs894278	G	0.074	0.103	0.06	4.00E-04	0.0146	0.1392
rs6532197	G	0.089	0.141	0.076	0.0027	0.001	0.0148

As expected, one large highly inter-correlated block of high LD was observed across the *MAPT* locus (Figure 2B). Available genotype data of the H1/H2 haplotypes in this region showed that the risk alleles of the associated SNPs are in LD with the H1 haplotype (r^2=0.761 with rs393152).

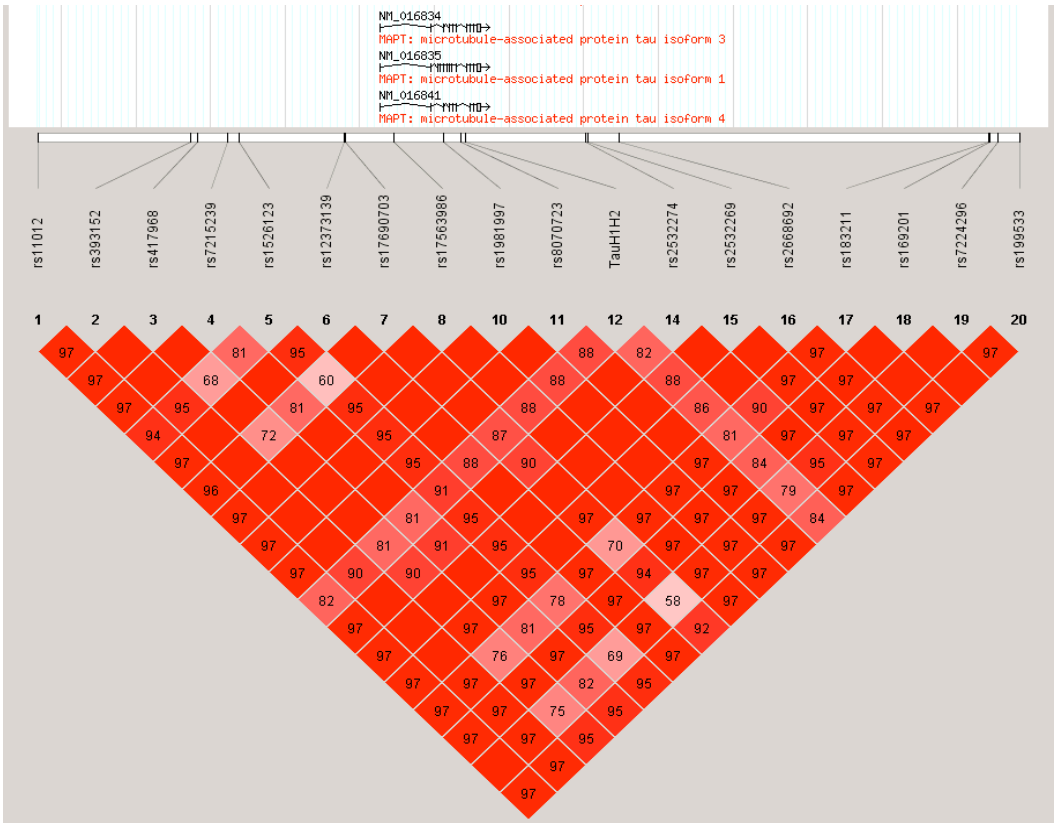


Figure 14: LD structure around H1/H2 polymorphism in the German cohort.

It is unclear from the current data whether the *MAPT* risk haplotype identified here corresponds to the subhaplotype associated with corticobasal degeneration (CBD) and progressive supranuclear palsy (PSP) [225-228]. Because of the LD structure we cannot rule out other genes at this locus as the pathogenically relevant risk genes; however, from the perspective of biological plausibility and the expression data discussed below, *MAPT* is the most likely candidate.

Mutations in both *SNCA* and *MAPT* have been associated with autosomal dominant forms of parkinsonism [6, 29, 31, 229]. Given this, it is particularly interesting that we observed association proximal to *LRRK2*, which also contains mutations causing autosomal dominant PD [10, 119]. In stage I 23 SNPs located upstream of *LRRK2*, and 12 SNPs within *LRRK2* were associated with PD (lowest p=5.03x10^{-6} in rs2896905, located in *SLC2A13*, 0.27Mb from *LRRK2*). Of these, only 3 SNPs surpassed our p value threshold for replication and were analyzed in stage II. Only one single SNP, located 0.17Mb upstream of *LRRK2,* remained associated with PD after stage II (rs1491923, p=7.12x10^{-3}). While this did not surpass our threshold for multiple testing, the combined stage I and II p values revealed a compelling association (p=2.10x10^{-5}). Interestingly, the other 2 replicated SNPs at this locus were also nominally associated with PD after combining stage I and stage II datasets (p values of 2.76x10^{-5} and 3.30x10^{-3} respectively for rs11564612 and rs2896905). Notably, data from the Asian cohort also revealed a significant association with PD at this locus [209]. *SLC2A13*, the

neighboring gene of *LRRK2*, cannot be ruled out as the gene of effect at this locus, however, *LRRK2* is clearly a more plausible candidate.

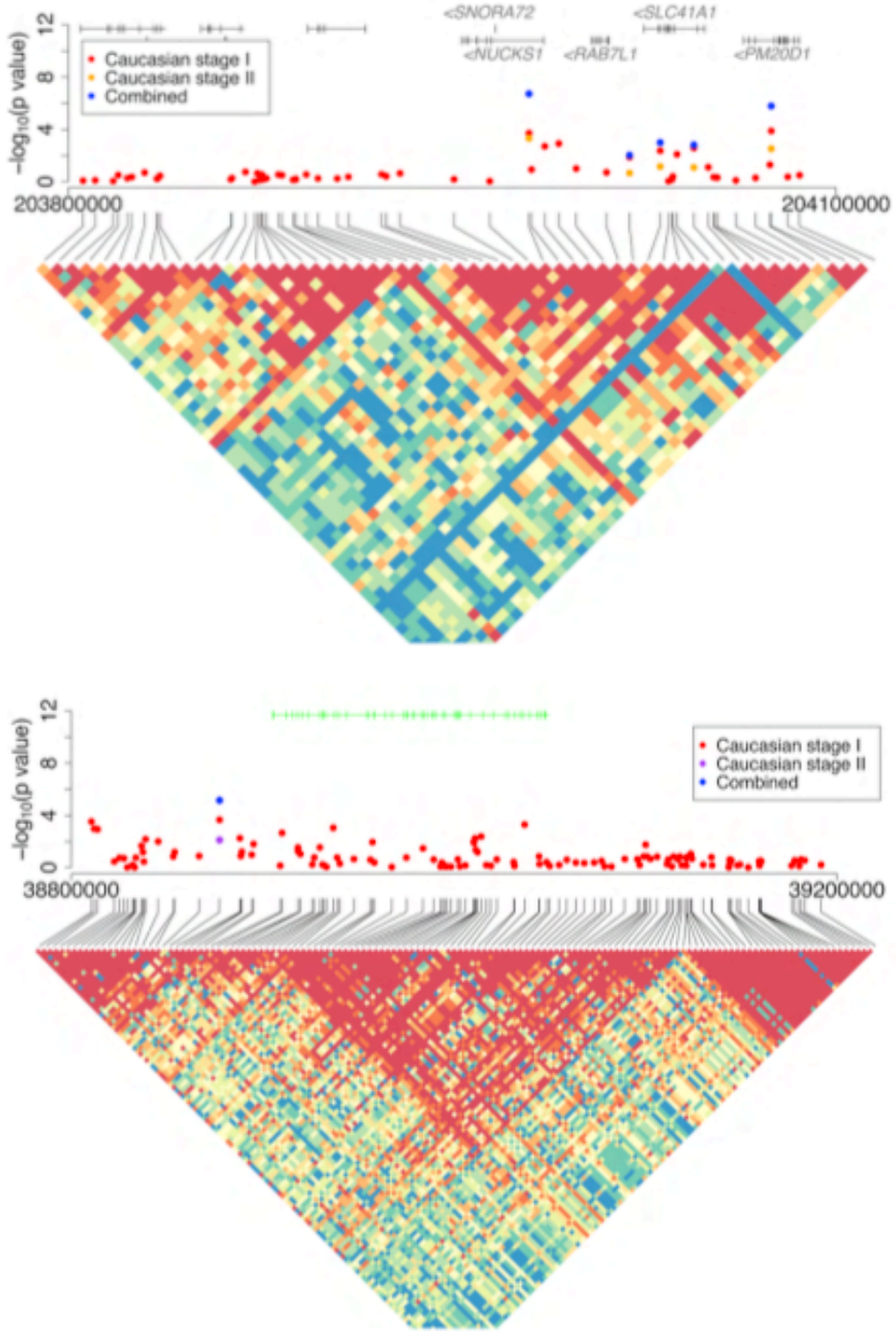


Figure 15: Association and LD across *PARK16* and *LRRK2*. -log$_{10}$ p values are shown for stage I and II analyses, annotated transcripts are shown across the top of each plot. 0.3Mb across the *PARK16* locus at chromosome 1q32, which illustrates significant association signal across a single block of LD; five transcripts are identified across the region of association; 0.4Mb on chromosome 12 including *LRRK2*, illustrating a significant association signal 5' to *LRRK2* in an LD block distinct from this gene.

Following data exchange with our colleagues running a PD GWAS in Japan we chose to study two loci implicated in Asian PD on chromosomes 1q32 and 4p15. After reassessing our stage I data, the most significant *p* values at these 12 SNPs were $1.3x10^{-4}$ and $6.5x10^{-3}$ (1q32 and 4p15.3 respectively). The signal at 1q32 would have been sufficiently significant to carry through to stage II replication, but this SNP had been excluded from analysis because of the low minor allele frequency in controls (0.03). Genotyping of these 12 SNPs was performed in an available subset of our replication cohort comprising 2,909 cases and 3,500 controls. The signal on chromosome 1q32 was replicated in this cohort (rs823128, $p=5.01x10^{-3}$; Supplementary Table 5). While this failed to surpass Bonferroni correction, the *p* value across stages was highly significant (rs823128, $p=1.32x10^{-7}$) and it is worth noting that the significance improved for all SNPs at this locus when combining stage I and II results (Figure 15A). For these reasons and because the association at this locus was consistently detected in the Asian cohort [209] and across both our stages, we are confident this signal represents a true association and this has been designated *PARK16*. Although we failed to replicate the signal on chromosome 4p15, which included only one gene, *BST1*, the low minor allele frequency of the associated SNPs in individuals of Caucasian genetic background may have affected our ability to observe association. Further studies will help to clarify the role of this locus in modulating the risk for PD in individuals of European ancestry.

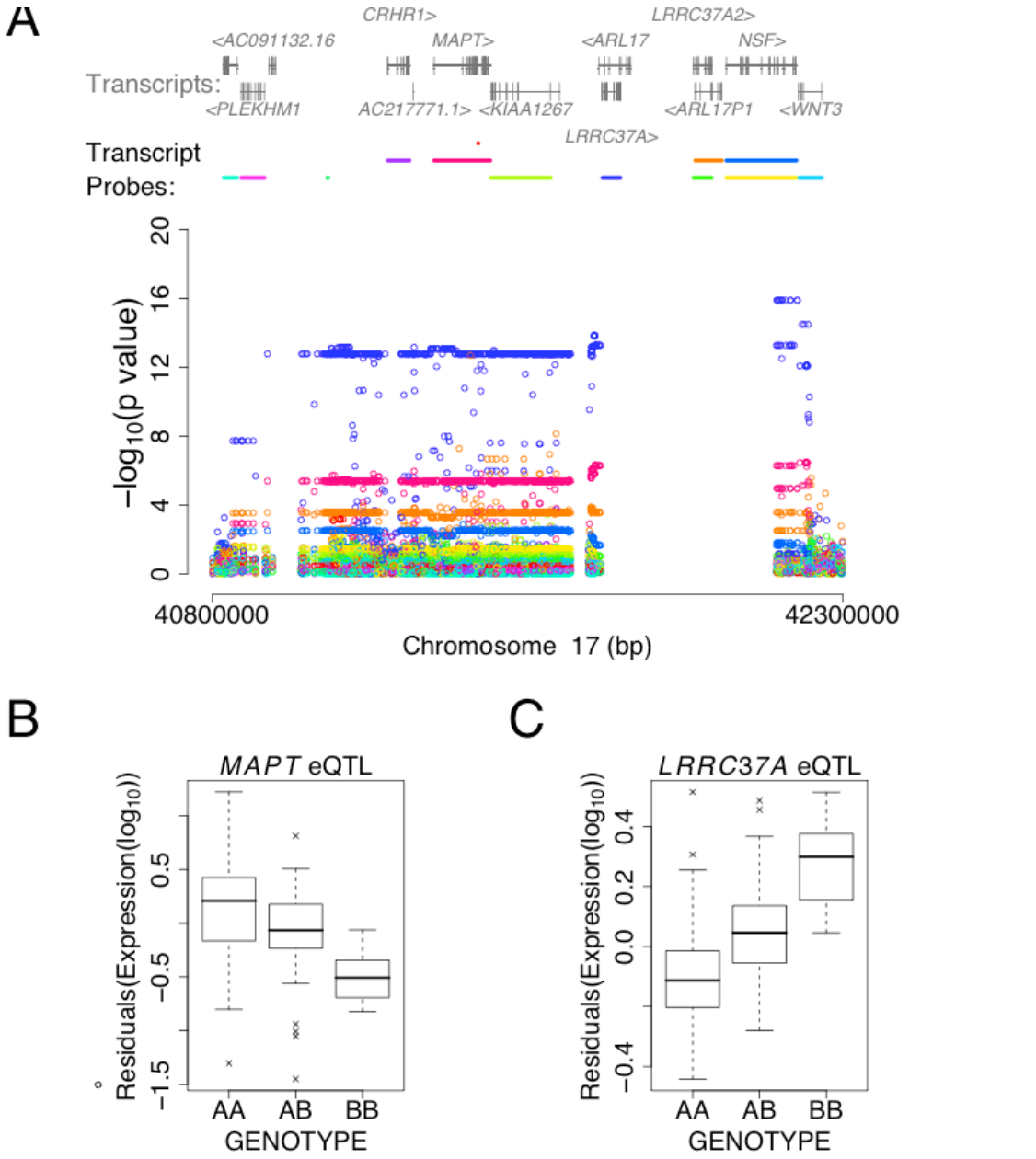


Figure 16: Expression quantitative trait loci across the MAPT locus measured in 133 human frontal cortex samples; panel A shows association between genotypes and transcript levels across the MAPT locus. In this analysis the allelic load at genotyped polymorphism across the locus is tested for association with transcript levels of each gene across the locus. The results of the analysis are shown as log transformed p values color-coded to match the transcript of interest.

In an attempt to define a biological consequence of risk variants, we mined data produced within our laboratory as a part of an expression quantitative trait-mapping project. In this work genome wide genotyping and expression profiling of >22,000 transcripts was performed in 133 human frontal cortex samples; thus allowing us to determine SNPs where genotype is significantly

associated with expression level. These data revealed a strong association between genotype at the risk alleles of the *MAPT* locus and expression levels of both *MAPT* and *LRRC37A*, but did not reveal association between risk SNPs and expression levels of proximal genes at the *SNCA*, *LRRK2* or *PARK16* loci.

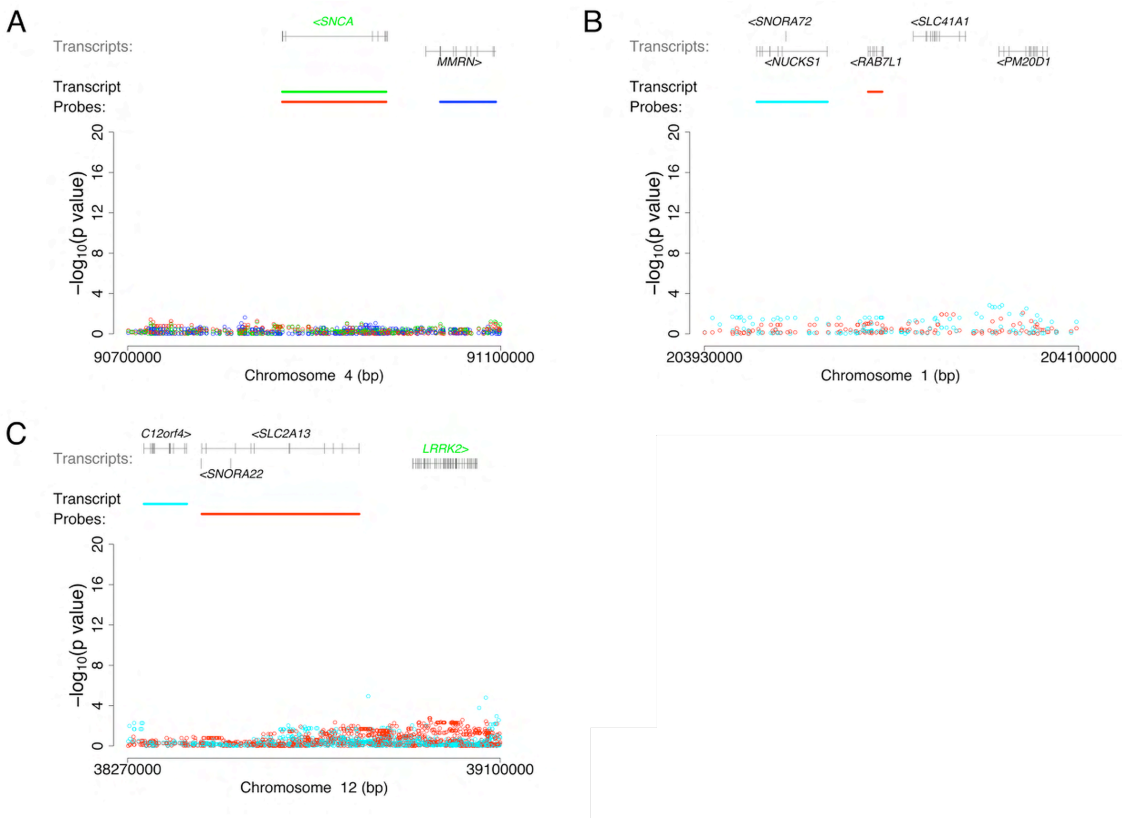


Figure 17: Expression quantitative trait loci across the *SNCA, PARK16* and *LRRK2* loci (A, B and C respectively), measured in 133 human frontal cortex samples; each plot A shows association between genotypes and transcript levels across the locus. In this analysis the allelic load at genotyped polymorphism across the locus is tested for association with transcript levels of each gene across the locus. The results of the analysis are shown as log transformed *p* values color-coded to match the transcript of interest.

Notably the alleles at the *MAPT* locus associated with increase risk of PD are associated with increased expression of *MAPT* in the human brain, suggesting that *MAPT* levels are etiologically important in the pathogenesis of this disease.

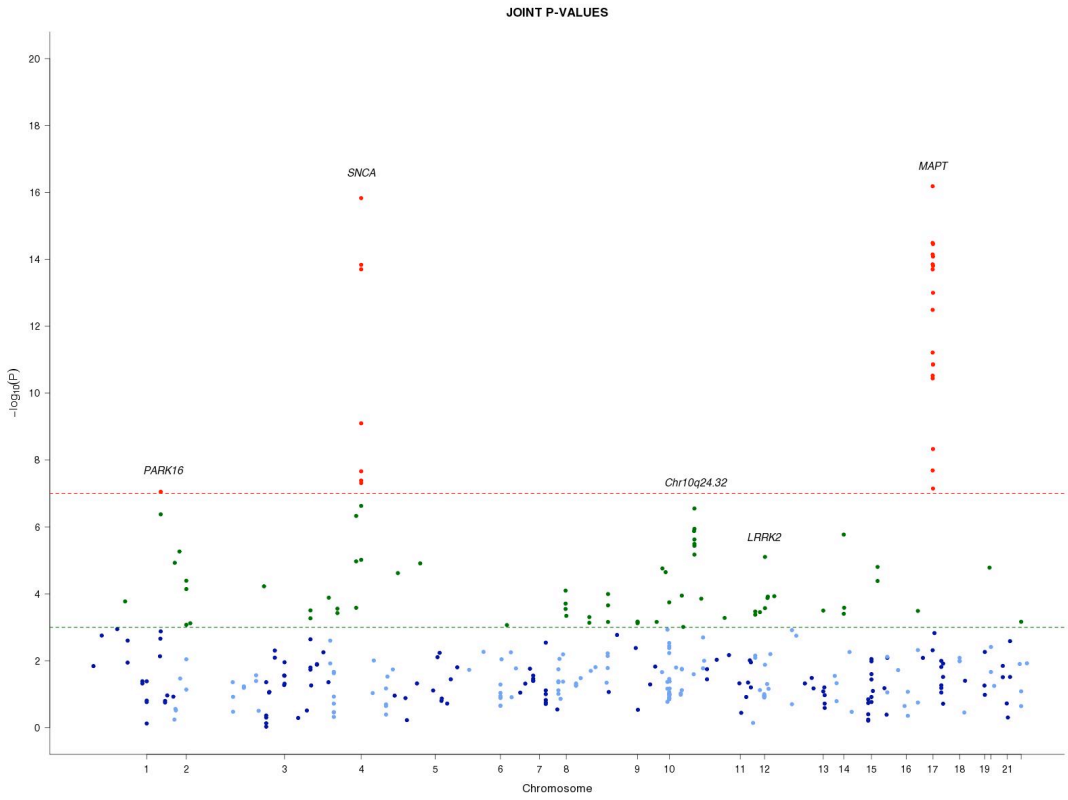


Figure 18: Graphical representation of p-values generated after combining stage I and stage II. In red SNPs surpassing Bonferroni correction; in green SNPs suggestive to be associated with PD.

Figure 18 shows the association results across stages and for the combined cohort, for the 345 nominally significant SNPs that successfully surpassed our quality control procedures. Notably, we observed clusters of SNPs showing improved association signals when combining our stage I and stage II datasets. Although some of these SNPs are within loci that contain biologically plausible candidate genes for the development of PD, they do not reach genome-wide significance and thus we have resisted drawing too many conclusions from these data; however, of particular note is a cluster of 7 SNPs in chromosome 10q24.32, with *p* values below $1x10^{-3}$.

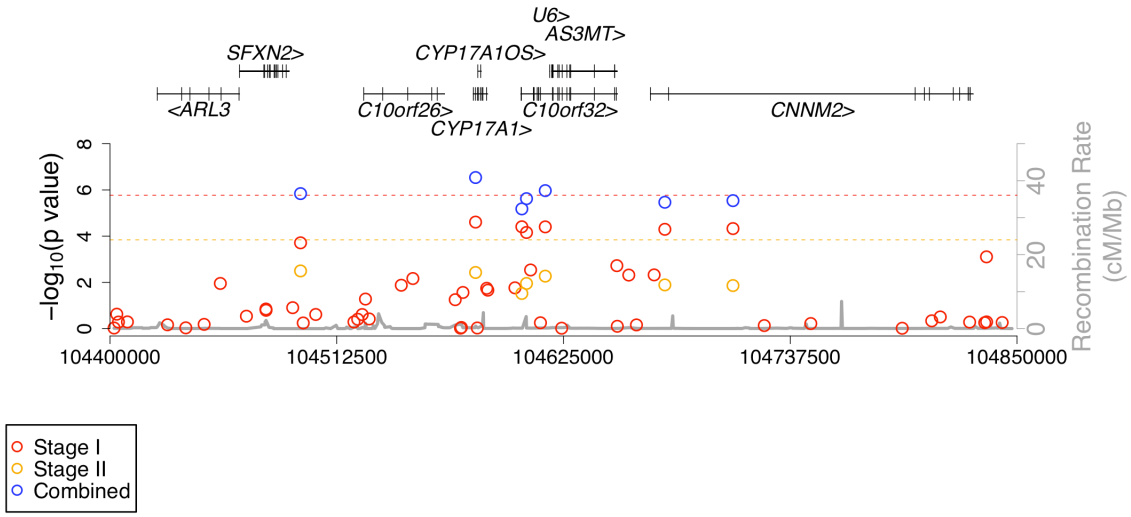


Figure 19: graphical representation of association p-values and recombination rates at the Chr10q24.32 showing suggestive association with PD. Individual stages and joint p-values are shown.

These and other variants that show a consistent but moderate association across the stages warrant independent replication.

Discussion

Although mutations and copy number variants of *SNCA* are the cause of rare familial forms of PD [6, 8], association of common variants has been more controversial. The present study provides unequivocal evidence that variation in *SNCA* contributes to the etiology of sporadic PD. The clustering of associated SNPs in the 3' UTR suggests that the causal variant might affect post-transcriptional RNA processing or RNA stability, possibly mediated by miRNA binding sites [230] or by influencing alternative splicing.

A strong association of the H1 haplotype at the *MAPT* locus with PSP and CBD has been described and repeatedly replicated [225-227]; however,

association studies of variants at *MAPT* in PD have produced conflicting results [34, 231, 232]. Again, our data provide unequivocal evidence for an association of a haplotype block containing the *MAPT* gene with sporadic PD. This is surprising given the classic separation of synucleinopathies and tauopathies, although a cross-talk between molecular pathways characterized by different aggregating proteins has been repeatedly suggested on multiple levels [233]. While there are additional genes at the *MAPT* locus, a role of *MAPT* itself in neurodegenerative diseases is well established and this association is biologically plausible. We further provide compelling evidence for an association of PD with variability proximal to *LRRK2* and at a novel locus at 1q32 (*PARK16*). Both of these insights open exciting avenues for research. The kinase activity of Lrrk2 has become an attractive therapeutic target; the current data suggests that this protein is also relevant to the etiology in sporadic PD patients without frank mutations. Finally, the *PARK16* locus spans 5 transcripts, *SNORA72*, *NUCKS1*, *RAB7L1*, *SLC41A1* and *PM20D1*. Clearly it will be crucial to fine map and define the immediate biological consequences of all four risk loci identified here. It is notable that three of the most significant loci identified here contain genes known to be mutated in Mendelian forms of parkinsonism. This not only supports the notion that rare familial disease is etiologically related to typical sporadic PD, but also that genes that contain common risk variants are excellent candidates to contain rare disease causing mutations. One might also predict that deep sequencing of these loci will reveal rare mutations that alter risk for, rather than cause, disease. It is also interesting that two of the four loci

discussed here, are risk factors for other neurodegenerative diseases, including MSA (*SNCA*), PSP and CBD (*MAPT*).

The combined population attributable risk associated with the identified loci, considering the genotypic counts of those most associated SNPs in our Caucasian cohort, is approximately 25%. Since our study was a retrospective case-control study and the frequency of the risk variants detected might not reflect the frequencies of the true causal variants, these values should be interpreted with caution [234].

In summary we show for the first time a clear role for common genetic variability in the risk of developing PD. Further we describe population specific genetic heterogeneity in this disorder, an observation that has potential implications for the analysis of complex traits across populations; such genetic heterogeneity, particularly at minor risk loci, has the potential to mask true associations when analyses are performed across populations. With the discovery of the *PARK16* locus in the Asian population, this highlights the power of comparing GWAS across different populations. A further increase in the number and size of cohorts for GWAS in PD will likely reveal additional common genetic risk loci and these in turn will improve understanding and ultimately treatment of this devastating disorder.

Conclusions

It is now clear that genetics plays an important role in the pathogenesis of PD. Specifically in the Portuguese population, we have found a reasonable number of mutations.

The c.6055G > A; p.G2019S *LRRK2* mutation presents one of the highest frequencies in Europe: 4.3% of early onset cases and 9.2% of late-onset cases.

In addition to the common *LRRK2* variant, we have also found that 8.7% of Portuguese early-onset PD cases are attributable to *PRKN* mutations. Similar to other reports we found *PINK1* and *SNCA* mutations to be a rare cause of disease in our population.

We also report the first genetic screening for PD genes in a sub-Saharan population. We found no pathogenic mutations in the genes most commonly known to cause PD in European North American, or North African populations. Although the cohort studied is clearly small and definite conclusions regarding frequencies are unachievable, a trend for different genetic basis of PD in this sub-Saharan population is, in our opinion, noteworthy.

Given the possible role for iron metabolism in PD pathogenesis, we show that genetic variability within the *HFE locus* may be a risk factor for PD. However the low frequency of the variants limits the statistical power of the analysis and, thus, studies in larger samples and in diverse cohorts are needed to further clarify the relation between variability in *HFE* and PD.

A recent study suggested that mutations in *GIGYF2* were associated with PD. We attempted to replicate these findings by performing a detailed analysis of the genetic variability in the *GIGYF2* gene and its association with PD in two large sets of Cases and age-matched Controls, from two geographically distinct populations. Although we cannot rule out a small genetic contribution of GIGYF2 to PD, our data strongly suggest that the pathogenic variants previously published are rare polymorphisms. We support this statement based on the fact that two of such mutations were found in our Control groups and that SNPs across GIGYF2 did not show any association with PD. Thus, in our extended dataset, *GIGYF2* is not a PD gene.

We also demonstrate that *GBA* mutations are significantly more common in patients with PD than in neurologically normal controls in the Portuguese cohort. We have expanded this work two-fold: first we have analyzed *GBA* in a separate population of British origin were we found that the frequency of *GBA* mutations was also significantly increased in PD when compared to controls; secondly we performed a meta-analysis of 16 worldwide populations were we also found such an increase. This clearly shows that *GBA* plays a role in PD.

We also performed a GWAS in 1,713 Caucasian patients with Parkinson's disease and 3,978 controls. After replication in 3,513 cases and 4,710 controls, two strong association signals were observed: in the alpha-synuclein gene and at the *MAPT* locus. We exchanged data with colleagues performing a GWAS in Asian PD cases. Association at *SNCA* was replicated in the Asian GWAS, confirming this as a major risk locus across populations. The association at *MAPT* was absent in this cohort, indicating population specific

genetic heterogeneity in this complex disease. We were able to replicate the effect of a novel locus detected in the Asian cohort (*PARK16*) and provide evidence supporting the role of common variability around *LRRK2* in modulating risk for PD. These data demonstrate an unequivocal role for common genetic variability in the etiology of typical PD. These results are of significant importance, not only for the research field, in that they allow for a better understanding of the pathobiological events in PD, but also for clinicians who search for aids in diagnosing a complex disease and facillitate genetic counseling to families.

Together, the results detailed here provide a genetic basis for about 40% of all PD cases in Caucasian samples.

Reflections and future steps

The implementation of GWAS has inarguably revolutionized genetic investigation of disease. Although there were (and still continue to be) some critics of GWAS, the identification of several hundred loci for complex diseases (http://www.genome.gov/gwastudies/) [235] has proven that this method successfully finds common risk variants for disease and traits. The identification of such risk loci adds weight to the “common disease common variant” hypothesis; clearly, however, this is not the only mode by which genetic variability confers risk for disease. It is likely that in addition to disease causing mutations and common risk alleles, rare risk-conferring variants will also exist [236]. The search for rare risk variants has been limited by technical feasibility and adequate sample size. Large sections of the genome, including intronic epigenic regions, are required to be analyzed in large sample series because such variants are by definition rare. The technology now exists to perform ultra-high throughput DNA sequencing on target-enriched samples. We plan to sequence several Mb of DNA in thousands of PD patients and neurologically normal controls to unequivocally define the role of rare risk variants at these loci in PD. We have provided convincing evidence for the association of 4 loci with risk for disease by GWAS; 3 of these loci are known to contain rare mutations that cause autosomal dominant PD/parkinsonism [6, 10, 29, 119] (further all 3 genes unequivocally linked to autosomal dominant parkinsonism were also found by us to be risk genes for PD). We argue therefore that the other locus identified in this screen (*PARK16*) and future loci identified by GWAS, are excellent candidates as genes that contain mutations

causing parkinsonism. Likewise parsimony would suggest that all these genes are excellent candidates for screening for rare risk (rather than causative) variants. Such a screening will require analysis of both coding and non-coding portions of these genes, as low-risk variants may exert their effects through expression or splicing rather than through protein sequence changes.

As the genetic architecture of PD gains clarity a critical next step will involve investigation of the biological effects of risk variants. The relationship between genetic variability and epigenetic alterations or gene expression is one that has been largely and necessarily confined to observations at single loci and transcripts in individual cell types or tissues. The development of genome-scale technologies provides unprecedented opportunities to expand upon these experiments. The integration of genetic, epigenetic and expression data promises to provide general observations regarding the relationship between genetic variation and expression. These data can be readily mined to unravel the network of effects associated with genomic variants. This may reveal some of the rather cryptic intermediate events that occur between DNA risk variant and clinical phenotype. The generation of maps of the genetic control of expression within the human brain has been performed by us and others [79, 237, 238]. We plan to expand upon these data to create a clear picture of the immediate effects of risk variants.

We believe the immediate goals of PD research should follow three main approaches: 1) continue to pool genome-wide association data among collaborators already performing their own GWAS in PD, in order to achieve a large enough number of samples that allows us to detect low effect sizes (OR <1.2); 2) within the *loci* found in the previous approach, perform ultra-deep re-

sequencing to detect all variants present in these *loci* in large cohorts; 3) attempt to understand the immediate biological effects of risk variability, specifically effects on local DNA methylation and gene expression. This can be accomplished by performing eQTL and methQTL mapping of these variants in series of control brains. This will provide a direct biological consequence to PD risk variants and mechanistic insight into the underlying disease process.

References

1. Forno, L.S., *Neuropathology of Parkinson's disease.* J Neuropathol Exp Neurol, 1996. **55**(3): p. 259-72.
2. Tanner, C.M., *Is the cause of Parkinson's disease environmental or hereditary? Evidence from twin studies.* Adv Neurol, 2003. **91**: p. 133-42.
3. Dick, F.D., et al., *Environmental risk factors for Parkinson's disease and parkinsonism: the Geoparkinson study.* Occup Environ Med, 2007. **64**(10): p. 666-72.
4. Taylor, K.S., et al., *Screening for undiagnosed parkinsonism among older people in general practice.* Age Ageing, 2005. **34**(5): p. 501-4.
5. Hughes, A.J., et al., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.* J Neurol Neurosurg Psychiatry, 1992. **55**(3): p. 181-4.
6. Polymeropoulos, M.H., et al., *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.* Science, 1997. **276**(5321): p. 2045-7.
7. Spillantini, M.G., et al., *Alpha-synuclein in Lewy bodies.* Nature, 1997. **388**(6645): p. 839-40.
8. Singleton, A.B., et al., *alpha-Synuclein locus triplication causes Parkinson's disease.* Science, 2003. **302**(5646): p. 841.
9. Fuchs, J., et al., *Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication.* Neurology, 2007. **68**(12): p. 916-22.
10. Paisan-Ruiz, C., et al., *Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease.* Neuron, 2004. **44**(4): p. 595-600.
11. Simon-Sanchez, J., et al., *Parkinson's disease due to the R1441G mutation in Dardarin: a founder effect in the Basques.* Mov Disord, 2006. **21**(11): p. 1954-9.
12. Kachergus, J., et al., *Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations.* Am J Hum Genet, 2005. **76**(4): p. 672-80.
13. Di Fonzo, A., et al., *A common missense variant in the LRRK2 gene, Gly2385Arg, associated with Parkinson's disease risk in Taiwan.* Neurogenetics, 2006. **7**(3): p. 133-8.
14. Ross, O.A., et al., *Analysis of Lrrk2 R1628P as a risk factor for Parkinson's disease.* Ann Neurol, 2008. **64**(1): p. 88-92.
15. Greggio, E., et al., *Kinase activity is required for the toxic effects of mutant LRRK2/dardarin.* Neurobiol Dis, 2006. **23**(2): p. 329-41.
16. Kitada, T., et al., *Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.* Nature, 1998. **392**(6676): p. 605-8.
17. Lucking, C.B., et al., *Association between early-onset Parkinson's disease and mutations in the parkin gene.* N Engl J Med, 2000. **342**(21): p. 1560-7.
18. Bonifati, V., et al., *Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism.* Science, 2003. **299**(5604): p. 256-9.
19. Miller, D.W., et al., *L166P mutant DJ-1, causative for recessive Parkinson's disease, is degraded through the ubiquitin-proteasome system.* J Biol Chem, 2003. **278**(38): p. 36588-95.
20. Valente, E.M., et al., *Hereditary early-onset Parkinson's disease caused by mutations in PINK1.* Science, 2004. **304**(5674): p. 1158-60.
21. Rogaeva, E., et al., *Analysis of the PINK1 gene in a large cohort of cases with Parkinson disease.* Arch Neurol, 2004. **61**(12): p. 1898-904.

22. Hedrich, K., et al., *Clinical spectrum of homozygous and heterozygous PINK1 mutations in a large German family with Parkinson disease: role of a single hit?* Arch Neurol, 2006. **63**(6): p. 833-8.
23. Ramirez, A., et al., *Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase.* Nat Genet, 2006. **38**(10): p. 1184-91.
24. Shojaee, S., et al., *Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays.* Am J Hum Genet, 2008. **82**(6): p. 1375-84.
25. Di Fonzo, A., et al., *FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome.* Neurology, 2009. **72**(3): p. 240-5.
26. Laman, H., *Fbxo7 gets proactive with cyclin D/cdk6.* Cell Cycle, 2006. **5**(3): p. 279-82.
27. Kruger, R., et al., *Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype.* Ann Neurol, 1999. **45**(5): p. 611-7.
28. Conrad, C., et al., *Genetic evidence for the involvement of tau in progressive supranuclear palsy.* Ann Neurol, 1997. **41**(2): p. 277-81.
29. Hutton, M., et al., *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature, 1998. **393**(6686): p. 702-5.
30. Haugarvoll, K., Z.K. Wszolek, and M. Hutton, *The genetics of frontotemporal dementia.* Neurol Clin, 2007. **25**(3): p. 697-715, vi.
31. Spillantini, M.G., et al., *Mutation in the tau gene in familial multiple system tauopathy with presenile dementia.* Proc Natl Acad Sci U S A, 1998. **95**(13): p. 7737-41.
32. Fung, H.C., et al., *Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts.* Neurodegener Dis, 2006. **3**(6): p. 327-33.
33. Kwok, J.B., et al., *Tau haplotypes regulate transcription and are associated with Parkinson's disease.* Ann Neurol, 2004. **55**(3): p. 329-34.
34. Skipper, L., et al., *Linkage disequilibrium and association of MAPT H1 in Parkinson disease.* Am J Hum Genet, 2004. **75**(4): p. 669-77.
35. Winkler, S., et al., *Role of ethnicity on the association of MAPT H1 haplotypes and subhaplotypes in Parkinson's disease.* Eur J Hum Genet, 2007. **15**(11): p. 1163-8.
36. Zabetian, C.P., et al., *Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease.* Ann Neurol, 2007. **62**(2): p. 137-44.
37. Caffrey, T.M., et al., *Haplotype-specific expression of exon 10 at the human MAPT locus.* Hum Mol Genet, 2006. **15**(24): p. 3529-37.
38. Myers, A.J., et al., *The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts.* Neurobiol Dis, 2007. **25**(3): p. 561-70.
39. Tobin, J.E., et al., *Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study.* Neurology, 2008. **71**(1): p. 28-34.
40. Winfield, S.L., et al., *Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: implications for Gaucher disease.* Genome Res, 1997. **7**(10): p. 1020-6.
41. Wafaei, J.R. and F.Y. Choy, *Glucocerebrosidase recombinant allele: molecular evolution of the glucocerebrosidase gene and pseudogene in primates.* Blood Cells Mol Dis, 2005. **35**(2): p. 277-85.
42. Sidransky, E., *Heterozygosity for a Mendelian disorder as a risk factor for complex disease.* Clin Genet, 2006. **70**(4): p. 275-82.
43. Hruska, K.S., et al., *Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA).* Hum Mutat, 2008. **29**(5): p. 567-83.

44. Neudorfer, O., et al., *Occurrence of Parkinson's syndrome in type I Gaucher disease.* Qjm, 1996. **89**(9): p. 691-4.
45. Lwin, A., et al., *Glucocerebrosidase mutations in subjects with parkinsonism.* Mol Genet Metab, 2004. **81**(1): p. 70-3.
46. Aharon-Peretz, J., H. Rosenbaum, and R. Gershoni-Baruch, *Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews.* N Engl J Med, 2004. **351**(19): p. 1972-7.
47. Clark, L.N., et al., *Pilot association study of the beta-glucocerebrosidase N370S allele and Parkinson's disease in subjects of Jewish ethnicity.* Mov Disord, 2005. **20**(1): p. 100-3.
48. Sato, C., et al., *Analysis of the glucocerebrosidase gene in Parkinson's disease.* Mov Disord, 2005. **20**(3): p. 367-70.
49. Eblan, M.J., et al., *Glucocerebrosidase mutations are also found in subjects with early-onset parkinsonism from Venezuela.* Mov Disord, 2006. **21**(2): p. 282-3.
50. Toft, M., et al., *Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population.* Neurology, 2006. **66**(3): p. 415-7.
51. Clark, L.N., et al., *Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease.* Neurology, 2007. **69**(12): p. 1270-7.
52. Bras, J., et al., *Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Portugal.* Neurobiol Aging, 2007.
53. Balducci, C., et al., *Lysosomal hydrolases in cerebrospinal fluid from subjects with Parkinson's disease.* Mov Disord, 2007. **22**(10): p. 1481-4.
54. Zimprich, A., et al., *The PARK8 locus in autosomal dominant parkinsonism: confirmation of linkage and further delineation of the disease-containing interval.* Am J Hum Genet, 2004. **74**(1): p. 11-9.
55. Nichols, W.C., et al., *Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease.* Lancet, 2005. **365**(9457): p. 410-2.
56. Di Fonzo, A., et al., *A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease.* Lancet, 2005. **365**(9457): p. 412-5.
57. Gilks, W.P., et al., *A common LRRK2 mutation in idiopathic Parkinson's disease.* Lancet, 2005. **365**(9457): p. 415-6.
58. Bras, J.M., et al., *G2019S dardarin substitution is a common cause of Parkinson's disease in a Portuguese cohort.* Mov Disord, 2005. **20**(12): p. 1653-5.
59. Ozelius, L.J., et al., *LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews.* N Engl J Med, 2006. **354**(4): p. 424-5.
60. Lesage, S., et al., *LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs.* N Engl J Med, 2006. **354**(4): p. 422-3.
61. Mata, I.F., et al., *Lrrk2 pathogenic substitutions in Parkinson's disease.* Neurogenetics, 2005. **6**(4): p. 171-7.
62. Fung, H.C., et al., *A common genetic factor for Parkinson disease in ethnic Chinese population in Taiwan.* BMC Neurol, 2006. **6**: p. 47.
63. Farrer, M.J., et al., *Lrrk2 G2385R is an ancestral risk factor for Parkinson's disease in Asia.* Parkinsonism Relat Disord, 2007. **13**(2): p. 89-92.
64. Funayama, M., et al., *Leucine-rich repeat kinase 2 G2385R variant is a risk factor for Parkinson disease in Asian population.* Neuroreport, 2007. **18**(3): p. 273-5.
65. Chan, D.K., et al., *LRRK2 Gly2385Arg mutation and clinical features in a Chinese population with early-onset Parkinson's disease compared to late-onset patients.* J Neural Transm, 2008. **115**(9): p. 1275-7.
66. An, X.K., et al., *LRRK2 Gly2385Arg variant is a risk factor of Parkinson's disease among Han-Chinese from mainland China.* Eur J Neurol, 2008. **15**(3): p. 301-5.

67. Li, C., et al., *The prevalence of LRRK2 Gly2385Arg variant in Chinese Han population with Parkinson's disease.* Mov Disord, 2007. **22**(16): p. 2439-43.
68. Lu, C.S., et al., *The LRRK2 Arg1628Pro variant is a risk factor for Parkinson's disease in the Chinese population.* Neurogenetics, 2008. **9**(4): p. 271-6.
69. Tan, E.K., et al., *LRRK2 R1628P increases risk of Parkinson's disease: replication evidence.* Hum Genet, 2008.
70. Strauss, K.M., et al., *Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease.* Hum Mol Genet, 2005. **14**(15): p. 2099-111.
71. Gray, C.W., et al., *Characterization of human HtrA2, a novel serine protease involved in the mammalian cellular stress response.* Eur J Biochem, 2000. **267**(18): p. 5699-710.
72. Park, H.J., et al., *Alzheimer's disease-associated amyloid beta interacts with the human serine protease HtrA2/Omi.* Neurosci Lett, 2004. **357**(1): p. 63-7.
73. Jones, J.M., et al., *Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice.* Nature, 2003. **425**(6959): p. 721-7.
74. Martins, L.M., et al., *Neuroprotective role of the Reaper-related serine protease HtrA2/Omi revealed by targeted deletion in mice.* Mol Cell Biol, 2004. **24**(22): p. 9848-62.
75. Simon-Sanchez, J. and A.B. Singleton, *Sequencing analysis of OMI/HTRA2 shows previously reported pathogenic mutations in neurologically normal controls.* Hum Mol Genet, 2008.
76. Khan, N.L., et al., *Dopaminergic dysfunction in unrelated, asymptomatic carriers of a single parkin mutation.* Neurology, 2005. **64**(1): p. 134-6.
77. Couzin, J. and J. Kaiser, *Genome-wide association. Closing the net on common disease genes.* Science, 2007. **316**(5826): p. 820-2.
78. Manolio, T.A., L.D. Brooks, and F.S. Collins, *A HapMap harvest of insights into the genetics of common disease.* J Clin Invest, 2008. **118**(5): p. 1590-605.
79. Myers, A.J., et al., *A survey of genetic human cortical gene expression.* Nat Genet, 2007. **39**(12): p. 1494-9.
80. Riess, O., R. Kruger, and J.B. Schulz, *Spectrum of phenotypes and genotypes in Parkinson's disease.* J Neurol, 2002. **249 Suppl 3**: p. III/15-20.
81. Beutler, E., *Gaucher disease: new molecular approaches to diagnosis and treatment.* Science, 1992. **256**(5058): p. 794-9.
82. Mata, I.F., et al., *Glucocerebrosidase gene mutations: a risk factor for Lewy body disorders.* Arch Neurol, 2008. **65**(3): p. 379-82.
83. Dooling, E.C., W.C. Schoene, and E.P. Richardson, Jr., *Hallervorden-Spatz syndrome.* Arch Neurol, 1974. **30**(1): p. 70-83.
84. Arawaka, S., et al., *Lewy body in neurodegeneration with brain iron accumulation type 1 is immunoreactive for alpha-synuclein.* Neurology, 1998. **51**(3): p. 887-9.
85. Neumann, M., et al., *Alpha-synuclein accumulation in a case of neurodegeneration with brain iron accumulation type 1 (NBIA-1, formerly Hallervorden-Spatz syndrome) with widespread cortical and brainstem-type Lewy bodies.* Acta Neuropathol, 2000. **100**(5): p. 568-74.
86. Galvin, J.E., et al., *Neurodegeneration with brain iron accumulation, type 1 is characterized by alpha-, beta-, and gamma-synuclein neuropathology.* Am J Pathol, 2000. **157**(2): p. 361-8.
87. Hortnagel, K., H. Prokisch, and T. Meitinger, *An isoform of hPANK2, deficient in pantothenate kinase-associated neurodegeneration, localizes to mitochondria.* Hum Mol Genet, 2003. **12**(3): p. 321-7.
88. Hannun, Y.A. and L.M. Obeid, *Principles of bioactive lipid signalling: lessons from sphingolipids.* Nat Rev Mol Cell Biol, 2008. **9**(2): p. 139-50.

89. Morgan, N.V., et al., *PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron.* Nat Genet, 2006. **38**(7): p. 752-4.
90. Hayashi, S., et al., *An autopsy case of late infantile and juvenile neuroaxonal dystrophy with diffuse Lewy bodies and neurofibrillary tangles.* Clin Neuropathol, 1992. **11**(1): p. 1-5.
91. Tang, J., et al., *A novel cytosolic calcium-independent phospholipase A2 contains eight ankyrin motifs.* J Biol Chem, 1997. **272**(13): p. 8567-75.
92. Lei, X., et al., *The group VIA calcium-independent phospholipase A2 participates in ER stress-induced INS-1 insulinoma cell apoptosis by promoting ceramide generation via hydrolysis of sphingomyelins by neutral sphingomyelinase.* Biochemistry, 2007. **46**(35): p. 10170-85.
93. Vanier, M.T. and G. Millat, *Niemann-Pick disease type C.* Clin Genet, 2003. **64**(4): p. 269-81.
94. Tamura, H., et al., *Niemann-Pick type C disease: novel NPC1 mutations and characterization of the concomitant acid sphingomyelinase deficiency.* Mol Genet Metab, 2006. **87**(2): p. 113-21.
95. Saito, Y., et al., *Aberrant phosphorylation of alpha-synuclein in human Niemann-Pick type C1 disease.* J Neuropathol Exp Neurol, 2004. **63**(4): p. 323-8.
96. Dawkins, J.L., et al., *Mutations in SPTLC1, encoding serine palmitoyltransferase, long chain base subunit-1, cause hereditary sensory neuropathy type I.* Nat Genet, 2001. **27**(3): p. 309-12.
97. Hornemann, T., et al., *Cloning and initial characterization of a new subunit for mammalian serine-palmitoyltransferase.* J Biol Chem, 2006. **281**(49): p. 37275-81.
98. Mandell, A.J. and C.K. Smith, *Hereditary sensory radicular neuropathy.* Neurology, 1960. **10**: p. 627-30.
99. Dyck, P.J., P.A. Low, and J.C. Stevens, *"Burning feet" as the only manifestation of dominantly inherited sensory neuropathy.* Mayo Clin Proc, 1983. **58**(7): p. 426-9.
100. Najim al-Din, A.S., et al., *Pallido-pyramidal degeneration, supranuclear upgaze paresis and dementia: Kufor-Rakeb syndrome.* Acta Neurol Scand, 1994. **89**(5): p. 347-52.
101. Paxinou, E., et al., *Induction of alpha-synuclein aggregation by intracellular nitrative insult.* J Neurosci, 2001. **21**(20): p. 8053-61.
102. Lee, H.J., et al., *Clearance of alpha-synuclein oligomeric intermediates via the lysosomal degradation pathway.* J Neurosci, 2004. **24**(8): p. 1888-96.
103. Li, W., et al., *Stabilization of alpha-synuclein protein with aging and familial parkinson's disease-linked A53T mutation.* J Neurosci, 2004. **24**(33): p. 7400-9.
104. Cuervo, A.M., et al., *Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy.* Science, 2004. **305**(5688): p. 1292-5.
105. Vogiatzi, T., et al., *Wild type a-synuclein is degraded by chaperone mediated autophagy and macroautophagy in neuronal cells.* J Biol Chem, 2008.
106. Settembre, C., et al., *A block of autophagy in lysosomal storage disorders.* Hum Mol Genet, 2008. **17**(1): p. 119-29.
107. Kruger, R., et al., *Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease.* Nat Genet, 1998. **18**: p. 106-108.
108. Ibanez, P., et al., *Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease.* Lancet, 2004. **364**(9440): p. 1169-71.
109. Chartier-Harlin, M.C., et al., *Alpha-synuclein locus duplication as a cause of familial Parkinson's disease.* Lancet, 2004. **364**(9440): p. 1167-9.
110. Zarranz, J.J., et al., *The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia.* Ann Neurol, 2004. **55**(2): p. 164-73.

111. Ostrerova-Golts, N., et al., *The A53T alpha-synuclein mutation increases iron-dependent aggregation and toxicity.* J Neurosci, 2000. **20**(16): p. 6048-54.
112. Golovko, M.Y., et al., *Alpha-synuclein gene deletion decreases brain palmitate uptake and alters the palmitate metabolism in the absence of alpha-synuclein palmitate binding.* Biochemistry, 2005. **44**(23): p. 8251-9.
113. Patil, S., J. Melrose, and C. Chan, *Involvement of astroglial ceramide in palmitic acid-induced Alzheimer-like changes in primary neurons.* Eur J Neurosci, 2007. **26**(8): p. 2131-41.
114. van Ham, T.J., et al., *C. elegans model identifies genetic modifiers of alpha-synuclein inclusion formation during aging.* PLoS Genet, 2008. **4**(3): p. e1000027.
115. Bras, J., et al., *G2019S Dardarin Substitution is a Common Cause of Parkinson's Disease in a Portuguese Cohort.* Mov Disord, 2005. **20**(12): p. 1653-1655.
116. Greggio, E., et al., *The Parkinson Disease-associated Leucine-rich Repeat Kinase 2 (LRRK2) Is a Dimer That Undergoes Intramolecular Autophosphorylation.* J Biol Chem, 2008. **283**(24): p. 16906-14.
117. Aasly, J.O., et al., *Clinical features of LRRK2-associated Parkinson's disease in central Norway.* Ann Neurol, 2005. **57**(5): p. 762-5.
118. Funayama, M., et al., *An LRRK2 mutation as a cause for the parkinsonism in the original PARK8 family.* Ann Neurol, 2005. **57**(6): p. 918-21.
119. Zimprich, A., et al., *Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology.* Neuron, 2004. **44**(4): p. 601-7.
120. Gwinn-Hardy, K., et al., *Spinocerebellar ataxia type 3 phenotypically resembling parkinson disease in a black family.* Arch Neurol, 2001. **58**(2): p. 296-9.
121. Subramony, S.H., et al., *Ethnic differences in the expression of neurodegenerative disease: Machado-Joseph disease in Africans and Caucasians.* Mov Disord, 2002. **17**(5): p. 1068-71.
122. Okubadejo, N.U., *An analysis of genetic studies of Parkinson's disease in Africa.* Parkinsonism Relat Disord, 2008. **14**(3): p. 177-82.
123. Gelb, D.J., E. Oliver, and S. Gilman, *Diagnostic criteria for Parkinson disease.* Arch Neurol, 1999. **56**(1): p. 33-9.
124. Berg, D., et al., *Alpha-synuclein and Parkinson's disease: implications from the screening of more than 1,900 patients.* Mov Disord, 2005. **20**(9): p. 1191-4.
125. Hatano, Y., et al., *Novel PINK1 mutations in early-onset parkinsonism.* Ann Neurol, 2004. **56**(3): p. 424-7.
126. *The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations.* Mov Disord, 2003. **18**(7): p. 738-50.
127. Hoehn, M.M. and M.D. Yahr, *Parkinsonism: onset, progression, and mortality. 1967.* Neurology, 2001. **57**(10 Suppl 3): p. S11-26.
128. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* J Psychiatr Res, 1975. **12**(3): p. 189-98.
129. Simon-Sanchez, J., et al., *Genomewide SNP assay reveals mutations underlying Parkinson disease.* Hum Mutat, 2008. **29**(2): p. 315-22.
130. Dachsel, J.C., et al., *Digenic parkinsonism: investigation of the synergistic effects of PRKN and LRRK2.* Neurosci Lett, 2006. **410**(2): p. 80-4.
131. Paisan-Ruiz, C., et al., *Comprehensive analysis of LRRK2 in publicly available Parkinson's disease cases and neurologically normal controls.* Hum Mutat, 2008. **29**(4): p. 485-90.

132. Lesage, S., et al., *LRRK2 haplotype analyses in European and North African families with Parkinson disease: a common founder for the G2019S mutation dating from the 13th century.* Am J Hum Genet, 2005. **77**(2): p. 330-2.
133. Ng, P.C. and S. Henikoff, *Accounting for human polymorphisms predicted to affect protein function.* Genome Res, 2002. **12**(3): p. 436-46.
134. Lucking, C.B., et al., *Coding polymorphisms in the parkin gene and susceptibility to Parkinson disease.* Arch Neurol, 2003. **60**(9): p. 1253-6.
135. Lincoln, S.J., et al., *Parkin variants in North American Parkinson's disease: cases and controls.* Mov Disord, 2003. **18**(11): p. 1306-11.
136. Wood-Kaczmar, A., S. Gandhi, and N.W. Wood, *Understanding the molecular causes of Parkinson's disease.* Trends Mol Med, 2006. **12**(11): p. 521-8.
137. Paisan-Ruiz, C., et al., *Characterization of PLA2G6 as a locus for dystonia-parkinsonism.* Ann Neurol, 2009. **65**(1): p. 19-23.
138. Leroy, E., et al., *The ubiquitin pathway in Parkinson's disease.* Nature, 1998. **395**(6701): p. 451-2.
139. Simon-Sanchez, J. and A.B. Singleton, *Sequencing analysis of OMI/HTRA2 shows previously reported pathogenic mutations in neurologically normal controls.* Hum Mol Genet, 2008. **17**(13): p. 1988-93.
140. Bomford, A., *Genetics of haemochromatosis.* Lancet, 2002. **360**(9346): p. 1673-81.
141. Le Gac, G. and C. Ferec, *The molecular genetics of haemochromatosis.* Eur J Hum Genet, 2005. **13**(11): p. 1172-85.
142. Dekker, M.C., et al., *Mutations in the hemochromatosis gene (HFE), Parkinson's disease and parkinsonism.* Neurosci Lett, 2003. **348**(2): p. 117-9.
143. Buchanan, D.D., et al., *The Cys282Tyr polymorphism in the HFE gene in Australian Parkinson's disease patients.* Neurosci Lett, 2002. **327**(2): p. 91-4.
144. Pankratz, N., et al., *Significant linkage of Parkinson disease to chromosome 2q36-37.* Am J Hum Genet, 2003. **72**(4): p. 1053-7.
145. Pankratz, N., et al., *Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families.* Hum Mol Genet, 2003. **12**(20): p. 2599-608.
146. Pankratz, N., et al., *Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations.* Am J Hum Genet, 2002. **71**(1): p. 124-35.
147. Prestel, J., et al., *PARK11 is not linked with Parkinson's disease in European families.* Eur J Hum Genet, 2005. **13**(2): p. 193-7.
148. Maraganore, D.M., et al., *High-resolution whole-genome association study of Parkinson disease.* Am J Hum Genet, 2005. **77**(5): p. 685-93.
149. Giovannone, B., et al., *Two novel proteins that are linked to insulin-like growth factor (IGF-I) receptors by the Grb10 adapter and modulate IGF-I signaling.* J Biol Chem, 2003. **278**(34): p. 31564-73.
150. Lautier, C., et al., *Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease.* Am J Hum Genet, 2008. **82**(4): p. 822-33.
151. Berlin, D., et al., *Evaluation of HFE (hemochromatosis) mutations as genetic modifiers in sporadic AD and MCI.* Neurobiol Aging, 2004. **25**(4): p. 465-74.
152. Saunders, A.M., et al., *Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease.* Neurology, 1993. **43**(8): p. 1467-72.
153. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
154. Borie, C., et al., *Association study between iron-related genes polymorphisms and Parkinson's disease.* J Neurol, 2002. **249**(7): p. 801-4.
155. Cox, T.M. and J.P. Schofield, *Gaucher's disease: clinical features and natural history.* Baillieres Clin Haematol, 1997. **10**(4): p. 657-89.

156. Capablo, J.L., et al., *Neurological evaluation of patients with Gaucher disease diagnosed as type 1.* J Neurol Neurosurg Psychiatry, 2008. **79**(2): p. 219-22.
157. Sidransky, E., *Gaucher disease: complexity in a "simple" disorder.* Mol Genet Metab, 2004. **83**(1-2): p. 6-15.
158. Horowitz, M., et al., *The human glucocerebrosidase gene and pseudogene: structure and evolution.* Genomics, 1989. **4**(1): p. 87-96.
159. Lachmann, R.H., et al., *Twin pairs showing discordance of phenotype in adult Gaucher's disease.* QJM, 2004. **97**(4): p. 199-204.
160. Beutler, E., et al., *Gaucher disease: gene frequencies in the Ashkenazi Jewish population.* Am J Hum Genet, 1993. **52**(1): p. 85-8.
161. Grabowski, G.A., *Gaucher disease: gene frequencies and genotype/phenotype correlations.* Genet Test, 1997. **1**(1): p. 5-12.
162. Machaczka, M., et al., *Parkinson's syndrome preceding clinical manifestation of Gaucher's disease.* Am J Hematol, 1999. **61**(3): p. 216-7.
163. Bembi, B., et al., *Gaucher's disease with Parkinson's disease: clinical and pathological aspects.* Neurology, 2003. **61**(1): p. 99-101.
164. Tayebi, N., et al., *Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism?* Mol Genet Metab, 2003. **79**(2): p. 104-9.
165. Wong, K., et al., *Neuropathology provides clues to the pathophysiology of Gaucher disease.* Mol Genet Metab, 2004. **82**(3): p. 192-207.
166. Goker-Alpan, O., et al., *Parkinsonism among Gaucher disease carriers.* J Med Genet, 2004. **41**(12): p. 937-40.
167. Gan-Or, Z., et al., *Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset.* Neurology, 2008. **70**(24): p. 2277-83.
168. Goker-Alpan, O., et al., *Glucocerebrosidase mutations are an important risk factor for Lewy body disorders.* Neurology, 2006. **67**(5): p. 908-10.
169. Ziegler, S.G., et al., *Glucocerebrosidase mutations in Chinese subjects from Taiwan with sporadic Parkinson disease.* Mol Genet Metab, 2007. **91**(2): p. 195-200.
170. Stone, D.L., et al., *Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease.* Hum Mutat, 2000. **15**(2): p. 181-8.
171. Braak, H., et al., *Staging of brain pathology related to sporadic Parkinson's disease.* Neurobiol Aging, 2003. **24**(2): p. 197-211.
172. McKeith, I.G., et al., *Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium.* Neurology, 2005. **65**(12): p. 1863-72.
173. Gibb, W.R. and A.J. Lees, *The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease.* J Neurol Neurosurg Psychiatry, 1988. **51**(6): p. 745-52.
174. Mtsui, J., et al., *Mutations for Gaucher disease confer a high susceptibility to Parkinson disease.* Arch Neurol, 2009. **In Press.**
175. Woolf, B., *On estimating the relation between blood group and disease.* Ann Hum Genet, 1955. **19**(4): p. 251-3.
176. Team, R.D.C., *R: A language and environment for statistical computing.* 2005, R Foundation for Statistical Computing.: Vienna, Austria.
177. Rozenberg, R., et al., *Detection of 12 new mutations in Gaucher disease Brazilian patients.* Blood Cells Mol Dis, 2006. **37**(3): p. 204-9.
178. Park, J.K., et al., *The E326K mutation and Gaucher disease: mutation or polymorphism?* Clin Genet, 2002. **61**(1): p. 32-4.
179. Lacerda, L., et al., *Gaucher disease: N370S glucocerebrosidase gene frequency in the Portuguese population.* Clin Genet, 1994. **45**(6): p. 298-300.
180. Amaral, O., et al., *Gaucher disease: expression and characterization of mild and severe acid beta-glucosidase mutations in Portuguese type 1 patients.* Eur J Hum Genet, 2000. **8**(2): p. 95-102.

181. Abou-Sleiman, P.M., et al., *The role of pathogenic DJ-1 mutations in Parkinson's disease.* Ann Neurol, 2003. **54**(3): p. 283-6.
182. Abou-Sleiman, P.M., et al., *A heterozygous effect for PINK1 mutations in Parkinson's disease?* Ann Neurol, 2006. **60**(4): p. 414-9.
183. Zimran, A., et al., *High frequency of the Gaucher disease mutation at nucleotide 1226 among Ashkenazi Jews.* Am J Hum Genet, 1991. **49**(4): p. 855-9.
184. Aarsland, D., J. Zaccai, and C. Brayne, *A systematic review of prevalence studies of dementia in Parkinson's disease.* Mov Disord, 2005. **20**(10): p. 1255-63.
185. Wooten, G.F., et al., *Are men at greater risk for Parkinson's disease than women?* J Neurol Neurosurg Psychiatry, 2004. **75**(4): p. 637-9.
186. Williams, D.R. and A.J. Lees, *Visual hallucinations in the diagnosis of idiopathic Parkinson's disease: a retrospective autopsy study.* Lancet Neurol, 2005. **4**(10): p. 605-10.
187. Harding, A.J., G.A. Broe, and G.M. Halliday, *Visual hallucinations in Lewy body disease relate to Lewy bodies in the temporal lobe.* Brain, 2002. **125**(Pt 2): p. 391-403.
188. Schmitz, M., et al., *Impaired trafficking of mutants of lysosomal glucocerebrosidase in Gaucher's disease.* Int J Biochem Cell Biol, 2005. **37**(11): p. 2310-20.
189. Vogiatzi, T., et al., *Wild type alpha-synuclein is degraded by chaperone-mediated autophagy and macroautophagy in neuronal cells.* J Biol Chem, 2008. **283**(35): p. 23542-56.
190. Beutler E, G.G., *Gaucher Disease*, in *The Metabolic & Molecular Bases of Inherited Disease*, B.A. Scriver C, Sly W, Valle D, Editor. 2001, McGraw-Hill: New York. p. 3635-3668.
191. Wan, L., et al., *Mutation analysis of Gaucher disease patients in Taiwan: high prevalence of the RecNciI and L444P mutations.* Blood Cells Mol Dis, 2006. **36**(3): p. 422-5.
192. Nichols, W.C., et al., *Mutations in GBA are associated with familial Parkinson disease susceptibility and age at onset.* Neurology, 2008.
193. Clark, L.N., et al., *Glucocerebrosidase Mutations are Associated with Dementia with Lewy Bodies.* Arch Neurol, 2009. **In Press**.
194. Farrer, M.J., et al., *Glucosidase-beta variations and Lewy body disorders* Park and related disorders 2008. **In Press**.
195. Segarane, B., et al., *Glucocerebrosidase mutations in 108 neuropathologically confirmed cases of multiple system atrophy.* Neurology, 2009. **72**(13): p. 1185-6.
196. DePaolo, J., et al., *The association between mutations in the lysosomal protein glucocerebrosidase and parkinsonism.* Mov Disord, 2009. **In Press**.
197. Bras, J., et al., *Emerging pathways in genetic Parkinson's disease: Potential role of ceramide metabolism in Lewy body disease.* FEBS J, 2008. **275**(23): p. 5767-73.
198. Zhong, H. and R.L. Prentice, *Correcting "winner's curse" in odds ratios from genomewide association findings for major complex human diseases.* Genet Epidemiol. **34**(1): p. 78-91.
199. Satake, W., et al., *Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease.* Nat Genet, 2009. **41**(12): p. 1303-7.
200. Simon-Sanchez, J., et al., *Genome-wide association study reveals genetic risk underlying Parkinson's disease.* Nat Genet, 2009. **41**(12): p. 1308-12.
201. Latourelle, J.C., et al., *Genomewide association study for onset age in Parkinson disease.* BMC Med Genet, 2009. **10**: p. 98.

202. Pankratz, N., et al., *Genomewide association study for susceptibility genes contributing to familial Parkinson disease.* Hum Genet, 2009. **124**(6): p. 593-605.
203. Holle, R., et al., *KORA--a research platform for population based health research.* Gesundheitswesen, 2005. **67 Suppl 1**: p. S19-25.
204. Wichmann, H.E., C. Gieger, and T. Illig, *KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes.* Gesundheitswesen, 2005. **67 Suppl 1**: p. S26-30.
205. Krawczak, M., et al., *PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships.* Community Genet, 2006. **9**(1): p. 55-61.
206. *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
207. Fung, H.C., et al., *Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data.* Lancet Neurol, 2006. **5**(11): p. 911-6.
208. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
209. Toda, T., *Identification of risk loci in Asian Parkinson's disease.* Submitted, 2009.
210. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
211. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome.* Science, 2002. **296**(5576): p. 2225-9.
212. Maraganore, D.M., et al., *Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease.* JAMA, 2006. **296**(6): p. 661-70.
213. Baker, M., et al., *Association of an extended haplotype in the tau gene with progressive supranuclear palsy.* Hum Mol Genet, 1999. **8**(4): p. 711-5.
214. Scholz, S., et al., *SNCA variants are associated with increased risk of multiple system atrophy.* Annals of Neurology, 2009. **in press**.
215. van der Brug, M.P., et al., *RNA binding activity of the recessive parkinsonism protein DJ-1 supports involvement in multiple cellular pathways.* Proc Natl Acad Sci U S A, 2008. **105**(29): p. 10244-9.
216. Lee, W.C., *Detecting population stratification using a panel of single nucleotide polymorphisms.* Int J Epidemiol, 2003. **32**(6): p. 1120.
217. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
218. Price, A.L., et al., *Discerning the ancestry of European Americans in genetic association studies.* PLoS Genet, 2008. **4**(1): p. e236.
219. Devlin, B. and K. Roeder, *Genomic control for association studies.* Biometrics, 1999. **55**(4): p. 997-1004.
220. Cordell, H.J. and D.G. Clayton, *A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes.* Am J Hum Genet, 2002. **70**(1): p. 124-41.
221. Mueller, J.C., et al., *Multiple regions of alpha-synuclein are associated with Parkinson's disease.* Ann Neurol, 2005. **57**(4): p. 535-41.
222. Mizuta, I., et al., *Multiple candidate gene analysis identifies alpha-synuclein as a susceptibility gene for sporadic Parkinson's disease.* Hum Mol Genet, 2006. **15**(7): p. 1151-8.
223. Chiba-Falek, O. and R.L. Nussbaum, *Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system.* Hum Mol Genet, 2001. **10**(26): p. 3101-9.
224. Scholz, S.W., et al., *SNCA variants are associated with increased risk of multiple system atrophy.* Annals of Neurology, 2009. **in press**.

225. Pittman, A.M., et al., *Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration.* J Med Genet, 2005. **42**(11): p. 837-46.
226. Cruchaga, C., et al., *5'-Upstream variants of CRHR1 and MAPT genes associated with age at onset in progressive supranuclear palsy and cortical basal degeneration.* Neurobiol Dis, 2009. **33**(2): p. 164-70.
227. Webb, A., et al., *Role of the tau gene region chromosome inversion in progressive supranuclear palsy, corticobasal degeneration, and related disorders.* Arch Neurol, 2008. **65**(11): p. 1473-8.
228. Rademakers, R., et al., *High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy.* Hum Mol Genet, 2005. **14**(21): p. 3281-92.
229. Hong, M., et al., *Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17.* Science, 1998. **282**(5395): p. 1914-7.
230. Rademakers, R., et al., *Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia.* Hum Mol Genet, 2008. **17**(23): p. 3631-42.
231. Johansson, A., et al., *TAU haplotype and the Saitohin Q7R gene polymorphism do not influence CSF Tau in Alzheimer's disease and are not associated with frontotemporal dementia or Parkinson's disease.* Neurodegener Dis, 2005. **2**(1): p. 28-35.
232. Healy, D.G., et al., *Tau gene and Parkinson's disease: a case-control study and meta-analysis.* J Neurol Neurosurg Psychiatry, 2004. **75**(7): p. 962-5.
233. Arvanitakis, Z. and Z.K. Wszolek, *Recent advances in the understanding of tau protein and movement disorders.* Curr Opin Neurol, 2001. **14**(4): p. 491-7.
234. Kraft, P., et al., *Beyond odds ratios--communicating disease risk based on genetic profiles.* Nat Rev Genet, 2009. **10**(4): p. 264-9.
235. Hindorff, L.A., et al. *A Catalog of Published GWAS*. 2009 [cited 2009 March 1st]; Available from: www.genome.gov/26525384.
236. Hardy, J. and A. Singleton, *Genomewide Association Studies and Human Disease.* N Engl J Med, 2009. **360**(17): p. 1759-1768.
237. Stranger, B.E., et al., *Population genomics of human gene expression.* Nat Genet, 2007. **39**(10): p. 1217-24.
238. Veyrieras, J.B., et al., *High-resolution mapping of expression-QTLs yields insight into human gene regulation.* PLoS Genet, 2008. **4**(10): p. e1000214.