February, 2010

The Promise of Genome-Wide SNP Genotyping: from Population Genetics to Disease Gene Identification

Sonja Waltraud Scholz, MD

This thesis is submitted to the University College London for the degree of

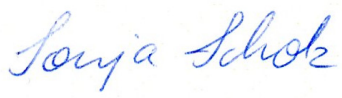
Doctor of Philosophy.

Institute of Neurology
University College London

Laboratory of Neurogenetics
National Institutes of Health

Declaration:

I, Sonja Waltraud Scholz, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



"Indeed, I find the process of doing science, of exploring biological mysteries on a day-to-day basis, deeply rewarding, not only intellectually but also emotionally and socially. Doing experiments gives me the thrill of discovering anew the wonders of the world."

Erich Kandel, In Search of Memory

Abstract

Advances in single nucleotide polymorphism (SNP) genotyping technologies have revolutionised our ability to scrutinise the human genome. My PhD research focuses on using these new technologies to catalogue genetic variability in a collection of diverse populations from around the globe, and to determine the role of genetic variants in neurological diseases. First, I present and discuss the analysis of genome-wide SNP data in individuals from ethnically and geographically diverse human populations uncovering the diversity of genotype, haplotype and copy number variation in these populations. Second, I will describe an autozygosity mapping approach in three Brazilian dystonia-parkinsonism families which lead to the identification of a novel disease-segregating mutation in the gene *PRKRA*. Third, I will report on a large genome-wide association study in Parkinson's disease, uncovering genetic variability at the *SNCA* and *MAPT* loci that are strongly associated with risk for developing disease. Forth, I provide compelling evidence that genetic variants at the *SNCA* locus are also significantly associated with risk for developing multiple system atrophy. This finding represents the first reproducible risk gene for this devastating disorder, and causally links this condition to the more common neurodegenerative disorder Parkinson's disease. Finally, I present the results of a comprehensive mutational screening study investigating the frequency and spectrum of sequence and copy number mutations in the parkinsonism genes *PRKN* and *PINK* in individuals with early-onset Parkinson's disease, in multiple system atrophy patients and in normal controls. In summary, the data presented in this thesis emphasise the critical role that genetic variability plays in the pathogenesis of neurological disorders.

Acknowledgement

I am very grateful to my supervisors Professor Andrew Lees and Professor John Hardy for their excellent mentorship, support, advice and encouragement.

I am deeply indebted to Dr. Andrew Singleton, chief of the laboratory of neurogenetics, who supported and encouraged me in my research aspirations. His patience, tireless enthusiasm, excellent leadership and support enabled me to succeed with my projects.

Thanks to all the present and former members of the laboratory of neurogenetics; special thanks to Dr. Coro Paisan-Ruiz, Dr. Maria del Mar Matarin, Dr. Javier Simon-Sanchez, Janet Brooks, Sampath Arepalli, Angela Britton, Dr. Hon-Chung Fung, Dr. Sarah Camargos, Dr. Jennifer Schymick, Dena Hernandez, Ansha Islam, Dr. Joyce van de Leemput, Dr. Jinhui Ding, Raphael J. Gibbs, Dr. Michael Nalls, Ryan Zonozi, Rita Guerreiro, Jose Bras, Dr. Patrick Lewis, Cynthia Crews and Dr. Mark Cookson.

Furthermore, I owe gratitude to the members of the Department of Molecular Neuroscience and Reta Lila Weston Laboratories, University College London, UK. I am particularly thankful to Dr. Henry Houlden for his support and his friendly, collegial and pragmatic approach towards science. Without his persistent support in the background much of the multiple system atrophy research would have not been feasible. Special thanks also to Abi Li, Reema Paudel, Badmavady Segarane, Dr. Anna Melchers and Dr. Lee Stanyer for their valuable help and contributions.

I am grateful to Professor Tamas Revesz, Dr. Janice Holton and Linda Parsons for providing tissue samples from the Queen Square brain bank.

For support and advice, I want to thank Dr. Rohan de Silva and Professor Nick Woods.

Thanks also to Dr. Noah Rosenberg and his team from the University of Michigan for a marvelous collaboration and inspiring discussions.

Special thanks to our collaborators from the Hertie Institute for Clinical Brain Research in Tuebingen, Germany (Principal Investigator: Dr. Thomas Gasser) and the European Multiple System Atrophy Study Group (EMSA).

For discussions, corrections, advice and patience, I thank Dr. Bryan Traynor.

Finally, I would like to thank my family for their ongoing support and encouragement.

Table of Contents

Abbreviations

A	Adenine
A	Alanine
A	Allelic model
A_1	Minor allele
A_2	Major allele
AACT	*α-1-antichymotrypsin*
ABI	Applied Biosystems
AD	Alzheimer's disease
AD	Autosomal dominant
ADH7	*Alcohol dehydrogenase 7*
Ala	Alanine
Ann.	Annealing
APOE	*Apolipoprotein E*
AR	Autosomal recessive
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ASPE	Allele-specific primer extension
ATF3	*Activating transcription factor 3*
ATF4	*Activating transcription factor 4*
ATP13A2	*ATPase type 13 A2*
ATP1A3	*ATPase, Na+/K+ transporting, alpha-3 polypeptide*
ATXN1	*Ataxin 1*
ATXN3	*Ataxin 3*

BAF	B allele frequency
BDNF	*Brain-derived neurotrophic factor*
bp	Base pairs
C	Cysteine
C	Cytosine
c.	c.DNA sequence
CA	California
CARS	*Cysteinyl t-RNA synthetase*
CBD	Corticobasal degeneration
cDNA	Complementary deoxyribonucleic acid
CEBPB	*CCAAT/enhancer-binding protein-β*
CEPH	Centre d'Etude du Polymorphisme Humain
CEU	European American people with northern and western European ancestry
CGEMS	Cancer Genetics Markers of Susceptibility
CHB	Chinese individuals from Beijing
CHOP	*CCAAT/enhancer-binding protein homologous protein*
Chr.	Chromosome
CI	Confidence interval
CNTF	*Ciliary neurotrophic factor*
CNV	Copy number variant
cPAR %	Combined population attributable risk percent
CT	Computed tomography
Ct	Cycle threshold
CYP1A1	*Cytochrome P450, 1A1*
CYP2D6	*Cytochrome P450, 2D6*
Cys	Cysteine

D	Aspartic acid
D	Dominant model
DAT	*Dopamine transporter 1*
DBH	*Dopamine β-hydroxylase*
Den.	Denaturing
df	Degrees of freedom
DLB	Dementia with Lewy bodies
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
E	Glutamic acid
EDTA	Ethylenediaminetetraacetic acid
EIF4EBP1	*Eukaryotic translation initiation factor 4E-binding protein*
ENCODE	Encyclopedia of DNA elements
et al.	*et alia*
Ext.	Extension
F	Phenylalanine
FAM	6-carboxy-fluorescein
FBXO7	*F-box only protein 7*
FDR	False discovery rate
FGF20	*Fibroblast growth factor 20*
FL	Florida
FMR1	*Fragile X mental retardation 1*
FTD	Frontotemporal dementia
FTDP-17	Frontotemporal dementia with parkinsonism linked to chromosome 17
G	Genotypic model

G	Glycine
G	Guanine
GBA	*Glucocerebrosidase*
GC	Genomic control
GCH1	*GTP cyclohydroxylase 1*
GCI	Glial cytoplasmic inclusion
GD	Gaucher disease
gDNA	Genomic deoxyribonucleic acid
GIGYF2	*Grb10-interacting GYF protein-2*
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GSTM1	*Glutathione S-transferase, mu 1*
GT	Genotype
GTP	Guanosine triphosphate
GWA	Genome-wide association study
H	Histidine
het.	Heterozygote
HGDP	Human genome diversity panel
His	Histidine
HLA	*Human leukocyte antigen*
hom.	Homozygote
I	Isoleucine
IBD	Identical by descent
IBR	In-between RING domain
IBS	Identical by state
ICAM-1	*Intercellular adhesion molecule 1*

IGF1	*Insulin-like growth factor 1*
IL10	*Interleukin 10*
IL1A	*Interleukin 1A*
IL1B	*Interleukin 1B*
IL1RA	*Interleukin 1RA*
IL6	*Interleukin 6*
IL8	*Interleukin 8*
Ile	Isoleucine
IRB	Institutional review board
IVS	Intronic variants
JPT	Japanese subjects from Tokyo
K	Lysine
K	Thousand years
kb	Kilobase(s)
kg	Kilogram(s)
KORA	Cooperative Health Research in the Region of Augsburg
KRS	Kufor-Rakeb syndrome
L	Leucine
LB	Lewy body
LB Broth	Luria-Bertami Broth
LCT	*Lactase*
LD	Linkage disequilibrium
L-DOPA	Levo-3,4-dihydroxyphenylalanine
Leu	Leucine
LRR	Log R ratio
LRRK2	*Leucine rich repeat kinase 2*

LRT	Likelihood ratio test
Lys	Lysine
M	Methionine
MAF	Minor allele frequency
MAPT	*Microtubule-associated protein tau*
Mb	Megabase(s)
MD	Maryland
MDS	Multi-dimensional scaling
Met	Methionine
MI	Michigan
MPTP	1-methyl 4-phenyl 1,2,3,6-tetrahydropyridine
MR1	*Myofibrillogenesis regulator 1*
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MSA	Multiple system atrophy
MSA-C	Multiple system atrophy with predominant cerebellar features
MSA-P	Multiple system atrophy with predominant parkinsonism features
MTS	Mitochondrion transit sequence
N	Asparagine
n	Number of samples
N/A	Not available or not applicable
NAT2	*N-acetyltransferase 2*
NBIA2	Neurodegeneration with brain iron accumulation 2
NC	North Carolina
NCBI	National Center for Biotechnology Information
ng	Nanogram(s)
NINDS	National Institute for Neurological Disorders and Stroke
NJ	New Jersey

nM	Nanomolar
NR4A2	*Nuclear receptor subfamily 4, group A, member 2*
NY	New York
OMIM	Online Mendelian Inheritance in Man
OPCA	Olivopontocerebellar ataxia
OR	Odds ratio
p	*p* value (a statistical measure that estimates the probability of observing a test statistic under the null hypothesis)
p	Prevalence of the risk allele
P	Proline
P	Protective allele
p.	Protein sequence
PAR	Population attributable risk
PCR	Polymerase chain reaction
PD	Parkinson's disease
PDC	paroxysmal dystonic choreoathetosis
PET	Positron emission tomography
PGRN	*Progranulin*
Phe	Phenylalanine
PINK1	*PTEN-induced putative kinase 1*
PLA2G6	*Phospholipase A2, calcium-independent, group IV, A*
POLG	*DNA polymerase gamma*
P-P plot	Probability-probability plot
PPS	Parkinsonian-pyramidal syndrome
PRKN	*Parkin*
PRKRA	*Protein kinase, interferon-inducible double-stranded RNA-dependent activator*

Pro	Proline
PSP	Progressive supranuclear palsy
Q	Glutamine
QC	Quality control
R	Arginine
R	Recessive model
R	Risk allele
Ref.	Reference(s)
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
rs-number	Reference number
S	Serine
S.O.C medium	Super-optimal catabolite repressed medium
SCA 1	*Spinocerebellar ataxia 1*
SCA 2	*Spinocerebellar ataxia 2*
SCA 3	*Spinocerebellar ataxia 3*
SCA 6	*Spinocerebellar ataxia 6*
SCA 7	*Spinocerebellar ataxia 7*
SCA 8	*Spinocerebellar ataxia 8*
SCA 12	*Spinocerebellar ataxia 12*
SCA 17	*Spinocerebellar ataxia 17*
SD	Standard deviation
Ser	Serine
SGCE	*Sarcoglycan epsilon*
SLC1A4	*Solute carrier family 1A4*
SLC2A1	*Solute carrier family 2, member1*
SN	Substantia nigra

SNCA	*α-synuclein*
SNP	Single nucleotide polymorphism
SQSTM1	*Sequestosome 1*
T	Threonine
T	Thymine
T	Trend model
TAF1	*TATA box-binding protein-associated factor 1*
TE	Tris-EDTA buffer
TGFB1	*Transforming growth factor β-1*
TH	*Tyrosine hydroxylase*
THAP1	*Thanatos-associated protein domain-containing apoptosis-associated protein 1*
Thr	Threonine
TNF	*Tumor necrosis factor α*
TOR1A	*Torsin-A*
Trp	Tryptophan
Tyr	Tyrosine
UBL	Ubiquitin-like domain
UCH-L1	*Ubiquitin carboxyl-terminal hydrolase L1*
UK	United Kingdom
US	United States
USA	United States of America
UTR	Untranslated Region
V	Valine
v.	version
Val	Valine

VIC	2’-chloro-7’-phenyl-1,4-dichloro-6-carboxy-fluorescein
VNTR	Variable number of tandem repeats
W	Tryptophan
XR	X-chromosomal recessive
Y	Tyrosine
YRI	Yoruba people of Ibadan, Nigeria
µg	Microgram(s)
µl	Microliter(s)
µM	Micromolar

Publications

Parts of this thesis have been published in the following manuscripts:

Fung HC*, **Scholz S***, Matarin M*, Simon-Sanchez J*, Hernandez D, Britton A, Gibbs JR, Langefeld C, Stiegert ML, Schymick J, Okun MS, Mandel RJ, Fernandez HH, Foote KD, Rodriguez RL, Peckham E, De Vrieze FW, Gwinn-Hardy K, Hardy JA, Singleton A. **Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data.** Lancet Neurology 2006, 5(11):911-916.

Camargos S*, **Scholz S***, Simon-Sanchez J*, Paisan-Ruiz C*, Lewis P, Hernandez D, Ding J, Gibbs JR, Cookson MR, Bras J, Guerreiro R, Oliveira CR, Lees A, Hardy J, Cardoso F, Singleton AB. **DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress response protein prkra.** Lancet Neurology 2008, 7(3):207:215.

Jakobsson M*, **Scholz SW***, Scheet P*, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez D, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB. **Genotype, haplotype and copy number variation in worldwide human populations.** Nature 2008, 451(7181):998-1003.

Scholz S, Singleton A. **Susceptibility Genes in Movement Disorders.** Movement Disorders 2008, 23(7):927-934.

Brooks J, Ding J, Simon-Sanchez J, Paisan-Ruiz C, Singleton A, **Scholz SW.** ***Parkin*** **and** ***PINK1*** **mutations in early-onset Parkinson's disease: comprehensive screening in publicly available case-control cohorts**. Journal of Medical Genetics 2009, 46(6):375-381.

Scholz SW*, Houlden H*, Schulte C*, Sharma M*, Li A, Berg D, Melchers A, Paudel R, Gibbs JR, Simon-Sanchez J, Paisan-Ruiz C, Bras J, Ding J, Chen H, Traynor BJ, Arepalli S, Zonozi RR, Revesz T, Holton J, Wood N, Lees A, Oertel W, Wüllner U, Goldwurm S, Pellecchia MT, Illig T, Riess O, Fernandez HH, Rodriguez RL, Okun MS, Poewe W, Wenning GK, Hardy JA, Singleton AB, Gasser T. ***SNCA*** **variants are associated with increased risk for multiple system atrophy.** Annals of Neurology 2009, 65(5):610-614.

* these authors contributed equally

Scholz SW. ***PINK1*** **mutations: does the dosage make the poison?** Human Mutation 2009, 30(11):v.

Simon-Sanchez J*, Schulte C*, Bras JM*, Sharma M*, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, **Scholz SW**, Hernandez DG, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA, Illig T, Gieger C, Houlden H, Steffens M, Okun MS, Racette BA, Cookson M, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D, Deuschl G, Chen H, Riess O, Hardy JA, Singleton AB, Gasser T. **Genome-wide association study reveals genetic risk underlying Parkinson's disease.** Nature Genetics 2009, 41(12):1308-1312.

Brooks JA*, Houlden H*, Melchers A, Islam AJ, Ding J, Li A, Paudel R, Revesz T, Holton JL, Wood N, Lees A, Singleton AB, **Scholz SW**. **Mutational analysis of** ***parkin*** **and** ***PINK1*** **in multiple system atrophy.** Neurobiology of Aging, *in press.*

* these authors contributed equally

Tables

Figures

1 Introduction

1.1 Specific Aims of this Thesis

During the last decade, genomic research has made substantial progress. The availability of the human genome sequence, as well as detailed data on genomic variation and advances in biotechnology have provided unprecedented opportunities to investigate genetic factors involved in human health and disease. The ability to rapidly test several hundred thousands of single nucleotide polymorphisms (SNPs) at a relatively affordable price revolutionised the genomic landscape. My doctoral thesis is a reflection of this exciting advancement.

My thesis is designed to study two related areas of genomic research. In the first instance, I will use genome-wide genotype data to study individuals with diverse ethnic background in an attempt to examine genomic patterns of variation and to uncover human population origins. The availability of such genome-wide data provides an essential resource for genetic studies in diverse worldwide populations. Secondly, I will use whole genome SNP data in an attempt to uncover loci that cause or contribute to neurological diseases, especially movement disorders. These two areas of research are complimentary, as it is only through a firm understanding of the normal genetic architecture of ethnically diverse populations that we can successfully dissect the genetic aetiology of human disease.

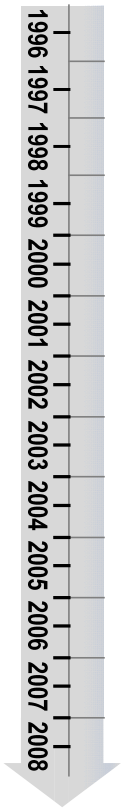
1.2 The Promise of Genome-Wide SNP Genotyping

1.2.1 Background

Over the past several years, progress in genomic technology, research concepts and efforts to catalogue genetic variation have laid the foundation for genome-wide association (GWA) studies as a novel approach to identify genetic variability underlying

risk for disease. The basic tenet is to study genome-wide SNP genotype information in large cohorts of cases with a specific trait, in particular a disease trait, and controls. At its most basic level, χ^2-statistics are then calculated for each SNP to identify significant allele or genotype frequency differences between cases and controls.

Figure 1. Timeline of Genomic Milestones



1996	Risch N. and Merikangas K. demonstrate that genome-wide association tests have a higher power than linkage studies to detect common variants with low odds ratios[1]
2001	Publication of the first drafts of the human genome[2, 3]
2002	Launch of the International HapMap Project
2003	Completion of the Human Genome Project
2005	Using a whole genome association approach, Klein et al. discover a polymorphism in complement factor H (CFH) that causes age-related macular degeneration[4]
2005	Completion of phase I of the International HapMap Project[5]
2007	Publication of the pilot phase of the ENCODE Project[6]
2007	Completion of phase II of the International HapMap Project[7]
2008	Launch of the 1000 Genomes Project[8]
2008	Launch of the pilot phase of the Genotype-Tissue Expression (GTEx) project[9]

The concept of genome-wide SNP genotyping was first proposed by Neil Risch and Kathleen Merikangas as a mechanism to discover common disease-associated variants in complex disorders (figure 1). This idea is based on the 'common disease – common variant' hypothesis, which suggests that common modest-risk alleles confer susceptibility to common disease. Risch and Merikangas demonstrated that GWA studies have a higher power than linkage studies to detect common variants with low effect size.[1]

Before GWA studies became feasible, strategies for identifying disease genes focused on mapping microsatellite markers in familial traits following classic Mendelian modes of inheritance. This approach is called linkage analysis and is still incredible successful in mapping disease-related loci of familial disorders. However, there are considerable limitations of this approach: (1) genes with low penetrance are likely to be missed, (2) the genes found in linkage studies are not always relevant to the common, complex form of disease, and (3) rare families are difficult to ascertain. GWA studies, in contrast, offer a novel strategy to identify disease-related genes because large cohorts of unrelated cases and controls are studied and therefore sample collection is remarkably simplified. In addition, GWA studies do not rely on *a priori* knowledge on the pattern of inheritance and furthermore the same sample size yields a higher power in detecting genes with modest effects.

As an alternative strategy, but less successful, candidate gene association studies were widely applied in the search for disease genes. Candidate gene studies are based on testing variants in genes that have been implicated with disease based on pathobiological understanding. The advantage of GWA studies over candidate studies is that no knowledge on the pathophysiological links with disease is necessary and even variants in non-coding, regulatory regions of the genome, which would most likely be missed in a candidate gene approach, are tested in a genome-wide manner.

Ambitious efforts to catalogue the human genome sequence and common genetic variants were a necessary prerequisite for genome-wide SNP genotyping. In 1990, the Human Genome Project, an international collaboration to sequence the human genome, was launched and the first genome drafts were published in 2001. This project represented a major step forward in decoding the cellular programs that determine human health and disease and as former US-president Bill Clinton described it at the White House ceremony:

"Without a doubt, this is the most important, most wondrous map ever produced by humankind."

Bill Clinton, June 26, 2000

This comprehensive sequence map paved the way to study genetic variants that determine differences in cellular processes, drug responses and treatment options. As a consequence, the International HapMap Consortium was founded in 2002 to determine common patterns of DNA sequence variation in the human genome, by characterising sequence variants, their frequencies, and correlations between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe.[10] One of the other specific aims of the HapMap Project was to stimulate technology to make SNP genotyping faster, more reliable, and above all cheaper. This has spurred technological advancement so that multiple different platforms now exist (Illumina, Affymetrix, Nimblegen, Perlegen, Agilent). The net result of this is that the cost of SNP genotyping has fallen to <0.05 cents per SNP and continues to fall. Three years later, phase I of the HapMap project, cataloguing three million common SNPs in the human genome, was published and GWA testing became a reality in distinguished research laboratories.[5]

One of the first published GWA studies demonstrated the power of this novel approach to find disease genes. In 2005, Klein et al. reported the results of a GWA scan testing around 100,000 SNPs in only 96 patients with macular degeneration, a common cause of blindness in the elderly population, and 50 controls. Despite the small sample size of this study, significantly associated SNPs in the complement factor H gene were discovered and subsequent replication studies confirmed this finding.[4, 11, 12] The success of this study was surprising because the effect sizes of the discovered variants were much higher than the expected effect sizes in most complex diseases. For the risk allele of the SNP with the lowest *p* value the odds ratio (OR) for homozygous carriers was 7.4, for heterozygous carriers 4.6. As a comparison, the ORs in most complex disorders, such as hypertension, diabetes, obesity and others, are between 1.1 and 1.5.[13-16] For most

complex disorders a much higher sample size, involving hundreds or thousands of samples, is therefore required to yield sufficient power.

While the Human Genome Project and the HapMap project were major scientific achievements, many critical questions cannot be answered without a clear understanding of functional elements regulating transcriptional activity, DNA structure and function. The Encyclopedia of DNA Elements (ENCODE) project was initiated by the National Human Genome Research Institute, part of the National Institutes of Health, to do just that. Computational and experimental analyses of the ENCODE pilot phase analyzing 1% of functional elements in the human genome were published in 2007.[6] This study provided intriguing insights into the functional organization of the genome and demonstrated that regulatory mechanisms are far more complicated and challenging than researchers had thought.

"This impressive effort has uncovered many exciting surprises and blazed the way for future efforts to explore the functional landscape of the entire human genome."

Francis Collins, press release on ENCODE pilot phase

The rapid pace of progress in genomic research continues. Biotechnological advances provide cheaper and faster solutions for genotyping and sequencing and, as a consequence, projects of unprecedented scales, such as the 1000 Genomes Project and the Genotype-Tissue Expression (GTEx) project, are underway.[8, 9] In addition, a growing bioinformatics community offers sophisticated tools to parse, share and analyse enormous genome-wide datasets. These ongoing developments provide novel insights and better approaches to understand the genetic determinants of human health and disease.

In the next section, I will briefly outline some challenges and possible alternative applications for using genome-wide genotype data.

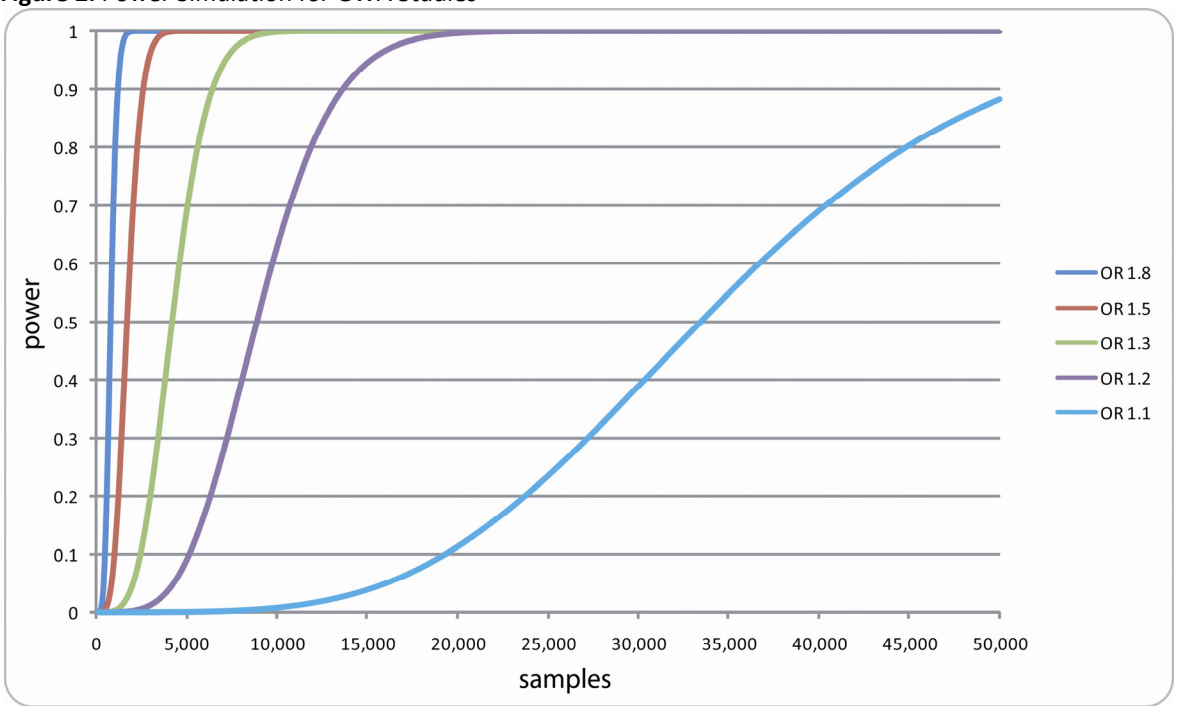
1.2.2 Challenges and Pitfalls of GWA Studies

Much of the discussion surrounding GWA studies is centered on the interpretation of association test results. To understand the challenges, limitations and possible false interpretations, several critical issues have to be considered:

Sample Size and Effect Size

The number of required cases and controls is determined by the expected effect size of a genetic variant underlying a specific trait and by the minor allele frequency (MAF) of this variant (figure 2).[17]

Figure 2. Power Simulation for GWA Studies



This graph shows the relationship between sample size and power for five different effect sizes. For this simulation a minor allele frequency of 0.1 and a *p* value of 1×10^{-7}, based on the Bonferroni corrected significance threshold for testing 500,000 SNP markers, were chosen under an additive model. With respect to the sample size, an equally matched case-control cohort was assumed. For example, this plot shows that a GWA study with a sample size of 5,000 subjects (2,500 cases and 2,500 controls), testing 500,000 SNP markers in each individual, has an 80% power to detect variants with an odds ratio of 1.3.

Common disorders are thought to be caused by a combination of common risk factors with modest effects. Successful GWA studies, such as studies in type 2 diabetes and breast cancer, have used thousands of samples to yield sufficient power for the

detection of risk factors with modest effects (OR 1.2-2).[16, 18] For most diseases it is therefore realistic to aim at a minimum sample size of 1,000 cases and 1,000 controls.

Correction for Multiple Testing and Interpretation of GWA Study Results

Much of the confusion about the interpretation of GWA study results stems from the fact that hundreds of thousands of tests are performed. It is well known that the possibility of chance findings increases dramatically with the number of tests performed (i.e. false positive findings due to multiple testing). For example, if 1,000,000 SNPs are tested for association with a specific trait at a *p* value of 0.05, it is expected that approximately 50,000 SNPs will be associated purely by chance. Consequently, stringent corrections for multiple testing have to be applied. Bonferroni correction, the most common correction method, is calculated by dividing the pre-determined significance value α with the number of independent tests performed (e.g. if 500,000 tests are performed the Bonferroni-corrected α' would be 1×10^{-7} [two-sided α of 0.05 divided by 500,000 SNPs tested]). The problem with Bonferroni correction is that it is overly conservative as it assumes that all tests are independent and therefore does not take linkage disequilibrium (LD) between neighboring SNPs or multiple adjacent genes implicated in given metabolic pathways into account.[19] This correction method leads to a high false negative error rate (type II errors) and decreased power to identify truly disease-associated SNPs. Alternative correction methods, such as the false discovery rate (FDR) method and permutation testing have been devised to minimise false positive and false negative errors. Briefly, the FDR method estimates the expected proportion of false rejections of the null hypotheses, whereas permutation testing is a simulation-based resampling technique that compares observed *p* values with *p* values calculated from simple repeated perturbations of the data.[19] Both methods have the advantage of being less conservative than Bonferroni corrections but on the negative side are more computationally intensive and require programming skills. With growing expertise in the GWA studies research community and the availability of advanced user-

friendly bioinformatics tools, these corrections are now becoming the gold standard in GWA scans.

Phenotype Characterisation

An important aspect for GWA studies is the availability of well-characterised, large cohorts of cases and controls. International efforts embarking on sample collection, detailed phenotypic characterisation involving biochemical, physiological and clinical tests, as well as public availability of these samples are necessary to ensure a maximum benefit from genome-wide genotyping and sequencing projects. It is furthermore critical that data on well phenotyped control samples are publicly available to the entire research community to make GWA studies more cost effective.

Allelic Heterogeneity

Allelic heterogeneity poses an additional problem, which is when different disease causing mutations occur in the same gene. Each allele might only contribute a small effect size and, unless those alleles are in LD with each other, large sample numbers are therefore required to detect them.

Population Stratification

It is well known that population substructure (i.e. allele frequency differences between cases and controls due to systematic ancestral differences) can produce spurious associations in disease studies.[20] A number of tools have been devised to deal with population stratification in association studies: (a) EIGENSTRAT,[20] which uses a principal component analysis to model ancestry differences between cases and controls along continuous axes of variation; (b) STRUCTURE,[21] an algorithm that assigns subpopulation clusters; and (c) genomic control (GC),[22] a method that adjusts for population stratification using a uniform inflation factor. Of those three methods, the GC method is the easiest and is therefore most commonly applied. However, GC decreases the power to detect disease-associated variants as it assumes that allele frequency differences

from ancestral populations are uniform across the genome. In populations with pronounced substratification, both STRUCTURE and EIGENSTRAT, although computationally more intense, are therefore preferred methods.

1.2.3 More than just Association Tests

The ability to determine SNP genotypes in a genome-wide manner has many potential applications. Most obviously SNP chips enable the detection of disease genes through association testing. Beyond this application, genome-wide SNP data can furthermore be analysed with respect to copy number variation, extended tracks of homozygosity (which is particularly relevant for recessive disorders), relatedness between samples, population structure and to reconstruct human evolutionary history. Through comparative genomics, we are furthermore able to extend our understanding on evolutionary mechanisms to different species. Moreover, pathway and network-based analyses enable the detection of variants with modest disease association influencing particular cellular pathways. Indeed, genome-wide SNP genotyping is already having a big impact on our understanding of the human evolution and the genomic mechanisms involved in disease, as I will describe in detail.

1.3 Genetic Research in Diverse Human Populations

Examination of human genetic diversity is a fundamental prerequisite for understanding the genetic mechanisms that contribute to human health and disease. Much has been learned about human evolution from complementary disciplines such as anthropology, linguistics, archeology, as well as studies of the demographic, geographic, cultural, and biological determinants that influenced human evolution. Recently, genomic research has provided crucial insights into the evolutionary events that gave rise to distinct genomic patterns of human population structure. Genetic relationships among populations can now be reconstructed by comparing allele frequencies from genome-wide SNP data; in addition, genomic selection signatures and copy number variants can be identified and thus provide important clues how humans adapted to changing

climatic and nutritional environments. Some of these adaptations have important medical relevance, as they can be linked to differential disease susceptibility.[23-25]

1.3.1 Evolutionary Forces Shape Phenotypic Diversity

It is well known that some traits, such as the blood group protein markers or certain diseases, occur at different frequencies in some ethnic groups. The two main evolutionary forces that shaped these traits are: (1) natural selection and (2) random genetic drift. A good example for the selection phenomenon is sickle cell anemia in African populations. Individuals who are heterozygous for the hemoglobin S allele have a greater resistance to malaria and therefore a survival advantage over individuals carrying two wild-type hemoglobin genes. In contrast, persons who are homozygous for hemoglobin S develop severe sickle cell anemia. This results in a disproportionately high frequency of the hemoglobin S gene in malaria-infested regions, but, because homozygous hemoglobin S carriers have serious complications, the hemoglobin S allele is balanced by wild-type hemoglobin.[26] Random genetic drift on the other hand is a random change in allele frequencies that results from random mutations. Perhaps a good example for genetic drift is the high frequency of Tay-Sachs disease, a fatal autosomal recessive lipid storage disorder, among Ashkenazi Jewish people; however, some researchers have suggested that this might also be explained by selection.[27] They hypothesised that heterozygous risk allele carrier are resistant to tuberculosis infections, although statistical analyses do not support this theory.[28, 29]

1.3.2 Interpreting Evolutionary Signatures

There is growing realisation that using genomic information to detect signatures of human evolutionary origins also has a number of potentially useful biomedical applications. One important application is to study information on non-random association of variants, called linkage disequilibrium (LD), for making inferences on human migration, selection and recombination events. The basic idea behind this kind of research is that disequilibrium of associated variants decays over time, with the

number of recombinational events being directly correlated to the age of a founding population. For instance, the average length of autosomal LD is markedly lower in African populations than in non-African populations, suggesting that non-African populations represent a small genetic subset of the Africans.[30-34] By comparing the patterns of LD and the age of populations we can therefore make inferences on migrations and population expansions. Of particular interest in biomedical research are regions in the human genome with high interpopulation differentiation; these regions commonly represent signatures of selection and adaptation and have important implications for disease-mapping.

1.3.2 Resources for Human Genome Diversity Research

Over the past few decades, human genetic research has made enormous leaps forward. Many valuable insights into the genetic underpinnings influencing human health and disease, response to drugs and environmental factors have been gained. However, for the most part, participants from Caucasian populations have been studied. The bias against non-Caucasian populations raises many concerns as ethnic background can influence the disease phenotype and treatment outcome. To understand the implications of genes in human health and disease, it is thus imperative for future genetic research to catalogue the similarities and differences in the human genome across diverse human populations. The International HapMap Project and the Human Genome Diversity Panel (HGDP) are important resources for this type of research. Both projects aim at studying the genetic diversity within and among various human populations and are complementary in their goals. The HGDP consists of 1063 cultured lymphoblastoid cell lines from individuals in 51 different world populations and is stored at the Centre d'Etude du Polymorphisme Humain in Paris, France (www.cephb.fr/en/hgdp). This resource was created to provide unlimited supplies of DNA for studies of sequence diversity and history of modern human populations to non-profit research laboratories.[35, 36] The main aim of the International HapMap Project is to develop a detailed haplotype map of the human genome, the HapMap, describing the

common patterns of human DNA sequence variation (www.hapmap.org).[5] Both projects, however, raise a number of ethical concerns. The principal issues revolve around informed consent, the confidentiality for study participants, the inclusion of individuals from indigenous populations and the possibility to fuel discrimination based on distinct genetic findings. Each concern has its legitimate basis and mandates a diligent ethical oversight. Detailed information that is culturally sensitive, voluntary participation and informed consent following strict ethical guidelines are required for research in genetically diverse populations.[37] It is important to highlight that first results from the HapMap and the HGDP projects clearly demonstrate that human genetic diversity is small between populations and represent the strongest argument that there is no scientific basis for racism.[38]

1.4 Genetics of Selected Movement Disorders

In this thesis, I demonstrate the utility of genome-wide SNP data for the discovery of genomic loci implicated in movement disorders and for alternative applications. In the next section, I will briefly introduce the current status of genetic research in three movement disorders, which I have studied for this thesis: dystonia, Parkinson's disease and multiple system atrophy.

1.4.1 Genetics of Dystonia

Dystonia is defined as a neurological disorder characterised by sustained involuntary muscle contraction, resulting in twisting and abnormal posture. Dystonic disorders are a complex group of syndromes that are divided into two main groups: (a) primary dystonias (which are defined as dystonic disorders that cannot be explained by environmental insults to the brain), and (b) secondary dystonias (which are defined as dystonic disorders that develop as the result of environmental factors, such as brain trauma, tumor or infection, that affect the brain).[39]

The pathogenesis of primary dystonia is poorly understood, although family studies have lead to the identification of nine disease genes (table 1) and 18 disease-associated loci.

DYT1: TOR1A Mutations

The dystonia gene *TOR1A*, encoding for the protein torsin A, was identified in 1997 by Ozelius and colleagues.[40] Most patients with DYT1-related dystonia have a 3bp in-frame GAG deletion in exon 5 resulting in the loss of a glutamic acid in a conserved region of torsinA. This mutation was first described in Ashkenazi Jewish patients. Screening studies demonstrated that this mutation is also found in other populations.[41-43] Two more pathogenic variants in the *TOR1A* gene – an 18bp deletion,[44] and a 4bp deletion[45] – have been described. The mode of inheritance is autosomal dominant with a penetrance of approximately 30%. Interestingly, the polymorphism Asp216His in *TOR1A* has been reported to modify disease penetrance with the histidine allele in *trans* being protective whereas the aspartic acid in *cis* seemingly required for disease penetrance.[46]

DYT3: TAF1 Mutations

Mutation of the gene *TATA box-binding protein-associated factor 1 (TAF1)* underlies X-linked recessive dystonia-parkinsonism in Filipino men (this disease is also known as Lubag). All patients harbor a 2.6kb retrotransposon insertion in intron 32 of *TAF1*.[47]

DYT5a: GCH1 Mutations

Mutations in *GTP cyclohydroxylase 1 (GCH1)* lead to autosomal dominantly inherited, L-DOPA-responsive dystonia.[48] The phenotype of patients with *GCH1* mutations is characterised by childhood-onset dystonia with a tendency to generalise, marked diurnal fluctuations with worsening of disease symptoms in the evening, and a female predominance. Men are often asymptomatic carriers. This might be explained by a higher GCH1 enzyme activity in males that is thought to be protective.[48]

DYT5b: TH Mutations

Tyrosine hydroxylase (TH) mutations have been identified in families with autosomal recessive dystonia-parkinsonism.[49] The phenotype is characterised by a disease onset during infancy including hypomimia, oculogyric crisis, tremor, truncal hypotonia, and limb dystonia. The symptoms show marked response to L-DOPA treatment.[50]

DYT6: THAP1 Mutations

Mutations in the *THAP domain-containing protein 1 gene (THAP1)* have recently been identified in three Amish-Mennonite families and a German family with adolescence onset primary torsion dystonia.[51] Subsequent screening studies analysing *THAP1* in familial dystonia patients have shown that *THAP1* mutations are found in a substantial proportion of dystonia cases, in particular in individuals with European non-Jewish ancestry.[52,53] Female preponderance, reduced lifetime disease penetrance and involvement of the craniocervical muscles appear to be common features of *THAP1*-related dystonia.[52, 54]

DYT8: MR1 Mutations

In families with autosomal dominant, paroxysmal dystonic choreoathetosis (PDC), genetic linkage has been demonstrated with a locus on chromosome 2q.[55, 56] Using candidate gene sequencing Rainier and colleagues discovered nonsynonymous disease-segregating mutations (p.Ala9Val and p.Ala7Val) in the *myofibrillogenesis regulator 1 (MR) gene*.[57] These findings have subsequently been confirmed in other PDC families.[58, 59] Haplotype analysis suggests that the mutations arose independently in various families and are not attributed to a common founder.[60]

DYT11: SGCE Mutations

Mutations in the gene *sarcoglycan epsilon (SGCE)* have been demonstrated to cause myoclonus-dystonia, an autosomal dominant dystonic syndrome characterised by

bilateral myoclonic jerks, dystonia and psychiatric disturbances.[61, 62] The symptoms improve on ingestion of alcohol.

DYT12: ATP1A3 Mutations

In families with autosomal dominantly inherited rapid-onset dystonia-parkinsonism de Carvalho Aguiar and colleagues discovered six different disease-segregating missense mutations in *ATP1A3*.[63] Rapid-onset dystonia-parkinsonism is characterised by a sudden onset (hours to a few weeks) of dystonic spasms and parkinsonism which do not improve with L-DOPA therapy.[64]

DYT18: SLC2A1 Mutations

Mutations in the *SLC2A1* gene encoding the glucose transporter GLUT1 have been demonstrated to cause childhood-onset paroxysmal exertion-induced dystonia.[65] Additional signs may include hemolytic anemia, migraine, epilepsy, developmental delay and ataxia. The symptoms improve with intravenously administered glucose and with permanent ketogenic diet.

Table 1. Monogenic Forms of Dystonia

Gene	Locus	Chromosome position	Inheritance	Mutation(s)	Phenotype	Ref.
TOR1A	DYT1	9q34	AD	deletion	usually childhood/adolescence-onset focal limb dystonia, often generalises	[40]
TAF1	DYT3	Xq13	XR	insertion	male patients from Philippines with segmental or generalised dystonia and often parkinsonism	[47]
GCH1	DYT5a	14q22.1-q22.2	AD	point mutations deletion	L-DOPA responsive dystonia with diurnal fluctuations	[48] [49]
TH	DYT5b	11p15.5	AR	point mutations deletion	L-DOPA responsive infantile parkinsonism	
THAP1	DYT6	8p11.21	AD	insertion-deletion point mutations	adolescence-onset primary torsion dystonia, craniocervical dystonia, dysarthria, dysphagia	[51, 52]
MR1	DYT8	2q35	AD	point mutations	paroxysmal dystonic choreoathetosis	[57]
SGCE	DYT11	7q21	AD	point mutations deletions	alcohol-responsive myoclonus-dystonia	[62]
ATP1A3	DYT12	19q12-q13.2	AD	point mutations	acute or subacute onset of dystonia-parkinsonism	[63]
SLC2A1	DYT18	1p35-p31.3	AD	point mutations deletion	childhood-onset paroxysmal exertion-induced dystonia, epilepsy, improvement with ketogenic diet	[65]

Parts of this table have been published elsewhere (see reference [66]). AD, autosomal dominant; AR, autosomal recessive; *ATP1A3, ATPase, Na+/K+ transporting, alpha-3 polypeptide; GCH1, GTP cyclohydroxylase 1; MR1, myofibrillogenesis regulator 1; PLA2G6, phospholipase A2 group 6; SGCE, sarcoglycan epsilon; SLC2A1, solute carrier family 2, member 1; TAF1, TATA box-binding protein-associated factor 1*; *TH, tyrosine hydroxylase; THAP1, thanatos-associated protein domain-containing apoptosis-associated protein 1; TOR1A, Torsin-A;* XR, X-chromosomal recessive.

1.4.2 Genetics of Parkinson's Disease

Parkinson's disease (PD; OMIM #168600) is the second most common neurodegenerative disorder after Alzheimer's disease, affecting approximately 1% of the population over the age of 65 years.[67] PD is characterised clinically by bradykinesia, resting tremor, rigor, postural instability and marked response to levo-dopamine (L-DOPA). Pathologically, brains from PD patients show α-synuclein-positive neuronal inclusions in the brainstem (known as Lewy bodies [LBs]) and degeneration of dopaminergic neurons in the substantia nigra pars compacta.[68]

The aetiology of PD is only incompletely understood. Of the known risk factors for PD, age is the main factor. Other risk factors include: environmental factors (such as exposure to pesticides, herbicides, or 1-methyl 4-phenyl 1,2,3,6-tetrahydropyridine [MPTP]) and genetic factors (tables 2 and 3).[69, 70] Mendelian forms of PD are rare and usually have an earlier disease manifestation than sporadic PD. Twin studies have been used to estimate the genetic contribution to the pathogenesis and several of them showed low concordance rates in monozygotic and dizygotic twins.[71-73] These results were controversial, the major criticism being that the cross-sectional study designs used did not exclude the possibility of a later disease onset in unaffected siblings. This obstacle has been overcome by using positron emission tomography studies (PET), which is sufficiently sensitive to identify pre-symptomatic subjects by detecting decreased striatal [18F]6-fluorodopa as a metric for decreased striatonigral dopaminergic function.[74, 75] Based on PET scan data, the concordance rate was significantly higher for monozygotic twins than for dizygotic twins (55% versus 18%), suggesting a substantial genetic contribution to the PD pathogenesis.[75]

To date, 15 PD-associated loci (PARK1-15) have been described (table 2). In PARK1 (PARK4), PARK2, PARK5, PARK6, PARK7, PARK8, PARK9, PARK14 and PARK15 the underlying Mendelian inheritance patterns and their corresponding gene mutations

have been identified (table 2). In addition, a growing number of susceptibility genes rather than disease-causing genes have been reported (see table 3).

PARK1 and PARK4: SNCA Mutations

Polymeropoulos and colleagues identified the first PD gene in 1997.[76] In a large family of Italian ancestry (Contursi kindred) with autosomal dominant PD, genetic linkage was identified on chromosome 4q21-23.[77] Sequence analysis of a candidate gene in this region, *α-synuclein (SNCA)*, revealed a disease-segregating missense mutation in exon 4 (p.A53T) in this family and in three unrelated Greek PD families.[76] Two other *SNCA* mutations - p.A30P and p.E46K - were later identified in a German and a Spanish PD family respectively.[78, 79] Soon after the discovery of *SNCA* mutations in PD, abnormally phosphorylated fibrillar α-synuclein has been found to be the primary structural component of LBs.[80] More recently, it has been shown that *SNCA*-related PD can also be caused by a dosage effect, whereby the entire *SNCA* gene is either duplicated or triplicated.[81, 82] The identification of this first PD gene had many implications on the PD genetics field. It provided the basis for most of the animal- and cell-based work that has been performed over the last years and furthermore it demonstrated that studying rare familial forms of disease had direct relevance to typical sporadic PD. There is growing evidence showing association of *SNCA* promoter variants with sporadic PD.[83-85] It is therefore likely that a disturbed regulation of *SNCA* plays a critical role in sporadic PD.

PARK2: PRKN Mutations

Autosomal recessive early-onset PD with linkage to chromosome 6q was first recognised in Japanese families and subsequently mutations in *PRKN*, encoding the protein parkin, have been identified by positional cloning.[86, 87] *PRKN* is one of the largest genes in the human genome, spanning approximately 1.4Mb, and is located in a highly unstable genomic region.[88] Screening studies demonstrated that *PRKN* mutations (including point mutations, insertions and exon rearrangements) are common worldwide and

probably cause up to 50% of early-onset PD cases (disease onset before the age of 50).[89-91]

PARK5: UCHL1 Mutations

Using a candidate gene approach, Leroy et al. described a small German PD family with a disease-segregating missense mutation (p.I93M) in *ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1)*. Subsequent replication efforts have failed to confirm *UCH-L1* as a disease-causing gene in familial PD.[92, 93] The small size of the original family and the lack of replication indicate that this *UCH-L1* mutation is either a rare cause of PD or is not a disease-causing mutation.

PARK6: PINK1 Mutations

Linkage to a region on chromosome 1p36 was first described in European families with autosomal recessive early-onset PD.[94, 95] Candidate gene sequencing in the linked region led to the identification of a disease-segregating missense mutation (p.G309D) in *PTEN-induced putative kinase 1 (PINK1)*.[96] Subsequent screening studies demonstrated that *PINK1* mutations represent the second most frequent cause of early-onset Parkinson's disease after *PRKN* mutations.[97, 98]

PARK7: DJ1 Mutations

Loss of function mutations in the gene *DJ1* are a rare cause of PD affecting approximately 1% of autosomal recessive early-onset cases.[99] To date, only three families and a few reports on sporadic cases with mutations in *DJ1* have been described.[99-101]

PARK8: LRRK2 Mutations

Autosomal dominant PD with linkage to 12p11.2-q13.1 was first reported in a large eight-generational Japanese family (Sagamihara family).[102] Further studies in two Caucasian families confirmed linkage to this novel locus.[103] Positional cloning

experiments at the PARK8 locus lead to the identification of missense mutations in the *leucine-rich repeat kinase 2 gene (LRRK2)*.[104, 105] Several screening studies demonstrated that mutations in *LRRK2* are the most common known causes of PD. The frequency of *LRRK2* mutations differs between populations. The most common mutation in European populations is the G2019S mutation and is found in about 2% of sporadic and 5% of familial PD cases.[106-108] By contrast, the frequency of this mutation is much higher in Ashkenazi Jews (13% in sporadic cases, 30% in familial cases) or North African Arabs (41 % in sporadic cases, 37% in familial cases).[109, 110] In Asian populations, the G2385R and R1628P mutations appear to be common risk factors for sporadic and familial PD.[111-113] These mutation are not pathogenic per se, as they are also commonly observed in healthy controls, but represent important susceptibility factors that are associated with a two-fold increased risk for developing PD.

PARK9: ATP13A2 Mutations

ATP13A2 mutations underlie the autosomal recessive neurodegenerative disorder Kufor-Rakeb syndrome, a rare juvenile-onset disorder characterised by parkinsonism, dementia and supranuclear gaze palsy.[114]

PARK14: PLA2G6 Mutations

Mutations in the gene *PLA2G6*, encoding for calcium-independent group VI phospholipase A2, are found in families with autosomal-recessively inherited dystonia-parkinsonism.[115] Brain iron accumulation is usually present in affected individuals, but may be absent in some cases. The symptoms are L-DOPA responsive.[116]

PARK15: FBXO7 Mutations

FBXO7 mutations were identified through linkage mapping and subsequent candidate gene sequencing in an Iranian family with autosomal recessive juvenile-onset parkinsonian-pyramidal syndrome.[117] Patients with *FBXO7* mutations show variable degrees of L-DOPA responsiveness.[118]

Susceptibility Genes: Omi/HTRA2, POLG1, NR4A2, GBA, MAPT, FGF20

During the past few years, a number of susceptibility genes have been identified (table 3) and add further complexity to the understanding of the molecular mechanisms underlying parkinsonism.

PARK13: Omi/HTRA2

Omi/HTRA2 has been implicated with parkinsonism based on observations in mice in which targeted deletion of this gene lead to striatal degeneration, astrogliosis and a parkinsonian phenotype.[119] In a candidate gene approach, Strauss et al. performed a mutational analysis of *Omi/HTRA2* in German PD samples and normal controls. Significant association of the missense mutations p.G399S and p.A141S with increased risk of PD was identified.[120] However, two replication studies tested the identified variants for association with Caucasian PD cases, but failed to detect a significant association.[121, 122] The relevance of *Omi/HTRA2* variants for the pathogenesis of PD therefore remains questionable.

POLG1

Several lines of evidence have linked mitochondrial dysfunction with neurodegeneration. Increased oxidative stress and mitochondrial deletions in dopaminergic neurons of the substantia nigra are commonly observed in aged people and in PD patients.[123, 124] These deletions are thought to be caused by impaired mitochondrial DNA replication and mutations in *POLG1*, encoding the mitochondrial DNA polymerase gamma, have been associated with a variety of phenotypes including ataxia, ophthalmoplegia, neuropathy, migraine, premature menopause and parkinsonism.[125-128] Luoma et al. screened a sample of 140 Finnish sporadic PD patients and 127 matched controls for *POLG1* mutations and found that rare variants of a CAG-repeat in *POLG1* were significantly associated with increased risk for PD.[129] Further research is necessary to clarify the relationship of *POLG1* variants and the risk for developing PD.

NR4A2 (also known as NURR1)

NR4A2, encoding a member of the nuclear receptor superfamily, has been demonstrated to be critical for the formation and survival of dopaminergic midbrain neurons.[130-134] Based on a candidate gene study in 201 PD patients and 221 controls, two heterozygous mutations in exon 1 (-291Tdel, -245T>G) have been implicated with increased risk for PD.[135] Subsequently, three replication studies have investigated this relationship in independent cohorts, but failed to identify a significant association with PD.[136-138] It remains unclear whether *NR4A2* truly is a susceptibility gene for PD.

GBA

A role of mutations in *GBA*, encoding for the lysosomal enzyme glucocerebrosidase, and the development of parkinsonism is increasingly appreciated. Homozygous mutations in *GBA* cause Gaucher's disease, a rare lysosomal storage disease most commonly observed in persons with Ashkenazi Jewish background. The phenotypic presentation of Gaucher's disease is variable and based on the characteristic clinical features three main types are distinguished: type I or chronic nonneuronopathic type is the most common presentation which is characterised by liver, spleen and bone marrow involvement; type II (acute infantile neuronopathic type) and type III (subacute neuronopathic type) are rare manifestations with neurological deficits and early disease onset.[139] Rarely, patients with Gaucher's disease can present with parkinsonism indistinguishable from idiopathic PD.[140] An increased frequency of heterozygous or homozygous *GBA* mutations in PD patients of different ethnicities has been reported and underlines the importance of *GBA* mutations as susceptibility factors involved in the pathogenesis of parkinsonism.[141-144]

MAPT

Microtubule-associated protein tau (MAPT) mutations have been associated with a number of neurodegenerative disorders including progressive supranuclear palsy (PSP),

corticobasal degeneration (CBD), frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17), Alzheimer's disease (AD) and PD.[145-147] *MAPT* is part of a large conserved region on the long arm of chromosome 17. Two main extended haplotypes (H1 and H2) that have diverged about 3 million years ago and stretch about 900kb in size have been described at this locus. These two haplotypes are inverted relative to each other and do not recombine.[148, 149] The H1 haplotype has been reported to be significantly associated with late-onset PD by a number of independent studies.[146, 147, 150] It is unclear, however, how the H1 haplotype at the *MAPT* locus influences risk for developing PD.

FGF20

Fibroblast growth factor 20, encoded by the gene *FGF20*, is a neurotrophic factor that has been implicated to play a critical role in dopaminergic cell survival.[151, 152] In a North-American candidate gene study van der Walt et al. discovered strong associations of variants at the *FGF20* locus with increased risk for developing PD.[153] Subsequently, replication studies have reported conflicting results. Clarimon et al. tested the proposed risk variants in a sample of sporadic PD cases and controls from Greece, and in a separate case-control sample from Finland, but failed to identify a significant association with PD.[154] In contrast, Satake et al. confirmed the association in a Japanese cases-control study.[155] Further replication studies are therefore still required to draw a conclusion on the disease relevance of *FGF20* variants.

Linkage studies have implicated the following disease-associated loci in which the underlying risk genes are not known as of yet: PARK3 (located on chromosome 2p13), PARK10 (located on chromosome 1p32) and PARK11 (located on chromosome 2q37.1).[156-158] Mutations in the gene *GIGYF2*, encoding Grb10-interacting GYF protein-2, have been suggested to be the disease-causing variants at the PARK11 locus based on a candidate gene study in which seven heterozygous missense mutations were found in familial PD patients but not in controls; however, three independent replication studies

did not confirm this observation.[159-162] These findings suggest that *GIGYF2* variants are not responsible for parkinsonism linked to the PARK11 locus.

Table 2. Disease-Causing Parkinsonism Genes

Gene	Locus	Position	Inheritance	Mutation(s)	Phenotype		Ref.
					Disease type	Clinical features	
SNCA	PARK1	4q21	AD	point mutations	PD	parkinsonism, dementia, autonomic failure	[76]
	PARK4			duplication triplication	PD, DLB, MSA		[163] [82]
PRKN	PARK2	6q25.2-q27	AR	point mutations deletions duplications	PD	juvenile- or young-onset parkinsonism, foot dystonia, slow progression	[86]
PINK1	PARK6	1p36	AR	point mutations	PD	early-onset parkinsonism, hyperreflexia, dystonia, L-DOPA induced dyskinesia, slow progression	[96]
DJ1	PARK7	1p36	AR	point mutations deletion	PD	early-onset parkinsonism, psychiatric features, slow progression	[99]
LRRK2	PARK8	12q12	AD	point mutations	PD	late onset parkinsonism, good response to L-DOPA	[104, 105]
ATP13A2	PARK9	1p36	AR	point mutations duplication	KRS	juvenile-onset parkinsonism, spasticity, hallucinations, dementia, supranuclear gaze paresis	[114]
PLA2G6	PARK14	22q13.1	AR	point mutations deletions	NBIA2, PD	adult-onset L-DOPA responsive dystonia-parkinsonism	[115]
FBXO7	PARK15	22q12-q13	AR	point mutation	PPS	early-onset parkinsonism, pyramidal signs, variable response to L-DOPA	[118]
ATXN2	SCA2	12q24	AD	triplet repeat expansion	PD, SCA	phenotypic presentation is dependent on the ethnic background: parkinsonism in ethnic Chinese and Africans; predominantly spinocerebellar ataxia in Caucasians	[164, 165]
ATXN3	SCA3	14q24.3-q31	AD	triplet repeat expansion	PD, SCA	phenotypic presentation is dependent on the ethnic background: parkinsonism in ethnic Chinese and Africans; predominantly spinocerebellar ataxia in Caucasians	[166-168]

Parts of this table have been published elsewhere (see reference [66]). AD, autosomal dominant; AR, autosomal recessive; *ATP13A2, ATPase type 13 A2*; DLB, dementia with Lewy bodies*; FBXO7, F-box only protein 7*; KRS, Kufor-Rakeb syndrome; *LRRK2, leucine-rich repeat kinase 2*; MSA, multiple system atrophy; NBIA2, neurodegeneration with brain iron accumulation type 2; *PINK1, PTEN-induced putative kinase 1; PLA2G6, phospholipase A2, calcium-independent, group IV, A; PRKN, parkin*; PPS, parkinsonian-pyramidal syndrome; Ref., reference(s); *SNCA, α-synuclein.*

Table 3. Susceptibility Genes Involved in Parkinsonism

Gene	Locus	Position	Mutation(s)	Phenotype		Ref.
				Disease type	Clinical features	
Omi/HtrA2	PARK13	2p12	point mutations	PD	late onset L-DOPA responsive parkinsonism	[120]
POLG1	-	15q25	point mutations	PD	early-onset parkinsonism, peripheral neuropathy, ophthalmoplegia, ataxia	[125]
NR4A2	-	2q22-q23	point mutations	PD	adult-onset L-DOPA responsive parkinsonism	[135]
GBA	-	1q21	point mutations	PD, DLB, GD	adult-onset L-DOPA responsive parkinsonism, cognitive changes, depression	[141, 169]
MAPT	-	17q21.1	point mutations deletions	PD, PSP, CBD, FTDP-17	parkinsonism, frontotemporal dementia, supranuclear gaze palsy	[146, 147, 170, 171]
FGF20	-	8p22-p21.3	N/A	PD	late onset parkinsonism	[153]

CBD, corticobasal degeneration; DLB, dementia with Lewy bodies; *FGF20, fibroblast growth factor 20;* FTDP-17, frontotemporal dementia with parkinsonism linked to chromosome 17; *GBA, glucocerebrosidase;* GD, Gaucher disease; *MAPT, microtubule-associated protein tau;* N/A, not applicable; *NR4A2, nuclear receptor subfamily 4 group A member 2; POLG1, DNA polymerase gamma;* PSP, progressive supranuclear palsy; Ref., reference(s).

1.4.3 Genetics of Multiple System Atrophy

Multiple system atrophy (MSA) is an adult-onset neurodegenerative disorder which is characterised clinically by parkinsonism, ataxia, and autonomic failure, and pathologically by deposition of abnormally phosphorylated fibrillar α-synuclein.[172] Relentlessly progressive in its disease course and poorly responsive to L-DOPA therapy, MSA is invariably fatal with an average survival time of eight years after symptom onset.[173] Pathogenic mechanisms and disease aetiology are unknown. MSA is a rare disease with an incidence of about three new cases per 100,000 per year and a prevalence of 1.9 - 4.4 per 100,000.[174-176] The disease affects both men and women with the first disease symptoms usually manifesting in the sixth decade of life.[173] Based on the predominant initial presentation, two main types of MSA are distinguished: MSA-P, which primarily presents with parkinsonian features, and MSA-C, which is characterised by predominant cerebellar ataxia.[172, 177] Furthermore, MSA is divided into three main diagnostic categories: (a) definite MSA, which is characterised by neuropathological confirmation of α-synuclein positive glial cytoplasmic inclusions and striatonigral or olivopontocerebellar neurodegeneration, (b) probable MSA, which is defined as a sporadic progressive adult-onset disorder featuring rigorously defined autonomic failure and poorly L-DOPA responsive parkinsonism or cerebellar ataxia, and (c) possible MSA, which is defined as a sporadic progressive adult-onset disease including parkinsonism or cerebellar ataxia and at least one feature suggesting autonomic failure plus one other feature that may be a clinical or a neuroimaging abnormality.[177]

MSA is commonly considered a non-genetic, sporadic disorder; however, a number of recent observations challenge this common school of thought. The first suggestions that genetic factors could play a pivotal role in the disease pathogenesis were based on observations that neurological signs of parkinsonism are more common in relatives of MSA patients.[178, 179] Furthermore, in a rare form of familial PD, caused by triplication or duplication of *SNCA*, individuals presenting with a clinical or pathological phenotype similar to MSA have been reported.[81, 82, 180] In addition, reports of rare pathology-proven

familial MSA exist, – one German family in which a mother and her daughter were affected and four Japanese families with apparently autosomal recessively inherited MSA.[181-183]

A number of candidate genes have been tested for association with disease (for a summary see table 4), but no convincingly reproducible risk genes have been identified. Three main obstacles explain why performing genetic research in MSA poses a major challenge. Firstly, MSA typically occurs sporadically in the community and therefore genetic approaches designed to identify disease genes in familial diseases, such as linkage studies, are not feasible in MSA. Secondly, the low incidence of the disease makes it difficult to recruit a large enough cohort of affected individuals required for sufficiently powered studies. Thirdly, diagnostic uncertainty in clinically diagnosed patients resulting in a high false positive rate negatively impacts the power of studies using clinically ascertained cases.[184, 185] Some of these challenges can now be tackled using modern high-throughput SNP genotyping technologies as well as publicly available DNA and tissue banks, as I will demonstrate in detail in section 3.3 of my thesis.

Table 4. Candidate Gene Studies in MSA

Gene (abbreviation)	Investigated variant(s)/region†	Population	Number of cases / controls	Finding	Ref.
AACT	Ala-15Thr (= A/T transversion)	Japanese	105/105	A/A genotype associated with disease	186
ADH7	sequence of all exons	S. Korean	50/50	no significant association	187
APOE	alleles e2,e3,e4	Caucasian	47/72	no significant association	188
	alleles e2,e3,e4	Caucasian	22/66	no significant association	189
ATF3	SNPs at the ATF 3 gene locus	Japanese	119/123	no significant association	190
ATF4	SNPs at the ATF 4 gene locus	Japanese	119/123	no significant association	190
ATXN1	trinucleotide repeat	Caucasian	80/80	repeats in normal range	191
ATXN3	trinucleotide repeat	Caucasian	80/80	repeats in normal range	191
BDNF	-240C/T, 480G/A	Japanese	122/275	no significant association	192
CARS	SNPs at the CARS gene locus	Japanese	119/123	no significant association	190
CEBPB	SNPs at the CEBPB gene locus	Japanese	119/123	no significant association	190
CHOP	SNPs at the CHOP gene locus	Japanese	119/123	no significant association	190
CNTF	null mutation	Caucasian	80/80	no significant association	191
CYP1A1	alleles *2A, *2B	Caucasian	38/38	no significant association	193
CYP2D6	G->A transition at the junction	Caucasian	91/145	no significant association	194
	between intron 3/exon 4;	Caucasian	38/38	no significant association	193
	1 base-pair deletion alleles *3, *4				
DAT1	repeat polymorphism	Caucasian	38/543	no significant association	193
DBH	C-1021T, G259A, IVS1+2T>C, C300A,	Caucasian,	39/88	no significant association	195
	IVS3+8C->T, G991A, IVS10+415A->G	1 Korean			
DM2	CCTG repeat in intron 1	N/A	1/0	patient had one abnormal allele	196
EIF4EBP1	3 SNPs at the EIF4EBP1 gene locus	Japanese	119/123	significant haplotype association	190
FMR1	trinucleotide repeat	Chinese	15/200	no significant repeat expansion	197
	trinucleotide repeat	Mixed	65/0	no abnormal repeat expansion	198
	trinucleotide repeat	Japanese	77/0	no abnormal repeat expansion	199
	trinucleotide repeat	Caucasian	507/622	no significant repeat expansion	200
GSTM1	homozygous deletion	Caucasian	38/38	no significant association	193
Potassium Channel hiGIRK2	sequence of H5 pore region	Caucasian	80/80	no significant association	191

table continued on page 55

Table 4 (continued). Candidate Gene Studies in MSA

Gene (abbreviation)	Investigated variant(s)/region†	Population	Cases / controls	Finding	Ref.
HLA	subtype A32	Caucasian	80/80	no significant association	191
	16 HLA-A loci, 23 HLA-B loci	N/A	12/280	subtype A32 is protective	201
	subtype A32	Colombian	18/0	no significant increase in frequency	202
IGF1	trinucleotide repeat	Caucasian	80/80	no significant association	191
IL1A	-889C/T	Caucasian	30/110	allele 2 associated with disease	203
	-889C/T	Japanese	111/160	no significant association	204
IL1B	-511C/T, +3953C/T	Japanese	111/160	genotype 2/2 in -511C/T SNP is protective	204
IL1RA	VNTR in intron 2	Japanese	111/160	no significant association	204
IL10	-592A/C	Japanese	122/277	no significant association	205
IL6	-634C/G	Japanese	122/277	no significant association	205
IL8/ICAM-1	-251A/T (Glu469Lys)	Caucasian	41/93	T allele increases disease risk (in particular in presence of ICAM-1 K/K genotype)	206
LRRK2	Gly2019Ser	Mixed	136/110	no significant association	207
	14 different mutations	Chinese	15/0	no pathogenic mutation	208
	Gly2385Arg	Taiwanese	57/313	no significant association	111
MAPT	intronic deletion on H2 haplotype	Caucasian	47/72	no significant association	188
Mitochondrial DNA	4,977-bp deletion in midbrain neurons	N/A	4/4	no significant difference	209
NAT2	alleles *5A/B, *5B/C, *13, *6A	Caucasian	38/38	no significant association	193
OPCA	trinucleotide repeat	Japanese	105/105	repeats in normal range	186
PGRN	sequence of all exons	Caucasian	13/0	no mutations detected	210
SCA 1-3, 6-8, 12, 17	trinucleotide repeat	Japanese	105/105	repeats in normal range	186
SLC1A4	7 SNPs at the SLC1A4 gene locus	Japanese	119/123	significant haplotype association	190
SNCA	microsatellites in promoter and intron 4	Caucasian	47/72	no significant association	188
	sequence of all exons	Japanese	11/5	no mutation	211
	sequence of all exons	N/A	7/0	no mutation	212
	gene dosage measurement of exon 5	N/A	58	no gene dosage change	213
	expression analysis	Japanese	11/14	no significant expression alteration	214
	expression analysis	Caucasian	8/8	no significant expression alteration	215
	haplotype tagging	Caucasian	457/1472	no significant association	216
SQSTM1	6 SNPs at the SQSTM1 gene locus	Japanese	119/123	significant haplotype association	190
Synphilin	A/T polymorphism 3' to exon 3	Caucasian	47/72	no significant association	188

table continued on page 56

Table 4 (continued). Candidate Gene Studies in MSA

Gene (abbreviation)	Investigated variant(s)/region†	Population	Cases / controls	Finding	Ref.
TGFB1	-509C/T	Japanese	122/277	no significant association	205
TNF	-1031C/T	Japanese	122/277	significant association with disease	205
	-850C/T	Caucasian	41/93	no significant association	191
UCHL1	haplotype tagging	Caucasian	457/1536	no significant association	217

†Variant nomenclature is shown according to each respective original publication. N/A, data not available.

AACT, α-1-antichymotrypsin; ADH7, alcohol dehydrogenase 7; APOE, apolipoprotein E; ATF3, activating transcription factor 3; ATF3, activating transcription factor 4; ATXN1, ataxin 1; ATXN3, ataxin 3; BDNF, brain-derived neurotrophic factor; CARS, cysteinyl t-RNA synthetase; CEBPB, CCAAT/enhancer-binding protein-β; CHOP, CCAAT/enhancer-binding protein homologous protein; CNTF, ciliary neurotrophic factor; CYP1A1, cytochrome P450, 1A1; CYP2D6, cytochrome P450, 2D6; DAT1, dopamine transporter 1; DBH, dopamine β-hydroxylase; EIF4EBP1, eukaryotic translation initiation factor 4E-binding protein; FRM1, fragile X mental retardation 1; GSTM1, glutathione S-transferase, mu1; HLA, human leukocyte antigen; IGF1, insulin-like growth factor 1; IL1A, interleukin 1A; IL1B, interleukin 1B; IL1RA, interleukin 1RA; IL10, interleukin 10; IL6, interleukin 6; IL8, interleukin 8; ICAM-1, intercellular adhesion molecule 1; LRRK2, leucine-rich repeat kinase 2; MAPT, microtubule-associated protein tau; DM2, myotonic dystrophy 2; NAT2, N-acetyltransferase 2; OPCA, olivopontocerebellar ataxia; PGRN, progranulin; SCA, spinocerebellar ataxia; SLC1A4, solute carrier family 1A4; SNCA, α-synuclein ; SQSTM1, sequestosome 1; TGFB1, transforming growth factor β-1; TNF, tumor necrosis factor α; UCHL1, ubiquitin carboxyl-terminal esterase L1.

2 Human Genome Diversity Project

STATEMENT OF CONTRIBUTIONS TO THIS RESEARCH:

I was involved in the conceptualisation and experimental design of this project. I performed most of the laboratory based genotyping experiments, I was responsible for the data management and parts of the data analyses, in particular quality control and copy number analyses. I drafted parts of the original publication that has been published in the journal Nature.

2.1 Introduction

Recent advances in genomic technology opened new opportunities to enhance our understanding of human evolutionary history and the genetic basis of human health and disease. Knowledge on genetic diversity across ethnically diverse populations is particularly important for understanding phenotype variability, disease susceptibility and drug response. Furthermore, such genomic data can be used to accurately reconstruct human demographic history by allowing us to follow migration patterns, as well as to detect population bottlenecks, selection and adaptation to new environments.

In an effort to identify patterns of genetic diversity in modern human populations, we performed a large-scale, genome-wide SNP genotyping study in an ethnically and geographically diverse selection of individuals from 29 worldwide populations. Here we report on genotype, haplotype and copy-number variation in these ethnically diverse human populations.

2.2 Materials and Methods

Samples

We selected 513 samples from 29 populations (table 5) from the Human Genome Diversity Cell Line Panel (HGDP; CEPH Institute, Paris, France). These samples represented a geographically and ethnically diverse subset, including all HGDP African populations (figure 3). We also included data from 112 individuals from the International HapMap Project for some analyses (www.hapmap.org). DNA was derived from Epstein-Barr virus immortalised lymphoblastoid cell lines. The study was approved by the institutional review board.

Figure 3. Geographic Location of Studied HGDP Populations

Table 5. HGDP and HapMap Individuals Included in the Diversity Project

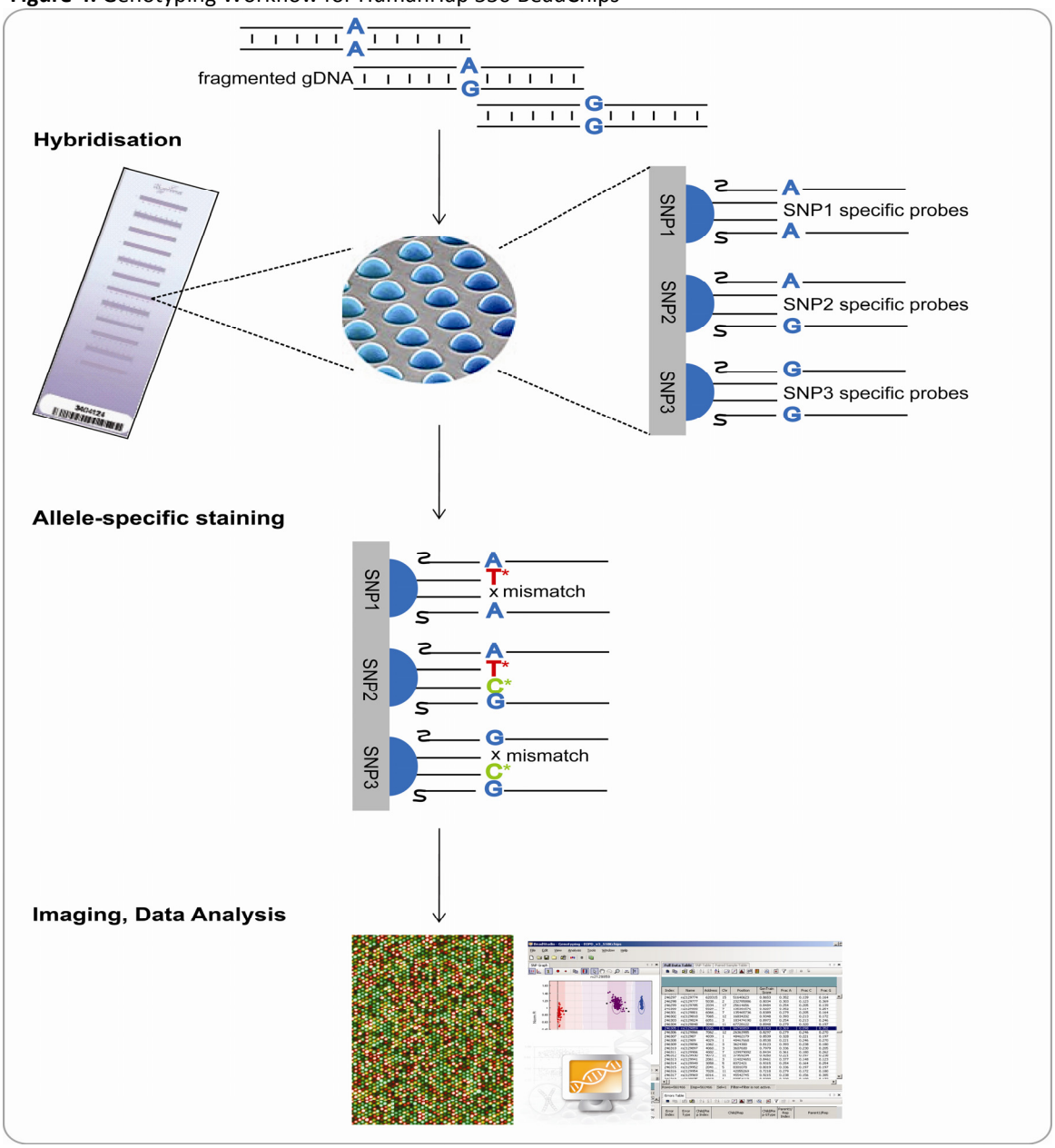
Geographic region	Population	Number of individuals	Number of females / males	Number of unrelated individuals
AFRICA	San	7	0 / 7	6
	Mbuti Pygmy	15	2 / 13	13
	Biaka Pygmy	32	2 / 30	23
	Bantu (Southern Africa)	8	0 / 8	8
	Bantu (Kenya)	12	1 / 11	11
	Yoruba	25	12 / 13	22
	YRI HapMap	36	14 / 22	24
	Mandenka	24	8 / 16	22
MIDDLE EAST	Mozabite	30	10 / 20	29
	Bedouin	47	19 / 28	45
	Palestinian	26	20 / 6	24
	Druze	43	30 / 13	38
EUROPE	Basque	13	4 / 9	13
	CEU HapMap	48	25 /23	32
	Russian	13	3 / 10	13
	Adygei	14	8 / 6	14
C/S ASIA	Balochi	15	0 / 15	15
	Kalash	18	4 / 14	16
	Burusho	7	1 / 6	7
	Uygur	10	2 / 8	10
EAST ASIA	Yakut	15	3 /12	15
	Daur	10	3 / 7	10
	Mongola	9	2 / 7	9
	JPT HapMap	16	9 / 7	16
	CHB HapMap	12	8 / 4	12
	Yi	10	1 / 9	10
	Lahu	8	1 / 7	8
	Cambodian	10	4 / 6	10
OCEANIA	Melanesian	17	11 / 6	11
	Papuan	16	4 / 12	16
AMERICA	Pima	11	5 / 6	8
	Maya	13	11 / 2	10
	Colombian	7	5 / 2	7

This table has been published elsewhere (see reference[218]). The numbers of participants from studied populations are listed. YRI, Yoruba people of Ibadan in Nigeria; CEU, European American people from Utah, USA; JPT, Japanese people from Tokyo; CHB, Han Chinese people from Beijing.

Genotyping

We performed genotyping on an Illumina platform using HumanHap550 BeadChips (Illumina, San Diego, CA, USA). Genotyping was performed as per the manufacturer's instructions (the workflow of this assay is summarised in figure 4).[219] Briefly, for each sample a total of 750ng of genomic DNA (gDNA), resuspended in TE solution (10mM Tris, 1mM EDTA) and normalised to 50ng/µl, was denatured in 15µl of 1N NaOH. After addition of the neutralisation solution (supplied by Illumina) and the amplification master mix (Illumina), a DNA amplification reaction at 37°C for 20-24 hours generated hundreds of micrograms of amplified DNA.[219] Following the amplification reaction, the DNA was fragmented in an endpoint enzymatic reaction at 37°C for 1 hour to an average size of around 300bp using the fragmentation master mix (Illumina), precipitated with 100% 2-propanol and resuspended in resuspend amp1 solution (RA1, Illumina). After an over-night hybridisation to the BeadChips, unhybridised and non-specifically hybridised DNA was washed off the chips using RA1 solution and followed by an automated allele-specific extension and staining reaction using a Tecan Freedom EVO robot (www.tecan.com, Männedorf, Switzerland). The BeadChips were then washed in PB1 solution (Illumina), coated with XC3 polymer (Illumina), vacuum-dried for 1 hour, and imaged using a two-color confocal laser system in a BeadArray Reader (Illumina).

Figure 4. Genotyping Workflow for HumanHap 550 BeadChips



The genotyping assay workflow of the Illumina Infinium platform is shown. For each individual, amplified and fragmented DNA is hybridised to the surface of a BeadChip, containing microbeads that are covered with SNP-specific oligonucleotide capture probes. After removal of unhybridised DNA, an allele-specific extension and staining reaction is performed. Following washing, coating and drying of the labelled chips, the array is scanned on a BeadArray reader. The color signal emitted from each microbead is used for automated genotype calling.

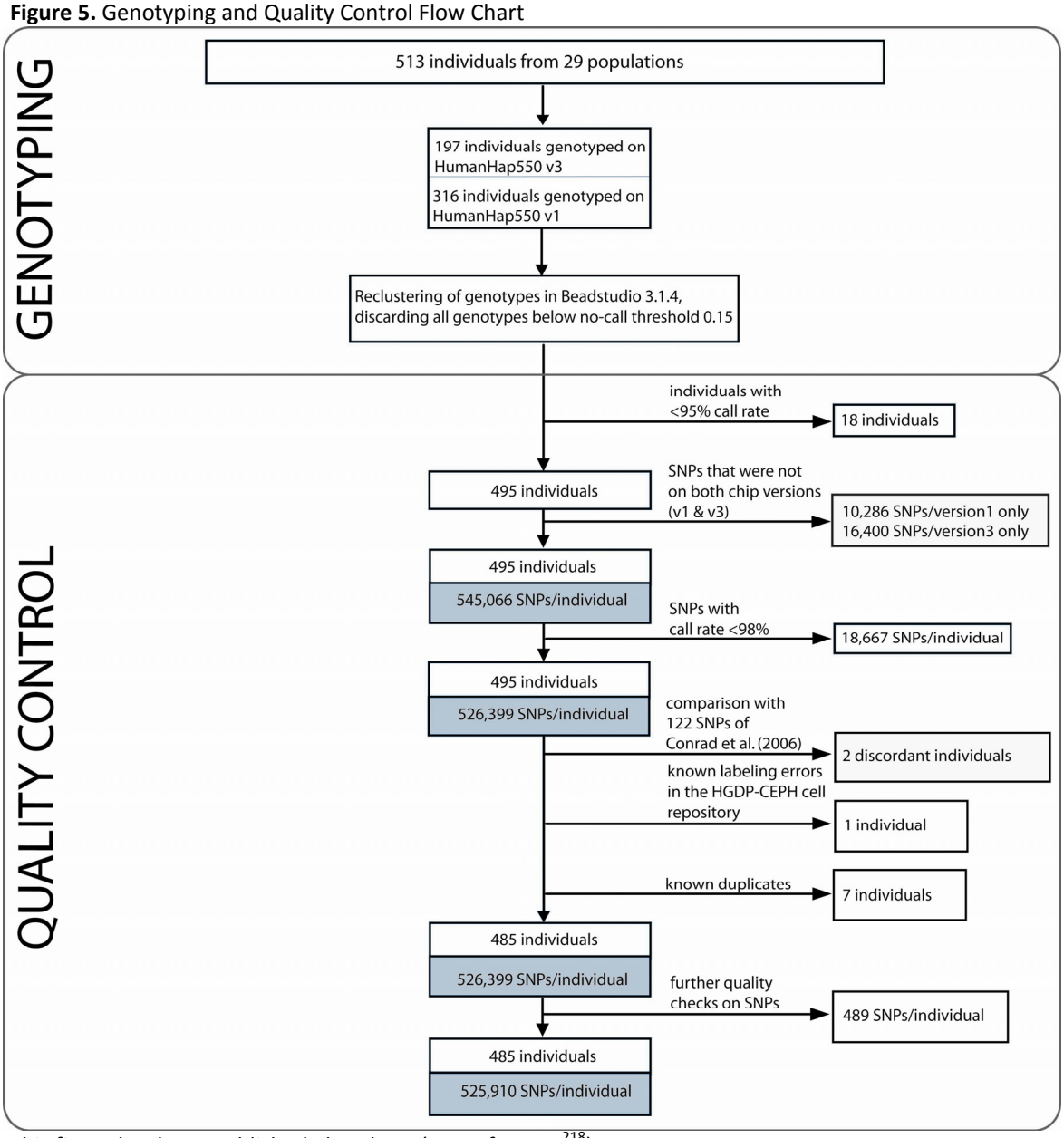
Three-hundred and sixteen individuals were genotyped on version 1 HumanHap 550 BeadChips and one-hundred and ninety-seven individuals were genotyped on version 3 HumanHap 550 BeadChips (figure 5). Only SNPs that were in common of both chip

versions were included in the analyses (545,066 SNPs per individual). For each chip version the bead intensities were extracted using the GenCall software (Illumina) and genotypes were calculated using the GenTrain genotype clustering algorithm implemented in BeadStudio (version 3.1.4, Illumina); raw genotype data for each chip version were processed in separate BeadStudio project files. Reclustering of SNP genotypes was performed using a no-call threshold of 0.15. The reclustered datasets were then used for further quality control analyses (figures 6 and 7).

Quality Control in Individuals

After reclustering, we excluded 18 individuals based on a call rate <95% (call rates, allele frequencies and quality control measures of the remaining samples are shown in figure 6 and figure 7). Of the remaining 495 individuals, we excluded seven duplicate samples and one sample (Biaka Pygmy 980) which had previously been reported as a labeling error in the HGDP repository.[220, 221] To exclude a possible sample mix-up during genotyping, we first verified that the reported sex matched with the sex inferred on the basis of X-chromosomal genotypes using the sex estimation function in BeadStudio software. Second, we compared 122 autosomal SNP genotypes with the same genotypes from a HGDP dataset published by Conrad et al.[222] Four individuals were not genotyped by Conrad et al. (Adygei 1383, Adygei 1384, Biaka Pygmy 980, and Russian 890) and were consequently excluded from this particular quality check. On the basis of this analysis we discovered two discordant individuals which we excluded from our study. The final cohort therefore consisted of 485 distinct individuals (figure 5).

Figure 5. Genotyping and Quality Control Flow Chart



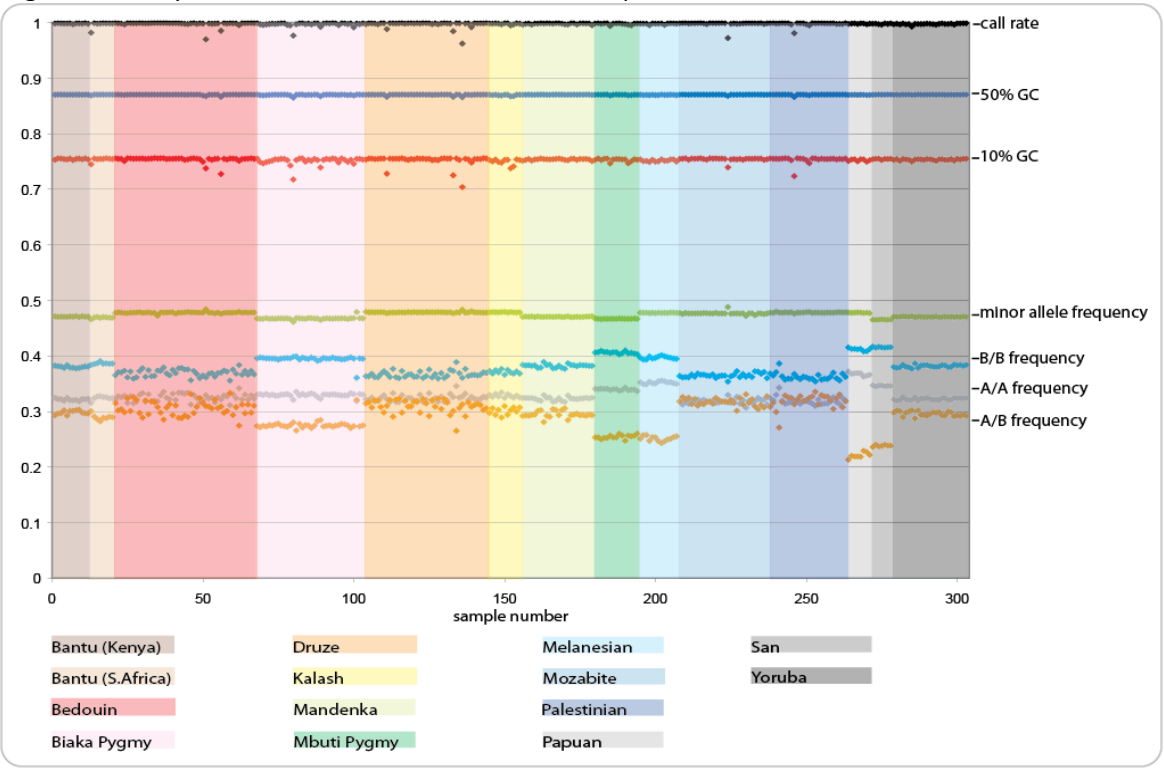
This figure has been published elsewhere (see reference[218]).

Quality Control of SNPs

After removal of SNPs that were not in common between version 1 and version 3 BeadChips (10,286 SNPs in version 1 chips; 16,400 SNPs in version 3 chips), we excluded SNPs that were called in less than 98% of individuals (18,667 SNPs). We excluded additional 42 monomorphic SNPs, 161 SNPs with a high overall missing rate, 135 SNPs with considerable missing data in at least one population, 198 SNPs with Hardy-

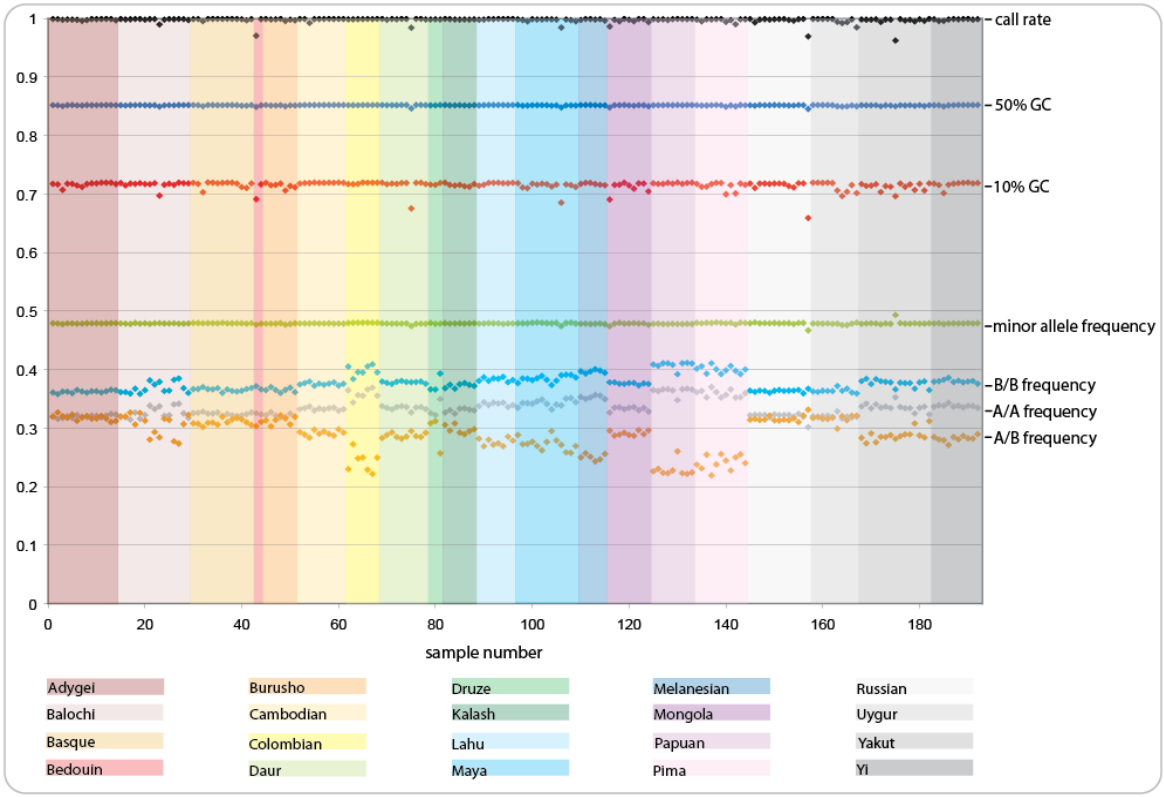
Weinberg disequilibrium, 2 SNPs with high discordance rates between duplicates and 26 SNPs with Mendelian incompatibility. Taking SNPs into account that failed in more than one of the quality checks, a total of 489 additional SNPs were excluded from the dataset. The final number of SNPs that were used in the analyses was 525,910 SNPs (512,762 autosomal SNPs; 13,052 X-chromosomal SNPs; 9 Y-chromosomal SNPs; 15 pseudoautosomal SNPs; 72 mitochondrial SNPs).

Figure 6. Quality Control Plots of 550K Version 1 BeadChips



This figure has been published elsewhere (see reference [218]). Quality control data for 303 individuals genotyped on HumanHap550 version 1 BeadChips are shown. Only individuals with a genotype call rate >95% are depicted. Each background color represents a distinct population and within each population individuals are sorted by the HGDP identification number. Black diamonds indicate call rates, green diamonds indicate the average minor allele frequency across all SNPs genotyped in that individual, turquoise diamonds indicate the frequency of B/B calls, gray diamonds indicate the frequency of A/A calls, and orange diamonds indicate the frequency of heterozygote calls. The blue diamonds indicate the 50% GenCall (GC) score (median GenCall score across SNPs); red diamonds indicate the 10% GC score (tenth percentile of the ranked GenCall scores). For a given individual and SNP, the GenCall score is a measure of data quality that takes into account the fit of the individual genotype to defined genotype clusters. 10% GC scores above 0.7 indicate high-quality genotypes and scores below 0.7 indicate low-quality genotypes. Genotype frequencies show population-specific patterns.

Figure 7. Quality Control Plots of 550K Version 3 BeadChips



This figure has been published elsewhere (see reference [218]). Quality control data for 192 individuals genotyped on HumanHap550 version 3 BeadChips are shown. Only individuals with a genotype call rate >95% are depicted. Each background color represents a distinct population and within each population individuals are sorted by the HGDP identification number. Black diamonds indicate call rates, green diamonds indicate the average minor allele frequency across all SNPs genotyped in that individual, turquoise diamonds indicate the frequency of B/B calls, gray diamonds indicate the frequency of A/A calls, and orange diamonds indicate the frequency of heterozygote calls. The blue diamonds indicate the 50% GenCall (GC) score (median GenCall score across SNPs); red diamonds indicate the 10% GC score (tenth percentile of the ranked GenCall scores). For a given individual and SNP, the GenCall score is a measure of data quality that takes into account the fit of the individual genotype to defined genotype clusters. 10% GC scores above 0.7 indicate high-quality genotypes and scores below 0.7 indicate low-quality genotypes. Genotype frequencies show population-specific patterns.

Haplotype Analysis

For haplotype analysis we used fastPHASE version 1.3. This algorithm estimates clusters of haplotypes using a hidden Markov model allowing both block-like patterns of linkage disequilibrium (LD) and a gradual decline of LD with distance.[223] We set the default number of haplotype clusters to 20. To increase accuracy during model fitting procedures, each sample was labelled according to geographic region as described in table 5. After removal of related individuals, a total of 443 HGDP samples and 84 HapMap samples were studied for haplotype analysis. Haplotype phase was estimated

for autosomes and the pseudoautosomal region. For the X chromosome all males were treated as having known haplotype phase.

To measure LD (a nonrandom association of alleles from different loci), we calculated pairwise r^2 and HR^2 statistics for all autosomal SNP pairs within a 70.5 kb distance. To adjust for a possible influence of sample size on r^2, we used a resampling method. For each SNP pair in each population, we randomly chose ten haplotypes that were then used to perform LD computations. We excluded SNP pairs that were monomorphic at one or both SNPs in the chosen haplotypes. The remaining SNP pairs were placed in overlapping bins, and the mean was taken for each bin. For LD estimates, haplotype length and frequency, a 'best guess' estimate of haplotype structure was used.

Copy Number Variant Analysis

We used the PennCNV algorithm to detect copy number variants (CNVs) (this program is available at www.neurogenome.org).[224] PennCNV software infers copy number variation by combining normalised signal intensity data, the physical SNP coordinates in the genome (NCBI build 35), and SNP allele frequencies with a hidden Markov model. For each sample, genotype and signal intensity data (precomputed Log R ratio [LRR] and B allele Frequency [BAF]) were exported from BeadStudio software version 3.1.4 (Illumina, CA, USA). The LRR is an indirect measure of copy number of each SNP derived by calculating the ratio of observed to expected hybridisation intensity normalised to the reference population. The BAF is the proportion of times an allele is called A or B at a given genotype. Based on previously validated quality control criteria,[224] we excluded samples with a LRR standard deviation of >0.28, a median BAF of >0.55 or <0.45, or a BAF drift of >0.002. After application of these quality control criteria, 42 HGDP samples were excluded from the analysis. Only autosomal and X-chromosomal SNPs were included to generate CNV calls. To reduce the number of possible false-positive CNV calls, we excluded CNVs that spanned less than 10 consecutive SNPs. Of the remaining CNV calls we excluded variants in regions of known V(D)J-type recombination (on

chromosome regions 2p.11, 14q11.2, 14q32.33, 22q11.22) and CNVs larger than 1Mb in size, which are likely to be cell line artifacts as described previously.[225] In total, we discovered 3,552 CNVs in 1,428 non-overlapping loci (table 6).

Table 6. CNVs in 443 HGDP Samples from 29 Populations

Geographic region	Population	Number of individuals in CNV dataset	Number of CNVs	Number of deletions / duplications	Number of total CNV loci	Number of new CNV loci
AFRICA	San	7	38	22 / 16	33	6
	Mbuti Pygmy	14	101	49 / 52	70	21
	Biaka Pygmy	31	214	107 / 107	112	24
	Bantu (Kenya)	12	99	79 / 20	71	16
	Bantu (S. Africa)	7	42	25 / 17	36	6
	Yoruba	25	147	85 / 62	79	11
	Mandenka	24	148	98 / 50	96	28
MIDDLE EAST	Mozabite	29	159	88 / 71	80	12
	Bedouin	43	247	147 / 100	131	28
	Palestinian	25	182	126 / 56	121	31
	Druze	40	262	143 / 119	126	35
EUROPE	Basque	11	80	28 / 28	57	16
	Russian	13	153	117 / 36	110	37
	Adygei	13	80	61 / 19	63	21
C/S ASIA	Balochi	14	70	39 / 31	55	8
	Kalash	13	278	258 / 20	147	39
	Burusho	6	42	18 / 24	39	7
	Uygur	9	39	15 / 24	33	10
EAST ASIA	Yakut	12	86	58 / 28	71	24
	Mongola	9	53	27 / 26	45	9
	Daur	10	60	38 / 22	56	18
	Yi	9	36	21 / 15	34	12
	Cambodian	10	44	18 / 26	41	13
	Lahu	8	38	19 / 19	28	8
OCEANIA	Melanesian	11	332	289 / 43	178	46
	Papuan	12	246	200 / 46	163	42
AMERICA	Pima	8	70	55 / 15	52	26
	Maya	11	169	148 / 21	138	59
	Colombian	7	37	20 / 17	28	7

This table has been published elsewhere (see reference [218]). C/S ASIA, Central-South Asia. New CNVs are variants that have not been previously reported in the Database of Genomic Variation (http://projects.tcag.ca/variation/) version hg18.v3.

To investigate CNVs in unrelated HGDP samples only, we excluded 38 samples from known relatives. In the remaining 405 samples, we observed 3,069 CNVs in 1,333 loci.

We then filtered the data excluding CNVs that had been observed only once, thereby yielding 2,118 CNVs in 396 copy-number variable loci (1,470 deletions at 296 loci and 648 duplications at 134 loci).

Analysis of Population Structures

We inferred population structures using a Bayesian clustering algorithm implemented in the software STRUCTURE. SNP data from all 443 unrelated HGDP individuals were studied in this analysis. An admixture model (F model), assuming that individuals share ancestry with multiple populations, was applied. To estimate cluster membership coefficients, the data were processed using 20,000 iterations followed by 10,000 iterations and the results were graphically displayed using the program DISTRUCT.[226] To avoid studying markers in LD, we selected four interspersed subsets of about 1% of autosomal SNPs each. For each subset and cluster value K, we performed ten replicate analyses in STRUCTURE. To identify shared modes among replicates, we used the LargeKGreedy algorithm incorporated in CLUMPP software with 10,000 random permutations.[227]

Analysis of population structures using haplotype data from 443 unrelated HGDP samples was performed similar to the SNP analysis described above. Ten imputations of cluster memberships were used. For each imputed dataset, two subsets of SNP data were selected containing about 1% of the autosomal SNPs. For each cluster value K, 40 replicate analyses were performed in STRUCTURE; analysis of these results was proceeded with CLUMPP analysis.

Inferences on population structures using CNV data was performed in STRUCTURE and CLUMPP similar to the analyses of SNP and haplotype data. In this analysis, we used data from 405 unrelated individuals and the 396 non-singleton CNV loci. Again, 40 replicate STRUCTURE analyses were carried out; this computation was followed by CLUMPP analysis.

Analysis of Population Relationships: Neighbor-Joining Trees

Estimates on pairwise allele-sharing distances were used to construct neighbor-joining trees of population relationships. Autosomal SNP data from all 443 unrelated HGDP individuals were included in this analysis. To calculate confidence intervals, we applied a bootstrap sampling method with 1,000 resamples across loci. Bootstrap calculations of bootstrap distances were performed in microsat software (http://hpgl.stanford.edu/projects/microsat/); next, the neighbor-joining consensus tree was constructed using the neighbor, consense and drawtree modules implemented in PHYLIP software (http://evolution.genetics.washington.edu/phylip.html).

Analysis of population relationships using haplotype cluster data from 443 unrelated HGDP samples was performed similar to the SNP analysis described above. For inferences on population relationships using CNV data, we constructed neighbour-joining trees with data from the 396 autosomal non-singleton CNV loci determined in 405 unrelated individuals similar to the analyses of SNP and haplotype data.

Analysis of Population Relationships: Multi-dimensional Scaling Analysis

Multi-dimensional scaling (MDS) analysis was performed to determine population distances and to infer group memberships using cmdscale algorithm in R software (http://www.r-project.org/). For SNP data, a pairwise distance matrix was constructed using autosomal SNP data from 443 unrelated HGDP samples. Between-individual distances were obtained using the allele-sharing distances; for the calculation of the overall distance between individuals, we used the average across loci.

To generate a genetic distance matrix for haplotype clusters, we calculated haplotype distances between all pairs of individuals. MDS analysis then was performed similar to the SNP analysis described above.

MDS analysis in CNV data used the 405 unrelated individuals in the CNV dataset and the 396 autosomal non-singleton CNV loci to generate a genetic distance matrix. MDS analysis was performed similar to the analyses of SNP and haplotype data.

Genetic and Geographic Relationship Analysis

Using a linear regression analysis, we investigated the relationship between LD, expressed as mean HR^2 values, and geographic distance from Addis Ababa (latitude 9°N, longitude 38°E; East Africa). To determine the geographic distance of each population from East Africa, waypoint routes described by Ramachandran et al. have been used.[228] Coordinates of the geographic location of each studied population are shown in table 7.

Table 7. Coordinates for Geographic Relationship Analysis

Geographic region	Population	Latitude	Longitude
AFRICA	San	-21°	20°
	Mbuti Pygmy	1°	29°
	Biaka Pygmy	4°	17°
	Bantu (Kenya)	-26°	24°
	Bantu (S. Africa)	-3°	37°
	Yoruba	8°	5°
	Mandenka	12°	-12°
MIDDLE EAST	Mozabite	32°	3°
	Bedouin	31°	35°
	Palestinian	32°	35°
	Druze	32°	35°
EUROPE	Basque	43°	0°
	Russian	61°	40°
	Adygei	44°	39°
C/S ASIA	Balochi	31°	67°
	Kalash	36°	72°
	Burusho	37°	74°
	Uygur	44°	81°
EAST ASIA	Yakut	63°	130°
	Mongola	49°	124°
	Daur	45°	111°
	Yi	28°	103°
	Cambodian	22°	100°
	Lahu	12°	105°
OCEANIA	Melanesian	-6°	155°
	Papuan	-4°	143°
AMERICA	Pima	29°	-108°
	Maya	19°	-91°
	Colombian	3°	-68°

This table has been published elsewhere (see reference [218]). C/S ASIA, Central-South Asia.

2.3 Results

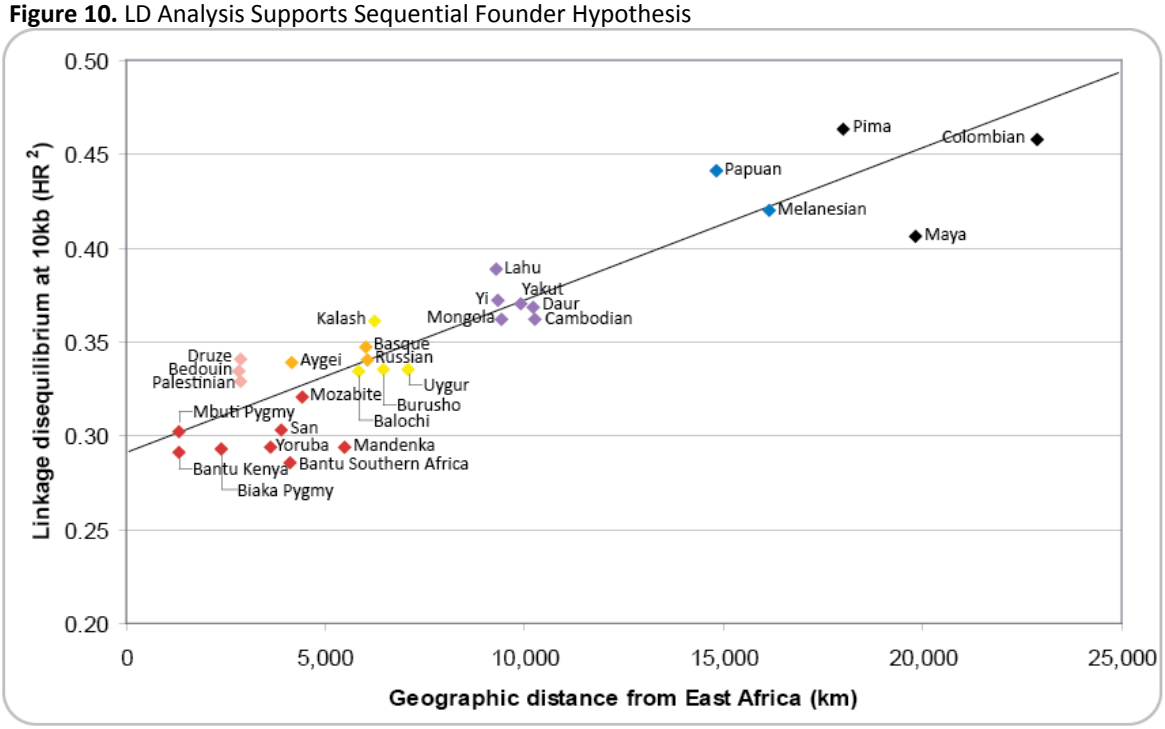
Allele Frequencies

Allele frequencies were corrected for sample size difference across different populations. We observed 81.2% of all SNP alleles in each of the five main geographic regions (Africa, Eurasia, South-East Asia, Oceania, America). Private alleles for each geographic region were rare: only 0.9% of alleles were exclusively observed in African samples, 0.8% in Eurasian samples, and nearly 0.0% in other regions (figure 8).

Haplotype Analysis

We observed a linear relationship between LD and distance from East Africa (figure 10). LD increased as a function of physical distance from East Africa. This observation supports the sequential founder hypothesis, which suggests that increase of LD at each expansion step is observed in each founder population.[229, 230]

Figure 10. LD Analysis Supports Sequential Founder Hypothesis



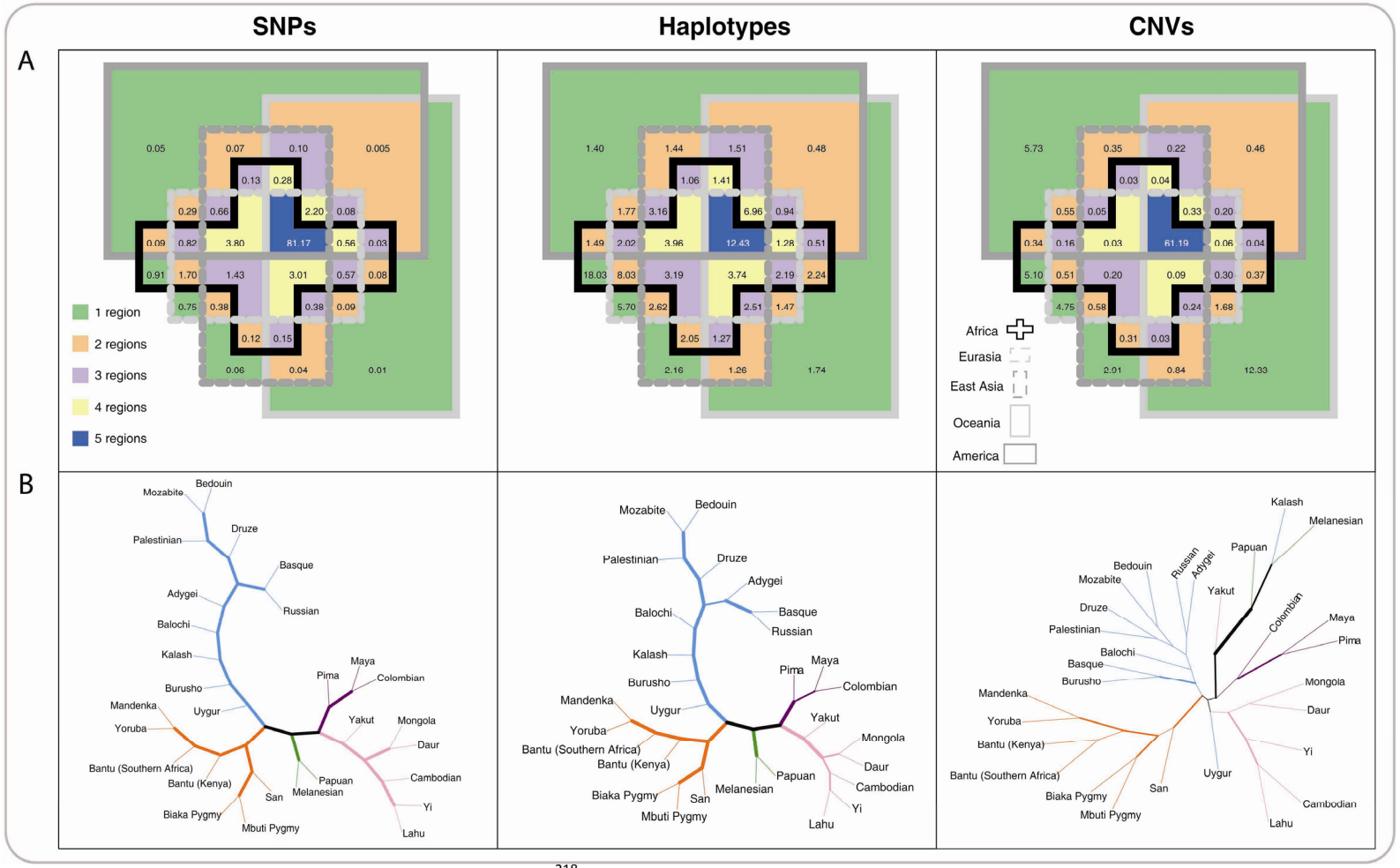
This figure has been published elsewhere (see reference [218]). We observed a linear relationship between LD and the geographic distance from Addis Ababa (Ethiopia). The color legend for the seven main geographic groups is as follows: red = African populations, pink = Middle Eastern populations, orange = European populations, yellow = Central/South Asian populations, purple = East Asian populations, blue = Oceanian populations, black = Native American populations.

To study a genomic region where natural selection is known to have occurred, we visualised the haplotype structure at the *lactase (LCT)* locus. In contrast to non-Caucasian populations, we observed increased homozygosity and one predominant haplotype cluster in Caucasian samples (figure 11). These observations are in line with previous findings according to which a selective sweep at the *LCT* locus occurred in populations that were dependent on diary product consumption. The selected *LCT* haplotype is known to be associated with high intestinal lactase activity that persists into adulthood and therefore allows a carrier individual to digest lactose after maturation. This expansion in the nutritional spectrum provided early agricultural societies that specialised in animal domestication and milk production with a survival advantage.[231, 232]

Copy Number Analysis

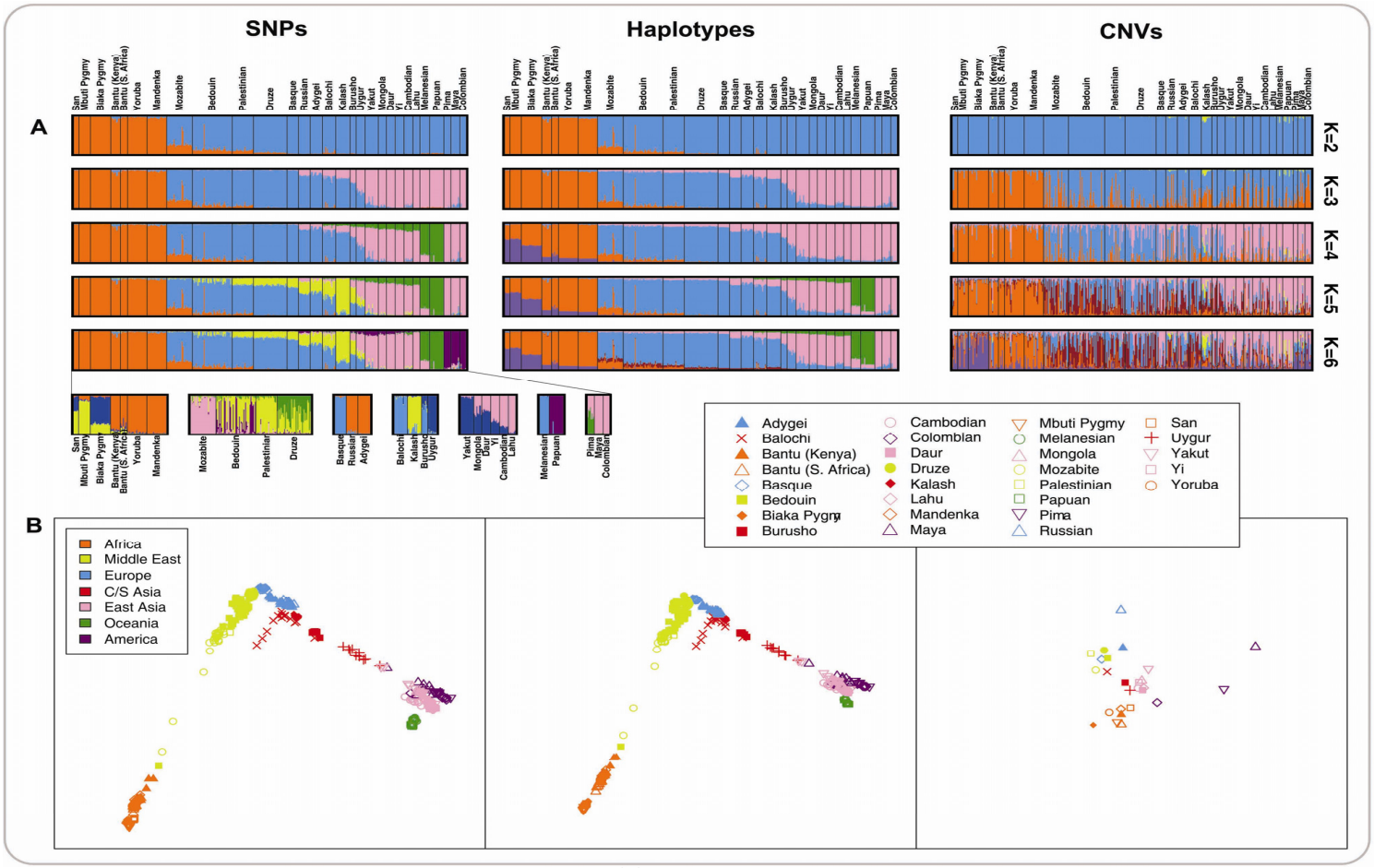
We detected 3,552 CNVs (2,398 deletions, 1,154 duplications) in 1,428 non-overlapping loci (table 6). Of these loci, 507 loci have not been previously described in the database of genomic variants (version hg18.v3).[233] Analysis of CNVs in 405 unrelated HGDP samples, revealed 3,069 CNVs in 1,333 loci. Of these loci, 2,118 CNVs in 396 loci (1,470 deletions at 296 loci and 648 duplications at 134 loci) have been observed in more than one individual. Worldwide CNV frequencies were low with only one CNV on chromosome 6 exceeding a frequency of 10%. The average size of deletions was 82.7kb (median 58.5kb, range 2 - 934kb). Duplications had a mean size of 130.4kb (median 81.1kb, range 5.6 - 998 kb). 62.2% of deletions and 68.5% of duplications were within or across genes.

Figure 8. SNPs, Haplotypes, CNVs in the HGDP Series



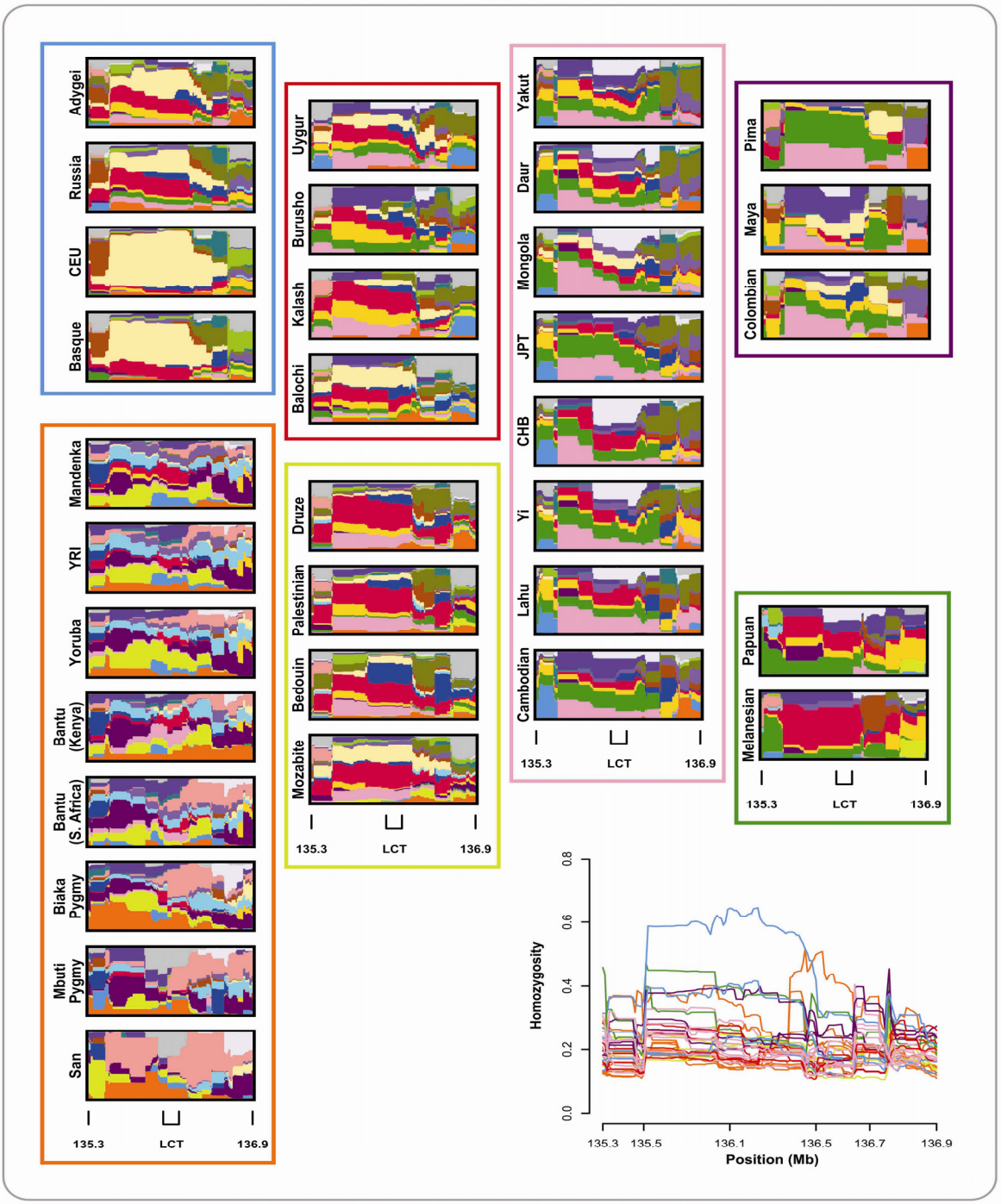
This figure has been published elsewhere (see reference [218]). Figure A shows the Venn diagrams of allele frequencies according to five different geographic regions (Africa, Eurasia, East Asia, Oceania, America). Population relationships are reconstructed in the neighbor-joining trees (figure B). Thick lines represent branches with more than 95% bootstrap support; thin lines are branches with more than 50% bootstrap support. CNVs, copy number variants. These figures have been constructed by Dr. Noah Rosenberg's group.

Figure 9. Population Structure in the HGDP Series



This figure has been published elsewhere (see reference [218]). Subpopulation structure is depicted in figure A. Each individual is indicated by a vertical line. The colors are inferred according to the number of K clusters assigned. For example, with two clusters we can differentiate the African populations from the non-African populations. In the bottom panel the subpopulation clusters for each major geographic region are shown. In figure B, the genetic population distances inferred by MDS analysis are shown. CNVs, copy number variants; C/S, Central/South Asia; K, number of clusters. These figures have been constructed by Dr. Noah Rosenberg's group.

Figure 11. Haplotype Structure at the LCT Locus Demonstrates Natural Selection



This figure has been published elsewhere (see reference [218]). This figure shows the haplotype cluster frequencies for 156 consecutive SNPs at the LCT locus across populations. Vertical lines indicate the haplotype cluster frequencies at each SNP position. Each of the 20 different haplotype clusters has a separate color. In the Caucasian populations, (blue box) one predominant haplotype cluster, suggesting selection, is observed. In support of this notion, we observed increased homozygosity at the LCT locus in the Caucasian populations (bottom panel). This figure has been constructed by Dr. Paul Scheet.

Population Relationships

Neighbor-joining trees inferred phylogenetic relationships (figure 8). The five main geographic regions (Africa, Eurasia, East Asia, Oceania, America) separated into five different branches when the SNP and haplotype data were used, however, with the CNV data the results were less reliable. This observation can be explained by the lower frequency of CNVs compared to SNPs, a bias against the detection of common CNVs due to SNP signal reclustering, as well as cell culture and CNV calling artifacts. It is worth noting that Native American populations clustered closely to the Yakut population from Siberia, supporting the hypothesis that Native Americans are the descendents of a Siberian population that migrated to the American continent via the frozen Bering Strait (figure 12).[234] Population distances, inferred by MDS analyses, clearly demonstrate that geographic distance is a primary determinant of human genetic differentiation (figure 9).

2.4 Discussion

Here we report the analysis of genotype, haplotype and copy-number variation in 29 worldwide human populations. Raw genotype data (over 250 million genotypes) are publicly available at the Gene Expression Omnibus homepage (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10331) hosted by the National Center for Biotechnology Information.

Our study indicates that genotypic variability within and across human populations is limited. Over 80% of all investigated SNP alleles were shared between all five main geographic population groups (figure 8). However, these estimates have to be interpreted with caution, as the selection of SNPs on the Illumina platform was biased towards the Caucasian populations. More specifically, the genomic coverage of the Illumina HumanHap 550K chip, defined as the fraction of the common SNPs that are tagged by the SNPs on the chip, is best in Caucasians (genomic coverage in the Caucasian CEU HapMap population = 87%), intermediate for Asian populations (genomic

coverage for the Asian JPT & CHB HapMap population = 83%) and moderate for African populations (genomic coverage in African YRI HapMap population = 50%).[235, 236] Thus, it is likely that the frequency of common private alleles is higher than our estimates, particularly in the genetically more diverse African populations.

In previous studies, population structure and relationship inferences have been drawn using microsatellite markers, short insertion-deletion polymorphisms, SNPs as well as mitochondrial and Y-chromosomal markers.[34, 221, 237-243] Clustering according to major geographic groups and increased genotypic diversity in African populations have been consistently described. Concordant with these observations, we demonstrated increasing LD with geographic distance from East Africa supporting the serial founder hypothesis for an out-of-Africa spread of human populations.[229, 230]

Figure 12. Human Migration Paths

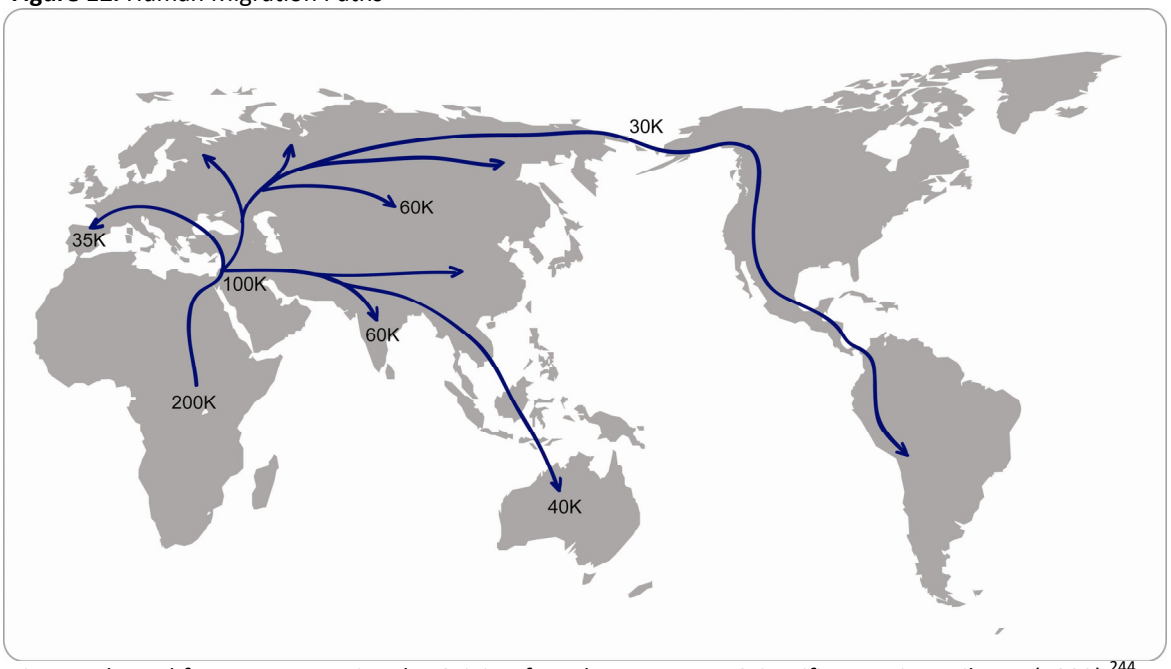


Figure adapted from Roger Lewin, The Origin of Modern Humans, Scientific American Library (1993).[244] K, thousand years ago.

One other intriguing finding was the observation that geographical distance is the primary determinant of human genetic differentiation. MDS plots, arranged in a geographic orientation, clearly show a North-South axis determining genetic diversity

among African and Middle Eastern populations and an East-West axis explaining the genetic diversity among other populations (figure 9). This path of genetic diversity indicates routes of human population migrations, starting from the African continent and spreading via the Middle East to other continents. For comparison, figure 12 shows the presumed human migration paths based on previous archeological, anthropological and genetic research.[244]

We furthermore were able to reconstruct a high-resolution map of population relationships, with five main geographic subdivisions (Africa, Eurasia, East Asia, Oceania, America). Interestingly, in the African populations we observed a separation of Sub-Saharan African hunter-gatherers (San, Mbuti, Biaka) from other African populations (figures 8 and 9). In addition, we noted that Native American populations (Maya, Pima, Colombian) grouped closely to the Yakut population from Siberia indicating a close relatedness between Siberian and Native American populations. This observation supports the hypothesis that Native Americans are the descendents of a Siberian population that migrated to the American continent via the frozen Bering Strait.[234]

One very important application of genome-wide datasets in genetically diverse populations is to investigate regions of interest for selection signatures. As an example, we studied the LCT locus on chromosome 2q21. Our analysis suggests a selective sweep at the LCT locus in European populations, particularly the western European CEU population, as indicated by a high frequency of uninterrupted haplotypes and high homozygosity levels (figure 11). The most common haplotype identified in Europeans is well known to be associated with high lactase activity that persists into adulthood, - a selection for the ability to digest lactose from milk of domesticated animals.[245, 246]

Accumulating evidence emphasises the importance of cataloguing copy number variable loci in the human genome. CNVs form a novel type of genetic diversity of which little is known. Many CNVs probably have clinical relevance as they commonly embrace genes.

Recent CNV studies have observed significant associations of CNVs with neuroblastoma, autism, schizophrenia, and autoimmune diseases.[247-254] More CNVs predisposing individuals for disease are likely to be identified over the next years and thus, detailed CNV catalogues constitute a valuable resource for this kind of research. In this study, we observed 3,552 CNVs in 1,428 loci, of which 507 loci have not been previously described. Over 60% of CNVs were within or across genes and are therefore likely to contribute to phenotypic variability within or across populations. What is currently not clear is whether rare or common CNVs are more likely to contribute to disease. Conrad et al. demonstrated that CNVs were less frequently observed in RefSeq genes compared to random permutations, indicating that purifying selection over time acts against coding CNVs and therefore makes it more likely that rare CNVs, rather than common CNVs, contribute to complex disease.[327] In our study, deletions were found to be about twice as common as duplications, an observation which was also described by Conrad et al.[327] All CNVs have been published in the publicly accessible Database of Genomic Variants (http://projects.tcag.ca/variation/).[233]

In this study, we demonstrated the utility of genome-wide SNP genotyping for population genetic research, generated a genome-wide catalogue of CNVs in ethnically and geographically diverse populations, and drew inferences about population relationships and inter-population variation. These publicly available data will serve as a valuable genomic resource to the research community.

3 Neurogenomics of Movement Disorders

3.1 Autozygosity Mapping in Brazilian Dystonia-Parkinsonism Families

STATEMENT OF CONTRIBUTIONS TO THIS RESEARCH:

Dr. Sarah Camargos, Dr. Javier Simon-Sanchez, Dr. Coro Paisan-Ruiz and I contributed equally to the laboratory based sequencing and genotyping experiments. I was involved in the data management and statistical analysis. I drafted parts of the original manuscript that has been published in the journal Lancet Neurology.

3.1.1 Background

We identified three apparently unrelated Brazilian families with dystonia and parkinsonism (pedigrees are shown in figure 13). Consanguinity in parents of affected family members suggested an autsosomal recessive mode of inheritance. To identify regions of autozygosity (i.e. homozygous regions identical by descent due to a common recent ancestor) across all affected family members, we performed genome-wide SNP genotyping, yielding approximately 550,000 SNPs per individual. We identified a homozygous region on chromosome 2 that encompassed 277 contiguous homozygous markers in all affected family members. Sequence analysis of all genes in this region revealed a single, nonsynonymous, disease-segregating mutation in *protein kinase, interferon-inducible double-stranded RNA-dependent activator (PRKRA)*.

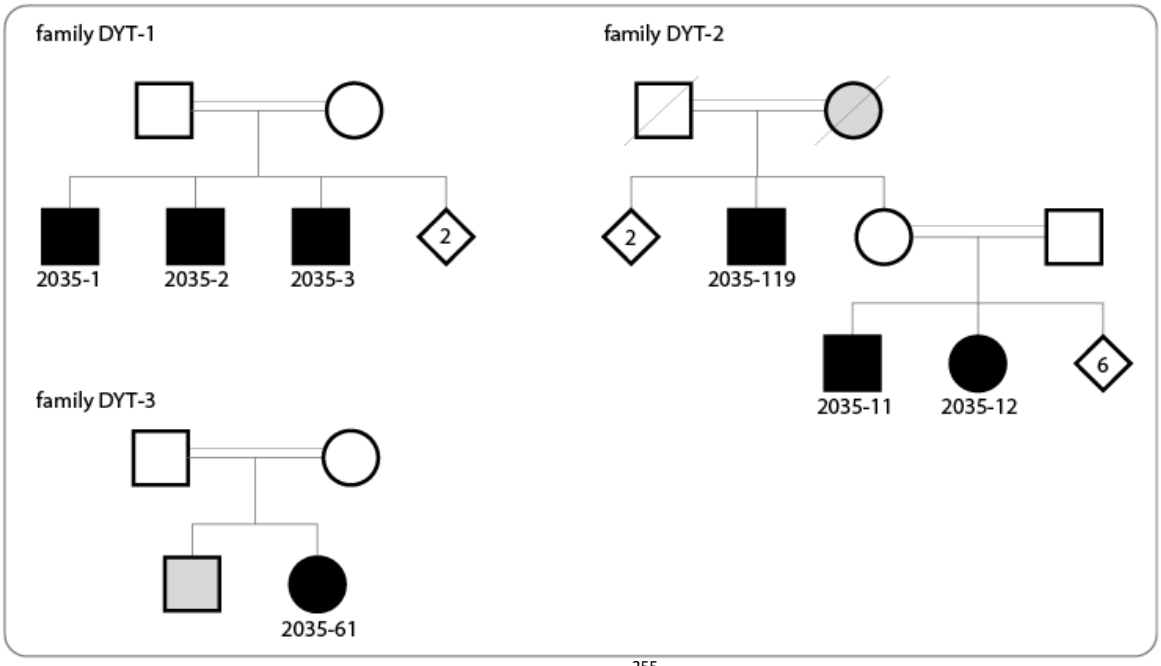
3.1.2 Materials and Methods

Samples

DNA samples from three dystonia-parkinsonism families (DYT-1, DYT-2, DYT-3; pedigrees are shown in figure 13), five sibling pairs with dystonia, 11 unrelated young-onset dystonia patients (age at onset 8 - 42 years), 45 young-onset PD patients (age at onset 18 - 40 years), and 83 neurologically normal controls (age at sampling 25 - 50 years)

were collected at the Movement Disorders Clinic at the Federal University of Minas Gerais, Brazil, by Dr. Sarah Camargos and Dr. Francisco Cardoso.

Figure 13. Pedigrees from Three Brazilian Dystonia-Parkinsonism Families



This figure has been published elsewhere (see reference [255]). A square represents a male person, a circle represents a female person and a diamond represents a person of unspecified sex. A double horizontal line indicates parental consanguinity. Black symbols indicate affected members in whom a neurological exam was performed; grey symbols indicate individuals who were affected based on the family history, but no neurological exams were performed; blank symbols show unaffected family members. A diagonal line marks deceased individuals.

We furthermore included 426 Portuguese neurologically normal controls (age at sampling 29 - 85 years) which were collected at the Coimbra University Hospital in Portugal. Additional 439 North American Caucasian neurologically normal controls (age at sampling 55 - 95 years) and 249 North American Caucasian young-onset PD patients (age at onset 7 - 52 years) were selected from the NINDS neurogenetics cell line repository at the Coriell Institute for Medical Research, NJ, USA (http://ccr.coriell.org/). We also included 738 ethnically diverse samples from the Human Genome Diversity Panel (CEPH Institute in Paris, France).

All participants gave written informed consent. Each local ethics board approved the study.

Phenotypic Characterisation in Dystonia Families

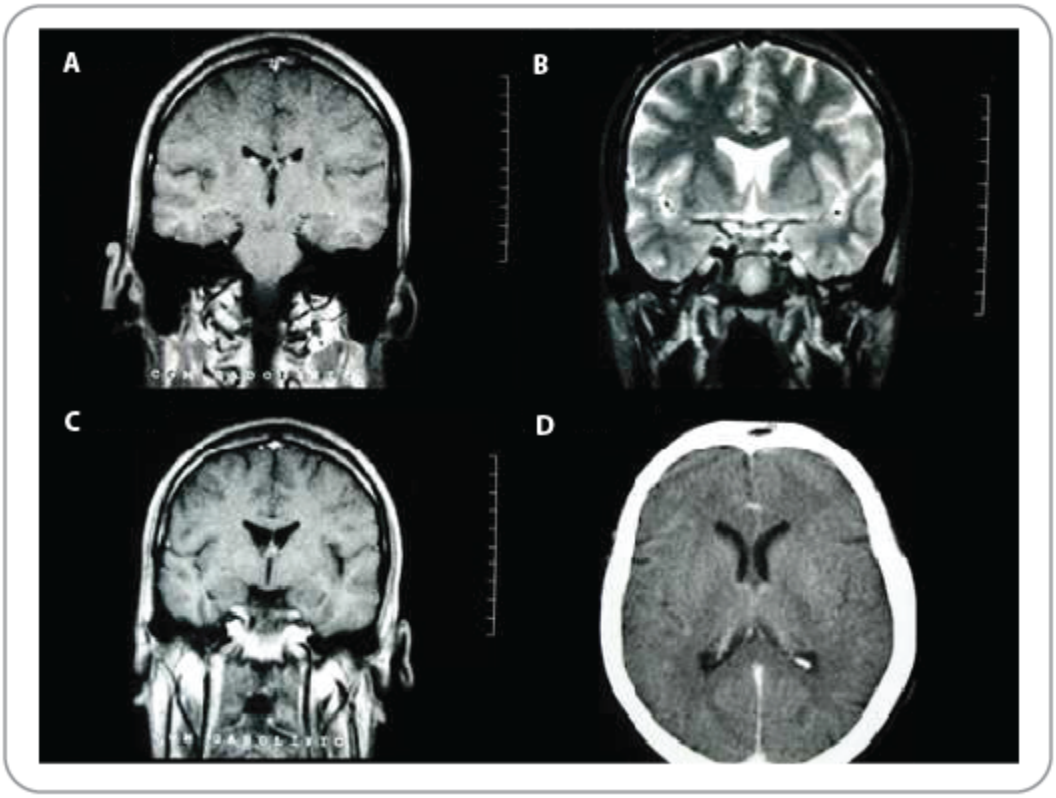
Family members from families DYT-1, DYT-2 and DYT-3 were examined by movement disorders specialists Dr. Sarah Camargos and Dr. Francisco Cardoso. Clinical features of affected family members are summarised in table 8. Patients from families DYT-1 and DYT-2 underwent computer tomography (CT) scanning of the head and in patients 2035-1 from family DYT-1 and 2035-11 from family DYT-2 magnetic resonance imaging (MRI) was performed on a 0.2T system (Magneton P8, Siemens, Erlanger, Germany) (figure 14).

Table 8. Clinical Characteristics of Affected Family Members

Patient ID	Family ID	AAO	First sign	Generalised dystonia	BFM scale score	Parkinsonism	UPDRS motor score	Pyramidal signs
2035-1	DYT-1	11	LLD[4]	severe	96	absent	0	absent
2035-2	DYT-1	12	LLD[4]	moderate	58	present	21	present
2035-3	DYT-1	2	LLD[4]	moderate	36	absent	0	present
2035-11	DYT-2	11	ULD[5]	severe	64	present	21	absent
2035-12	DYT-2	2	SD[6]	slight	14	present	30	absent
2035-119	DYT-2	18	LLD[4]	moderate	44	present	18	absent
2035-61	DYT-3	7	ULD[5]	severe	103	absent	19	present

This table has been published elsewhere (see reference [255]).
AAO, age at onset in years; BFM, Burke-Fahn-Marsden; UPDRS, Unified Parkinson's disease rating scale; LLD, lower limb dystonia; ULD, upper limb dystonia; SD, spasmodic dysphonia

Figure 14. CT and MRI Scans of Affected Family Members



These images have been published elsewhere (see reference [255]). Coronal T1 weighted (A, B) and T2 weighted (C) MRI scans after gadolinium injection in patient 2035-11 and a CT scan (D) in patient 2035-12 show no specific abnormalities.

Candidate Gene Screening

To rule out known genetic causes of dystonia or parkinsonism, affected family members from families DYT-1 and DYT-2 have been previously screened for mutations in the following candidate genes: *PRKN, ATP1A3* and *GCH1*. Dr. Sarah Camargos has carried out the candidate screening experiments and analyses; therefore, I only briefly outline the methods for this candidate screening. All exons and exon-intron boundaries were PCR amplified. After purification of the PCR products with PCR Cleanup Filter Plates (Millipore, MA, USA), direct dye terminator sequencing (BigDye v3.1; Applied Biosystems, CA, USA) was performed as per the manufacturer's instructions (see chapter 3.4 for a more detailed description). The resulting reactions were cleaned with PCR Cleanup Filter Plates (Millipore, MA, USA), processed on a 3730xl DNA Analyzer

(Applied Biosystems, CA, USA) and analysed with Sequencher™ software (version 4.1.4; Gene Codes Corporation, MI, USA). Furthermore, all affected family members were screened for the 3-bp ΔGAG mutation in exon 5 of *TOR1A* using direct sequencing as described above. In addition, copy number variation in *PRKN* was investigated by real-time polymerase chain reaction (PCR) using the protocol described by Dogu et al.[256] Also, one patient from family DYT-2 had previously been negatively screened for missense mutations in *PINK1, LRRK2* and *SNCA*, and for copy number variation in *SNCA*.[257]

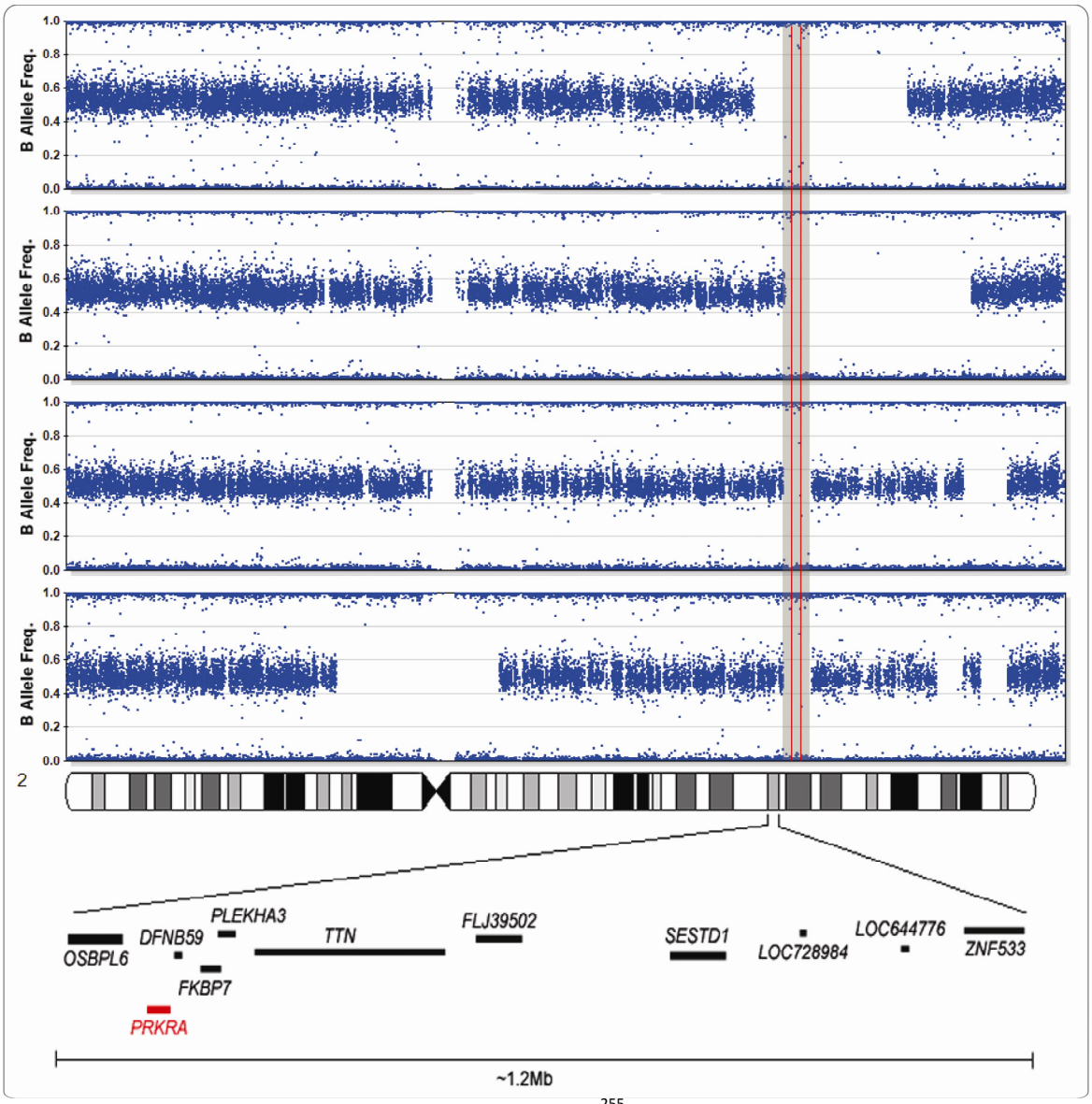
Identity by State Analysis

We estimated the proportion of the genome that was identical by state (IBS) between affected family members, to investigate whether families DYT-1, DYT-2 and DYT-3 were related due to a common recent ancestor and to estimate the levels of inbreeding. This analysis was performed in PLINK software (v.1.04)[258] using the --genome command.

Autozygosity Mapping

To study regions of homozygosity that are shared between affected individuals, we performed genome-wide SNP genotyping in two affected family members of families DYT-1 and DYT-2, five sibling pairs with dystonia, 11 unrelated young-onset dystonia patients and 45 early-onset parkinsonism cases. Each individual was assayed on a HumanHap550 SNP chip (Illumina, CA, USA) yielding approximately 550,000 SNP genotypes per individual. Genotyping was performed as per the manufacturer's instructions (see chapter 2.2 for further details).

Figure 15. Homozygous Disease-Segregating Locus on Chromosome 2



This figure has been published elsewhere (see reference [255]). B allele frequency (BAF) metrics are shown in four affected family members (2035-1 and 2035-2 from family DYT-1, 2035-11 and 2035-12 from family DYT-2). The BAF estimates the probability that a genotype is homozygous for the alternative form of an allele (B allele). Specifically, a SNP that is homozygous for the B allele has a BAF of 1.0, whereas a SNP that is homozygous for the A allele has a BAF of 0.0. Heterozygous SNPs cluster at a BAF of 0.5. Each blue dot represents the BAF of a SNP. The x-axis of the BAF plot indicates the physical position of each SNP on chromosome 2. Shared homozygous regions are highlighted in grey. The red lines mark the extend of the largest homozygous region. In the bottom panel the positions of all genes in this region are shown.

We analysed raw genotype data in BeadStudio (version 3.1.4; Illumina) and determined disease-segregating contiguous tracks of homozygosity in Tracker software (version 0.99; an in-house software tool developed at the Laboratory of Neurogenetics, National

Institute on Aging, MD, USA). We set the minimal size threshold for a homozygous track at more than 50 contiguous SNPs that had to be identical by state.

Next, we then visualised homozygous tracks in the Genome Viewer module within BeadStudio software. We noticed that patients 2035-1 from family DYT1 and 2035-11 from DYT-2 shared three adjacent homozygous regions of chromosome 2 that were identical by state. Next, we performed genome-wide SNP genotyping in eight additional family members from DYT-1 and DYT-2 ($n_{affected}$ = 4 members, $n_{unaffected}$ = 4 members). Our analysis revealed that all affected family members shared the three homozygous regions on chromosome 2 (figure 15). Furthermore, we observed that one unrelated dystonia patient also carried these homozygous regions. Re-examination of the family history of this case revealed that this patient likely has a brother with focal dystonia, but unfortunately, this brother was not available for neurological examination. We labeled this additional dystonia family as DYT-3 (the pedigree is shown in figure 13) and included this patient into further analyses. None of the other screened samples shared homozygosity at the chromosome 2 locus.

Sequence Analysis of Disease-Segregating Homozygous Regions

Autozygosity mapping revealed three homozygous disease-segregating regions on chromosome 2 spanning 2.44 Mb. A total of 12 genes or predicted transcripts were located in these regions: *OSBPL6, PRKRA, DFNB59, FKBP7, PLEKHA3, TTN, FLJ39502, SESTD1, LOC728984, LOC644776, ZNF533,* and *LOC729001*. We sequenced all exons and flanking exon-intron boundaries in patients 2035-3 from family DYT-1 and 2035-11 from family DYT-2 using BigDye terminator sequencing (Applied Biosystems). Primer sequences and the PCR thermocycling are listed in the supplemental tables S1-S2. The PCR reaction mix consisted of 25ng genomic DNA, 10pM forward primer, 10pM reverse primer and 6µl of FastStart PCR Master mix (Roche, IN, USA). Unincorporated dNTPs, primers, salts and polymerase were removed using PCR Cleanup Filter Plates (Millipore, MA, USA) as per the manufacturer's instructions; next,

Sanger-sequencing was performed using the following reaction mix: 2µl of cleaned PCR product, 0.5µl of BigDye (v.3.1, Applied Biosystems), 1µl of 10pM of primer, 2µl 5X Sequencing Buffer (Applied Biosystems), and 5µl distilled and deionised molecular grade water (Mediatech. Inc., VA, USA). The sequencing cycling conditions are listed in supplemental table S3. Following sequencing cleanup similar to the PCR purification using Cleanup Filter Plates (Millipore), sequencing reactions were processed on a 3730xl DNA Analyzer (Applied Biosystems) and analysed in Sequencher™ software (version 4.1.4; Gene Codes Corporation).

P222L Mutation Screening

We identified a nonsynonymous mutation, c.655C>T (p.P222L), in exon 7 of *PRKRA* (figure 16). This mutation has not been previously reported in the Human Gene Mutation Database, PubMed, or dbSNP. Consequently, we sequenced *PRKRA* in all family members from families DYT-1, DYT-2 and DYT-3 and in 11 sporadic Brazilian dystonia cases. Sequence analysis revealed that the P222L mutation segregated with disease in all three families (figure 17), but was not detected in the sporadic cases. To determine whether the P222L mutation is present in neurologically normal controls, we furthermore sequenced exon 7 of *PRKRA* in 439 North American Caucasian controls, 426 Portuguese Caucasian controls, 83 Brazilian controls and 738 ethnically diverse samples from the Human Genome Diversity Panel (CEPH Institute, Paris, France). Moreover, we screened 249 North American Caucasian young-onset PD patients and 45 Brazilian young-onset PD patients for the P222L mutation.

3.1.3 Results

Using autozygosity mapping followed by direct sequencing of genes and transcripts in identified homozygous tracks, we discovered a disease-segregating mutation (P222L) in *PRKRA* in three Brazilian families with autosomal recessive dystonia and parkinsonism (figure 16 and figure 17).

Initially, we excluded known genetic causes of dystonia and parkinsonism. Screening for mutations in *PRKN, LRRK2, SNCA, GCH1, TOR1A, ATP1A3* and *PINK1* was negative. In addition, measurements of serum and urine copper and ceruloplasmin ruled out Wilson's disease. CT and MRI scans showed no specific abnormalities of the basal ganglia (figure 14).

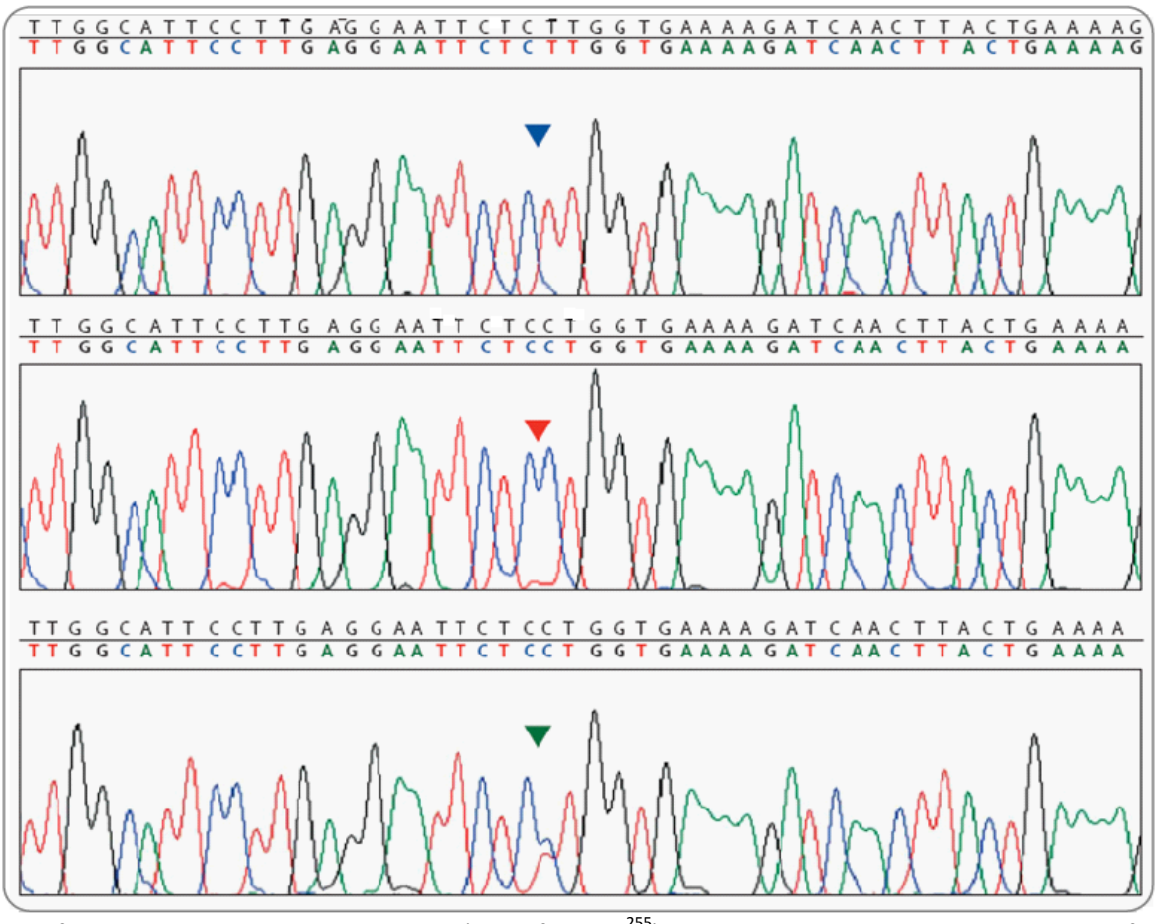
IBS analysis revealed that families DYT-1, DYT-2 and DYT-3 did not share a recent common ancestor. Consanguinity was confirmed within families DYT-1 and DYT-2, as reflected by an increased proportion of ancestrally shared genomic sequence (PI_HAT = 0.63 between siblings 2035-1 and 2035-2 from family DYT-1; PI_HAT = 0.68 between siblings 2035-11 and 2035-12 in family DYT-2). In family DYT-3 only the proband participated in the study. Therefore, consanguinity testing was not performed in this family.

Consanguinity in the parents of affected family members suggested a primary dystonia with an autosomal recessive mode of inheritance. We therefore performed autozygosity mapping in these families to identify regions that were shared between affected individuals. We used genome-wide SNP genotyping and computational analyses to identify contiguous homozygous tracts shared by all affected family members. We discovered three regions on chromosome 2 that fulfilled these criteria. These three regions were in close proximity to each other and spanned a total of 2.4Mb (figure 15). Region A contained 78 homozygous SNPs and spanned 0.4Mb, region B contained 67 homozygous SNPs and spanned 0.1Mb and region C contained homozygous 278 SNPs encompassing 1.2Mb. In region C, we observed a single SNP (rs4897088) for which affected family members from family DYT-1 were homozygous AA whereas affected family members in family DYT-2 were homozygous GG. To exclude a genotyping error, we confirmed this finding using direct sequencing. While this finding was surprising, the large extend of this homozygous track that was identical by state in 277 out of 278 SNPs

in all affected family member still implied that this region contained a disease-causing mutation.

Twelve genes or predicted transcripts were located in the identified homozygous regions (figure 15). Sequence analysis of all transcripts revealed a single, disease-segregating, homozygous mutation (c.655C>T, p.P222L) in exon 7 of *PRKRA*, the gene that encodes protein kinase, interferon-inducible double-stranded RNA-dependent activator (figure 16).

Figure 16. Electropherograms of the P222L Mutation in *PRKRA*



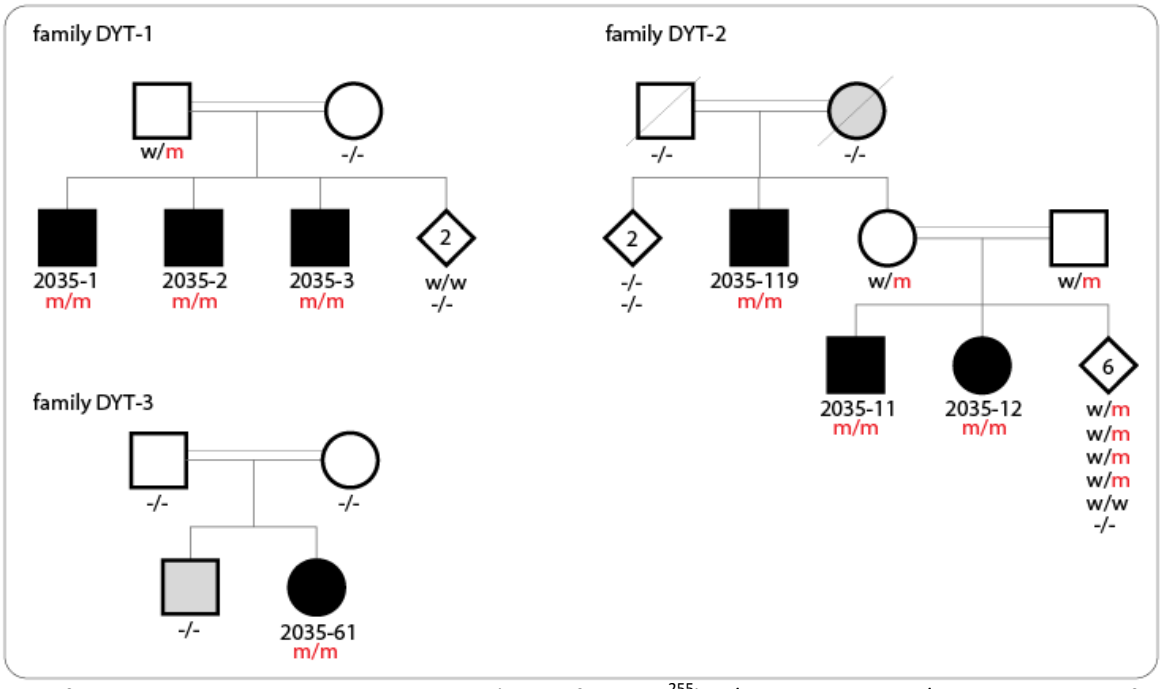
This figure has been published elsewhere (see reference [255]). Wildtype and mutant electropherograms for the P222L mutation in *PRKRA* are shown. The blue arrow in the top panel denotes a mutant homozygote, the red arrow in the middle panel shows the wildtype sequence in an unaffected family member and the green arrow in the bottom panel indicates a heterozygous asymptomatic mutation carrier.

Screening for the P222L mutation in 1,686 ethnically diverse controls was negative, indicating that the observed mutation is likely to be pathogenic. Furthermore, this mutation was not found in 294 patients with young-onset PD or in 11 patients with young-onset sporadic dystonia.

3.1.4 Discussion

Here we report the identification of a novel disease-segregating missense mutation (P222L) in *PRKRA* in three Brazilian families with autosomal recessive dystonia and parkinsonism (figure 17). Affected family members suffer from a childhood-onset, progressive, generalised dystonia, including opisthotonus, spasmodic dysphonia, and prominent orofacial dystonia. Parkinsonian features are present in some patients. These symptoms do not show improvement to L-DOPA therapy.

Figure 17. P222L Mutation in *PRKRA* Segregates with Disease



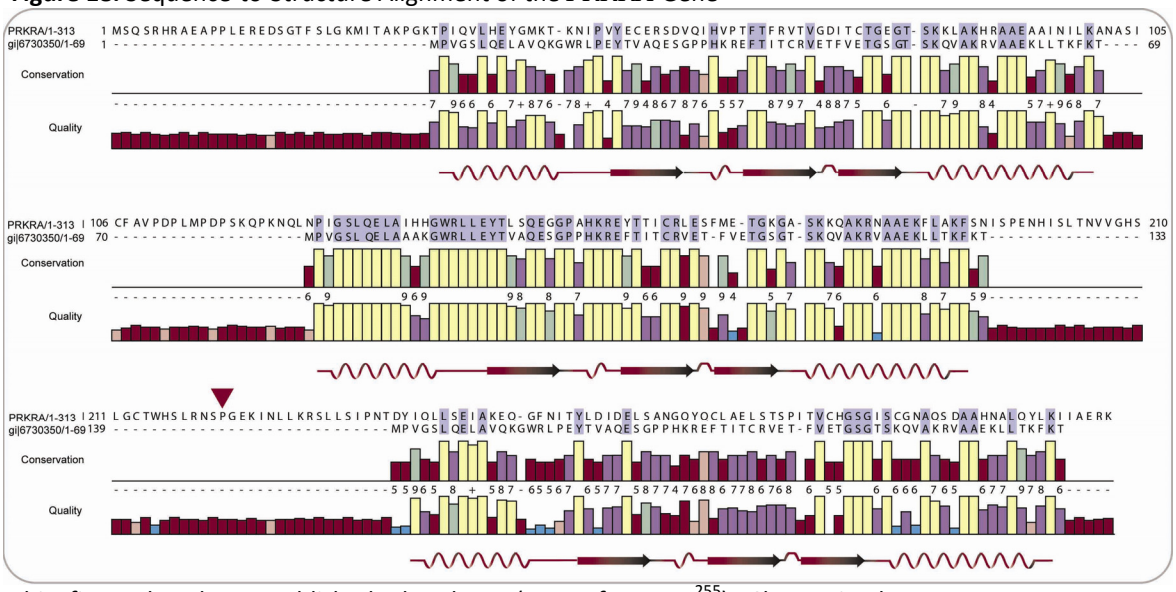
This figure has been published elsewhere (see reference [255]). -/-, not tested; w/m, heterozygous for wildtype and mutant allele in *PRKRA*; m/m, homozygous for mutant allele; w/w, homozygous for wildtype allele. Squares represent males and circles represent females. Open symbols describe unaffected individuals; in affected individuals the symbols are filled with black; for individuals in which medical history suggested disease affection but no neurological exam was available the symbols are filled with grey. Symbols with diagonal lines represent deceased subjects.

We performed autozygosity mapping in these families using dense genome-wide SNP chips, revealing three adjacent disease-segregating homozygous regions on chromosome 2 that were identical by state in all affected family members. Twelve genes and predicted transcripts were located within these regions. Detailed sequence analysis of all exons and exon-intron boundaries lead to the discovery of a single missense mutation in exon 7 of *PRKRA*. Affected family members were homozygous for the P222L mutation. Heterozygous mutation carriers were asymptomatic.

SNP genotypes surrounding the P222L were identical by state in all affected family member, pointing towards a founder effect. This mutation was absent in 1,686 ethnically diverse control samples, supporting the assertion that the P222L mutation is pathogenic. However, it is also possible that the P222L mutation is in LD with the actual disease-causing mutation. Description of the P222L mutation in other families with autosomal recessive generalised dystonia together with cell-based experiments is still required to establish *PRKRA* as a novel dystonia gene.

PRKRA has not previously been implicated with human disease and the mechanism by which the P222L mutation might cause disease is unclear. *PRKRA*, encoding protein kinase interferon-inducible double-stranded RNA-dependent activator, plays an important role in the regulation of mRNA translation. PRKRA protein binds to dsRNA (the dsRNA binding domains are shown in figure 18) and activates the latent protein kinase PKR.[259] Activated PKR then inhibits protein synthesis via inhibition of the eukaryotic translation initiation factor 2α (EIF2 α).[260] Another mechanism by which PRKRA regulates protein synthesis is through influencing RNA interference in the RNA-induced silencing complex (RISC), of which PRKRA is a component.[261]

Figure 18. Sequence-to-Structure Alignment of the *PRKRA* Gene



This figure has been published elsewhere (see reference [255]). Shown is the sequence-to-structure alignment between the human *PRKRA* sequence and the three-dimensional structure of a dsRNA-binding domain (gi:6730350, NCBI Structure database). The bar plots demonstrate sequence conservation and alignment quality. The red swirl ribbons and arrowed ribbons show the α-helical and β-strand structures within three conserved dsRNA-binding motifs. The newly identified nonsynonymous mutation P222L is located between the second and third dsRNA-binding domain (indicated by a red arrow). This mutation could alter the structure of *PRKRA* and/or affect substrate binding. This figure was constructed by Dr. Jinhui Ding.

A mouse model with a *PRKRA* null mutation has been described.[262] The phenotype of these mice includes craniofacial dysmorphia, hearing defect and reduced size. However, a movement disorder is not part of the phenotype. There are two explanations for the absence of dystonia or parkinsonism in these mice: (a) *PRKRA* could have a different function in mice, and (b) the deletion of the entire gene abolishes its natural function and thus, this might not be an appropriate model to make inferences on the effects of a missense mutation, which is likely to alter a protein's function rather than abolish it. More cell- and animal-based work is necessary to unravel the pathogenetic mechanisms of the P222L mutation in *PRKRA*.

After we published these findings, a German research team screened a series of sporadic and familial dystonia cases and neurologically normal controls for mutations in *PRKRA*.[263] They discovered a heterozygous frameshift mutation (c.266_267delAT;

p.H89fsX20) in exon 3 in a nine year old boy with generalised dystonia. The patient had no family history of dystonia. This mutation was absent in 126 dystonia patients and 189 neurologically normal controls. While our own research suggests an autosomal recessive mode of inheritance for *PRKRA*-associated dystonia, it is surprising that in this patient a heterozygous mutation seems to be sufficient to cause disease. It is possible that this patient has a gene dosage mutation or a mutation in a non-coding region of *PRKRA* that has not been tested in this study. Alternatively, the frameshift mutation, which is likely to cause protein truncation, might itself be pathogenic. It is hoped that more dystonia researchers will screen their patients for *PRKRA* mutations to understand the genetic mechanism of *PRKRA*-associated dystonia.

In summary, we identified a novel disease-segregating missense mutation in *PRKRA* in three Brazilian families with generalised early-onset dystonia and parkinsonism. Studying the pathophysiology of *PRKRA* mutations might reveal interesting insights into the pathogenesis of dystonia and parkinsonism.

3.2 Genome-Wide Association Study in Parkinson's Disease

STATEMENT OF CONTRIBUTIONS TO THIS RESEARCH:

I was involved in the conceptualisation and experimental design of this project. Dr. Javier Simon-Sanchez, Dr. Coro Paisan Ruiz, Jose Bras and I performed laboratory based genotyping experiments of the United States cohorts and parts of the German cohorts; I was responsible for parts of the data management, quality control procedures and descriptive data analysis. I drafted parts of the original publication that has been published in Nature Genetics and critically reviewed the manuscript.

3.2.1 Background

Over the last decade, genetic research in PD has taken enormous leaps forward. Linkage and positional cloning studies have successfully lead to the discovery of several causative genes in familial forms of PD and provided valuable insights into the pathophysiology of parkinsonism (see chapter 1.3.2 for a more detailed discussion). With advances in our understanding of the human genomic architecture and the advent of high-throughput SNP genotyping technologies, GWA testing became available and for the first time provided researchers with a powerful tool to unravel common genetic variation underlying common disease. However, the identification of risk genes for sporadically occurring PD is still challenging and attempts to define common genetic variants that confer modest risks for disease have largely failed.[150, 264, 265]

To identify common genetic variants that are significantly associated with risk for developing PD, we conducted the largest GWA study in PD to date, involving a total sample of 5,165 Caucasian PD patients and 8,151 Caucasian controls. Our analysis revealed that SNPs at the *SNCA* and *MAPT* gene loci were significantly associated with disease and clearly exceeded the conservative Bonferroni threshold for multiple testing (*SNCA*: rs2736990, *p* value = 1.89×10^{-16}; *MAPT*: rs393152, *p* value = 1.02×10^{-15}). Furthermore, we observed a single SNP (rs1491923), located 0.17Mb upstream of

LRRK2, with a low *p* value close to the significance threshold (*p* value = 7.02 x 10^{-6}), and we successfully replicated a novel risk locus on chromosome 1p32 (PARK16) which was originally identified by a Japanese GWA study.[266] Our data suggest that common genetic variability at these loci plays a critical role in the pathogenesis of PD.

3.2.2 Materials and Methods

The study consisted of two stages: (1) a screening stage and (2) a replication stage. Collection sites and characteristics of the studied cohorts are listed in table 9. Study design, sample size and quality control measures for both stages are summarised in figures 19 and 20.

Samples for the Screening Stage

The screening stage included 1,820 Caucasian PD cases and 4,047 Caucasian controls from Germany and from the United States of America (figure 19). Written informed consent was obtained for each individual and the study has been approved by the appropriate institutional review boards.

German samples: The study included 757 German PD cases from the Hertie Institute for Clinical Brain Research (Tübingen, Germany) and the University of Munich (Munich, Germany). The mean age at disease onset was 56 years (range 28 - 86 years). PD patients were diagnosed according to the United Kingdom Parkinson's Disease Society Brain Bank criteria for idiopathic PD by movement disorders specialists.[267] 20% of patients reported a positive family history for PD. Familial cases with an apparent autosomal dominant inheritance pattern were excluded. A total of 976 control samples were selected from the KORA survey (Cooperative Health Research in the Region of Augsburg, www.helmholtz-muenchen.de/kora; n = 488 samples) and the POPGEN initiative (www.popgen.de), a large German population genetics study designed to identify risk factors for various human diseases (n = 488 samples).

US samples: We selected 1,063 Caucasian PD cases (mean age at onset = 56 years, range = 7 - 98 years) and 3,071 Caucasian neurologically normal controls (mean age at sampling = 58 years, range 15 - 98 years). Specifically, 880 late-onset PD samples, 108 young-onset PD samples, and 828 neurologically normal control samples were obtained from the Coriell Institute for Medical Research, NJ, USA (http://ccr.coriell.org); 2,243 additional control samples were obtained from the CGEMS initiative (Cancer Genetics Markers of Susceptibility; http://cgems.cancer.gov) at the National Cancer Institute, Bethesda, MD, USA; and 75 young-onset PD samples were collected at the National Institute for Neurological Disorders and Stroke (Bethesda, MD, USA) by a neurologist specialised in movement disorders. All patients met the UK Brain Bank diagnostic criteria for idiopathic PD.[267] 17% of patients reported a first- or second-degree relative with parkinsonism. Patients with three or more relatives with parkinsonism or with an apparent Mendelian inheritance pattern were excluded from the study.

Samples for Replication Stage

An independent sample of 3,545 Caucasian PD cases and 4,311 Caucasian controls were selected for the replication stage (table 9, figure 20). These samples were collected in Europe (n = 1,100 cases and 2,168 controls from Germany; n = 93 cases and 99 controls from Serbia; n = 824 cases and 7 controls from the United Kingdom) and in North America (n = 1,528 cases and 2,037 controls). All cases were diagnosed according to the UK Brain Bank diagnostic criteria by movement disorders specialists.[267]

Table 9. Collection Sites and Characteristics of Studied Cohorts

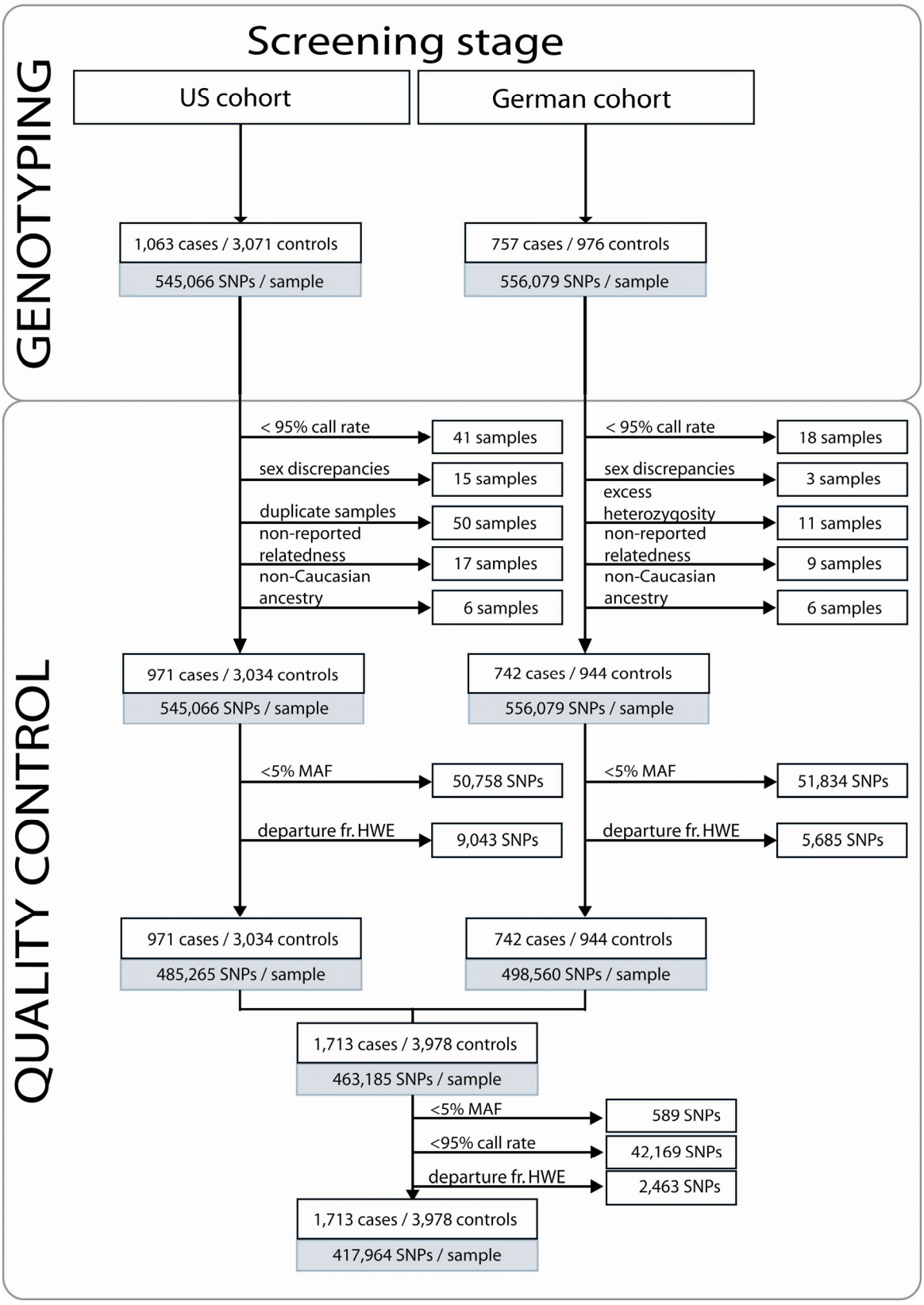
Collection site	Sample size	Sex (F/M)	Mean age at onset or age at sampling in years (range)
Screening stage			
Cases			
Hertie Institute for Clinical Brain Research, Tuebingen, Germany	757	303 / 454	56 (28 – 86)
Coriell Institute for Medical Research, NJ, USA	988	N/A	56 (7 – 98)
Laboratory of Neurogenetics, Bethesda, MD, USA	75	N/A	N/A
Controls			
KORA Initiative, Neuherberg, Germany[a]	488	N/A	N/A (25 – 74)
POPGEN survey controls, Kiel, Germany[b]	488	N/A	N/A
Coriell Institute for Medical Research, NJ, USA	828	481 / 347	59 (N/A)
CGEMS initiative, National Cancer Institute, Bethesda, MD, USA	2,243	1,142 / 1,101	N/A
Replication stage			
Cases			
Institute of Neurology, Queen Square, London, UK	824	183 / 641	59 (15 – 98)
Hertie Institute for Clinical Brain Research, Tuebingen, Germany	1,100	464 / 636	61 (N/A)
Department of Neurology, University of Belgrade, Serbia	93	29 / 64	53 (N/A)
Coriell Institute for Medical Research, NJ, USA	207	65 / 142	55 (16 – 80)
National Institute of Environmental Health Sciences, NC, USA	840	197 / 643	N/A
Washington University School of Medicine, St. Louis, MO, USA	481	182 / 299	N/A
Controls			
Institute of Neurology, Queen Square, London, UK	7	6 / 1	N/A
Hertie Institute for Clinical Brain Research, Tuebingen, Germany	2,168	902 / 1,266	57 (N/A)
Department of Neurology, University of Belgrade, Serbia	99	65 / 34	49 (N/A)
National Institute of Environmental Health Sciences, NC, USA	1,700	371 / 1,329	66 (N/A)
Washington University School of Medicine, St. Louis, MO, USA	337	219 / 118	49 (N/A)

[a] KORA Initiative (Cooperative Health Research in the Region of Augsburg, www.helmholtz-muenchen.de/kora): a population based study representative of the general population living in or near the region of Augsburg, Germany

[b] POPGEN survey controls (www.popgen.de): a cross-section epidemiological survey of regional German populations

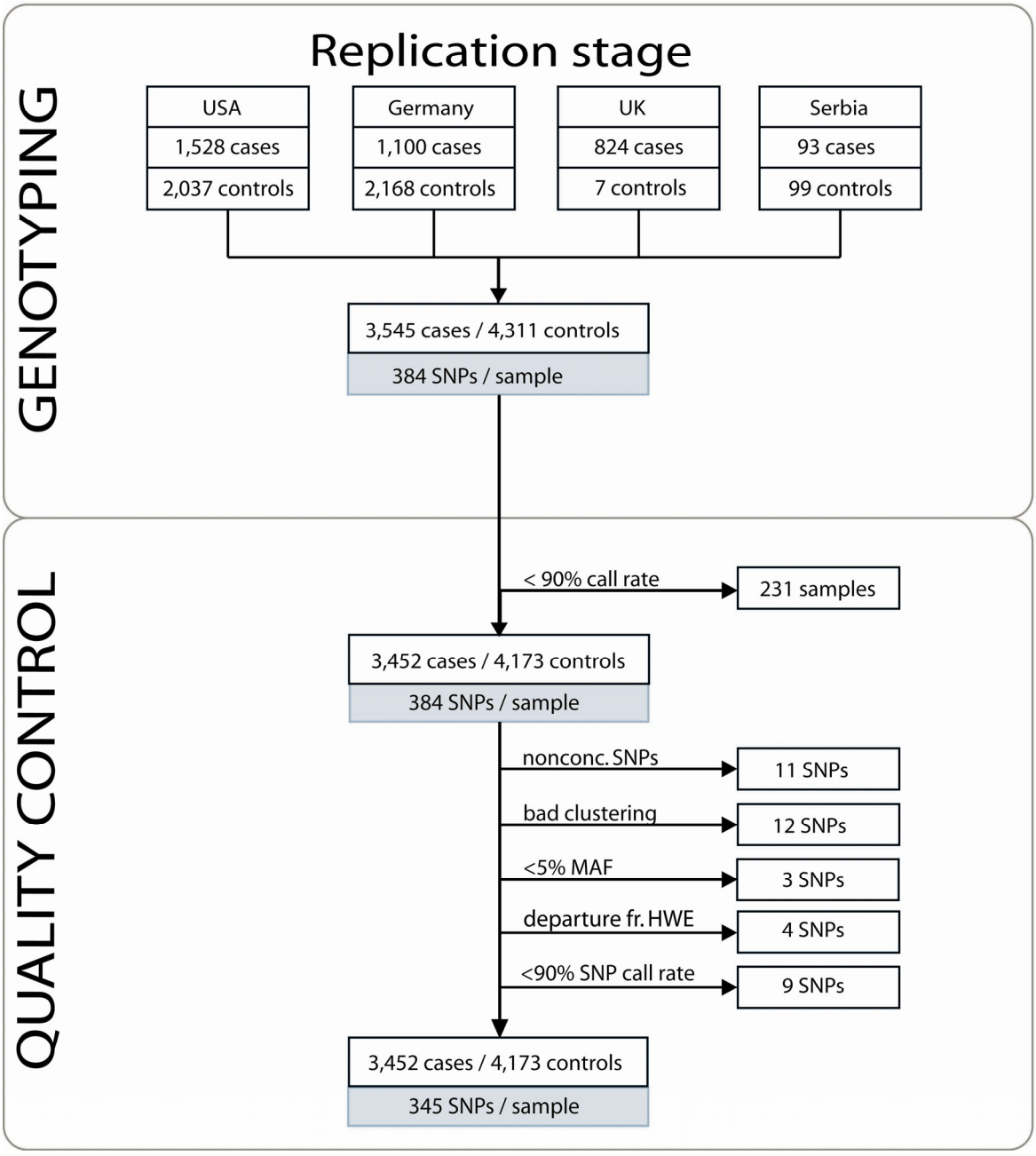
N/A, not available; F, female; M, male

Figure 19. Screening Stage: Genotyping and Quality Control Flow Chart



HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency

Figure 20. Replication Stage: Genotyping and Quality Control Flow Chart



HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; nonconc. SNPs, nonconcordant SNPs

Genotyping for the Screening Stage

Genotyping was performed on Illumina BeadChips (550K version 1 chips, 550K version 3 chips, 240S + 317K chips) as per the manufacturer's instructions (see chapter 2.2 for a more detailed description of the genotyping workflow). For each individual we

attempted to genotype 545,066 SNP genotypes that are in common of all BeadChip versions.

Genotyping for the Replication Stage

For the replication stage, we selected 384 SNPs from the screening stage for which we observed the strongest association with disease under the trend test model. Of the 384 SNPs, there were 12 SNPs for which GoldenGate genotyping assays could not be designed. These SNPs were substituted with the next 12 strongest associated SNPs from the trend model results. Rs-numbers and characteristics of the selected SNPs are listed in supplemental table S4.

Genotyping was performed on custom-made Illumina GoldenGate® Genotyping Assays on a BeadExpress platform as per the manufacturer's protocol. The workflow of this genotyping assay is summarised in figure 21. Briefly, for each subject 250ng of genomic DNA were normalised with 10mM Tris-HCl pH 8.0 and 1mM EDTA, and incubated with 5µl of MS1 reagent (supplied by Illumina) to make single-use DNA. After 2-propanol precipitation and resuspension in RS1 solution (supplied by Illumina), an allele-specific oligonucleotide hybridisation reaction was performed. A pool of custom-designed oligonucleotides, three for each SNP locus, was added to the resuspended DNA. For each SNP, one oligonucleotide specifically binds to the minor allele, a second oligonucleotide specifically hybridises to the major allele and a third oligonucleotide is a locus-specific tag designed to bind 1 - 20bp downstream of the SNP.[268] The two allele-specific oligonucleotides have each an allele-specific universal PCR sequence attached to the 5'-end and the 3'-end of the locus-specific oligonucleotide contains an universal PCR sequence and an IllumiCode sequence that is specific for the SNP locus.

Figure 21. GoldenGate SNP Genotyping Workflow

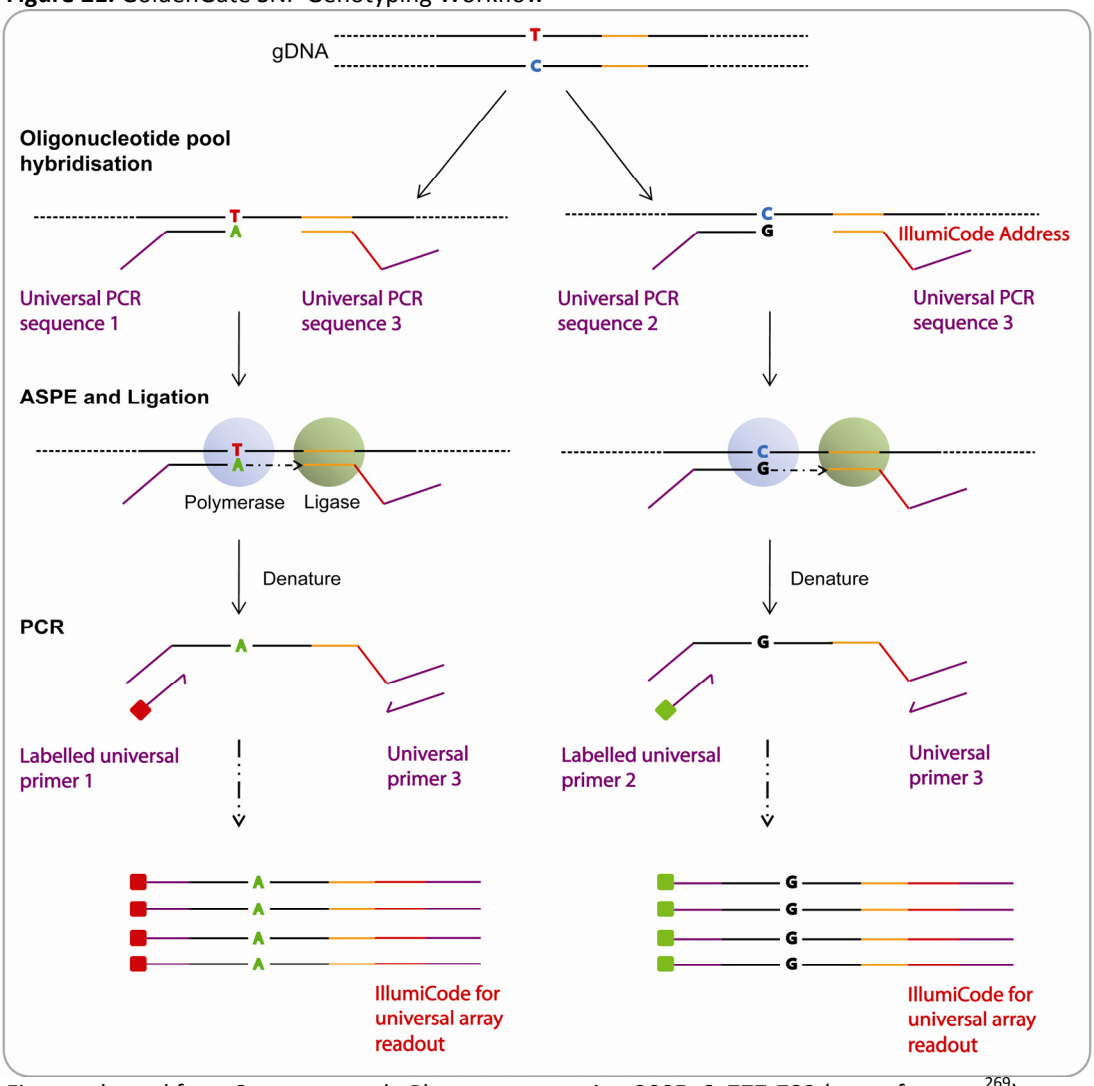


Figure adapted from Steemers et al., Pharmacogenomics, 2005, 6: 777-782 (see reference [269]).
ASPE, allele-specific primer extension; PCR, polymerase chain reaction.

After hybridisation, an allele-specific primer extension and ligation reaction was performed connecting the allele-specific oligonucleotide with the locus-specific oligonucleotide. Following this procedure, a PCR amplification of the extended and ligated oligonucleotides was carried out using three universal primers: (a) universal primer 1 binds specifically to the universal PCR sequence 1 and is labelled with fluorescent dye Cy3, (b) universal primer 2 binds to the universal PCR sequence 2 and is labelled with the fluorescent dye Cy5, and (c) universal primer 3 which binds to the universal PCR sequence 3. After this step, double-stranded PCR product was bound to

paramagnetic beads, washed, eluted and denatured with 0.1N NaOH. The IllumiCode sequence, which was part of the PCR product, was then hybridised to the VeraCode Bead Plate containing complementary sequences for each specific SNP IllumiCode address. The VeraCode Bead Plate was then scanned on an Illumina BeadExpress, raw data were generated in VeraScan software (v.1.1) and genotype calls were made in BeadStudio software (v.3.1.3) using the GenTrain clustering algorithm.

Quality Control for the Screening Stage

We excluded samples with a call rate <95% or with excess heterozygosity (>4 standard deviations from the sample mean) which could indicate genotyping errors or sample contamination. Call rates and heterozygosity rates were calculated in BeadStudio software (v.3.1.3, Illumina). To determine non-Caucasian ancestry, pairwise clustering of identity by state distances (IBS) of the genome-wide average allele proportions were computed using the IBS and multi-dimensional scaling (MDS) approach within PLINK software (v.1.04).[258] We also added publicly available data from the HapMap project (www.hapmap.org) to our dataset to visualise possible population stratification. We furthermore used the pairwise identity by descent (IBD) estimation approach implemented in PLINK to identify replicate samples or individuals with high unreported relatedness (PI-HAT value >0.2; exclusion of duplicate samples as well as first- or second-degree relatives). Genotypic sex was determined using BeadStudio software (Illumina) and samples with a discrepancy between reported and genotypic sex were excluded. Quality control filters for individual SNPs were applied in PLINK software. SNPs with a MAF <5%, SNPs with significant departure from Hardy-Weinberg equilibrium (HWE) in controls (*p* value <0.01), and SNPs with an individual genotyping rate <95% were excluded. Below, we provide a detailed explanation of the quality control measures performed in each cohort (see also figure 19).

German samples: We excluded 18 samples with a call rate <95% (n = 4 cases and 14 controls), three samples with sex mismatch (n = 2 cases and 1 control), 11 samples with

excess heterozygosity (n = 5 cases and 6 controls), six samples with non-Caucasian ancestry (n = 3 cases and 3 controls), and nine samples with high unreported relatedness (n = 1 case and 8 controls). The final cohort consisted of 742 German cases and 944 German controls. In these samples, we removed 51,834 SNPs with a MAF <5% and 5,685 SNPs with significant departure from HWE in controls (*p* value <0.01) yielding a total of 498,560 SNPs in the German cohort.

US samples: In the US cohort, we excluded 41 samples with a call rate <95% (n = 16 cases and 25 controls), 15 samples with sex mismatch (n = 11 cases and 4 controls), 50 duplicate samples (n = 49 cases and 1 control), six samples with apparently non-Caucasian ancestry (n = 5 cases and 1 control) and 17 samples with high unreported relatedness (n = 11 cases and 6 controls). The final US cohort consisted of 971 cases and 3,034 controls. In these samples, we removed 50,758 SNPs with a MAF <5% and 9,043 SNPs with significant departure from HWE in controls (*p* value <0.01).

Combined German-US samples: There were 463,185 SNPs in common between the German and US cohorts (n = 1,713 cases and 3,978 controls). After merging the datasets from both cohorts, we removed additional 589 SNPs with a MAF <5%, 42,169 SNPs with a genotyping rate <95% and 2,463 SNPs with significant deviation from HWE (*p* value <0.01). The final filtered dataset for the screening stage therefore consisted of 417,964 SNPs genotyped in 1,713 PD cases and 3,978 controls.

Quality Control for the Replication Stage

We first excluded 231 samples with a call rate <90% (n = 93 cases and 138 controls). In the remaining samples, we visualised the SNP genotype clusters in the Cartesian plots implemented in BeadStudio software (v.3.1.3, Illumina) and identified 12 SNPs with inaccurate clustering. After removal of these SNPs, we excluded additional three SNPs that had a MAF <5% in control samples, SNPs that showed significant departure from

HWE (n = 4 SNPs), or SNPs that were genotyped in less than 90% of samples (n = 9 SNPs).

In addition, we tested the concordance rate between the GoldenGate platform and the Illumina BeadChips. A total of 96 samples that were originally genotyped on BeadChips as part of the screening stage were re-genotyped using the GoldenGate assay; the concordance rate was found to be 99.2%. For 11 SNPs we observed inconsistent genotypes in more than nine samples and these SNPs have been excluded from the subsequent analyses.

The final dataset for the replication stage therefore consisted of 345 SNPs that were genotyped in a total of 3,452 cases and 4,173 controls.

Association Analysis

Screening stage: Association tests were performed using PLINK software (v.1.04).[258] For each SNP, we tested the trend test model and calculated the *p* value, OR, and 95% confidence interval (CI). Based on Bonferroni correction for multiple testing, a *p* value $\leq 1.2 \times 10^{-7}$ was considered significant (two-sided α of 0.05 divided by 417,964 SNPs tested).

Replication stage: Association tests were performed in PLINK software.[258] For all 345 SNPs that passed quality control filters, we calculated association tests under the trend model using a significance threshold of 1.5×10^{-4} (two-sided α of 0.05 divided by 345 SNPs tested).

Power Calculations

Power calculations were performed in Quanto© software (v1.2.3). This software can be downloaded from the University of Southern California website (http://hydra.usc.edu/gxe). Using the additive model, we simulated five different effect

sizes (OR = 1.1, OR = 1.2, OR = 1.3, OR = 1.5, OR = 1.8) to estimate the power for detecting a significant association in the screening stage or in the replication stage. We plotted the power to detect an association against different minor allele frequencies, considering a two-sided α of 1.2 x 10^{-7} for the screening stage and a two-sided α of 1.5 x 10^{-4} for the replication stage (genome-wide significance level after Bonferroni correction).

Linkage Disequilibrium Analysis

Computation of LD statistics at the *SNCA, MAPT, LRRK2* and PARK15 loci was performed in Haploview (v.4.1) (www.broad.mit.edu/mpg/haploview).[270] Pairwise measures of LD were calculated using the D' confidence limits algorithm described by Gabriel et al. and plotted with R software.[271]

Epistasis Test

To test possible epistatic interactions between variants at the *SNCA* and *MAPT* loci, we used a pairwise SNP x SNP epistasis algorithm implemented in PLINK software.[258]

Population Attributable Risk

The population attributable risk (PAR) was calculated with the following formula: PAR = (p[OR-1])/(p[OR-1]+1), where p represents the prevalence of the risk allele. Since epistasis analysis did not reveal evidence for an interaction between *SNCA* and *MAPT*, we calculated the combined PAR% as follows: cPAR% = [1-(1-PAR_{SNCA})*(1-PAR_{MAPT})]*100.

3.2.3 Results

Genotyping and Association Analysis in the Screening Stage

After stringent quality control filtering, we analysed 417,964 SNPs in 1,713 cases and 3,978 controls (~2.4 billion genotypes in total). The mean genotyping call rate in the filtered dataset was 99.8%. Under the trend model, we had an 80% power to detect significant associations of variants with an OR of 1.3 and a MAF of 0.1 (figure 22).

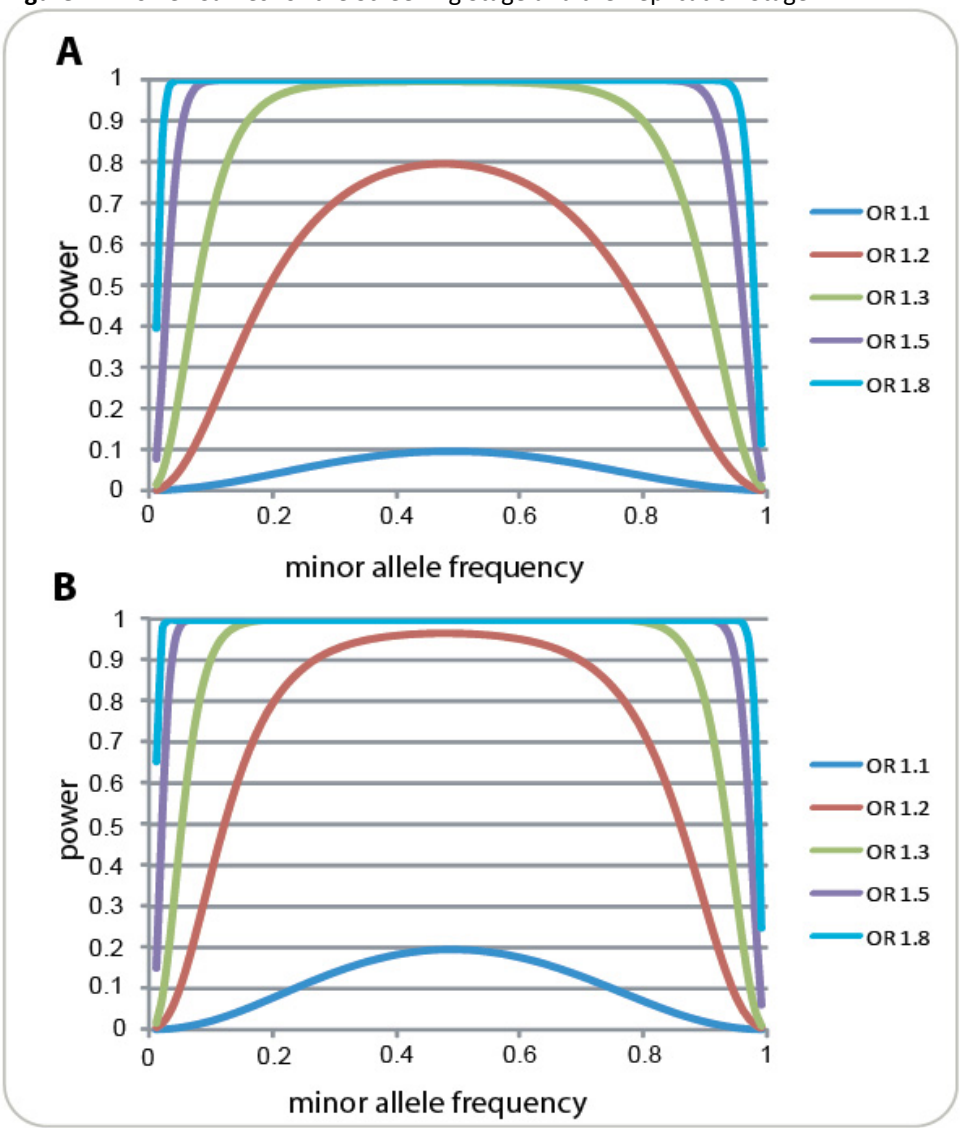
When we plotted the observed trend model *p* values against the expected *p* values (P-P plot, figure 23), we identified a significant deviation from the null hypothesis indicating disease associated loci. Furthermore, we observed mild population stratification in our data (λ = 1.2) (figures 24 and 25). We chose not to perform genomic control to correct for population stratification, as possible false positive signals would be controlled for in the replication stage.

We observed two strong association signals exceeding the conservative Bonferroni significance threshold (figure 26, table 10). The first signal was identified on chromosome 4q22 in SNPs that were located within or in close proximity to *SNCA* encoding the protein α-synuclein (rs2736990, *p* value = 5.7×10^{-9}). The second locus was found on chromosome 17q21 in a large LD block containing *microtubule-associated protein tau (MAPT)* (rs415430, *p* value = 4.5×10^{-8}).

Genotyping and Association Analysis in the Replication Stage

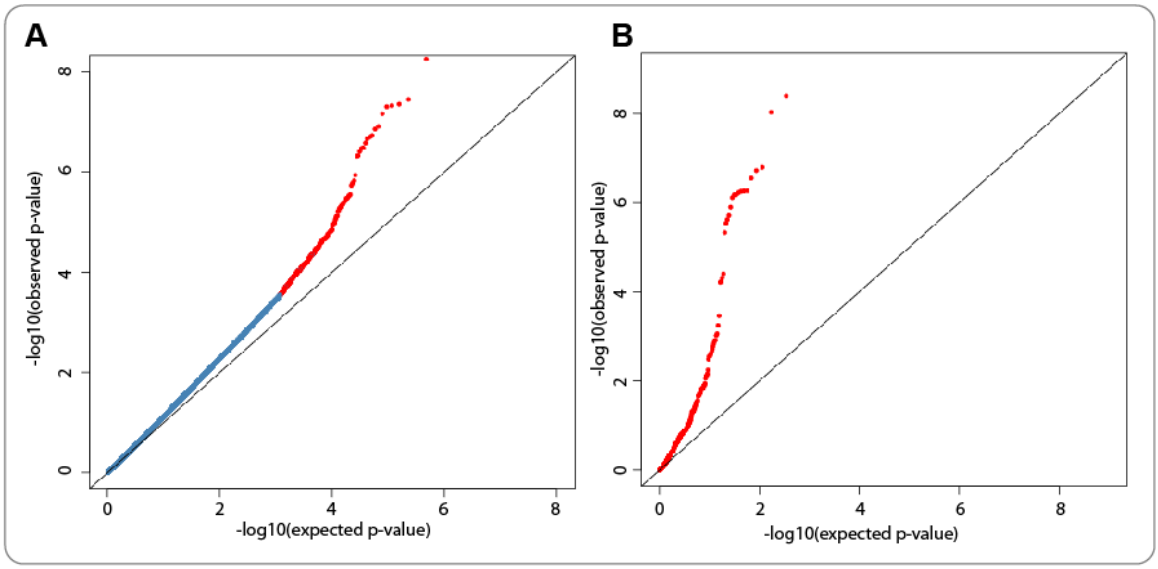
After quality control filtering, we analysed 345 SNPs in 3,452 PD patients and 4,173 controls (~2.6 million genotypes in total). Power calculations for the replication stage showed that we had 80% power to detect variants with an OR of 1.2 and a MAF of 0.2 (figure 22). We were able to replicate strong association with disease at the *SNCA* and *MAPT* loci (*SNCA* locus: rs2736990 *p* value = 4.0×10^{-9}, *MAPT* locus: rs393152 *p* value = 9.3×10^{-9}; figure 27, table 11).

Figure 22. Power Curves for the Screening Stage and the Replication Stage



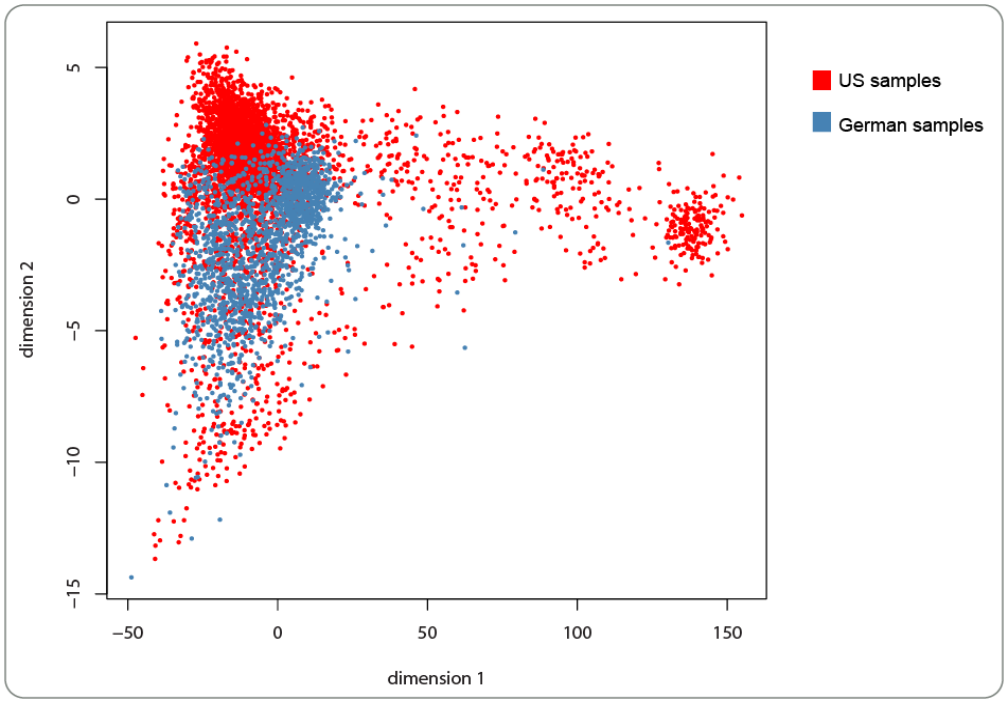
Power simulations under the trend model for the screening stage (A) and the replication stage (B) are shown. For the screening stage (n = 1,713 cases and 3,978 controls; 417,964 SNPs tested) a significance value of 1.2×10^{-7} was assumed. This simulation demonstrates that we had 80% power to detect risk variants with an OR of 1.3 and a MAF of 0.1. In the replication stage simulation (n = 3,452 cases and 4,173 controls; 345 SNPs tested) a significance threshold of 1.5×10^{-4} was used; these calculations showed that we had 80% power to detect variants with an OR of 1.2 and a MAF of 0.2.

Figure 23. P-P-plots in US-German Dataset



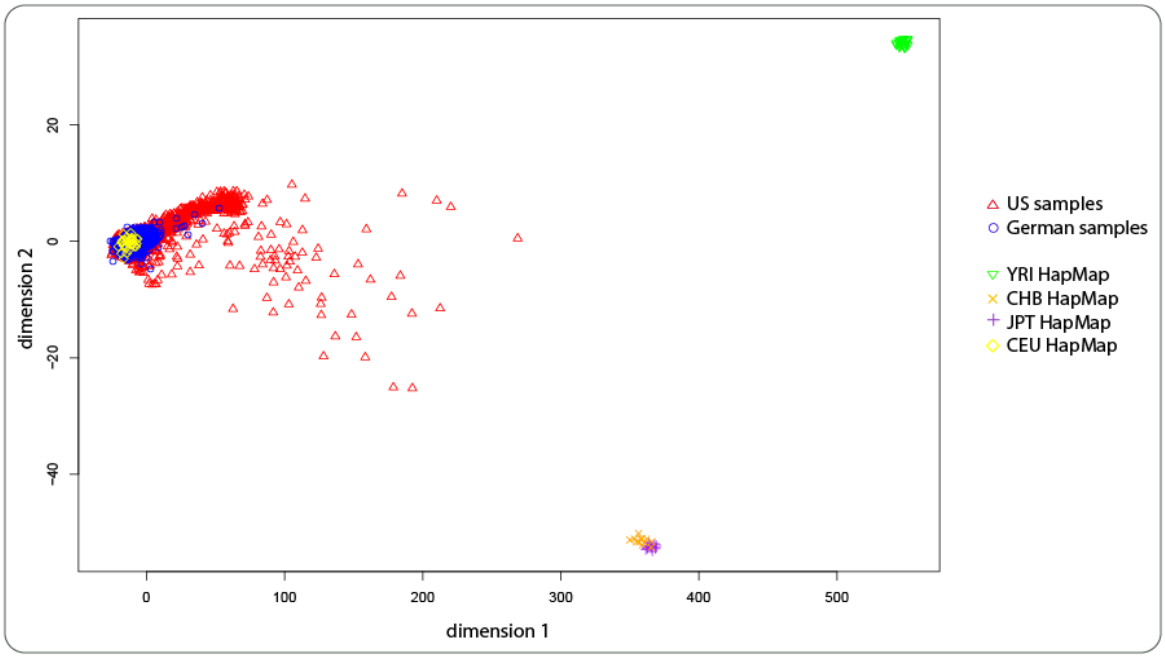
Observed *p* values of the Cochran-Armitage trend tests versus expected *p* values are shown. In figure A the trend test results of the screening stage are shown. Blue dots represent SNPs that have only been tested in the screening stage of the study; red dots represent SNPs that have been selected for the replication stage. Significant deviation from the null hypothesis is seen for the most extreme *p* values. Mild population stratification ($\alpha = 1.2$) was observed. Figure B represents the trend test results of the replication stage. The lowest observed *p* values deviate strongly from the null distribution indicating disease associated variants.

Figure 24. MDS Plot Shows Mild Population Stratification



MDS analysis shows that German und US samples cluster together, indicating that population stratification is not prominent.

Figure 25. MDS Plot of US-German Samples and HapMap Samples

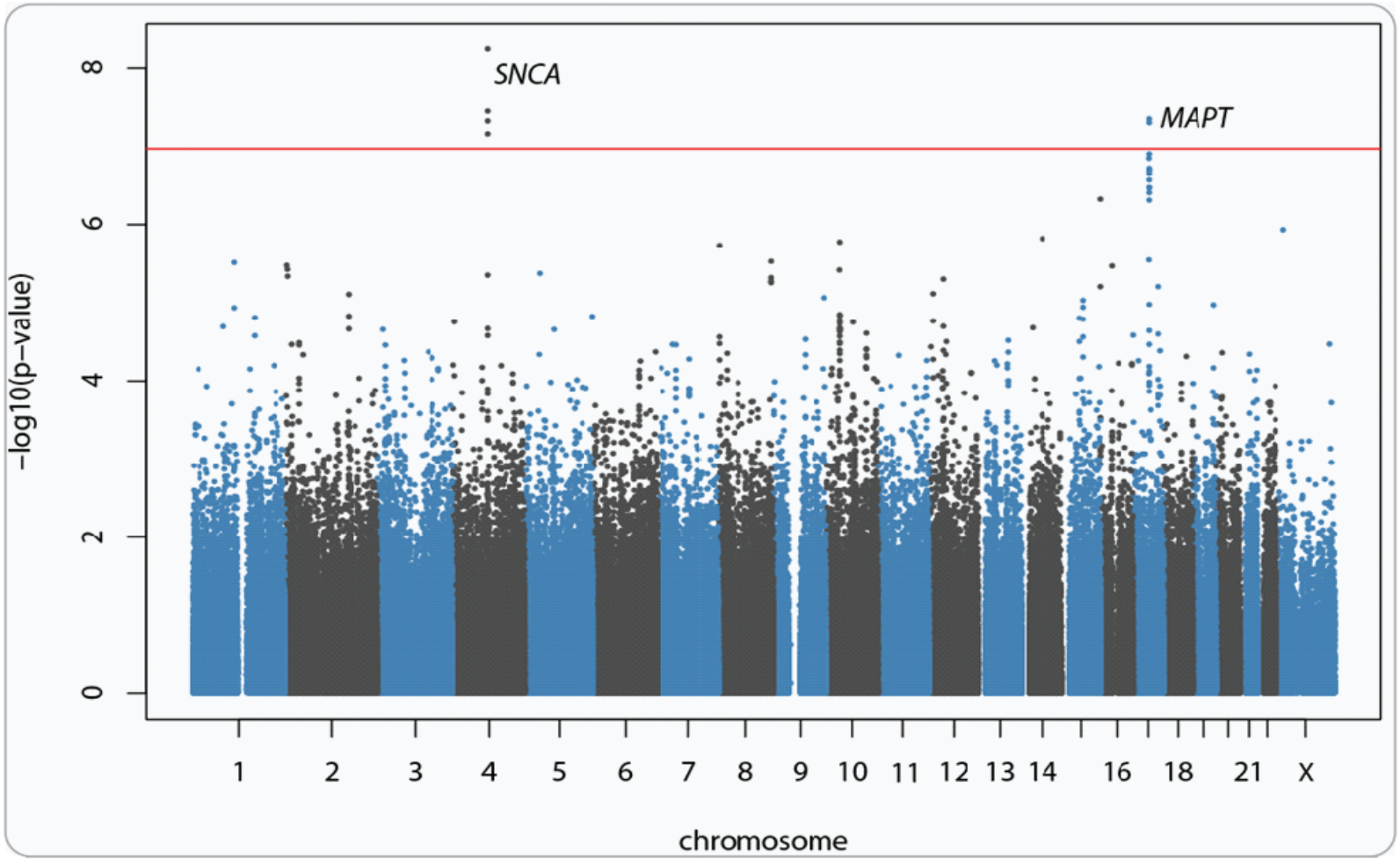


MDS analysis of the US-German samples and samples from four HapMap populations are shown. Our samples cluster nicely with the CEU HapMap population, but are clearly distinct from the YRI African and JPT/CHB Asian HapMap individuals.

Combined Analysis

The results of the combined screening and replication stage analysis are listed in table 12. We found significant association with disease at the *SNCA* locus and at a large LD block on chromosome 17q21 containing *MAPT* (*SNCA* locus: rs2736990, *p* value = 1.9 x 10^{-16}, OR = 1.2; *MAPT* locus: rs393152, *p* value = 1.0 x 10^{-15}, OR = 0.8).

Figure 26. Manhatten Plot of the Screening Stage Results: Significant Associations with *SNCA* and *MAPT*



This figure has been published elsewhere (see reference [272]). Shown are *p* values of trend tests in the cominbed US-German dataset (screening stage). Two loci, one on chromosome 4 at the *SNCA* locus and one on chromosome 17 at the *MAPT* locus, were significantly associated with risk for PD after correction for multiple testing. The Bonferroni threshold is indicated by a red line.

Table 10. Results of the Trend Tests in the Screening Stage

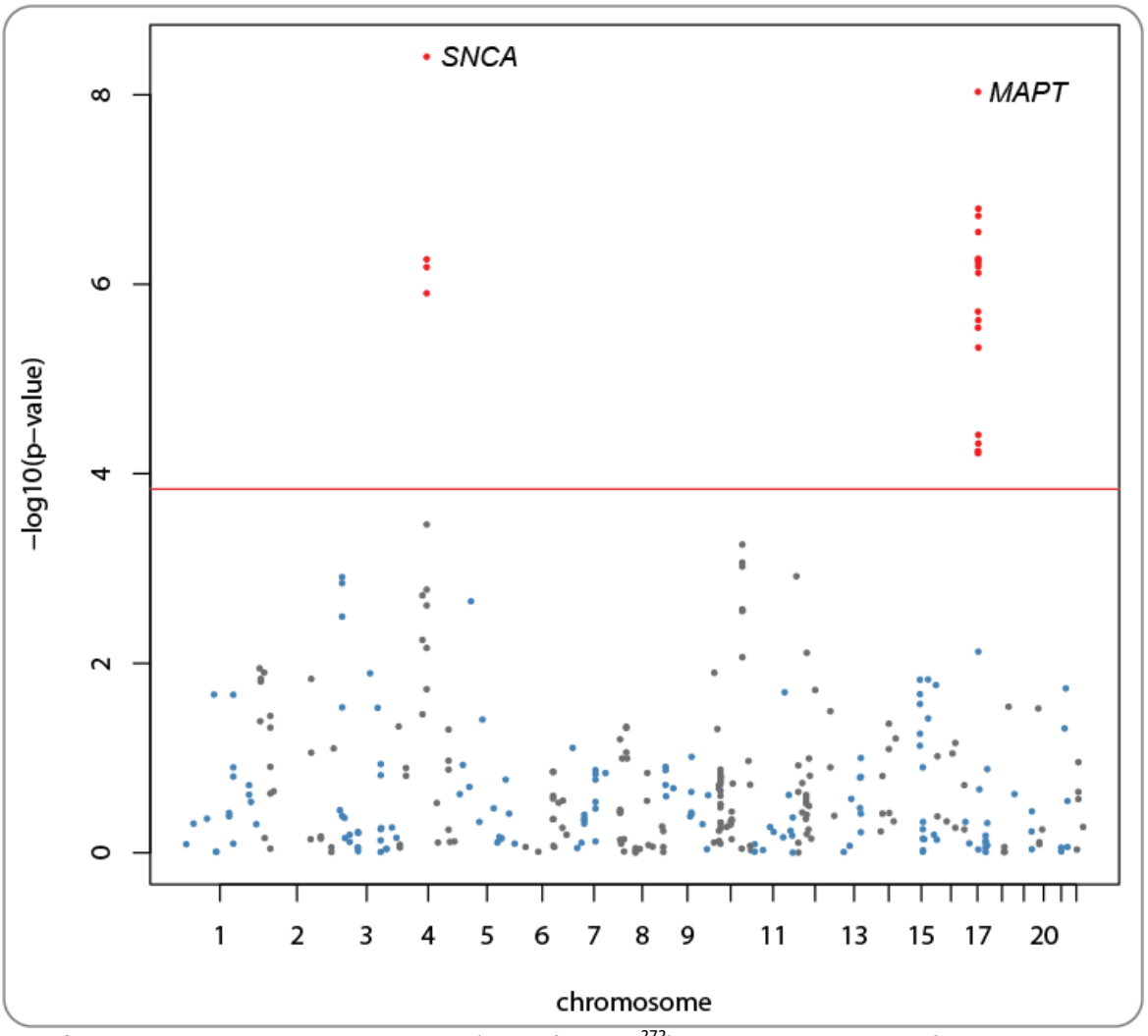
SNP	Chr.	Position[1]	Alleles (minor/major)	MAF		*p* value (trend test)	cOR	OR het. (95% CI)	OR hom. (95% CI)
				controls	cases				
rs356229	4	90,825,620	G/A	0.36	0.40	2.50E-04	1.16	1.20 (1.06-1.36)	1.34 (1.12-1.59)
rs11931074	4	90,858,538	T/G	0.07	0.10	4.78E-08	1.58	1.35 (1.15-1.58)	4.59 (2.44-8.62)
rs3857059	4	90,894,261	G/A	0.07	0.10	3.60E-08	1.58	1.35 (1.15-1.58)	4.59 (2.45-8.62)
rs2736990	4	90,897,564	C/T	0.46	0.52	5.69E-09	1.27	1.17 (1.02-1.34)	1.61 (1.37-1.88)
rs11012	17	40,869,224	A/G	0.18	0.15	2.85E-06	0.78	0.75 (0.66-0.85)	0.67 (0.46-0.98)
rs393152	17	41,074,926	G/A	0.22	0.18	1.42E-07	0.76	0.75 (0.66-0.85)	0.61 (0.44-0.83)
rs417968	17	41,084,159	C/T	0.26	0.22	8.00E-05	0.83	0.82 (0.73-0.93)	0.70 (0.54-0.90)
rs7215239	17	41,123,556	C/T	0.24	0.21	3.00E-05	0.82	0.79 (0.70-0.90)	0.71 (0.54-0.93)
rs1526123	17	41,139,123	G/A	0.47	0.43	8.00E-05	0.85	0.88 (0.77-1.00)	0.72 (0.61-0.84)
rs12373139	17	41,279,910	A/G	0.22	0.18	4.91E-07	0.78	0.75 (0.66-0.85)	0.64 (0.47-0.88)
rs17690703	17	41,281,077	T/C	0.25	0.22	7.00E-05	0.83	0.80 (0.71-0.90)	0.73 (0.56-0.95)
rs17563986	17	41,347,100	G/A	0.22	0.18	3.44E-07	0.77	0.75 (0.66-0.85)	0.63 (0.46-0.86)
rs1981997	17	41,412,603	A/G	0.22	0.18	2.02E-07	0.77	0.74 (0.65-0.84)	0.64 (0.47-0.88)
rs8070723	17	41,436,901	G/A	0.22	0.18	3.36E-07	0.77	0.74 (0.66-0.84)	0.64 (0.47-0.88)
rs2532274	17	41,602,941	C/T	0.22	0.18	2.22E-07	0.77	0.75 (0.66-0.84)	0.62 (0.45-0.85)
rs2532269	17	41,605,885	G/A	0.22	0.17	2.70E-07	0.77	0.74 (0.65-0.84)	0.64 (0.47-0.88)
rs2668692	17	41,648,797	T/C	0.22	0.17	3.94E-07	0.77	0.74 (0.65-0.84)	0.65 (0.47-0.89)
rs183211	17	42,143,493	A/G	0.24	0.20	1.00E-05	0.80	0.80 (0.71-0.90)	0.65 (0.49-0.86)
rs169201	17	42,145,386	G/A	0.20	0.16	1.25E-07	0.76	0.72 (0.64-0.82)	0.66 (0.47-0.92)
rs7224296	17	42,155,230	G/A	0.27	0.24	1.10E-04	0.84	0.82 (0.72-0.92)	0.72 (0.57-0.92)
rs199533	17	42,184,098	T/C	0.20	0.16	5.05E-08	0.76	0.72 (0.63-0.81)	0.64 (0.46-0.90)

This table has been published elsewhere (see reference [272]).

[1] Chromosome positions are shown according to NCBI genome build 36.3

MAF, minor allele frequency; CI, confidence interval; cOR, common odds ratio; OR, odds ratio; het., heterozygote; hom., homozygote

Figure 27. Replication of Strong Association Signals at the *SNCA* and *MAPT* Loci



This figure has been published elsewhere (see reference [272]). Shown are *p* values of trend tests in the replication dataset. SNPs within or in close proximity of *SNCA* and *MAPT* (red dots) clearly exceed the Bonferroni significance level (red horizontal line).

Table 11. Results of the Trend Tests in the Replication Stage

SNP	Chr.	Position[1]	Alleles (minor/major)	MAF		*p* value (trend test)	cOR	OR het. (95% CI)	OR hom. (95% CI)
				controls	cases				
rs356229	4	90,825,620	G/A	0.36	0.40	1.25E-06	1.17	1.20 (1.08-1.32)	1.37 (1.19-1.57)
rs11931074	4	90,858,538	T/G	0.08	0.10	6.59E-07	1.39	1.27 (1.12-1.44)	3.41 (1.80-6.45)
rs3857059	4	90,894,261	G/A	0.08	0.10	5.45E-07	1.41	1.26 (1.12-1.43)	3.80 (1.97-7.32)
rs2736990	4	90,897,564	C/T	0.46	0.51	3.97E-09	1.21	1.30 (1.17-1.45)	1.46 (1.28-1.66)
rs11012	17	40,869,224	A/G	0.17	0.15	6.00E-05	0.84	0.83 (0.75-0.92)	0.74 (0.57-0.97)
rs393152	17	41,074,926	G/A	0.21	0.18	9.31E-09	0.79	0.79 (0.72-0.87)	0.61 (0.48-0.78)
rs417968	17	41,084,159	C/T	0.25	0.22	1.95E-06	0.82	0.87 (0.79-0.95)	0.64 (0.52-0.78)
rs7215239	17	41,123,556	C/T	0.24	0.20	2.88E-06	0.82	0.87 (0.79-0.95)	0.63 (0.50-0.78)
rs1526123	17	41,139,123	G/A	0.47	0.44	6.00E-05	0.88	0.91 (0.82-1.01)	0.77 (0.67-0.87)
rs12373139	17	41,279,910	A/G	0.21	0.18	6.51E-07	0.81	0.84 (0.76-0.92)	0.61 (0.48-0.78)
rs17690703	17	41,281,077	T/C	0.24	0.21	1.60E-07	0.81	0.83 (0.75-0.91)	0.63 (0.51-0.79)
rs17563986	17	41,347,100	G/A	0.21	0.18	1.89E-07	0.80	0.83 (0.75-0.92)	0.59 (0.47-0.75)
rs1981997	17	41,412,603	A/G	0.21	0.18	5.36E-07	0.81	0.84 (0.76-0.92)	0.61 (0.48-0.77)
rs8070723	17	41,436,901	G/A	0.21	0.18	5.55E-07	0.81	0.84 (0.76-0.92)	0.62 (0.49-0.78)
rs2532274	17	41,602,941	C/T	0.22	0.19	4.00E-05	0.86	0.81 (0.74-0.90)	0.80 (0.65-0.99)
rs2532269	17	41,605,885	G/A	0.21	0.18	2.40E-06	0.81	0.86 (0.77-0.94)	0.60 (0.47-0.76)
rs2668692	17	41,648,797	T/C	0.21	0.18	2.81E-07	0.80	0.84 (0.76-0.92)	0.60 (0.47-0.76)
rs183211	17	42,143,493	A/G	0.23	0.20	4.66E-06	0.83	0.85 (0.77-0.94)	0.67 (0.54-0.83)
rs169201	17	42,145,386	G/A	0.20	0.17	5.90E-07	0.80	0.84 (0.76-0.93)	0.58 (0.45-0.75)
rs7224296	17	42,155,230	G/A	0.27	0.24	5.00E-05	0.86	0.87 (0.79-0.96)	0.72 (0.60-0.87)
rs199533	17	42,184,098	T/C	0.20	0.17	7.59E-07	0.80	0.84 (0.76-0.93)	0.59 (0.45-0.76)

This table has been published elsewhere (see reference [272]).

[1] Chromosome positions are shown according to NCBI genome build 36.3

MAF, minor allele frequency; CI, confidence interval; cOR, common odds ratio; OR, odds ratio; het., heterozygote; hom., homozygote

Table 12. Results of the Trend Tests in the Combined Analysis (Screening Stage + Replication Stage)

SNP	Chr.	Position[1]	Alleles (minor/major)	MAF		*p* value (trend test)	cOR	OR het. (95% CI)	OR hom. (95% CI)
				controls	cases				
rs356229	4	90,825,620	G/A	0.36	0.40	8.36E-10	1.17	1.19 (1.11-1.29)	1.36 (1.22-1.51)
rs11931074	4	90,858,538	T/G	0.07	0.10	4.38E-14	1.47	1.32 (1.20-1.45)	3.74 (2.39-5.85)
rs3857059	4	90,894,261	G/A	0.07	0.10	3.02E-14	1.48	1.32 (1.20-1.45)	3.94 (2.51-6.19)
rs2736990	4	90,897,564	C/T	0.46	0.51	1.89E-16	1.23	1.25 (1.14-1.36)	1.51 (1.37-1.67)
rs11012	17	40,869,224	A/G	0.18	0.15	3.04E-10	0.82	0.78 (0.72-0.85)	0.73 (0.59-0.91)
rs393152	17	41,074,926	G/A	0.22	0.18	1.02E-15	0.77	0.77 (0.71-0.83)	0.61 (0.50-0.74)
rs417968	17	41,084,159	C/T	0.26	0.22	1.96E-10	0.82	0.84 (0.78-0.91)	0.66 (0.56-0.77)
rs7215239	17	41,123,556	C/T	0.24	0.21	1.02E-10	0.82	0.83 (0.77-0.90)	0.65 (0.55-0.77)
rs1526123	17	41,139,123	G/A	0.47	0.43	1.33E-08	0.87	0.89 (0.82-0.96)	0.75 (0.68-0.83)
rs12373139	17	41,279,910	A/G	0.21	0.18	4.04E-13	0.79	0.79 (0.73-0.85)	0.63 (0.52-0.76)
rs17690703	17	41,281,077	T/C	0.25	0.21	3.61E-12	0.81	0.81 (0.75-0.87)	0.66 (0.55-0.78)
rs17563986	17	41,347,100	G/A	0.21	0.18	7.24E-14	0.79	0.79 (0.73-0.85)	0.61 (0.51-0.74)
rs1981997	17	41,412,603	A/G	0.21	0.18	1.69E-13	0.79	0.79 (0.73-0.85)	0.63 (0.52-0.76)
rs8070723	17	41,436,901	G/A	0.22	0.18	3.45E-13	0.79	0.79 (0.73-0.85)	0.63 (0.53-0.76)
rs2532274	17	41,602,941	C/T	0.22	0.19	2.46E-10	0.83	0.79 (0.73-0.85)	0.76 (0.64-0.90)
rs2532269	17	41,605,885	G/A	0.21	0.18	2.16E-12	0.79	0.80 (0.74-0.87)	0.62 (0.51-0.75)
rs2668692	17	41,648,797	T/C	0.21	0.18	3.78E-13	0.79	0.79 (0.74-0.86)	0.62 (0.51-0.75)
rs183211	17	42,143,493	A/G	0.23	0.20	9.68E-11	0.82	0.83 (0.77-0.89)	0.66 (0.56-0.78)
rs169201	17	42,145,386	G/A	0.20	0.16	2.03E-13	0.78	0.79 (0.73-0.85)	0.61 (0.50-0.75)
rs7224296	17	42,155,230	G/A	0.27	0.24	1.95E-08	0.85	0.85 (0.79-0.91)	0.73 (0.63-0.84)
rs199533	17	42,184,098	T/C	0.20	0.16	1.21E-13	0.78	0.78 (0.73-0.85)	0.61 (0.50-0.75)

This table has been published elsewhere (see reference [272]).

[1] Chromosome positions are shown according to NCBI genome build 36.3

MAF, minor allele frequency; CI, confidence interval; cOR, common odds ratio; OR, odds ratio; het., heterozygote; hom., homozygote

Table 13. Trend Test Results of SNPs in Known PD Genes

Gene	SNPs tested	Position[a]	Chr.	Minor allele	MAF cases	MAF controls	*p* value stage I	OR (95% CI)[b] stage I
SNCA	rs3857059	90,894,261	4q22.1	G	0.10	0.07	3.60E-08	1.5 (1.3 - 1.7)
PRKN	rs9458499	162,633,417	6q26	T	0.37	0.25	4.27E-05	1.2 (1.1 - 1.3)
PINK1	rs3121680	20,838,529	1q36.12	G	0.50	0.39	0.41	1.0 (0.9 - 1.1)
DJ1	rs161799	7,938,147	1p36.23	C	0.37	0.34	0.02	1.1 (1.0 - 1.2)
LRRK2	rs1491923	38,877,384	12q12	C	0.43	0.31	2.20E-04	1.2 (1.1 - 1.3)
ATP13A2	-	-	1p36	-	-	-	-	-
PLA2G6	rs11570734	36,848,568	22q13.1	C	0.60	0.43	0.02	1.1 (1.0 - 1.2)
FBXO7	rs5749450	31,216,477	22q12.3	G	0.42	0.40	0.04	1.1 (1.0 - 1.2)
Omi/HTRA2	-	-	2p13	-	-	-	-	-
GBA	rs9628662	153,472,965	1q22	G	0.28	0.28	0.97	1.0 (0.9 - 1.1)
MAPT	rs1981997	41,412,603	17q21.31	A	0.18	0.22	2.02E-07	0.8 (0.7 - 0.9)
FGF20	rs475738	16,735,019	8p22	A	0.29	0.31	0.01	0.9 (0.9 - 1.0)

Results for the trend tests in the screening stage data are shown. For each gene, the SNP with the lowest trend test *p* value at each respective locus is listed. SNPs at the *ATP13A2* and *Omi/HTRA2* loci did not meet quality control criteria. SNPs with low *p* values are highlighted in grey. Variants rs3857059 at the *SNCA* locus and rs1981997 at the *MAPT* locus surpassed the significance threshold for multiple testing.

[a] SNP positions are shown according to NCBI genome build 36.3.

[b] Odds ratios were calculated for the minor allele.

MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; Chr., chromosome.

Association Results at Known PD Loci

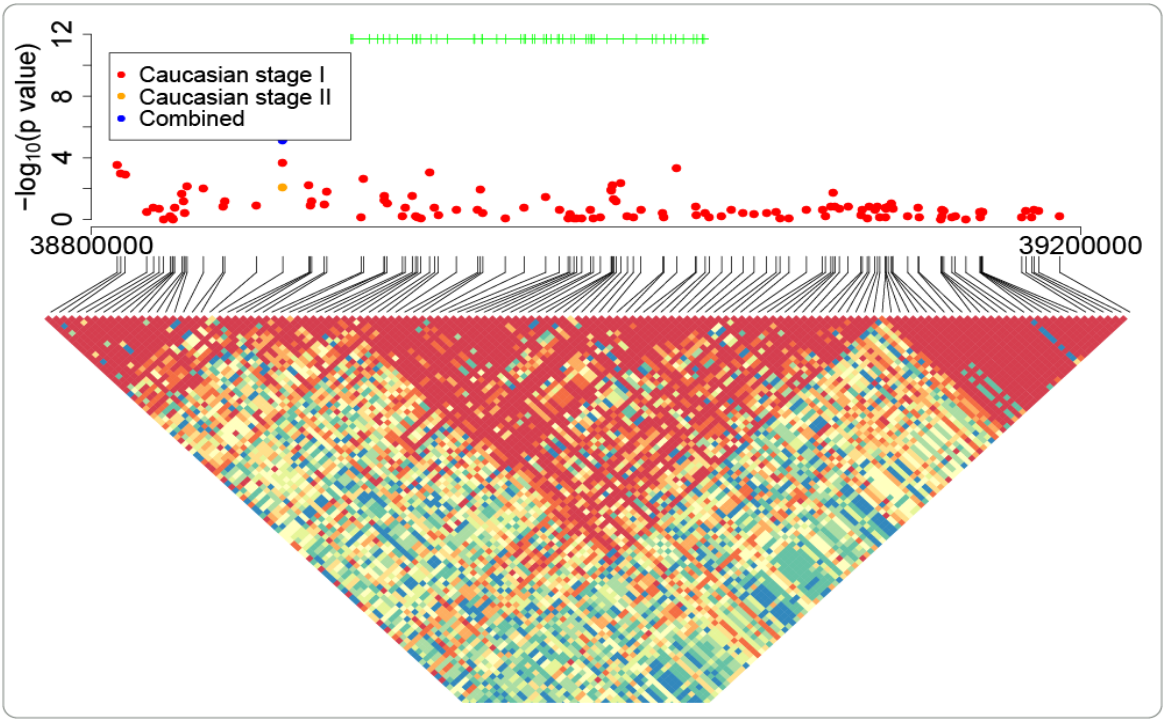
We studied the association results in the screening stage data at known PD loci (table 13). SNPs at four loci showed low *p* values (*SNCA* locus: rs3857059 *p* value = 3.6 x 10^{-8}; *MAPT* locus: rs1981997 *p* value = 2.0 x 10^{-7}; *LRRK2* locus: rs1491923 *p* value = 2.2 x 10^{-4}; *PRKN* locus: rs9458499 *p* value = 4.3 x 10^{-5}), but only variants at the *SNCA* and *MAPT* loci surpassed the conservative threshold for multiple testing.

SNPs at the *SNCA, MAPT, LRRK2* and *PRKN* loci were carried forward into the replication stage. As noted above, variants at the *SNCA* and *MAPT* loci were strongly associated with risk for developing PD. In the combined analysis, we noted that the SNP rs1491923 at the *LRRK2* locus had a low combined *p* value close to the significance threshold (*p* value = 7.02 x 10^{-6}, OR = 1.14). This SNP is located 0.17Mb upstream of *LRRK2* in a 5'-haplotype block distinct from a haplotype block containing *LRRK2* (figure 28). However, unlike the *LRRK2* locus the combined *p* value for the top SNP at the *PRKN* locus did not indicate disease association (rs9458499, uncorrected combined *p* value = 0.04). This observation could be explained by a different composition of the replication stage samples, by the exclusion of familial samples, by allelic heterogeneity, or - more importantly - by a false positive association in the screening stage. Taken together, our analyses indicate that common variants in known PD genes play a greater role in the pathogenesis of PD that previously appreciated.

We exchanged data with collaborators from Japan who performed a two-staged GWA study in 3,212 Japanese PD cases and 4,573 Japanese controls.[266] In this Asian GWA study, significant association with disease was identified for SNPs at the *SNCA* and *LRRK2* loci, but no association was detected at the *MAPT* locus on chromosome 17. Furthermore, a new protective locus on chromosome 1q32, designated as PARK16, was described in this GWA study. This locus spans five transcripts including *SNORA72, NUCKS1, RAB7L1, SLC41A1* and *PM20D1*. Hence, we reexamined the PARK16 locus in our study (figure 29) and found that SNPs at this locus were originally excluded from

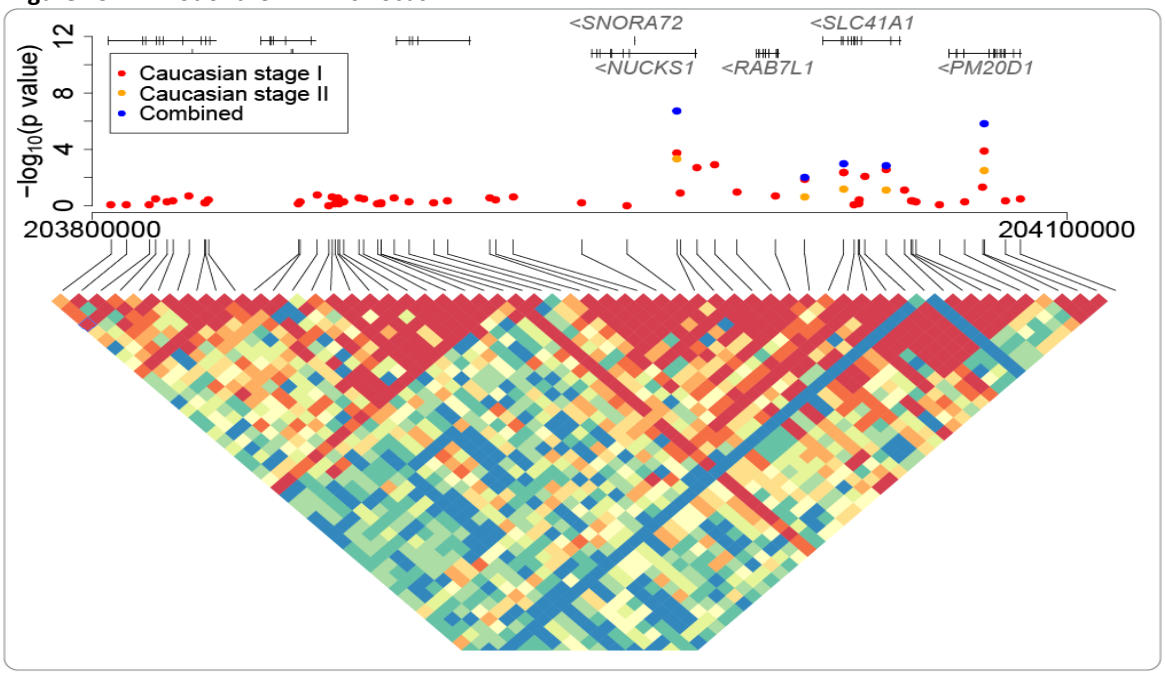
the analysis based on a low MAF in controls (MAF = 0.03). We therefore genotyped five SNPs at this locus (rs823128, rs947211, rs823156, rs708730, rs11240572) in a subset of our replication cohort comprising 2,909 cases and 3,500 controls using the matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry method (MassArray system, Sequenom, San Diego, CA, USA). In the combined analysis, we successfully replicated the protective effect of PARK16 (rs823128, *p* value = 1.86 x 10^{-7}, OR = 0.67) supporting the hypothesis that common variability at the PARK16 locus alters risk for disease in Caucasians as well as Asians.

Figure 28. LD Plot of the *LRRK2* Locus



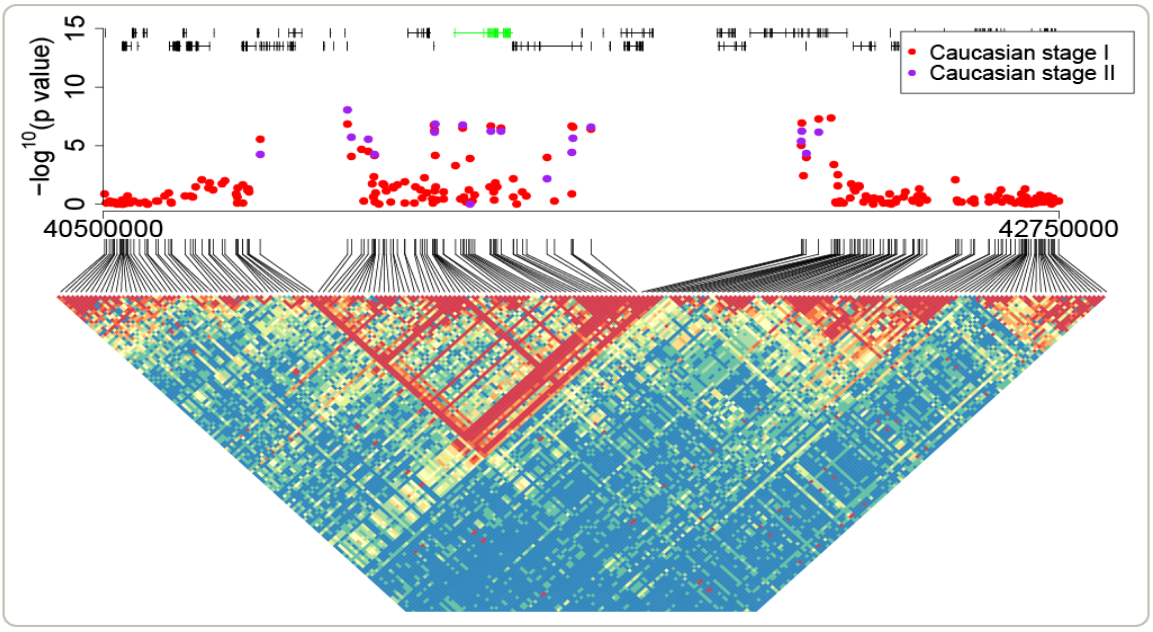
LD structure and association signals across the *LRRK2* locus on chromosome 12 are shown. We detected a low *p* value for a single SNP (rs1491923), located in the 5' region of *LRRK2*. Red dots indicate screening stage signals, yellow dots show replication stage results and blue dots indicate the results for the combined analysis.

Figure 29. LD Plot of the PARK16 Locus



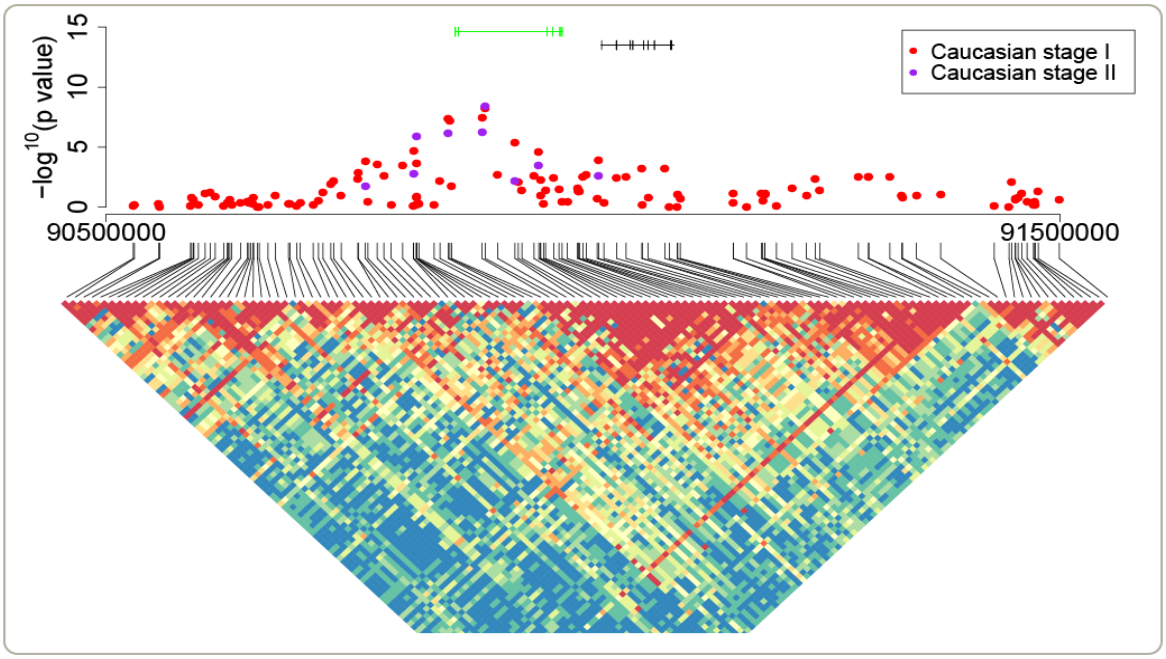
LD structure and association signals across the PARK16 locus are shown. This region contains a number of genes in a single haplotype block.

Figure 30. LD Plot of the *MAPT* Locus



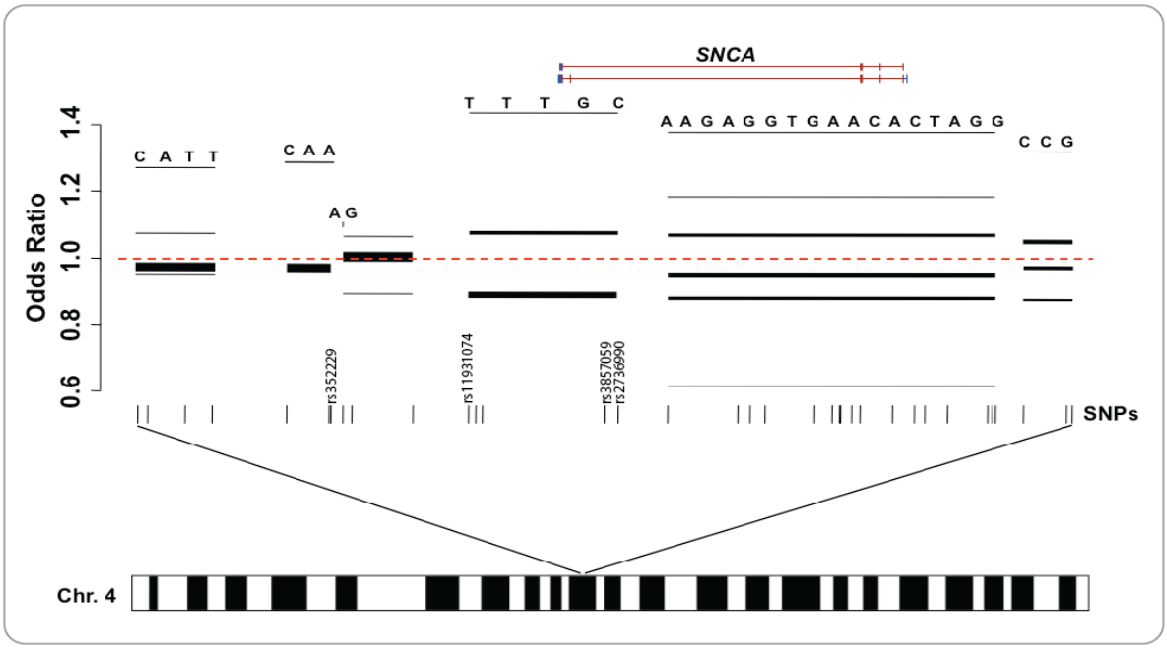
LD structure at the association signal on chromosome 17q21 illustrates a large haplotype block containing the *MAPT* gene (highlighted in green). Red dots indicate the association signals from the screening stage, purple dots indicate the results from the replication stage.

Figure 31. LD Plot of the *SNCA* Locus



LD structure and association signals at the *SNCA* locus are shown. *SNCA* is indicated in green. Red dots indicate the association signals from the screening stage, purple dots indicate the results from the replication stage.

Figure 32. Risk Haplotypes Identified at the *SNCA* Locus



This figure shows the risk haplotypes observed at the *SNCA* locus. The haplotype that is most significantly associated with risk for disease is located in the 3'region of the gene.

Linkage Disequilibrium Analysis

LD structure at the *MAPT* locus (figure 30) revealed a large single haplotype block, which significantly increased risk for developing PD. The risk alleles were observed on a block, stretching about 900kb, which is known as the H1 haplotype block. This LD block contains a number of genes including *MAPT*. It is unknown, however, whether specific H1 subhaplotypes are associated with disease. Detailed fine-mapping studies are required to address this issue.

LD structure at the *SNCA* locus (figure 31) revealed two haplotype blocks, a 3'-block that comprises the 3'-untranslated-region (UTR) and the 3'-region up to intron 4 of the *SNCA* gene and a 5'-block that extends to the promoter-enhancer region. We observed three of the four significantly associated SNPs in the 3'-block (figure 32) indicating that the pathogenic variant is located in this block. Further, many studies have been published detailing the role of a polymorphic microsatellite marker in the promoter region of *SNCA* known as the Rep1 marker. Specifically, expansion of the Rep1 polymorphism is associated with elevated disease risk, apparently mediated by altered gene expression.[85, 273] We genotyped the Rep1 marker in 1,774 US samples that were part of the screening stage; our results showed that Rep1 is in moderate LD with the 3'-risk haplotype block identified in this study (r^2 = 0.37 with rs3857059), hence the association between Rep1 and increased risk for PD could reflect a residual shared haplotype.

Epistasis Analysis and Population Attributable Risk

Our epistasis analysis did not detect a significant interaction between variants at the *SNCA* and *MAPT* loci, indicating that variants at each locus are independent risk factors for PD. The combined PAR% was estimated to be about 26%, with *SNCA* accounting for 8% and *MAPT* for 18%.

Comparison with Previous GWA Studies

To investigate SNPs that have been implicated with risk for developing PD in previous GWA studies, we reviewed the *p* values for the described risk SNPs in our screening stage data and summarised those findings in supplemental table S5. Although we observed low *p* values for some loci (*GAK/DGKQ*: rs11248060 *p* value = 2.3 x 10^{-4}; *SNCA/GPRIN3/MMRN1*: rs356229 *p* value = 2.5 x 10^{-4}; *GLT25D2*: rs1887279 *p* value = 0.02; *TMEM108*: rs1197313 *p* value = 0.04), the described risk variants did not reach significance in our study after correction for multiple testing.

3.2.4 Discussion

Here we report the results of a two-stage GWA study including a total of 5,165 Caucasian PD cases and 8,151 Caucasian controls. We detected two strong association signals in or near *SNCA* and *MAPT* that clearly surpassed the conservative Bonferroni correction level (combined *p* values and ORs of the top SNPs are listed in table 12). Furthermore, we exchanged results with Japanese collaborators who also performed a GWA study in PD; a new locus on chromosome 1q32.1, designated as PARK16, was implicated in this Asian GWA study and we confirmed the protective effect of variants at the newly described locus in our Caucasian sample (rs823128, *p* value = 1.86 x 10^{-7}, OR = 0.67).[266] In addition, we observed low *p* values in the known PD genes *LRRK2* and *PRKN* (table 13; *LRRK2*, rs1491923, *p* value = 2.2 x 10^{-4}; *PRKN*, rs9458499, *p* value = 4.3 x 10^{-5}) supporting the notion that common variability at these loci might modify susceptibility to develop PD.

The cPAR% for the *SNCA* and *MAPT* loci was about 26%, with *MAPT* attributing about 18% and *SNCA* contributing 8%. These numbers clearly underline the notion that variability at the *SNCA* and *MAPT* loci plays a crucial role in the pathogenesis of PD. However, these estimates have to be interpreted with caution, as our study design was a retrospective case-control study and as the levels of LD between the observed risk variants and the true causal variants are unknown.

SNCA and *MAPT* have been previously associated with PD. Copy number mutations and nonsynonymous sequence mutations in *SNCA* are known causes of autosomal dominant PD, and variants in the promoter region have been robustly associated with increased risk for PD in familial and sporadic cases.[76, 82, 84, 85] Furthermore, in a recent GWA study low *p* values for variants at the *SNCA* locus were observed, but none of the variants exceeded the correction threshold for multiple testing.[274] *MAPT* has been previously associated with a number of neurodegenerative diseases including progressive supranuclear palsy (PSP), Alzheimer's disease (AD), corticobasal degeneration (CBD), and frontotemporal dementia (FTD).[145] Reports on association of *MAPT* with PD have demonstrated consistent - albeit weak - effects of *MAPT* H1 haplotype clade.[275, 276] The H1 haplotype was also strongly associated with increased risk for PD in this study. A likely mechanism by which the H1 haplotype increases risk for PD is an increased gene expression of *MAPT* as demonstrated by Myers et al. in a large genotyping and expression study in normal control brain samples.[277] It is unclear, however, whether a specific subhaplotype within the H1 clade is driving the association with PD. Detailed fine-mapping studies are required to answer this question.

Previous studies testing a possible interaction between *SNCA* and *MAPT* have provided conflicting results.[278-283] *In vitro* studies have demonstrated an increased propensity for fibrillisation of α-synuclein in the presence of tau protein, suggesting a pathophysiological link between α-synuclein and tau.[280] This is of particular interest because variable co-occurrence of α-synuclein and tau inclusion pathology has been described in PD patients.[278, 279, 282] Goris et al. have performed a *MAPT-SNCA* genotype interaction analysis and found that risk conferred by the *SNCA* variant rs356219 was modified by stratification on *MAPT* genotypes. Individuals who were homozygous for the H1 *MAPT* haplotype and the G allele of rs356219 had markedly increased risk compared to the additive risks of individuals who were homozygous for just one locus.[281] However, this observation is not supported by another *MAPT-SNCA* epistasis study in which homozygosity for the H1 haplotype and for a *SNCA* risk variant

showed an increased risk compared to persons with neither, but the combined effect of the two genotypes was the same as for either of the genotypes alone.[283] In our epistasis analysis, we also have not observed evidence for an interaction of *MAPT* and *SNCA* risk variants as the combined effects of *SNCA-MAPT* risk genotypes were not significantly different from the additive risks conferred to by just one risk genotype. We therefore hypothesise, that the frequent co-occurrence of tau and α-synuclein protein aggregates is probable due to an aggregation-prone cellular state rather than a pathological interaction of α-synuclein and tau.

How do the results of this study compare to previously published GWA studies in PD? To date, three GWA studies in PD have been published.[150, 264, 265] The first study was published by Maraganore et al. and consisted of a two-tiered study in which about 200,000 SNPs were tested in 443 sibling pairs discordant for PD and followed by genotyping of the top associated SNPs in 332 additional case-control pairs. Thirteen SNPs were reported as associated with PD.[265] In our dataset, only one of those 13 SNPs was included in the screening stage analysis and this SNP was not significantly associated with disease (supplemental table S5; *p* value = 0.52, OR = 1.0, 95% CI = 0.8 - 1.1). Other studies also failed to confirm these findings in independent cohorts.[284-286] The second GWA study was published by us.[264] In this study, we genotyped about 400,000 SNPs in 267 PD patients and 270 neurologically normal controls. None of the tested SNPs reached significance after correction for multiple testing and we published our study merely as a preliminary analysis and public release of genome-wide data in PD rather than a gene discovery claim. Nonetheless, we reexamined the top SNPs of this study in our current GWA dataset, but we failed to detect a significant association (supplemental table S5). The third study was published in 2008 by Pankratz et al.; it included 857 familial PD cases and 867 controls that had been genotyped for about 370,000 SNPs and was augmented by the publicly available dataset from our second GWA study.[150] None of the tested SNPs met genome-wide significance after Bonferroni correction. The strongest association signals were observed for SNPs in the *GAK/DGKQ*

region on chromosome 4, for *SNCA* and *MAPT*. In our current GWA study, we failed to replicate the association of SNPs at the *GAK/DGKQ* region (rs11248060, *p* value = 2.28 x 10^{-4}, OR = 1.2, 95% CI = 1.1 – 1.4). This negative result could possibly be explained by a different composition of the investigated PD cohorts, containing predominantly sporadic cases in our study in contrast to the study reported by Pankratz et al., or by possible false positive results in the Pankratz dataset. Further replication studies are required before it is clear whether the *GAK/DGKQ* region can be truly established as a new risk locus for PD.

In summary, we report the results of the largest GWA study in PD to date. We demonstrated that common variability at the *SNCA* and *MAPT* loci plays a crucial role in the pathogenesis of PD. Furthermore, we replicated the protective effect of the newly discovered PARK16 locus on chromosome 1q32. This study provides valuable insights into the pathobiology of PD and it is hoped that fine-mapping of the identified loci as well as additional replication stages will improve our understanding of the pathophysiologic mechanisms leading to neurodegeneration.

3.3 Candidate SNP Association Study in Multiple System Atrophy

STATEMENT OF CONTRIBUTIONS TO THIS RESEARCH:

I designed the study, performed laboratory based experiments in the MSA samples and the statistical analysis. I wrote the original manuscript that was published in the journal Annals of Neurology.

3.3.1 Background

An increasing body of evidence indicates similar clinical and biochemical characteristics between PD and MSA. Both diseases are progressive neurodegenerative disorders characterised by abnormal deposition of fibrillar α-synuclein: in PD, the abnormal aggregates are found in neurons known as Lewy-bodies (LBs) and in MSA, the deposition occurs mainly in the form of glial cytoplasmic inclusions (GCIs).[80, 287, 288] On clinical grounds, MSA and PD are often indistinguishable, particularly in early disease stages, and as a consequence the misdiagnosis rate for MSA patients is high and varies between 24% and 75% depending of the disease stage, the experience of the examiners and the diagnostic criteria applied.[184, 185]

Since the mid 1990s, genetic studies in PD have been an ongoing success story. Several disease loci have been implicated in Mendelian forms of PD, but also an increasing number of susceptibility genes are known (see chapter 1.3.2 for a detailed discussion). The overlapping clinical and biochemical features in PD and MSA therefore raise the question whether genetic risk factors are shared between these two synucleinopathies.

To test the hypothesis that genetic risk factors associated with altered risk for PD also play a role in the pathobiology of MSA, we performed a candidate SNP association study in MSA, attempting to test the 384 strongest associated SNPs from the GWA study in PD (as described in chapter 3.2) as candidate risk factors for MSA. A total of 413 MSA cases and 3,974 controls have been genotyped using custom-made genotyping assays. After

stringent quality control procedures were applied, association tests for 339 SNPs in 400 MSA cases and 3,891 controls were performed. Replication of the ten strongest associated SNPs of this initial screening stage was then performed in an independent cohort of 108 MSA cases and 537 normal controls. Our analysis demonstrated that, similar to PD, genetic variants at the *SNCA* locus coding for α-synuclein were strongly associated with increased risk for developing MSA (combined *p* value under a recessive model = 5.5×10^{-12}, OR for homozygous risk allele carriers = 6.2, 95% CI = 3.4 - 11.2).

3.3.2 Materials and Methods

Study Subjects

The study consisted of two stages: a screening stage and a replication stage. Collection sites and characteristics of all investigated cohorts are summarised in tables 14 and 15.

Table 14. Characteristics of Studied Cohorts

Characteristics	MSA patients	Controls
Screening stage		
Number of individuals	413	3,974
Female / male / no sex information	156 / 155 / 102	1,336 / 2,629 / 9
MSA subtype†		
definite	99	-
probable	271	-
possible	43	-
Clinical subtype‡		
MSA-Parkinsonism	188	-
MSA-Cerebellar	105	-
N/A	120	-
Replication stage		
Number of individuals	108	537
Female / male	N/A	342 / 195
MSA subtype†		
definite	0	-
probable	108	-

This table has been published elsewhere (see reference [289]).
†subtypes defined by Gilman et al.[172]
‡individual-level data were available only for 136 MSA-Parkinsonism cases and 75 MSA-Cerebellar cases; only summary-level data were available for 52 MSA-P and 30 MSA-C cases
N/A, data not available

In the screening stage a total of 413 Caucasian MSA cases and 3,974 Caucasian normal controls were studied. These samples included 99 definite MSA cases and 314 possible or probable cases and were collected at collaborating centers in the United States, Germany, Italy, Austria and the United Kingdom. For the replication stage, an independent cohort of 108 possible Caucasian MSA cases and 537 Caucasian normal controls from the United Kingdom was studied. All cases were diagnosed by movement disorders specialists and the diagnosis was based on consensus criteria established by Gilman et al.[172] Institutional review board approval was obtained from each respective collaborating center. For all participants written informed consent was collected.

Table 15. Study Cohorts

Collection site	**Number of samples**
Screening stage	
Cases	
Queen Square Brain Bank, London, UK	87*
Brain and Tissue Bank for Developmental Disorders, University of Maryland, Baltimore, USA	2*
Human Brain and Spinal Fluid Resource Center, Los Angeles, CA, USA	2*
New York Brain Bank at Columbia University, NY, USA	8*
Coriell Cell repository, Camden, NJ, USA	21
University of Florida, Gainesville, FL, USA	6
National Institute on Neurological Disorders and Stroke, Bethesda, MD, USA	4
Hertie Institute for Clinical Brain Research, Tuebingen, Germany	52
Philipps-University of Marburg, Marburg, Germany	87
University Hospital of Bonn Medical Center, Bonn, Germany	49
Istituti Clinici di Perfezionamento, Milano, Italy	43
Medical University Innsbruck, Innsbruck, Austria	29
University Federico II, Napoli, Italy	23
Controls	
Institute of Neurology, Queen Square, London, UK	7
National Institute of Environmental Health Sciences, NC, USA	1,700
KORA Initiative, Neuherberg, Germany	1,323
Hertie Institute for Clinical Brain Research, Tuebingen, Germany	944
Replication stage	
Cases	
Institute of Neurology, Queen Square, London, UK	108
Controls	
Institute of Neurology, Queen Square, London, UK	537

This table has been published elsewhere (see reference [289]).
*pathology-proven samples

SNP Genotyping in the Screening Stage

Three hundred and eighty-four SNPs were genotyped in each subject. These SNPs were selected from our GWA study in PD (see chapter 3.2) and had the lowest *p* values based on the trend test model. Genotyping was performed on custom-made Illumina GoldenGate® Genotyping Assays on a BeadExpress platform as per the manufacturer's protocol (see chapter 3.2.2 for a detailed description of the genotyping workflow).

Quality Control in the Screening Stage

For quality control, samples with a call rate <90% were excluded (n = 13 cases and 83 controls). Of the genotyped 384 SNPs, 45 SNPs were removed from further analyses after applying standard quality control filters in PLINK software.[258] Specifically, three SNPs were excluded based on a MAF <0.01 in controls, 29 SNPs revealed significant departure from HWE in controls (*p* value <0.001), 26 SNPs had a missingness rate >5%, and two SNPs showed inaccurate genotype clustering in the Cartesian plots of the Genotyping module implemented in BeadStudio software (v.3.1.0). A total of 15 SNPs failed more than one quality control criterion.

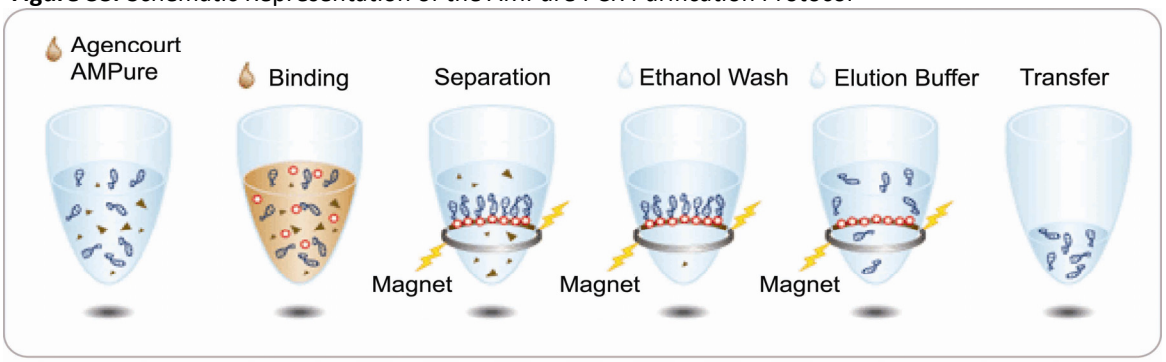
SNP Genotyping and Quality Control in the Replication Stage

Ten SNPs with the lowest *p* value (p_{min}) in the screening stage were selected for replication. Genotyping in MSA cases was performed by PCR amplification of a 150-400bp fragment surrounding each respective SNP followed by direct sequencing on an ABI3730xl DNA Analyzer (Applied Biosystems). The PCR reaction mix consisted of 15ng of genomic DNA, 10nM forward primer, 10nM reverse primer and 12µl of FastStart PCR Master (Roche, IN, USA). The PCR cycling conditions and primers are listed in supplemental tables S2 and S6.

To remove unincorporated dNTPs, primers, salts and DNA polymerase, we performed an automated PCR cleanup reaction using the AMPure purification kit (Agencourt Bioscience Corporation, MA, USA) on a Biomek FX robot (Beckman Coulter, CA, USA) as

per the manufacturer's protocol (figure 33). Briefly, 27µl of AMPure mix, containing paramagnetic particles, was added to 15µl of PCR product. After thorough mixing, the PCR-AMPure mix was incubated for 5 minutes for binding of the PCR amplicons to the paramagnetic beads. Next, the reaction plate was placed on the magnetic Agencourt SPRIPlate in order to separate the beads from the solution. The supernatant solution containing unbound primers and dNPTs was aspirated and discarded. Following two washing steps with 70% ethanol, the cleaned PCR amplicons were resuspended in 30µl of distilled and deionised molecular grade water and transferred into a separate 96-well PCR plate for further processing.

Figure 33. Schematic Representation of the AMPure PCR Purification Protocol



Source: http://www.agencourt.com/products/spri_reagents/ampure/

PCR reaction purification was followed by bi-directional direct dye-terminator sequencing using the BigDye chemistry (v.3.1, Applied Biosystems, CA, USA). The sequencing protocol was as follows: 2µl of cleaned PCR product, 0.5µl of BigDye (v.3.1), 1µl of 10nM of primer, 2µl 5X Sequencing Buffer (Applied Biosystems), and 5µl distilled and deionised molecular grade water (Mediatech. Inc., VA, USA). The sequencing cycling conditions are listed in supplemental table S3.

Excess fluorescent dye-terminator and contaminants were removed with CleanSEQ (Agencourt Bioscience Corporation, MA, USA) following the manufacturer's protocol. Briefly, 10µl of Agencourt CleanSEQ and 42µl of 85% ethanol were added to 10µl sequencing reaction. After thorough mixing to facilitate binding of sequencing

amplicons to the magnetic beads, the reaction plate was placed on the Agencourt SPRIPlate for 3 minutes in order to separate the bead-bound sequencing products from contaminants. The supernatant solution was then aspirated and discarded. After two washing steps with 85% ethanol, the cleaned sequencing products were eluted from the magnetic particles in 40µl of distilled and deionised molecular grade water and transferred into a clean 96-well semi-skirted reaction plate. Purified sequences were analysed on a 3730xl DNA Analyzer (Applied Biosystems, CA, USA) and electropherograms were visualised in Sequencher software (v.4.1.4; Gene Codes Corporation, MI, USA).

For the control samples, genotype data were extracted from 537 normal, Caucasian controls from the United Kingdom that had been previously genotyped on 610Y SNP chips (Illumina) as part of the Welcome Trust Case Control Consortium effort.

Statistical Analysis

Data were analysed in PLINK software (v1.04).[258] For the remaining 339 SNPs in the filtered screening stage data (n = 400 cases and 3,891 controls), association tests under different five models (allelic, genotypic, dominant, recessive, trend model) were calculated, and the lowest *p* value (p_{min}) was computed for each SNP. Based on Bonferroni correction for multiple testing a genome-wide significance threshold of 2.6×10^{-5} (two-sided α of 0.05 divided by [384 SNPs multiplied by 5 models]) was used.

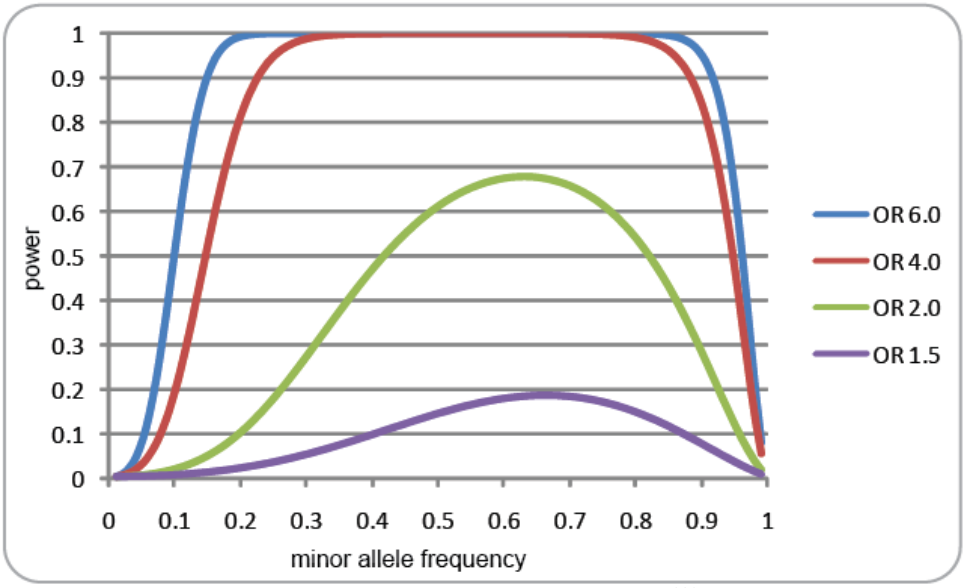
For the statistical analysis in the replication stage data, a recessive test model was applied after analysis of the screening stage data revealed best results under this model. To adjust for multiple testing, Bonferroni correction with a genome-wide significance threshold of <0.005 (two-sided α of 0.05 divided by 10 SNPs tested) was applied.

Quality control measures revealed significant departure from HWE in the controls (*p* value <0.01) for the SNP rs2856336, and the SNP rs10515822 was not polymorphic in the replication cohort. These two SNPs were excluded from further analyses.

Power Simulations

Power calculations were performed in Quanto© software (v1.2.3). We simulated four different effect sizes (OR = 6.0, OR = 4.0, OR = 2.0, OR = 1.5) to estimate the power for detecting significant associations under the recessive test model in the replication dataset. We plotted the power to detect association with different odds ratios against different minor allele frequencies, considering a two-sided α of 0.005 (genome-wide significance level after Bonferroni correction) (figure 34).

Figure 34. Power Simulations for the Replication Stage Data



This figure has been published elsewhere (see reference [289]). Power simulation to replicate associated loci under a recessive model for four different effect sizes in a cohort of 108 cases and 537 controls. A significance value of 0.005 and a population risk of 0.00001 were assumed. This simulation shows that we had 80% power to detect risk variants with an OR of 4.0 and a MAF of 0.2.

Simulations for sample size and power estimates to detect a modest SNCA gene expression alteration were performed in JMP software (version 7.0). We estimated that for the detection of a - for example - 10% increased gene expression a samples size of 82

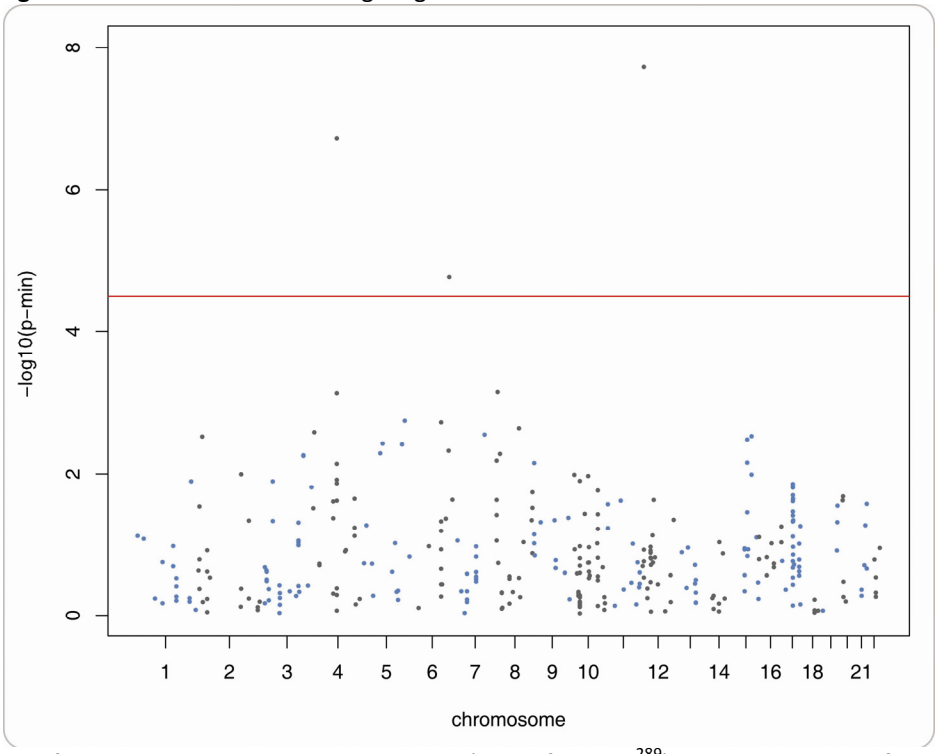
cases and 82 controls would be required to identify significantly altered levels (*p* value <0.05, error standard deviation 1.5, 90% power) (figure 37).

3.3.3 Results

Screening Stage

Of the 384 SNPs genotyped in 413 MSA cases and 3,974 controls, 339 SNPs passed quality control filtering in a final dataset of 400 cases and 3,891 controls. Association tests under allelic, genotypic, dominant, recessive and trend models were performed and the data were ranked according to the *p* values. Association signals with the lowest *p* values were observed under the recessive test model. The results of the screening stage are plotted in figure 35.

Figure 35. Results of the Screening Stage



This figure has been published elsewhere (see reference [289]). Manhatten plot of the association results for 339 SNPs in 400 MSA cases and 3,891 controls are shown. The –log10 of the p_{min}-value for each SNP is plotted against the genomic position in each chromosome. The Bonferroni correction threshold for multiple testing is indicated by a red horizontal line. Three SNPs (rs11931074 on chromosome 4, rs9480154 on chromosome 6, and rs2856336 on chromosome 12) exceeded the significance level.

Replication Stage

Ten SNPs that showed the strongest association in the screening stage were selected for replication in 108 additional MSA samples and 537 normal controls (table 16). Association tests under the recessive model revealed two SNPs (rs2856336, rs10515822) that did not meet our quality control criteria and these SNPs were excluded. Of the remaining eight SNPs, two SNPs showed strong association with increased risk for developing MSA (*p* value for rs11931074 = 1.6 x 10^{-4}; *p* value for rs3857059 = 1.3 x 10^{-6}; table 16). Both SNPs are located at the *SNCA* locus, coding for α-synuclein, and are in complete LD (r^2 = 1).

Combined Analysis

In the combined dataset, including the screening and the replication stage data, significant associations with increased risk for developing MSA were detected for the SNPs rs11931074 (*p* value = 5.5 x 10^{-12}, OR for homozygous risk allele carriers = 6.2, 95% CI: 3.4 - 11.2) and rs3857059 (*p* value = 2.1 x 10^{-10}, OR for homozygous risk allele carriers = 5.9, 95% CI: 3.2 - 10.9) (table 16). These two SNPs are located in intron 4 of *SNCA* (rs3857059) and in the 3'-region of *SNCA* (rs11931074) (figure 36).

Analysis of Definite MSA Cases

To exclude the possibility that clinically misdiagnosed PD patients were falsely driving the association with *SNCA*, we performed a subanalysis testing rs3857059 and rs11931074 in pathology-proven MSA cases and normal controls (n = 92 cases and 3,891 controls after quality control filtering). For both *SNCA* SNPs, we confirmed a strong association with increased risk for developing MSA (recessive model *p* value for rs3857059 = 4.9 x 10^{-6}; *p* value for rs11931074 = 1.4 x 10^{-11}; table 17).

Table 17. *SNCA* Risk Variants in Pathology-Proven MSA Cases

SNP ID	Minor allele	Test model	Genotypes in definite cases	Genotypes in controls	*p* value
rs11931074	T	recessive	6/20/66	22/564/3,303	1.41E-11
rs3857059	G	recessive	4/20/68	21/563/3,305	4.88E-06

This table has been published elsewhere (please see reference [289]).

Analysis of Clinical MSA Subtypes

We also tested the $SNCA$ variants rs3857059 and rs11931074 for association in the two clinical subgroups MSA-P and MSA-C. A total of 136 patients had MSA-P and 75 patients were classified as MSA-C cases. We did not detect a significant association between the two risk variants and increased risk for MSA (MSA-P: rs3857059 *p* value = 0.18 and rs11931074 *p* value = 0.19; MSA-C: rs3857059 *p* value = 0.07 and rs11931074 *p* value = 0.08; recessive model using Fisher's exact test). The lack of association in these clinical subgroups most likely reflects the low power of these cohorts, but also indicates that the association is not only present in one MSA subtype.

Figure 36. Significant Association Signals at the *SNCA* Locus in PD and in MSA

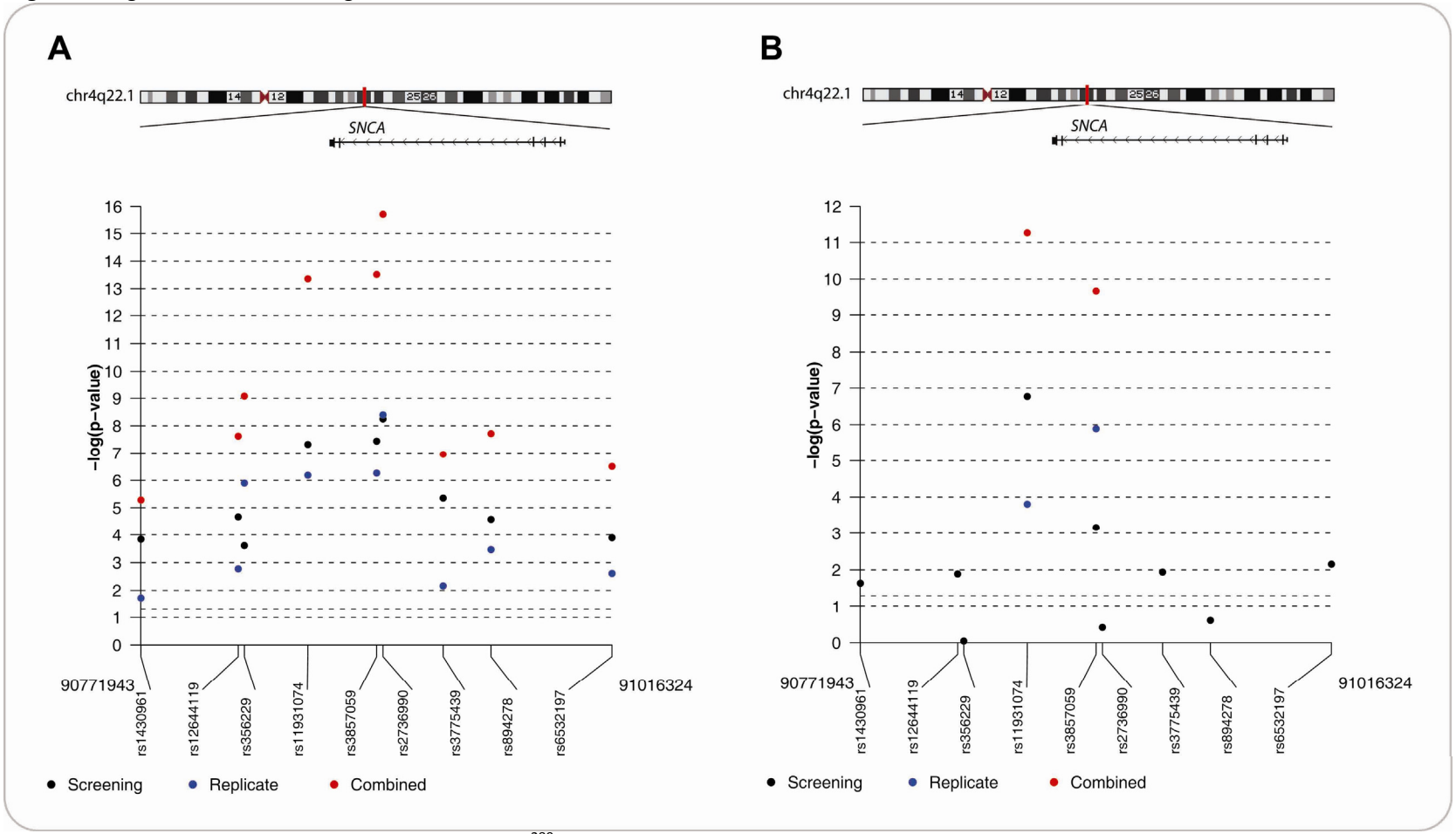


Figure B has been published elsewhere (see reference [289]). Shown are the association signals at the *SNCA* locus in PD and in MSA. In PD (A), the trend test results from our GWA study are depicted; *p* values from the screening stage (n =1,713 cases and 3,978 controls; black dots), the replication stage (n =3,452 cases and 4,173 controls; blue dots) and the combined analysis (red dots) are shown. In MSA (B), the results under the recessive model are shown for the screening stage (n = 413 cases and 3,974 controls; black dots), the replication stage (n =108 cases and 537 controls; blue dots) and the combined analysis (red dots).

Table 16. Ten Most Significant SNPs

SNP ID	Chr.	Gene	Risk allele	Screening stage		Replication stage		Combined	
				p_{min} (test model)	OR (95% CI) [RR vs (RP + PP)]	$p_{recessive}$	OR (95% CI) [RR vs (RP + PP)]	$p_{recessive}$	OR (95% CI) [RR vs (RP + PP)]
rs11931074	4	downstream of *SNCA*	T	1.7E-07 (rec.)†	5.4 (2.7 - 11.1)	1.6E-04‡	6.6 (2.15 - 19.93)	5.5E-12	6.2 (3.4 - 11.2)
rs3857059	4	*SNCA*	G	6.9E-04 (rec.)	3.8 (1.7 - 8.5)	1.3E-06‡	9.8 (3.20 - 29.78)	2.1E-10	5.9 (3.2 - 10.9)
rs9480154	6	downstream of *PPP1R14C*	A	1.6E-05 (rec.)†	5.0 (2.2 - 11.2)	0.99	1.0 (0.12 - 8.81)	1.3E-04	3.9 (1.8 - 8.2)
rs2794256	6	*LOC728727*	T	1.7E-03 (rec.)	1.7 (1.2 - 2.5)	0.17	1.6 (0.81 - 3.19)	4.0E-04	1.7 (1.3 - 2.4)
rs2042079	2	intergenic	A	2.7E-03 (rec.)	1.7 (1.2 - 2.5)	0.21	1.6 (0.77 - 3.18)	8.0E-04	1.7 (1.3 - 2.4)
rs13139027	4	upstream of *MSX1*	A	2.5E-03 (rec.)	3.9 (1.5 - 10.1)	0.53	1.5 (0.41 - 5.63)	1.8E-03	3.2 (1.5 - 6.9)
rs2515501	8	*MCPH1*	T	6.5E-04 (rec.)	2.4 (1.4 - 4.1)	0.45	0.6 (0.13 - 2.52)	7.0E-03	1.9 (1.2 - 3.2)
rs2896159	7	intergenic	T	3.0E-03 (rec.)	0.7 (0.5 - 1.1)	0.38	1.3 (0.73 - 2.26)	0.43	1.3 (1.1 - 1.6)
rs2856336	12	*ETV6*	C	1.6E-08 (rec.)†	4.6 (2.6 - 8.3)	0.12	-*	2.4E-05	3.1 (1.8 - 5.5)
rs10515822	5	downstream of *GABPB2*	A	1.7E-03 (rec.)	3.2 (1.5 - 6.7)	-**	-	-	-

This table has been published elsewhere (please see reference [289]). Significant SNPs at the *SNCA* locus are highlighted in grey.

*unable to calculate OR due to low allele frequency in cases

**not polymorphic in the replication stage

†exceeded Bonferroni significance threshold for multiple testing in the screening stage (i.e. α'= 0.05/[384*5]=2.6E-05)

‡exceeded Bonferroni significance threshold in the replication stage (i.e. α'=0.05/10=0.005)

OR, odds ratio; CI, confidence interval; HWE, Hardy-Weinberg equilibrium; rec., recessive; R, risk allele; P, protective allele

Comparison of SNCA Risk Genotypes in MSA, PD and Controls

We compared the allele frequencies of SNPs at the *SNCA* locus in MSA, PD and controls. For MSA we only used the screening stage genotyping results from definite cases (n = 92 after quality control filtering). For PD and control samples, the frequencies of *SNCA* variants genotyped in the replication stage of the PD GWA study have been included in the analysis (see chapter 3.2). Chi-squared tests were performed to detect significant differences between these three groups (table 18). We observed a significant difference in the allele frequency of rs2736990 which was found to be associated with PD but not with MSA. Furthermore, SNP rs11931074 was significantly associated with PD and with MSA, but homozygous risk allele carriers were significantly overrepresented in MSA.

Table 18. Comparison of *SNCA* Risk Genotypes in MSA, PD and Controls

SNP ID	Position[a]	Minor allele	MAF			*p* value[b]		
			MSA	PD	con.	MSA vs con.	PD vs con.	MSA vs PD
rs1430961	90,771,943	C	0.11	0.09	0.08	0.17	0.02	0.43
rs12644119	90,822,442	A	0.16	0.12	0.11	0.02	0.002	0.12
rs356229	90,825,620	G	0.36	0.40	0.36	0.89	1.12E-06	0.24
rs11931074	90,858,538	T	0.17	0.10	0.08	8.91E-07	7.87E-07	7.00E-04
rs3857059	90,894,261	G	0.15	0.10	0.08	1.00E-04	6.46E-07	0.02
rs2736990	90,897,564	T	0.47	0.51	0.46	0.74	2.90E-09	0.34
rs3775439	90,928,764	A	0.22	0.15	0.13	5.00E-04	0.01	0.01
rs894278	90,953,558	G	0.10	0.07	0.06	0.01	4.00E-04	0.14
rs6532197	91,016,324	G	0.14	0.09	0.08	0.001	0.003	0.01

This table has been published elsewhere (see reference [272]).

[a] SNP positions are shown according to NCBI genome build 36.3.

[b] *p* values for the allelic tests are shown.

MAF, minor allele frequency; con., controls.

3.3.4 Discussion

Here we demonstrate that variants at the *SNCA* locus are strongly associated with increased risk for developing MSA. This finding is important in that it proves for the first time that genetic determinants are involved in the pathogenesis of this devastating disorder. The identified risk variants are associated with increased risk for both MSA and

PD and add additional evidence for a pathobiological link between these two neurodegenerative synucleinopathies.

Significant associations with MSA were observed for the SNPs rs11931074 and rs3857059 under a recessive test model (table 16). Based on this finding, it would be tempting to hypothesise that MSA is an autosomal recessive disorder, especially as previously identified Japanese MSA families appeared to follow an autosomal recessive inheritance pattern.[181] However, our study used a relatively small sample size and it is possible that our sample was only powered to detect extreme individuals with two risk alleles but that an additive risk conferred by a single risk allele remained undetected. Our observation needs to be re-examined in a larger sample to determine whether a single copy of the risk allele is sufficient to increase risk for developing MSA.

Of the ten SNPs tested in the replication stage, none of the eight SNPs outside the *SNCA* locus reached significance. It is likely that these variants are not relevant to the pathogenesis of MSA; however, because of the small sample size of our replication case-control cohort, a weak association could have been missed. Re-examination of these variants in a larger sample is required to truly exclude those SNPs as risk variants for developing MSA.

Although *SNCA* is the most plausible risk gene for MSA, previous research failed to identify a significant association with this locus. Sequencing of the entire coding *SNCA* sequence, expression analyses, microsatellite testing, haplotype and gene dosage studies were negative (see table 3 for details).[188, 211-216] The failure to detect a significant association with disease can be explained by the low power of these studies and by the fact that none of the two identified risk variants has been previously tested. In our study, we identified strong associations that markedly exceeded the conservative Bonferroni threshold for multiple testing. A subanalysis in pathology-proven MSA

samples and controls confirmed this association (table 17) and excluded the possibility that misdiagnosed PD patients were falsely driving the association.

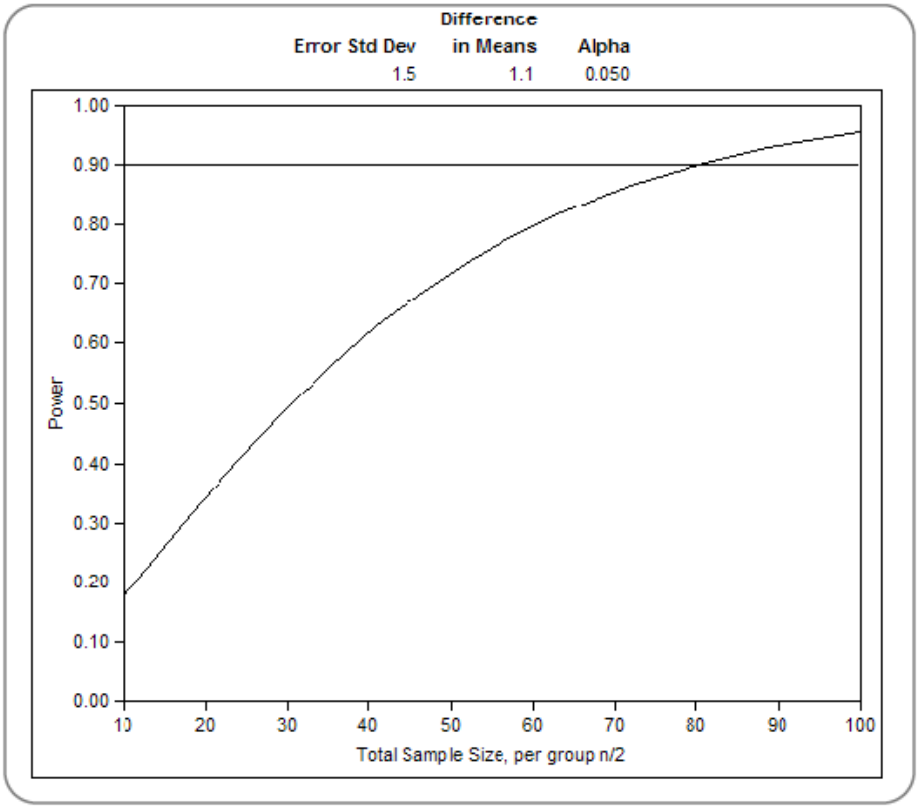
The identified risk variants were associated with both PD (see chapter 3.2) and MSA (figure 36). The odds ratio for SNP rs3857059 under the recessive model was 3.8 (95% CI: 2.4 - 5.9) in PD and 5.9 (95% CI: 3.2 - 10.9) in MSA. Surprisingly, the most significant *SNCA* SNP identified in the PD-GWAS (rs2736990) did not reach significance in our MSA study (*p* value under the recessive model for the screening stage data = 0.38). This observation could indicate a difference in the haplotype structure; a sufficiently powered haplotype fine-mapping study at the *SNCA* locus is necessary to address this important point. It is noteworthy that the H1 haplotype at the *MAPT* locus was observed to increase risk for PD, but no association with altered risk for MSA was noted. This finding is supported by previous research that also failed to identify significant association of the *MAPT* locus with MSA.[188]

Genetic confounding caused by population stratification was a concern in this study as DNA samples were collected in different parts of Europe and North America. To address this point, we studied the allele frequencies of the top SNPs in each population. No significant differences in the risk allele frequencies were observed between the populations. Furthermore, the association with the *SNCA* locus was confirmed in an independent homogeneous case-control cohort used in the replication stage, indicating that population stratification was unlikely to be falsely driving the association.

We did not detect a significant association between *SNCA* variants and increased risk for MSA in the clinical subgroups MSA-P and MSA-C. This failure to replicate our findings most likely reflects the small sample size in each subgroup and the moderate diagnostic accuracy of clinically diagnosed patients,[184, 185] but it also indicates that the association is not only present in one MSA subtype. Our observation would need to be re-examined in a larger sample to determine whether the identified risk variants are associated with a particular clinical subtype.

The mechanisms by which *SNCA* variants confer risk for developing MSA are unknown. Candidate gene studies have excluded the possibility that pathogenic mutations in the *SNCA* coding sequence or gene dosage changes cause disease. [211-215, 290] Furthermore, three independent expression studies have failed to detect abnormal gene expression levels.[214, 215, 290] However, the sample size and power of these expression studies were low (Langerveld: n = 8 cases and 5 controls; Vogt: n = 8 cases and 8 controls; Ozawa: n = 11 cases and 14 controls) and moderate alterations in *SNCA* expression could have been missed.

Figure 37. Power simulation to Detect Modest *SNCA* Expression Changes



This figure shows an estimation of the sample size and power required to detect a 10% increased *SNCA* expression. Approximately 82 cases and 82 controls are required for the detection of a 10% increased expression with a 90% power.

We performed an *in silico* simulation to estimate the required sample size to identify a modest alteration in gene expression (see figure 37). We estimated that for the detection of a - for example - 10% increased gene expression a sample size of at least 82

cases and 82 controls would be required to identify significantly altered levels (*p* value <0.05, error standard deviation 1.5, 90% power). Therefore, a modest change in gene expression could still be a plausible explanation for the strong association between *SNCA* variants and risk for MSA, in particular because duplication or triplication of *SNCA* can to lead to GCI formation in the brains of affected individuals.[82, 180] Alternatively, abnormal splicing, abnormal mRNA processing or other mechanisms could attribute for an altered disease risk.

Taken together, we identified strong associations of genetic variants at the *SNCA* locus with risk for developing MSA. This finding represents the first reproducible risk gene for this devastating disorder, and causally links this condition to the more common synucleinopathy PD. In keeping with observations from other neurodegenerative diseases,[291] we demonstrated that variants in the gene coding for the major pathologically deposited protein can be a risk factor for this particular disease. This discovery represents the first glimpse into the pathogenesis of MSA and opens a new avenue to understand the mechanisms leading to abnormal deposition of α-synuclein.

3.4 *PRKN* and *PINK1* Screening in Early-Onset Parkinson's Disease

STATEMENT OF CONTRIBUTIONS TO THIS RESEARCH:

I designed the study. I supervised the laboratory based sequencing experiments and performed gene-dosage experiments. I was responsible for the statistical analysis and I wrote the manuscript for the original publication in the journal Medical Genetics.

3.4.1 Background

Homozygous mutations in *PRKN* and *PINK1* can cause familial, autosomal recessively inherited forms of early-onset parkinsonism.[86, 96] To date, about 200 different *PRKN* mutations, including missense mutations, splice site mutations, frameshift mutations and copy number mutations, have been described.[292] For *PINK1*, about 80 mutations (mainly missense mutations, but also small insertions, deletions and a copy number mutation) have been reported.[292]

There is increasing evidence pointing towards a possible pathogenic role of heterozygous *PRKN* or *PINK1* mutations.[293] Some heterozygous mutations are thought to increase susceptibility to develop parkinsonism via haploinsufficiency.[294-297] Further screening studies investigating the frequency and spectrum of heterozygous mutations in *PRKN* and *PINK1* are therefore of central importance, in particular as heterozygous mutations are relatively common in the general population.[293]

To study the frequencies and spectrum of copy number and sequence mutations in *PRKN* and *PINK1*, we performed a comprehensive mutational screening study in a publicly available Caucasian cohort of 250 early-onset PD patients and 276 normal controls. We identified several copy number mutations and sequence variants. Our results demonstrated that the frequency of heterozygous pathogenic mutations was higher in cases compared to controls, but the difference between these two cohorts was not statistically significant. Furthermore, we were unable to detect a significant

decrease in age at disease onset in heterozygous mutation carriers. Our data do not support the notion that heterozygosity for pathogenic *PRKN* or *PINK1* mutations increases risk for developing PD.

3.4.2 Materials and Methods

Subjects

The study included 250 early-onset Caucasian PD cases and 276 neurologically normal, Caucasian controls (clinical and demographic characteristics of studied individuals are summarised in table 19). All samples were selected from the publicly available NINDS Neurogenetics Repository at the Coriell Institute for Medical Research (http://ccr.coriell.org/ninds).

Table 19. Clinical and Demographic Characteristics of Studied Subjects

Characteristics	PD patients	Controls
Number of individuals	250	276
Female / male individuals	97 / 153	199 / 77
Mean age at sampling (range)	54 years (22 - 87 years)	39 years (15 - 53 years)
Mean age at onset (range)	41 years (7 - 50 years)	-
Ethnicity		
Caucasian: Hispanic	7	3
Caucasian: not Hispanic	156	273
Caucasian: not specified	87	0
Positive family history	95	44
No information on family history	3	0

This table has been published elsewhere (see reference [298]).

The mean age at disease onset in PD patients was 41 years (range: 7 - 50 years) and the mean age at sampling in control subjects was 39 years (range: 15 - 53 years). 38% of patients and 16% of control individuals reported a positive family history for at least one of the following neurological disorders: parkinsonism, tremor, dementia, restless legs syndrome, dystonia. For each individual written informed consent was obtained by collaborating sites of the Coriell Institute. The study was approved by the Institutional Review Board.

PCR & Sequencing

Primers for all 12 exons in *PRKN* and 8 exons in *PINK1* were designed in Gene Runner (v.3.0.5, Hastings Software Inc., Hastings, NY, USA) using the GenBank reference sequences NM_004562.1 *(PRKN)* and NM_0032409.2 *(PINK1)*. We amplified all exons and exon-intron boundaries using the following PCR reaction mix: 15ng of genomic DNA, 10nM forward primer, 10nM reverse primer and 12µl of FastStart PCR Master (Roche, IN, USA). To amplify exon 1 of *PRKN*, we added 1µl of 5% dimethylsulfoxide (DMSO; American Bioanalytical, MA, and USA) to the PCR mix; for exon 1 of *PINK1* we added 12.4nM 7-deaza-GTP (New England Biolabs, MA, USA) to each reaction. Primer sequences and PCR thermo-cycling conditions are listed in supplemental tables S7 and S8. Following PCR cleanup using AMPure purification kit (Agencourt Bioscience Corporation, MA, USA; see chapter 3.2.2 for a detailed description), bidirectional sequencing of all exons and exon-intron boundaries was performed under the following protocol: 2µl of cleaned PCR product, 0.5µl of BigDye (v.3.1), 1µl of 10nM primer, 2µl 5X Sequencing Buffer (Applied Biosystems), and 5µl distilled and deionised molecular grade water (Mediatech. Inc., VA, USA). The sequencing cycling conditions are listed in supplemental table S3. After purification of the sequencing products with CleanSEQ purification mix (Agencourt Bioscience Corporation; see chapter 3.2.2 for details), sequences were analysed on a 3730xl DNA Analyzer (Applied Biosystems, CA, USA) and electropherograms were visualised in Sequencher software (v.4.1.4; Gene Codes Corporation, MI, USA).

Cloning

In sample ND00548 we observed a complex mutation in the sequence analysis of *PRKN* exon 2. To determine the sequence of each individual allele separately, we cloned PCR amplified exon 2 into the pCR®8/GW/TOPO® vector (Invitrogen, Carlsbad, CA, USA) and transformed One Shot® Chemically Competent *E.coli* cells (Invitrogen) as per the manufacturer's protocol. Then, the transformed cells were cultured in prewarmed S.O.C. medium (Invitrogen) at 37°C for 1 hour. After spreading 25µl of bacterial culture

on a Luria-Bertami (LB) agar plate containing 100µg/ml spectinomycin and an overnight incubation at 37°C, twelve colonies were randomly chosen for further cultivation in LB medium (KD Medical, MD, USA) containing spectinomycin. DNA from each of the twelve cultures was isolated using with PureLink™ Quick Plasmid Miniprep Kit (Invitrogen) as per the manufacturer's instructions and sequenced bi-directionally with direct dye-terminator chemistry as described above.

Gene Dosage Measurements

All individuals included in this study were also part of the screening stage in our GWA study in PD (see chapter 3.2 for details) and therefore genome-wide SNP genotyping data from the Infinium HumanHap 550K version 3 chips (Illumina) were readily available. To determine the gene dosage at the *PRKN* and *PINK1* loci, we visualised the signal intensity metrics 'B allele frequency' (BAF) and 'Log R ratio' (LRR) at these loci using the Genome Viewer tool within BeadStudio software (v.3.1.4, Illumina). The BAF estimates the probability that a genotype is homozygous for the alternative form of an allele (B allele), whereas the LRR is a metric for gene dosage. A heterozygous duplication was called if the LRR was increased by 50%, whereas a 50% decrease in the LRR in combination with an absence of heterozygous SNP calls was called a heterozygous deletion.

To confirm each observed copy number variant, we performed quantitative PCR assays using TaqMan chemistry (Applied Biosystems). Primers and probes were designed in Primer Express software (v.3.0; Applied Biosystems; supplemental tables S9 and S10). Quantitative PCR was performed on a 7900 HT Sequence Detection System (Applied Biosystems) using the following reaction mix: 25ng genomic DNA, 10µl of 2xTaqMan Universal PCR Master Mix (Applied Biosystems), 72nM primers, 10nM of the β-globin probe and 10nM test probe. The cycling conditions are listed in table S11. β-globin served as an endogenous reference. We replicated each measurement six times. Standard curves and Ct-values were calculated in S.D.S software (v.2.2.2; Applied

Biosystems); gene dosage of each exon relative to β-globin and normalised to control DNA was determined using the $2^{-\Delta\Delta Ct}$ method.[299] A heterozygous deletion was called if the $2^{-\Delta\Delta Ct}$-value was between 0.4 and 0.6, and a heterozygous duplication was called if the $2^{-\Delta\Delta Ct}$-value was between 1.3 and 1.6.

Bioinformatic Analysis

To determine the level of sequence conservation of newly identified variants, we performed multiple sequence alignments with paralogues and orthologues from the National Center for Biotechnology Information conserved domain database using ClustalW2 (www.ebi.ac.uk/Tools/clustalw2/index.html).

Variant Definitions

Any deviation from the reference sequences, NM_004562.1 *(PRKN)* and NM_0032409.2 *(PINK1)*, was called a variant or a mutation. Variants with a MAF ≥1% in control samples were classified as polymorphisms. Pathogenic mutations were defined as variants for which at least one of the following criteria applied: (a) stop mutation, (b) frameshift mutation, (c) gene dosage mutation, or (d) missense mutation that has been previously reported to be pathogenic.

Statistical Analyses

For each variant, we calculated the allele frequencies in cases and controls, tested for departures from HWE and performed Fisher's exact tests on allelic association. All computations were performed in PLINK software (version 1.04).[258] After Bonferroni correction for multiple testing, the significance threshold for allelic association was $\alpha' = 0.0012$. Furthermore, we compared the mean age at disease onset in individuals with or without pathogenic mutations using t-tests after Kolmogorov-Smirnov tests revealed a normal distribution of the sample (SPSS, v.12.0.1; SPSS Inc., IL, USA). For the age at onset analysis, a p value <0.05 was considered statistically significant. Schematic illustrations of parkin and PINK1 protein domains were drawn using an online tool

(http://ca.expasy.org/cgi-bin/prosite/mydomains/) and edited in Adobe Illustrator CS4 software (Adobe Systems Inc., CA, USA).

3.4.3 Results

A total of 250 early-onset PD patients and 276 normal controls were screened for mutations in *PRKN* and *PINK1*. We observed 41 sequence variants, including eight novel and 33 previously described variants (table 21, table 22; frameshift, stop and nonsynonymous mutations are illustrated in figure 38). In addition, we identified eight cases and three controls with heterozygous gene dosage mutations (supplemental figures S1-S3, table 20). Only 14 out of 250 cases (5.6%) and 5 out of 276 controls (1.8%) carried pathogenic mutations in *PRKN* or *PINK1* (table 20). None of the identified variants deviated from HWE or showed a significant association with disease using Fisher's exact tests for allelic association.

PRKN Analysis

Sequence analysis in *PRKN* revealed four frameshift mutations, one stop mutation, 13 missense mutations, two silent mutations and five intronic variants (table 21). Of these variants, five have not been previously reported: p.A38VfsX6, p.C166Y, p.Q171X, p.D243N, p.M458L (electropherograms of the mutant sequences are shown in figures 39 and 40).

Gene dosage analysis revealed four patients (1.6%) and two controls (0.7%) with a heterozygous exon deletion; in four cases (1.6%) and in one control (0.4%) we observed a heterozygous exon duplication (table 20, supplemental figures S1-S3). None of the studied individuals carried a homozygous copy number mutation.

Table 20. Individuals with Pathogenic Mutations in *PRKN* or *PINK1*

Sample ID	Mutation(s)[a]		AAO	Family history
Pathogenic *PRKN* mutations in cases				
ND00136	heterozygous for p.Q171X	heterozygous for p.R275W	38	present
ND00153	heterozygous deletion of exons 5-6		29	present
ND00187[b]	heterozygous duplication of exons 5-9		38	present
ND00429[b]	heterozygous duplication of exons 5-9		41	present
ND00548	heterozygous for p.A38VfsX6	heterozygous for p.Q34RfsX5	36	absent
ND01119	heterozygous deletion exon 2		50	absent
ND01122	heterozygous for p.R275W		47	absent
ND02639	homozygous for p.L112LfsX15		30	absent
ND02798	heterozygous duplication of exon 6		32	present
ND04581	heterozygous deletion of exon 2	heterozygous for p.Q34RfsX5	22	present
ND05921	heterozygous deletion exon 4		49	present
ND06330	heterozygous duplication exons 2-3		50	absent
ND06635	heterozygous for p.N52MfsX29		42	present
Pathogenic *PINK1* mutations in cases				
ND08471	heterozygous for p.M318L		47	absent
Pathogenic *PRKN* mutations in controls				
ND03967	heterozygous deletion exon 3		N/A	present
ND04990	heterozygous duplication exon 2		N/A	absent
ND08538	heterozygous deletion exons 2-4		N/A	absent
ND09912	heterozygous for p.R275W		N/A	absent
ND10272	heterozygous for p.T415N		N/A	present

This table has been published elsewhere (see reference [298]).

[a] Only frameshift mutations, stop mutations, gene dosage mutations that span the coding sequence, and missense
mutations that have been reported to be pathogenic are listed.

[b] ND00187 and ND00429 are siblings

AAO, age at disease onset; N/A, not applicable

Table 21. Variants in *PRKN* in 250 Early-Onset PD Cases and 276 Normal Controls

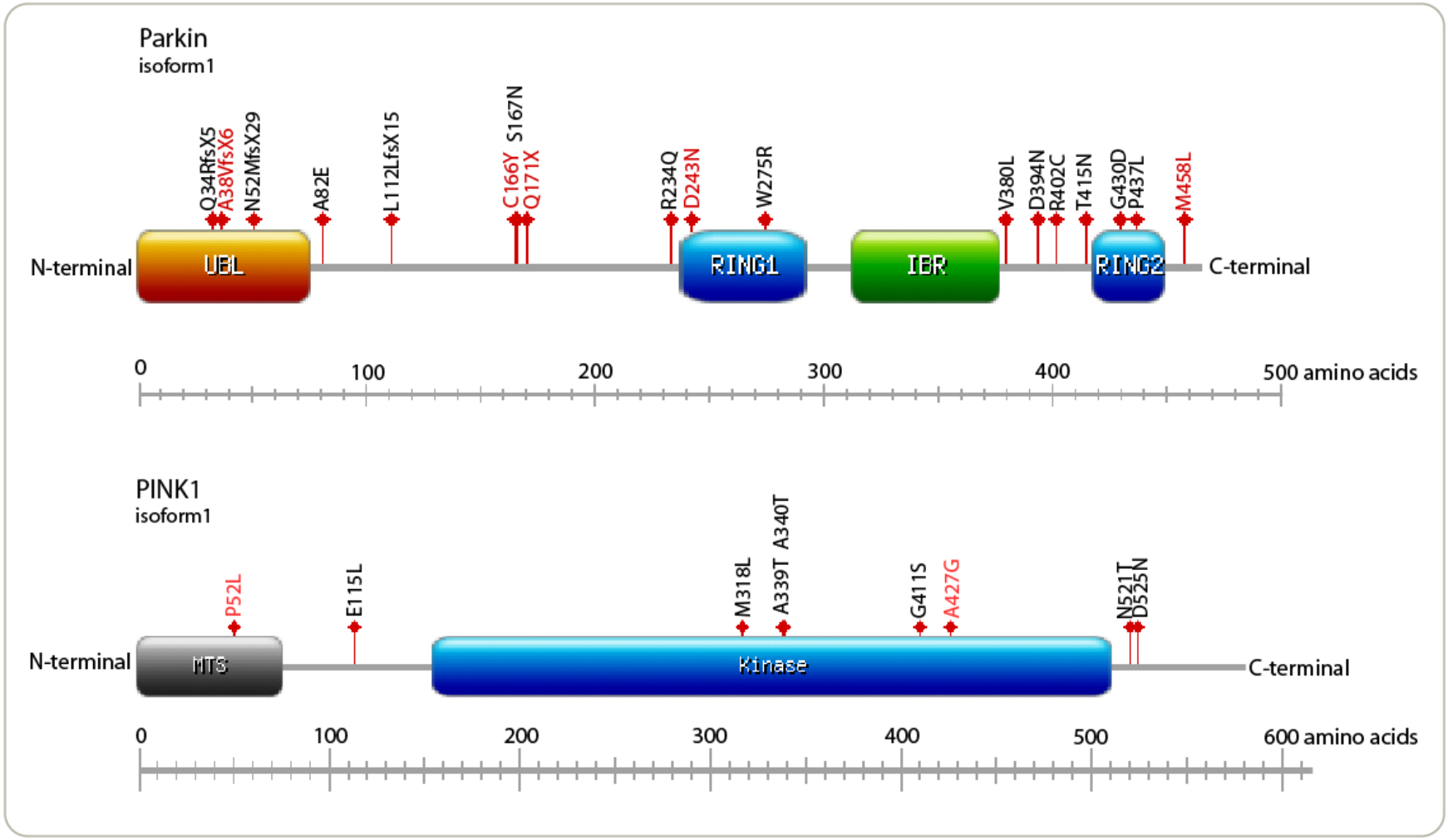
Nucleotide	Location/rs-number	Amino acid change	Mutation type	Genotypes in cases ($A_1A_1/A_1A_2/A_2A_2$)	Genotypes in controls ($A_1A_1/A_1A_2/A_2A_2$)	MAF cases - controls	Ref.
Novel mutations that are likely to be pathogenic							
c.[214delC+216G>T]	exon 2	p.A38VfsX6	frameshift	0/1/249	0/0/276	0.002 - 0.000	-
c.612C>T	exon 4	p.Q171X	stop	0/1/249	0/0/276	0.002 - 0.000	-
Novel mutations of unknown significance							
c.598G>A	exon 4	p.C166Y	missense	0/1/249	0/0/276	0.002 - 0.000	-
c.828G>A	exon 6	p.D243N	missense	0/0/250	0/1/275	0.000 - 0.002	-
c.1473A>C	exon 12	p.M458L	missense	0/1/249	0/0/276	0.002 - 0.000	-
Previously described mutations that are likely to be pathogenic							
c.202-203delAG	exon 2	p.Q34RfsX5	frameshift	0/2/248	0/0/276	0.004 - 0.000	89
c.256delA	exon 2	p.N52MfsX29	frameshift	0/1/249	0/0/276	0.002 - 0.000	90
c.437-477del	exon 3	p.L112LfsX15	frameshift	1/0/249	0/0/276	0.004 - 0.000	90
c.924C>T	exon 7/rs34424986	p.R275W	missense	0/2/248	0/1/275	0.004 - 0.002	89, 300
c.1345C>A	exon 11	p.T415N	missense	0/0/250	0/1/275	0.000 - 0.002	89
Previously described mutations with unknown significance							
c.802G>A	exon 6	p.R234Q	missense	0/1/249	0/0/276	0.002 - 0.000	301
c.1305C>T	exon 11	p.R402C	missense	0/0/250	0/1/275	0.000 - 0.002	302
c.1390G>A	exon 12	p.G430D	missense	0/1/249	0/0/276	0.002 - 0.000	91
c.1411C>T	exon 12	p.P437L	missense	0/3/247	0/3/273	0.006 - 0.005	295
Previously described mutations that are likely to be non-pathogenic							
c.273-18A	intron 2	-	intronic	0/4/246	0/2/274	0.008 - 0.004	303
c.346C>A	exon 3	p.A82E	missense	0/0 /250	0/2/274	0.000 - 0.004	304, 305
c.623C>T	exon 4	p.L174L	silent	0/1/249	0/2/274	0.002 - 0.004	303
c.884T>C	exon 7/rs9456711	p.L261L	silent	0/2/248	0/0/276	0.004 - 0.000	90, 306
Previously described polymorphisms							
c.272+25C	intron 2/rs2075923	-	intronic	13/83/154	10/95/171	0.218 - 0.208	89
c.514-20T	intron 3/rs4709583	-	intronic	1/35/214	1/42/233	0.074 - 0.080	89
c.601G>A	exon 4/rs1801474	p.S167N	missense	0/7/243	0/9/267	0.014 - 0.016	307
c.973-35G	intron 7/rs3765474	-	intronic	52/126/72	52/138/86	0.460 - 0.438	89
c.1034+48T	intron 8/rs10945756	-	intronic	16/87/147	17/98/161	0.238 - 0.239	303
c.1239G>C	exon 10/rs1801582	p.V380L	missense	6/78/166	5/89/182	0.180 - 0.179	89
c.1281G>A	exon 11/rs1801334	p.D394N	missense	0/16/234	0/14/262	0.032 - 0.025	89

Table 22. Variants in *PINK1* in 250 Early-Onset PD Cases and 276 Normal Controls

Nucleotide	Location/rs-number	Amino acid change	Mutation type	Genotypes in cases ($A_1A_1/A_1A_2/A_2A_2$)	Genotypes in controls ($A_1A_1/A_1A_2/A_2A_2$)	MAF cases - controls	Ref.
Novel mutation that is likely to be pathogenic							
c.1374C>A	exon 7	p.A427E	missense	0/2/248	0/0/276	0.004 - 0.000	-
Novel mutation of unknown significance							
c.249C>T	exon 1	p.P52L	missense	0/1/249	0/0/276	0.002 - 0.000	-
Novel mutation that is likely to be non-pathogenic							
c.1354G>A	exon 7	p.T420T	silent	0/1/249	0/0/276	0.002 - 0.000	-
Previously described mutation that is likely to be pathogenic							
c.1046A>T	exon 4	p.M318L	missense	0/1/249	0/0/276	0.002 - 0.000	[98, 298]
Previously described mutations with unknown significance							
c.1109G>A	exon 5/rs55831733	p.A339T	missense	0/2/248	0/0/276	0.004 - 0.000	[98]
c.1325G>A	exon 6/rs45478900	p.G411S	missense	0/0/250	0/1/275	0.000 - 0.002	[308]
c.1667G>A	exon 8	p.D525N	missense	0/1/249	0/0/276	0.002 - 0.000	[309]
c.1822A>G	exon 8	p.S576S	silent	0/1/249	0/0/276	0.002 - 0.000	[98]
Previously described mutations that are likely to be non-pathogenic							
c.1030G>A	exon 4/rs56200357	p.R312R	silent	0/1/249	0/0/276	0.002 - 0.000	[96]
c.1267T>C	exon 6/rs45499398	p.D391D	silent	0/1/249	0/0/276	0.002 - 0.000	[97, 98]
Previously described polymorphisms							
c.283C>T	exon 1	p.L63L	silent	9/81/160	5/78/193	0.198 - 0.159	[98]
c.438A>T	exon 1	p.Q115L	missense	0/17/233	0/32/244	0.034 - 0.058	[310]
c.482-7A	intron 1/rs2298298	-	intronic	2/48/200	3/63/210	0.104 - 0.125	[98]
c.1054-5G	intron 4/rs3131713	-	intronic	2/51/197	2/63/211	0.110 - 0.121	[98]
c.1112G>A	exon 5/rs3738136	p.A340T	missense	1/33/216	0/18/258	0.070 - 0.033	[98]
c.1656A>C	exon 8/rs1043424	p.N521T	missense	15/110/125	21/110/145	0.280 - 0.275	[98]

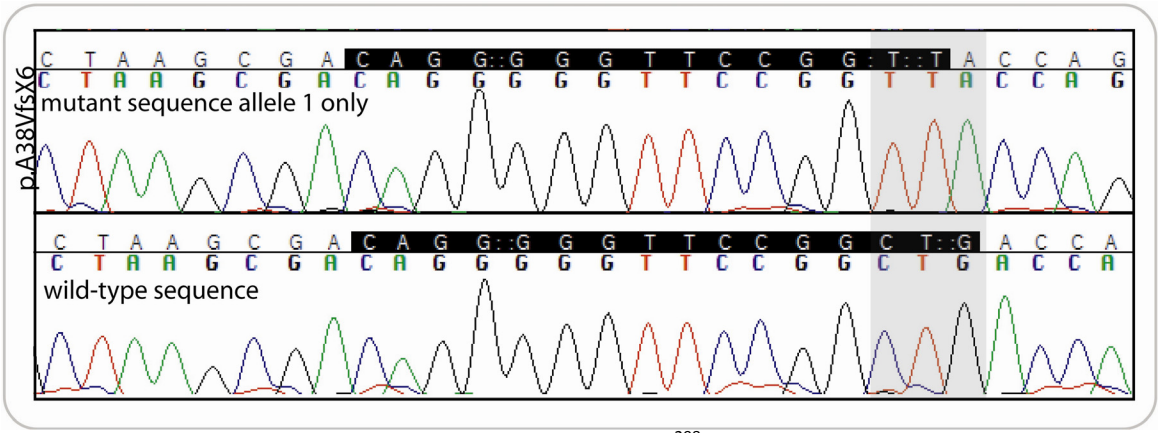
This table has been published elsewhere (see reference [298]). Genotype distribution of the major allele (A_2) and minor allele (A_1) is described for each variant. MAF, minor allele frequency.

Figure 38. Schematic Diagrams of Frameshift, Stop and Non-Synonymous *PRKN* and *PINK1* Mutations



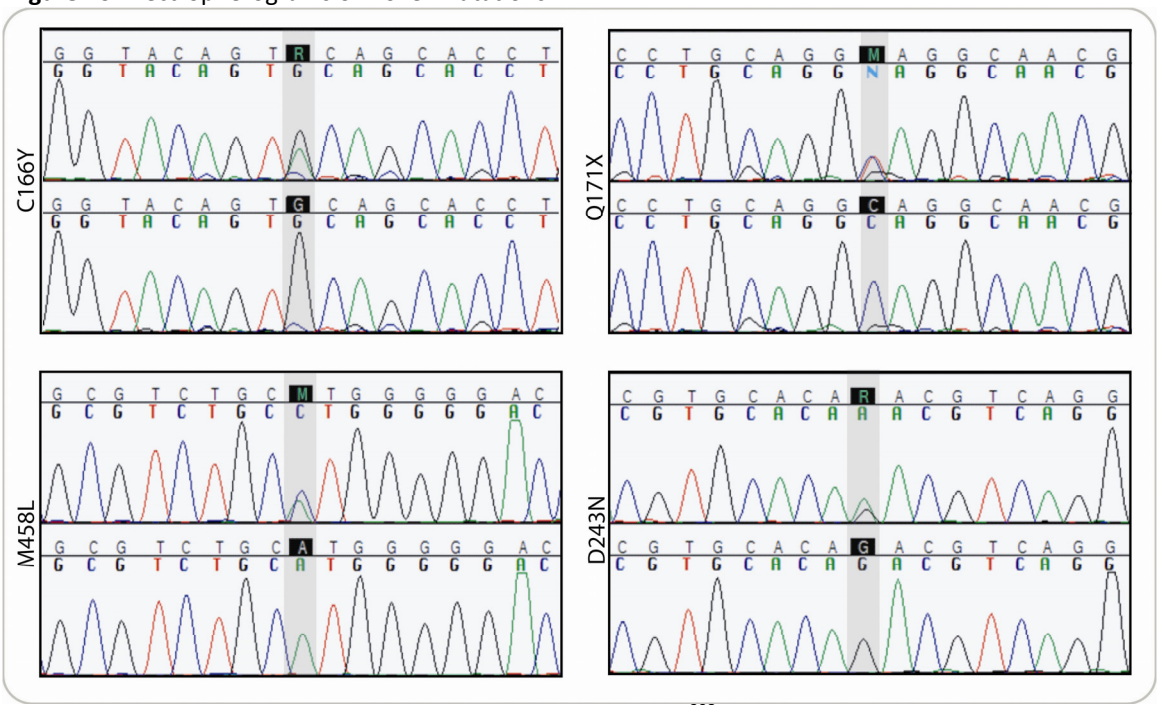
This figure has been published elsewhere (see reference [298]). Shown are frameshift, stop and non-synonymous mutations relative to the protein domains of parkin and PINK1. Novel mutations are highlighted in red. UBL, ubiquitin-like domain; IBR, in-between RING domain; MTS, mitochondrion transfer sequence.

Figure 39. Electropherogram of a Novel Frameshift Mutation in *PRKN*



This figure has been published elsewhere (see reference [298]). Chromatogram of a novel frameshift mutation (p.A38VfsX6) in patient ND00548 is shown. This individual carries two pathogenic mutations in *PRKN*, a previously described frameshift mutation (p.Q34RfsX5) on one allele (not shown) and a novel frameshift mutation (p.A38VfsX6) on the other allele. The mutated allele with the p.A38VfsX6 mutation is shown in the top panel, and the wild-type sequence is depicted in the bottom panel. The mutated nucleotides are highlighted in grey. All sequences are shown in the forward direction.

Figure 40. Electropherograms of Novel Mutations in *PRKN*



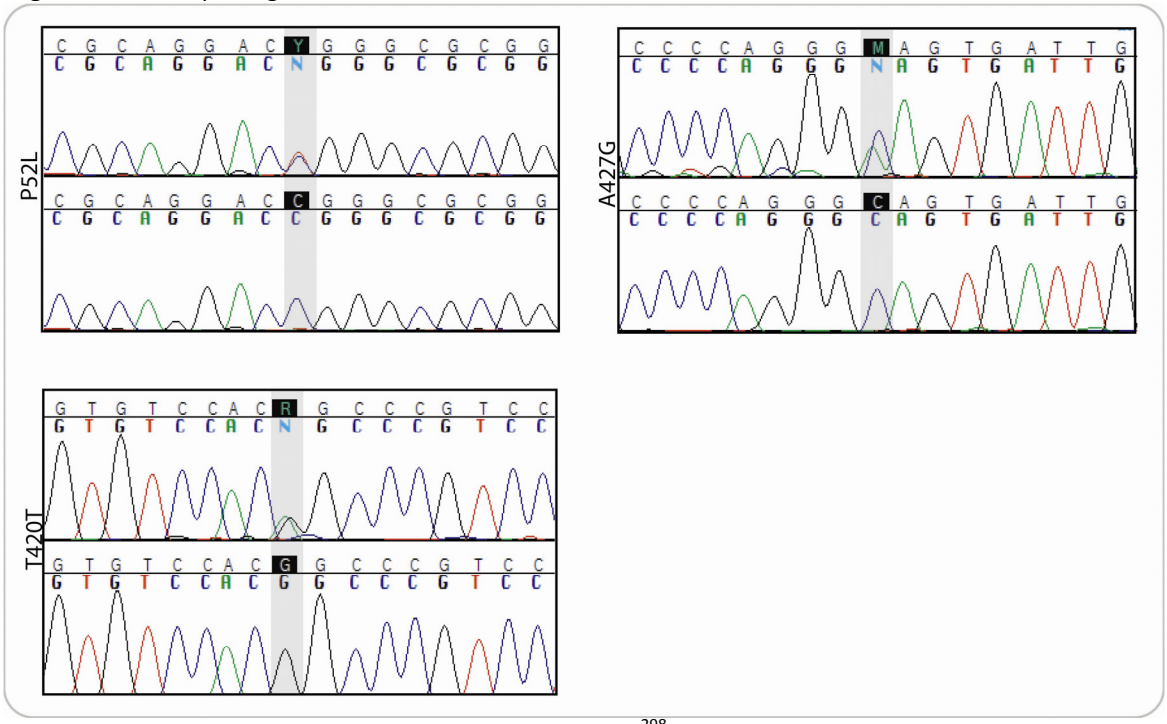
This figure has been published elsewhere (see reference [298]). Electropherograms of a novel stop mutation (p.Q171X) and three novel missense mutations (p.C166Y, p.M458L, p.D243N) are shown. The mutations p.C166Y, p.Q171X and p.M458L were observed in PD patients, whereas p.D243N was observed in a control sample. Wild-type sequences are depicted in the bottom panels; mutated sequences are shown in the top panels. Mutated nucleotides are highlighted in grey. All sequences are shown in the forward direction.

Heterozygous pathogenic *PRKN* mutations were found in nine patients (3.6%) and in five controls (1.8%), compound heterozygous mutations were present in three patients (1.2%) but were not found in controls, and one patient (0.4%) was found to be homozygous for a pathogenic mutation.

PINK1 Analysis

Sequence analysis in *PINK1* revealed the following variants: nine missense mutations, five silent mutations, and two intronic variants. Three of these mutations have not been previously reported (p.P52L, p.T420T, p.A427E; electropherograms of all novel variants are shown in figure 41). None of the studied individuals had a *PINK1* copy number mutation. Only one patient (0.4%) carried a heterozygous pathogenic mutation in *PINK1;* none of the control samples contained a pathogenic mutation.

Figure 41. Electropherograms of Novel Mutations in *PINK1*

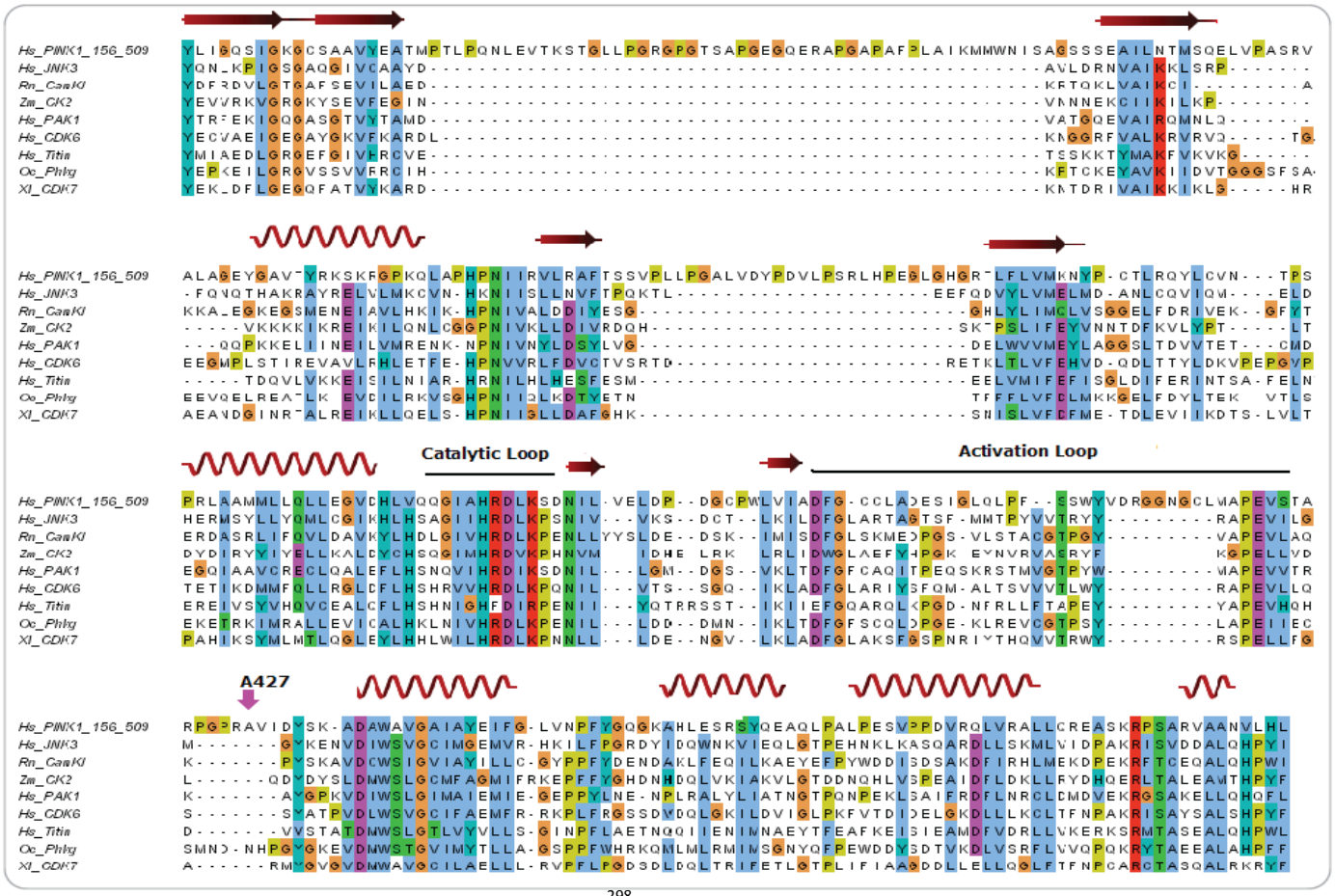


This figure has been published elsewhere (see reference [298]). Electropherograms of two novel missense mutations (p.P52L, p.A427G) and a novel silent mutation (p.T420T) are shown. All three novel mutations have been observed in PD patients. The wild-type sequences are depicted in the bottom panels; the mutated sequences are shown in the top panels. The mutated nucleotides are highlighted in grey. All sequences are shown in the forward direction.

Bioinformatic Analysis of Sequence Conservation

Sequence conservation analysis revealed that at least three of the newly described variants were probably pathogenic: variants p.A38VfsX6 and p.Q171X are frameshift mutations that lead to an early termination of translation; variant p.A427E is located in close proximity of the APE triple that is part of the highly conserved serine-threonine kinase activation loop. It is likely that mutations close to this crucial site interfere with substrate binding (figure 42).

Figure 42. Alignment of Human *PINK1* Protein Sequence with Conserved Domains



This figure has been published elsewhere (see reference [298]). Multiple sequence alignments of the *PINK1* kinase domain with several serine-threonine kinase orthologues and paralogues from the conserved domain database are shown. We observed a novel mutation (p.A427E) in *PINK1* (indicated with a purple arrow). This variant is located in close proximity to the highly conserved kinase activation loop and it is likely that mutations at this position interfere with substrate binding to this crucial domain. Red swirl ribbons indicate predicted α-helical structures and red arrowed ribbons stand for β-strands. This figure has been constructed by Dr. Jinhui Ding.

3.4.4 Discussion

We performed a comprehensive mutational screening study for *PRKN* and *PINK1* in 250 early-onset PD cases and 276 normal controls. We identified 41 sequence variants of which eight have not been previously reported (*PRKN*: p.A38VfsX6, p.C166Y, p.Q171X, p.D243N, p.M458L; *PINK1*: p.P52L, p.T420T, p.A427E). All variants are summarised in tables 20 and 21 and a schematic illustration is shown in figure 38. At least three of the newly identified mutations are likely to be pathogenic (*PRKN:* p.A38VfsX6, p.Q171X; *PINK1:* p.A427E). Variants p.A38VfsX6 and p.Q171X cause an early termination of protein translation, and variant p.A427E is located in close proximity to the crucial serine-threonine kinase domain in *PINK1* and is likely to interfere with substrate binding (figure 42). We furthermore observed eight cases and three controls with heterozygous copy number mutations (table 20). A total of 14 PD cases (5.6%) and five controls (1.8%) were found to carry one or more pathogenic mutation in PRKN or PINK1 (details are listed in table 20). Homozygous or compound heterozygous pathogenic mutations were observed in four PD patients (1.6%) but in none of the controls. The frequency of heterozygous pathogenic mutations was higher in PD cases (4.0%) compared to controls (1.8%) but the difference was not statistically significant (*p* value = 0.13). None of the patients were heterozygous for pathogenic mutations in both genes and therefore an epistasis analysis was not performed.

It is surprising that pathogenic mutations in *PRKN* or *PINK1* were found in only 5.6% of early-onset PD cases. This number is much lower than previous estimations have indicated.[91] One possible explanation for this finding could be that we applied conservative criteria for calling pathogenic mutations. We classified a variant as pathogenic only if at least one of the following criteria applied: frameshift mutation, stop mutation, copy number mutation or missense mutation reported as likely pathogenic. Furthermore, the case selection criteria in our study included early-onset PD patients with and without a family history of parkinsonism and the age at onset ranged from 7 - 50 years. It is also possible that population-specific variability in allele

frequencies exists, - a phenomenon that is well known for other PD genes such as *LRRK2, GBA* and *MAPT*.[109-112, 311-313]

This study failed to identify a significant overrepresentation of heterozygous pathogenic mutations in cases versus controls. This observation is of interest as a number of studies have reported that heterozygous mutations are significantly increased in parkinsonism cases suggesting that heterozygosity may act as a susceptibility factor for PD.[293, 314-319] In contrast to previous works, our study demonstrated only a trend towards an overrepresentation of heterozygous mutations in cases versus control (4.0% versus 1.8%), but the difference was not statistically significant (*p* value = 0.13). The negative result of this study and of a few other reports in literature can probably be explained by different case selection and mutation-calling criteria.[303, 320, 321] As mentioned above, we applied strict criteria for classifying a mutation as pathogenic and as a consequence the frequency of pathogenic mutations in our study was lower than in other studies. It is possible, that a larger cohort would have to be screened to detect significant differences in the frequencies of heterozygous pathogenic mutations.

Analysis of the relationship between age at disease onset and mutation status demonstrated that patients who were homozygous or compound heterozygous for pathogenic *PRKN* mutations had a significantly lower age at onset (mean age at onset = 32 years ± 7 years) than patients with a heterozygous pathogenic mutation (mean age at onset = 43 years ± 8 years) (mean difference = 11 years, *p* value = 0.03, 95% CI = 1.4 - 20.6). In contrast to previous reports, however, we failed to detect a significant decrease in age at onset in heterozygous patients compared to patients without a pathogenic mutation (43 ± 8 years vs. 41 ± 9 years; mean difference = 2 years, *p* value = 0.54, 95% CI= -3.7 - 7.0).[295, 297]

One limitation of this study was that genome-wide SNP chips used for the analysis of copy number mutations have only a moderate resolution. The average density of the

Illumina 550K chip was about 0.35 SNPs per kb at the *PRKN* locus and 0.39 SNPs per kb at the *PINK1* locus. Therefore, we could have missed small duplications or deletions and the measurement of 11 copy number mutations in our study could be an underestimation.

In summary, we report the results of comprehensive mutational screening of *PRKN* and *PINK1* in early-onset PD patients and normal controls. Eight out of a total of 41 sequence variants have not been previously described and at least three of these novel variants are likely to be pathogenic. Furthermore, our data indicate a trend towards an increased frequency of heterozygosity for pathogenic *PRKN* or *PINK1* mutations in patients compared to controls, but this difference was not significant.

3.5 *PRKN* and *PINK1* Screening in Multiple System Atrophy

STATEMENT OF CONTRIBUTIONS TO THIS RESEARCH:

I conceptualised and designed the study. I supervised laboratory based sequencing experiments and I performed gene-dosage measurements. I was responsible for the statistical analysis and I wrote the manuscript that was published in the journal Neurobiology of Aging.

3.5.1 Background

Multiple system atrophy (MSA) and Parkinson's disease (PD) are characterised as synucleinopathies based on abnormal accumulation of fibrillar α-synuclein in the brain. In MSA, α-synuclein is typically deposited in oligodendroglial cells (these deposits are known as glial cytoplasmic inclusions [GCIs]), whereas in PD α-synuclein is mainly found in neuronal aggregates in the form of Lewy bodies. The molecular processes that lead to pathologic deposition of α-synuclein in these two neurodegenerative disorders are largely unknown.

We recently discovered that common variability at the *SNCA* locus, coding for α-synuclein, is significantly associated with risk for developing MSA and PD (see chapters 3.2 and 3.3 for further details).[272, 289] In addition, reports on patients with *SNCA* multiplication indicate overlapping clinical and pathological features reminiscent of PD in some cases and of MSA in others.[81, 82, 180] These observations therefore raise the question whether additional genetic risk factors are shared between these two synucleinopathies.

Mutations in *PRKN* and *PINK1* have been associated with recessively inherited early-onset PD.[86, 96] Both, parkin and PINK1 protein have been described to co-localise with Lewy-bodies and variable immunoreactivity has also been observed in GCIs.[322-324] Therefore, we considered *PRKN* and *PINK1* as candidate risk genes for MSA. To test

the hypothesis that mutations in *PRKN* and *PINK1* are involved in the pathogenesis of MSA, we performed a mutational screening study in 87 autopsy-proven Caucasian MSA samples. We performed sequence analysis of all exons and exon-intron boundaries in *PRKN* and *PINK1*, and furthermore measured the gene dosage of all *PRKN* exons. All variants were compared to the results from a previously screened normal Caucasian control cohort (see chapter 3.4). We identified a total of 17 sequence variants including two novel silent mutations (*PINK1:* p.R337R, p.G189G), and four heterozygous exon deletions. None of the identified variants was significantly associated with disease. Our results indicate that mutations in *PRKN* and *PINK1* are not commonly associated with increased risk for developing MSA.

3.5.2 Materials and Methods

Samples

Eighty-seven pathology-proven Caucasian MSA cases from the Queen Square Brain Bank for Neurological Disorders, London, UK were included in the study. For association tests, data from 276 neurologically normal Caucasian controls were used. These controls have been described in chapter 3.4. The study was approved by the appropriate institutional review boards. Written informed consent was obtained for each patient.

DNA Extraction, PCR & Sequencing

We extracted the DNA from brain tissue using the DNeasy Tissue Kit (Qiagen, Hilden, Germany) according the manufacturer's instructions. PCR and sequencing of all exons and exon-intron boundaries in *PRKN* and *PINK1* were performed using the protocols described in chapter 3.4.

Gene Dosage Measurements

We measured the gene dosage for *PRKN* exons 1 - 12 using the $2^{-\Delta\Delta Ct}$ method as described in chapter 3.4. Each measurement was replicated three times.

Statistical Analysis

All computations were performed in PLINK software (version 1.04).[258] Fisher's exact tests were applied to test for departures from HWE and for allelic association of *PRKN* and *PINK1* variants. A Bonferroni corrected *p* value of 0.0029 was considered statistically significant (two-sided α of 0.05 divided by 17 tests performed). *PRKN* copy number mutations in MSA patients were compared with the number of copy number mutations in control samples by means of a χ^2-test (http://faculty.vassar.edu/lowry/newcs.html). Schematic diagrams of the *PRKN* and *PINK1* genes were drawn using an online graphical tool (http://warta.uni-muenster.de/cgi-bin/Tools/StrDraw.pl) and edited in Adobe Illustrator CS4 (Adobe Systems Inc.).

3.5.3 Results

Mutational Screening Results

Eighty-seven definite MSA patients were studied for mutations in *PRKN* and *PINK1* (table 23).

In *PRKN*, we identified eight sequence variants including six nonsynonymous mutations and two intronic variants. All variants have been found in a heterozygous state. Copy number variants were detected in four cases (4.6%) (figure 43).

PINK1 analysis revealed nine sequence variants of which four were silent mutations, two were intronic variants and three were common nonsynonymous polymorphisms (table 23, figure 43). Of the four silent variants, two have not been previously reported (c.661C>A = p.G189G, c.1105C>T = p.R337R; electropherograms are shown in Figure 44).

Table 23. *PRKN* and *PINK1* Sequence Variants in 87 Definite MSA Cases

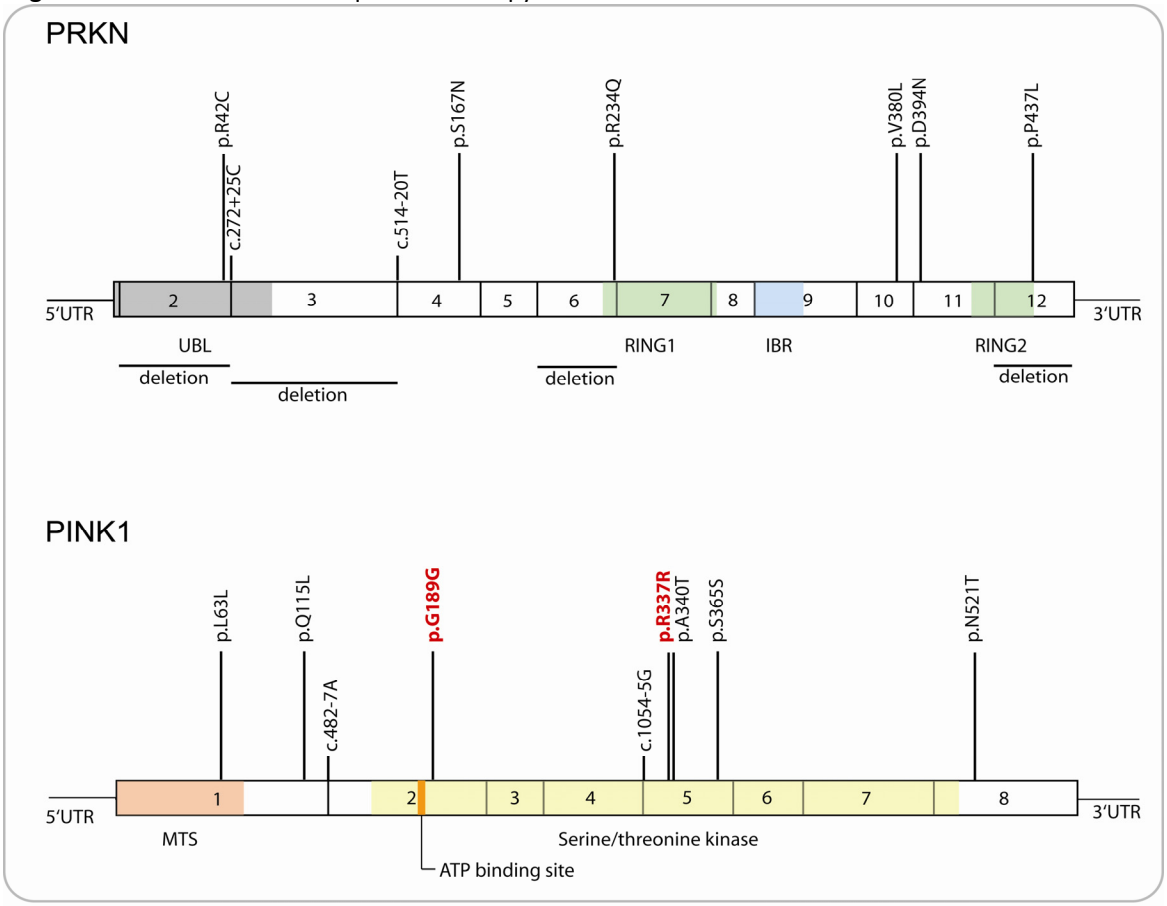
Nucleotide change	Amino acid change	Location	Mutation type	Alleles in cases	Alleles in controls[a]	*p* value	Ref.
PRKN							
c.258C>T	p.R42C	exon 2	missense	1 / 175	0 / 552	0.242	[325]
c.272+25C	-	intron 2	intronic	34 / 142	115 / 437	0.747	[298]
c.514-20T	-	intron 3	intronic	13 / 163	44 / 508	0.873	[298]
c.601G>A	p.S167N	exon 4	missense	2 / 174	9 / 543	1	[298]
c.835G>A	p.R234Q	exon 6	missense	1 / 175	0 / 552	0.242	[298]
c.1239G>C	p.V380L	exon 10	missense	33 / 143	99 / 453	0.823	[298]
c.1281G>A	p.D394N	exon 11	missense	12 / 162	14 / 538	0.016	[298]
c.1444C>T	p.P437L	exon 12	missense	1 / 175	0 /552	0.242	[298]
PINK1							
c.283C>T	p.L63L	exon 1	silent	36 / 130	88 / 464	0.101	[298]
c.438A>T	p.Q115L	exon 1	missense	12 / 154	32 / 520	0.467	[298]
c.482-7A	-	intron 1	intronic	21 / 147	69 / 483	1	[298]
c.661C>A	p.G189G	exon 2	silent	1 / 167	0 / 552	0.233	-
c. 1054-5G	-	intron 4	intronic	20 / 156	67 / 485	0.894	[298]
c.1105C>T	p.R337R	exon 5	silent	1 / 175	0 / 552	0.242	-
c.1112G>A	p.A340T	exon 5	missense	9 / 167	18 / 534	0.257	[298]
c.1189C>T	p.S365S	exon 5	silent	1 / 175	0 / 552	0.242	[309]
c.1656A>C	p.N521T	exon 8	missense	53 / 121	152 / 400	0.499	[298]

This table has been published elsewhere (see reference [326]).

[a] Mutational screening results of normal controls have been presented in chapter 3.4 of this thesis.

MAF, minor allele frequency; Ref., reference(s).

Figure 43. *PRKN* and *PINK1* Sequence and Copy Number Variants in 87 Definite MSA Cases

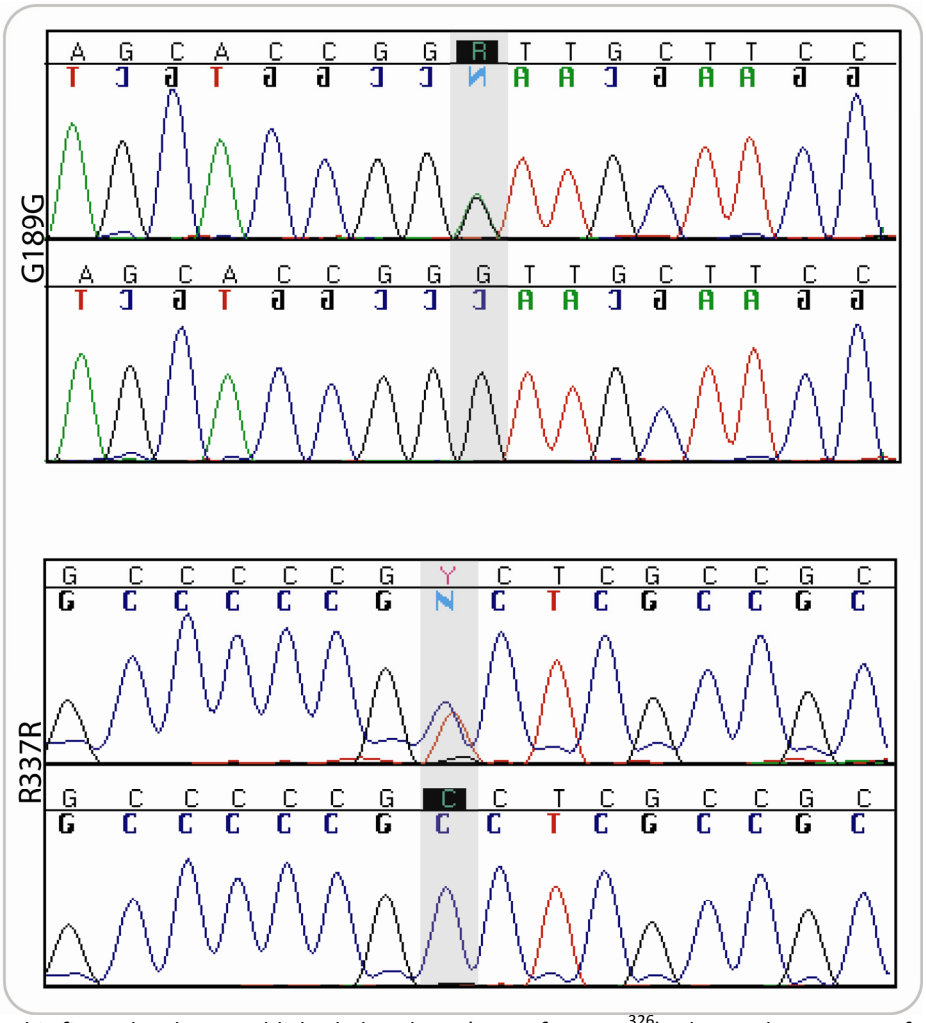


This figure has been published elsewhere (see reference [326]). Identified sequence variants and copy number mutations are shown relative to exons and corresponding protein domains of *PRKN* or *PINK1*. Novel variants are indicated in red. UBL, ubiquitin-like domain; RING1, zinc finger RING-type 1 domain; IBR, in-between RING domain; RING2, zinc finger RING-type 2; MTS, mitochondrion transit sequence.

Statistical Analysis

No deviations from Hardy-Weinberg equilibrium were detected for any identified variant in previously described control samples. None of the detected variants was significantly associated with risk for developing MSA (*p* values are shown in table 23). The frequency of copy number variants in *PRKN* was not significantly different in MSA cases compared to normal controls (*p* value = 0.08).

Figure 44. Novel *PINK1* Variants in MSA Cases



This figure has been published elsewhere (see reference [326]). Electropherograms of two novel silent mutations (p.G189G, p.R337R) are shown. The wild-type sequences are depicted in the bottom panels; the mutated sequence is shown in the top panels. The mutated nucleotides are highlighted in grey. All sequences are shown in the forward direction.

3.5.4 Discussion

In this study, we investigated the frequency and spectrum of mutations in *PRKN* and *PINK1* in 87 pathology-proven MSA cases. Our analysis revealed 17 sequence variants, including four intronic variants, four silent mutations and nine missense mutations (table 23). Two silent mutations in *PINK1* have not been previously reported (p.R337R and p.G189G; figure 44). Furthermore, we detected four heterozygous exon deletions in *PRKN* (figure 43).

None of the variants in *PRKN* or *PINK1* were significantly associated with MSA (Table 23). Further, the frequency of copy number variants observed in MSA patients did not differ significantly from the frequency observed in normal controls. All individuals who carried the possibly pathogenic *PRKN* sequence variants p.R42C, p.R234Q, p.P437L or copy number variants were heterozygous, and no carriers with homozygous pathogenic mutations were observed.

In summary, our data indicate that genetic variants in *PRKN* and *PINK1* are not commonly associated with MSA.

4 Conclusions and Future Directions

My PhD research focused on modern molecular genetic approaches to catalogue genetic variability in the human genome and to determine the role of genetic variants in neurological diseases. I demonstrated novel SNP chip applications in a GWA study in PD, an autozygosity mapping approach in Brazilian dystonia families, a candidate SNP association study in MSA and a population genetics study in ethnically and geographically diverse human populations. I furthermore applied SNP chip data to rapidly screen for copy number mutations in *PRKN/PINK1*. I will briefly summarise the major findings of these projects and point out future directions:

Human Genome Diversity Project

We studied an ethnically and geographically diverse selection of 513 individuals from 29 worldwide populations for genome-wide patterns of variation. We generated a publicly available resource on genotype, haplotype and copy-number variation in diverse populations. Our major findings included (a) the observation of increasing LD with increasing distance from East Africa supporting a serial founder effect for an out-of-Africa spread of human populations, (b) a detailed description of genome-wide copy number variation in diverse populations (of which many have not been previously described), and (c) new inferences on inter-population relationships and genomic variation. These data serve as a genomic resource for human genetic research studying genomic variation.

Autozygosity Mapping in Brazilian Dystonia-Parkinsonism Families

In three Brazilian families with dystonia-parkinsonism we performed autozygosity mapping to identify regions of ancestrally shared homozygosity. We identified a disease-segregating locus on chromosome 2 spanning 2.44 Mb. We sequenced all transcripts in this locus and observed a disease segregating missense mutation (p.P222L) in *PRKRA*.

This mutation was absent in 1,686 ethnically diverse control samples. Mutations in *PRKRA* have not been previously associated with human disease. Screening studies in familial dystonia patients together with cell- and animal-based work is still necessary to establish *PRKRA* as a novel dystonia gene.

Genome-Wide Association Study in PD

We performed a GWA study testing 417,964 SNPs in a total of 1,713 Caucasian PD cases and 3,974 Caucasian controls, and then attempted to replicate the 384 most associated of these SNPs in an independent cohort of 3,513 PD cases and 4,232 controls. Variants at the *SNCA* and *MAPT* loci were significantly associated with risk for developing PD (*SNCA*: rs2736990 p value = 1.9×10^{-16}, OR = 1.2; *MAPT:* rs393152 p value = 1.0×10^{-15}, OR = 0.8). Furthermore, we observed low *p* values at a newly designated risk locus on chromosome 1q32 (PARK16: rs823128 *p* value = 1.9×10^{-7}) and upstream of *LRRK2* (rs1491923 *p* value = 7.0×10^{-6}). Our research demonstrated that common genetic variants play a significant role in the pathogenesis of PD. Additional replication studies, fine-mapping of the identified risk loci and investigations of the functional effects are required to successfully dissect the mechanisms by which these variants increase disease risk.

Candidate SNP Association Study in MSA

We have shown that genetic variants at the *SNCA* locus are associated with increased risk of developing MSA. This finding represents the first reproducible risk gene identified in this devastating neurodegenerative disorder and might open a new avenue to understand the mechanisms that lead to abnormal deposition of α-synuclein. Further studies, including fine-mapping of the *SNCA* locus and epigenetic studies, will be required to determine whether genetic variation at the *SNCA* locus can differentiate between MSA and PD. Moreover, efforts to perform a GWA study in MSA are now under way to gain more insights into the genetic risk factors underlying MSA.

PRKN and PINK1 Screening in PD and Normal Controls

We screened 250 early-onset PD cases and 276 normal controls for mutations in *PRKN* and *PINK1*. We identified 41 sequence variants, including eight novel and 33 previously described variants. At least three of these novel variants (p.A38VfsX6, p.Q171X, p.A427E) are likely to be pathogenic. In addition, we observed eight cases and three controls with heterozygous copy number mutations. Our analysis showed that heterozygous pathogenic mutations in *PRKN* or *PINK1* were more common in cases than in controls, but the difference was not statistically significant. Further cell-based studies and functional imaging studies are required to investigate the effects of each observed sequence variant on dopaminergic dysfunction and cell death.

PRKN and PINK1 Screening in MSA

We screened 87 pathology-proven MSA cases for mutations in *PRKN* and *PINK1*. We observed four heterozygous deletions and 17 sequence variants, of which two silent variants (p.R337R and p.G189G) have not been previously reported. Comparison with results from normal control samples revealed that none of the identified variants was significantly associated with MSA or caused protein changes suggesting pathogenicity. Our data suggest that variants in *PRKN* and *PINK1* are not playing a critical role in the pathogenesis of MSA.

“Never, never, never give up.”

Winston Churchill

5 References

Web Resources

BLAST: http://blast.ncbi.nlm.nih.gov/

ClustalW2: http://www.ebi.ac.uk/Tools/clustalw2/index.html

Coriell Institute for Medical Research: http://www.coriell.org/

Database of Genomic Variants: http://projects.tcag.ca/variation/

dbSNP: http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp

Ensembl Genome Browser: http://www.ensembl.org/index.html

Entrez Nucleotide: http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore

ExonPrimer: http://ihg2.helmholtz-muenchen.de/ihg/ExonPrimer.html

ExPASy Prosite MyDomains: http://ca.expasy.org/cgi-bin/prosite/mydomains/

ENCODE Project: http://www.genome.gov/10005107

Fondation Jean Dausset CEPH: http://www.cephb.fr/

GenBank: http://www.ncbi.nlm.nih.gov/Genbank/

Haploview: http://www.broad.mit.edu/mpg/haploview/

Human Genome Mutation Database: http://www.hgmd.cf.ac.uk/ac/index.php

International HapMap Project: http://www.hapmap.org/

KORA Initiative: http://www.helmholtz-muenchen.de/kora

MapViewer: http://www.ncbi.nlm.nih.gov/mapview/

Microsat software: http://hpgl.stanford.edu/projects/microsat/

National Center for Biotechnology Information:

http://www.ncbi.nlm.nih.gov/sites/gquery

OMIM: http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim

PennCNV: http://www.neurogenome.org/cnv/penncnv/

PHYLIP: http://evolution.genetics.washington.edu/phylip.html

PLINK: http://pngu.mgh.harvard.edu/~purcell/plink/

POPGEN Survey: http://www.popgen.de

PubMed: http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed

Quanto: http://hydra.usc.edu/gxe/

R: http://www.r-project.org/

Recommendations for the description of DNA sequence variants (Human Genome Variation Society): http://www.genomic.unimelb.edu.au/mdi/mutnomen/recs-DNA.html

UCSC Genome Browser: http://www.genome.ucsc.edu/

VasserStats: http://faculty.vassar.edu/lowry/newcs.html

Articles

1. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273:1516-1517
2. Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921
3. Venter JC, Adams MD, Myers EW et al. The sequence of the human genome. Science. 2001;291:1304-1351
4. Klein RJ, Zeiss C, Chew EY et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308:385-389
5. A haplotype map of the human genome. Nature. 2005;437:1299-1320
6. Birney E, Stamatoyannopoulos JA, Dutta A et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447:799-816
7. Frazer KA, Ballinger DG, Cox DR et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:851-861
8. 1000 Genomes: a deep catalog of human genetic variation. (Accessed April 16, 2009, at http://www.1000genomes.org/page.php).
9. NIH Roadmap for Medical Research. Genotype-Tissue Expression (GTEx) project. National Institutes of Health, Bethesda, MD. (Accessed April 16, 2009, at nihroadmap.nih.gov/GTEx/).
10. The International HapMap Project. Nature. 2003;426:789-796
11. Souied EH, Leveziel N, Richard F et al. Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. Mol Vis. 2005;11:1135-1140
12. Zareparsi S, Branham KEH, Li M et al. Strong Association of the Y402H Variant in Complement Factor H at 1q32 with Susceptibility to Age-Related Macular Degeneration. The American Journal of Human Genetics. 2005;77:149-153
13. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661-678
14. Frayling TM, Timpson NJ, Weedon MN et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316:889-894

15. Herbert A, Gerry NP, McQueen MB et al. A common genetic variant is associated with adult and childhood obesity. Science. 2006;312:279-283
16. Sladek R, Rocheleau G, Rung J et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007;445:881-885
17. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet. 2005;6:109-118
18. Stacey SN, Manolescu A, Sulem P et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007;39:865-869
19. Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. Adv Genet. 2008;60:293-308
20. Price AL, Patterson NJ, Plenge RM et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904-909
21. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000;67:170-181
22. Devlin B, Bacanu SA, Roeder K. Genomic Control to the extreme. Nat Genet. 2004;36:1129-1130; author reply 1131
23. Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? Genome Res. 2006;16:702-712
24. Thompson EE, Kuttab-Boulos H, Witonsky D et al. CYP3A variation and the evolution of salt-sensitivity variants. Am J Hum Genet. 2004;75:1059-1069
25. Young JH, Chang YP, Kim JD et al. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. PLoS Genet. 2005;1:e82
26. Friedman MJ, Trager W. The biochemistry of resistance to malaria. Sci Am. 1981;244:154-155, 158-164
27. Rotter JI, Diamond JM. What maintains the frequencies of human genetic diseases? Nature. 1987;329:289-290
28. Spyropoulos B. Tay-Sachs carriers and tuberculosis resistance. Nature. 1988;331:666
29. Spyropoulos B, Moens PB, Davidson J, Lowden JA. Heterozygote advantage in Tay-Sachs carriers? Am J Hum Genet. 1981;33:375-380
30. Kidd JR, Pakstis AJ, Zhao H et al. Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. Am J Hum Genet. 2000;66:1882-1899
31. Kidd KK, Morar B, Castiglione CM et al. A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus. Hum Genet. 1998;103:211-227
32. Osier MV, Pakstis AJ, Soodyall H et al. A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. Am J Hum Genet. 2002;71:84-99
33. Payne R, Feldman M, Cann H, Bodmer JG. A comparison of HLA data of the North American black with African black and North American caucasoid populations. Tissue Antigens. 1977;9:135-147
34. Cavalli-Sforza LL, Feldman MW. The application of molecular genetic approaches to the study of human evolution. Nat Genet. 2003;33 Suppl:266-275
35. Cann HM, de Toma C, Cazes L et al. A human genome diversity cell line panel. Science. 2002;296:261-262
36. Cavalli-Sforza LL. Diversity project takes time but reaps rewards. Nature. 2004;428:467

37. Council for International Organizations of Medical Sciences. International ethical guidelines for biomedical research involving human subjects, National Research Council. Available from CIOMS, 1993:1-63
38. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. Nat Rev Genet. 2005;6:333-340
39. Fahn S, Bressman S. Dystonia. Merritt's Neurology, Lippincott Williams & Wilkins. 2005:816-826
40. Ozelius LJ, Hewett JW, Page CE et al. The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. Nat Genet. 1997;17:40-48
41. Hjermind LE, Werdelin LM, Sorensen SA. Inherited and de novo mutations in sporadic cases of DYT1-dystonia. Eur J Hum Genet. 2002;10:213-216
42. Ikeuchi T, Nomura Y, Segawa M et al. Multiple founder effects in Japanese families with primary torsion dystonia harboring the GAG deletion in the Tor1A (DYT1) gene. Neurogenetics. 2002;4:105-106
43. Kamm C, Castelon-Konkiewitz E, Naumann M et al. GAG deletion in the DYT1 gene in early limb-onset idiopathic torsion dystonia in Germany. Mov Disord. 1999;14:681-683
44. Leung JC, Klein C, Friedman J et al. Novel mutation in the TOR1A (DYT1) gene in atypical early onset dystonia and polymorphisms in dystonia and early onset parkinsonism. Neurogenetics. 2001;3:133-143
45. Kabakci K, Hedrich K, Leung JC et al. Mutations in DYT1: extension of the phenotypic and mutational spectrum. Neurology. 2004;62:395-400
46. Risch NJ, Bressman SB, Senthil G, Ozelius LJ. Intragenic Cis and Trans modification of genetic susceptibility in DYT1 torsion dystonia. Am J Hum Genet. 2007;80:1188-1193
47. Makino S, Kaji R, Ando S et al. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. Am J Hum Genet. 2007;80:393-406
48. Ichinose H, Ohye T, Takahashi E et al. Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase I gene. Nat Genet. 1994;8:236-242
49. Ludecke B, Dworniczak B, Bartholome K. A point mutation in the tyrosine hydroxylase gene associated with Segawa's syndrome. Hum Genet. 1995;95:123-125
50. Verbeek MM, Steenbergen-Spanjers GC, Willemsen MA et al. Mutations in the cyclic adenosine monophosphate response element of the tyrosine hydroxylase gene. Ann Neurol. 2007;62:422-426
51. Fuchs T, Gavarini S, Saunders-Pullman R et al. Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nat Genet. 2009;41:286-288
52. Bressman SB, Raymond D, Fuchs T et al. Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study. Lancet Neurol. 2009;8:441-446
53. Djarmati A, Schneider SA, Lohmann K et al. Mutations in THAP1 (DYT6) and generalised dystonia with prominent spasmodic dysphonia: a genetic screening study. Lancet Neurol. 2009;8:447-452
54. Saunders-Pullman R, Raymond D, Senthil G et al. Narrowing the DYT6 dystonia region and evidence for locus heterogeneity in the Amish-Mennonites. Am J Med Genet A. 2007;143A:2098-2105
55. Fink JK, Rainer S, Wilkowski J et al. Paroxysmal dystonic choreoathetosis: tight linkage to chromosome 2q. Am J Hum Genet. 1996;59:140-145
56. Matsuo H, Kamakura K, Saito M et al. Familial paroxysmal dystonic choreoathetosis: clinical findings in a large Japanese family and genetic linkage to 2q. Arch Neurol. 1999;56:721-726

57. Rainier S, Thomas D, Tokarz D et al. Myofibrillogenesis regulator 1 gene mutations cause paroxysmal dystonic choreoathetosis. Arch Neurol. 2004;61:1025-1029
58. Djarmati A, Svetel M, Momcilovic D et al. Significance of recurrent mutations in the myofibrillogenesis regulator 1 gene. Arch Neurol. 2005;62:1641
59. Lee HY, Xu Y, Huang Y et al. The gene for paroxysmal non-kinesigenic dyskinesia encodes an enzyme in a stress response pathway. Hum Mol Genet. 2004;13:3161-3170
60. Chen DH, Matsushita M, Rainier S et al. Presence of alanine-to-valine substitutions in myofibrillogenesis regulator 1 in paroxysmal nonkinesigenic dyskinesia: confirmation in 2 kindreds. Arch Neurol. 2005;62:597-600
61. Gasser T. Inherited myoclonus-dystonia syndrome. Adv Neurol. 1998;78:325-334
62. Zimprich A, Grabowski M, Asmus F et al. Mutations in the gene encoding epsilon-sarcoglycan cause myoclonus-dystonia syndrome. Nat Genet. 2001;29:66-69
63. de Carvalho Aguiar P, Sweadner KJ, Penniston JT et al. Mutations in the Na+/K+ -ATPase alpha3 gene ATP1A3 are associated with rapid-onset dystonia parkinsonism. Neuron. 2004;43:169-175
64. Brashear A, Farlow MR, Butler IJ et al. Variable phenotype of rapid-onset dystonia-parkinsonism. Mov Disord. 1996;11:151-156
65. Weber YG, Storch A, Wuttke TV et al. GLUT1 mutations are a cause of paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak. J Clin Invest. 2008;118:2157-2168
66. Scholz S, Singleton A. Susceptibility genes in movement disorders. Mov Disord. 2008;23:927-934
67. Twelves D, Perkins KS, Counsell C. Systematic review of incidence studies of Parkinson's disease. Mov Disord. 2003;18:19-31
68. Braak H, Del Tredici K. Invited Article: Nervous system pathology in sporadic Parkinson disease. Neurology. 2008;70:1916-1925
69. Langston JW, Ballard P, Tetrud JW, Irwin I. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. Science. 1983;219:979-980
70. Veldman BA, Wijn AM, Knoers N et al. Genetic and environmental risk factors in Parkinson's disease. Clin Neurol Neurosurg. 1998;100:15-26
71. Duvoisin RC, Eldridge R, Williams A et al. Twin study of Parkinson disease. Neurology. 1981;31:77-80
72. Marttila RJ, Kaprio J, Koskenvuo M, Rinne UK. Parkinson's disease in a nationwide twin cohort. Neurology. 1988;38:1217-1219
73. Ward CD, Duvoisin RC, Ince SE et al. Parkinson's disease in 65 pairs of twins and in a set of quadruplets. Neurology. 1983;33:815-824
74. Laihinen A, Ruottinen H, Rinne JO et al. Risk for Parkinson's disease: twin studies for the detection of asymptomatic subjects using [18F]6-fluorodopa PET. J Neurol. 2000;247 Suppl 2:II110-113
75. Piccini P, Burn DJ, Ceravolo R et al. The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. Ann Neurol. 1999;45:577-582
76. Polymeropoulos MH, Lavedan C, Leroy E et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science. 1997;276:2045-2047
77. Polymeropoulos MH, Higgins JJ, Golbe LI et al. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. Science. 1996;274:1197-1199
78. Kruger R, Kuhn W, Muller T et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet. 1998;18:106-108

79. Zarranz JJ, Alegre J, Gomez-Esteban JC et al. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol. 2004;55:164-173
80. Spillantini MG, Schmidt ML, Lee VM et al. Alpha-synuclein in Lewy bodies. Nature. 1997;388:839-840
81. Fuchs J, Nilsson C, Kachergus J et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology. 2007;68:916-922
82. Singleton AB, Farrer M, Johnson J et al. alpha-Synuclein locus triplication causes Parkinson's disease. Science. 2003;302:841
83. Farrer M, Maraganore DM, Lockhart P et al. alpha-Synuclein gene haplotypes are associated with Parkinson's disease. Hum Mol Genet. 2001;10:1847-1851
84. Holzmann C, Kruger R, Saecker AM et al. Polymorphisms of the alpha-synuclein promoter: expression analyses and association studies in Parkinson's disease. J Neural Transm. 2003;110:67-76
85. Maraganore DM, de Andrade M, Elbaz A et al. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. Jama. 2006;296:661-670
86. Kitada T, Asakawa S, Hattori N et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature. 1998;392:605-608
87. Matsumine H, Saito M, Shimoda-Matsubayashi S et al. Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. Am J Hum Genet. 1997;60:588-596
88. McAvoy S, Zhu Y, Perez DS et al. Disabled-1 is a large common fragile site gene, inactivated in multiple cancers. Genes Chromosomes Cancer. 2008;47:165-174
89. Abbas N, Lucking CB, Ricard S et al. A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. French Parkinson's Disease Genetics Study Group and the European Consortium on Genetic Susceptibility in Parkinson's Disease. Hum Mol Genet. 1999;8:567-574
90. Hedrich K, Marder K, Harris J et al. Evaluation of 50 probands with early-onset Parkinson's disease for Parkin mutations. Neurology. 2002;58:1239-1246
91. Lucking CB, Durr A, Bonifati V et al. Association between early-onset Parkinson's disease and mutations in the parkin gene. N Engl J Med. 2000;342:1560-1567
92. Healy DG, Abou-Sleiman PM, Casas JP et al. UCHL-1 is not a Parkinson's disease susceptibility gene. Ann Neurol. 2006;59:627-633
93. Lincoln S, Vaughan J, Wood N et al. Low frequency of pathogenic mutations in the ubiquitin carboxy-terminal hydrolase gene in familial Parkinson's disease. Neuroreport. 1999;10:427-429
94. Valente EM, Bentivoglio AR, Dixon PH et al. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36. Am J Hum Genet. 2001;68:895-900
95. Valente EM, Brancati F, Ferraris A et al. PARK6-linked parkinsonism occurs in several European families. Ann Neurol. 2002;51:14-18
96. Valente EM, Abou-Sleiman PM, Caputo V et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science. 2004;304:1158-1160
97. Bonifati V, Rohe CF, Breedveld GJ et al. Early-onset parkinsonism associated with PINK1 mutations: frequency, genotypes, and phenotypes. Neurology. 2005;65:87-95
98. Rogaeva E, Johnson J, Lang AE et al. Analysis of the PINK1 gene in a large cohort of cases with Parkinson disease. Arch Neurol. 2004;61:1898-1904
99. Bonifati V, Rizzu P, van Baren MJ et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science. 2003;299:256-259

100. Abou-Sleiman PM, Healy DG, Quinn N et al. The role of pathogenic DJ-1 mutations in Parkinson's disease. Ann Neurol. 2003;54:283-286
101. Hering R, Strauss KM, Tao X et al. Novel homozygous p.E64D mutation in DJ1 in early onset Parkinson disease (PARK7). Hum Mutat. 2004;24:321-329
102. Funayama M, Hasegawa K, Kowa H et al. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann Neurol. 2002;51:296-301
103. Zimprich A, Muller-Myhsok B, Farrer M 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;74:11-19
104. Paisan-Ruiz C, Jain S, Evans EW et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron. 2004;44:595-600
105. Zimprich A, Biskup S, Leitner P et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. 2004;44:601-607
106. Gilks WP, Abou-Sleiman PM, Gandhi S et al. A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet. 2005;365:415-416
107. Healy DG, Falchi M, O'Sullivan SS et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol. 2008;7:583-590
108. Johnson J, Paisan-Ruiz C, Lopez G et al. Comprehensive screening of a North American Parkinson's disease cohort for LRRK2 mutation. Neurodegener Dis. 2007;4:386-391
109. Lesage S, Durr A, Tazir M et al. LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. N Engl J Med. 2006;354:422-423
110. Ozelius LJ, Senthil G, Saunders-Pullman R et al. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. N Engl J Med. 2006;354:424-425
111. Di Fonzo A, Wu-Chou YH, Lu CS et al. A common missense variant in the LRRK2 gene, Gly2385Arg, associated with Parkinson's disease risk in Taiwan. Neurogenetics. 2006;7:133-138
112. Tan EK, Zhao Y, Skipper L et al. The LRRK2 Gly2385Arg variant is associated with Parkinson's disease: genetic and functional evidence. Hum Genet. 2007;120:857-863
113. Ross OA, Wu YR, Lee MC et al. Analysis of Lrrk2 R1628P as a risk factor for Parkinson's disease. Ann Neurol. 2008;64:88-92
114. Ramirez A, Heimbach A, Grundemann J et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nat Genet. 2006;38:1184-1191
115. Morgan NV, Westaway SK, Morton JE et al. PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. Nat Genet. 2006;38:752-754
116. Paisan-Ruiz C, Bhatia KP, Li A et al. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. Ann Neurol. 2009;65:19-23
117. Shojaee S, Sina F, Banihosseini SS et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. Am J Hum Genet. 2008;82:1375-1384
118. Di Fonzo A, Dekker MC, Montagna P et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology. 2009;72:240-245
119. Martins LM, Morrison A, Klupsch K et al. Neuroprotective role of the Reaper-related serine protease HtrA2/Omi revealed by targeted deletion in mice. Mol Cell Biol. 2004;24:9848-9862

120. Strauss KM, Martins LM, Plun-Favreau H et al. Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. Hum Mol Genet. 2005;14:2099-2111
121. Ross OA, Soto AI, Vilarino-Guell C et al. Genetic variation of Omi/HtrA2 and Parkinson's disease. Parkinsonism Relat Disord. 2008;14:539-543
122. Simon-Sanchez J, Singleton AB. Sequencing analysis of OMI/HTRA2 shows previously reported pathogenic mutations in neurologically normal controls. Hum Mol Genet. 2008;17:1988-1993
123. Bender A, Krishnan KJ, Morris CM et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. Nat Genet. 2006;38:515-517
124. Kraytsberg Y, Kudryavtseva E, McKee AC et al. Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. Nat Genet. 2006;38:518-520
125. Davidzon G, Greene P, Mancuso M et al. Early-onset familial parkinsonism due to POLG mutations. Ann Neurol. 2006;59:859-862
126. Hudson G, Chinnery PF. Mitochondrial DNA polymerase-gamma and human disease. Hum Mol Genet. 2006;15 Spec No 2:R244-252
127. Luoma P, Melberg A, Rinne JO et al. Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: clinical and molecular genetic study. Lancet. 2004;364:875-882
128. Mancuso M, Filosto M, Oh SJ, DiMauro S. A novel polymerase gamma mutation in a family with ophthalmoplegia, neuropathy, and Parkinsonism. Arch Neurol. 2004;61:1777-1779
129. Luoma PT, Eerola J, Ahola S et al. Mitochondrial DNA polymerase gamma variants in idiopathic sporadic Parkinson disease. Neurology. 2007;69:1152-1159
130. Le W, Conneely OM, He Y et al. Reduced Nurr1 expression increases the vulnerability of mesencephalic dopamine neurons to MPTP-induced injury. J Neurochem. 1999;73:2218-2221
131. Le W, Conneely OM, Zou L et al. Selective agenesis of mesencephalic dopaminergic neurons in Nurr1-deficient mice. Exp Neurol. 1999;159:451-458
132. Saucedo-Cardenas O, Quintana-Hau JD, Le WD et al. Nurr1 is essential for the induction of the dopaminergic phenotype and the survival of ventral mesencephalic late dopaminergic precursor neurons. Proc Natl Acad Sci U S A. 1998;95:4013-4018
133. Zetterstrom RH, Solomin L, Jansson L et al. Dopamine neuron agenesis in Nurr1-deficient mice. Science. 1997;276:248-250
134. Zetterstrom RH, Williams R, Perlmann T, Olson L. Cellular expression of the immediate early transcription factors Nurr1 and NGFI-B suggests a gene regulatory role in several brain regions including the nigrostriatal dopamine system. Brain Res Mol Brain Res. 1996;41:111-120
135. Le WD, Xu P, Jankovic J et al. Mutations in NR4A2 associated with familial Parkinson disease. Nat Genet. 2003;33:85-89
136. Healy DG, Abou-Sleiman PM, Ahmadi KR et al. NR4A2 genetic variation in sporadic Parkinson's disease: a genewide approach. Mov Disord. 2006;21:1960-1963
137. Karamohamed S, Golbe LI, Mark MH et al. Absence of previously reported variants in the SCNA (G88C and G209A), NR4A2 (T291D and T245G) and the DJ-1 (T497C) genes in familial Parkinson's disease from the GenePD study. Mov Disord. 2005;20:1188-1191
138. Nichols WC, Uniacke SK, Pankratz N et al. Evaluation of the role of Nurr1 in a large sample of familial Parkinson's disease. Mov Disord. 2004;19:649-655

139. Brady RO, Barton NW, Grabowski GA. The role of neurogenetics in Gaucher disease. Arch Neurol. 1993;50:1212-1224
140. Sidransky E. Gaucher disease and parkinsonism. Mol Genet Metab. 2005;84:302-304
141. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med. 2004;351:1972-1977
142. Neumann J, Bras J, Deas E et al. Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. Brain. 2009;13:13
143. Sato C, Morgan A, Lang AE et al. Analysis of the glucocerebrosidase gene in Parkinson's disease. Mov Disord. 2005;20:367-370
144. Spitz M, Rozenberg R, Pereira Lda V, Reis Barbosa E. Association between Parkinson's disease and glucocerebrosidase mutations in Brazil. Parkinsonism Relat Disord. 2008;14:58-62
145. Goedert M. Tau protein and neurodegeneration. Semin Cell Dev Biol. 2004;15:45-49
146. Martin ER, Scott WK, Nance MA et al. Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. Jama. 2001;286:2245-2250
147. Pastor P, Ezquerra M, Munoz E et al. Significant association between the tau gene A0/A0 genotype and Parkinson's disease. Ann Neurol. 2000;47:242-245
148. Baker M, Litvan I, Houlden H et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet. 1999;8:711-715
149. Stefansson H, Helgason A, Thorleifsson G et al. A common inversion under selection in Europeans. Nat Genet. 2005;37:129-137
150. Pankratz N, Wilk JB, Latourelle JC et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet. 2009;124:593-605
151. Ohmachi S, Mikami T, Konishi M et al. Preferential neurotrophic activity of fibroblast growth factor-20 for dopaminergic neurons through fibroblast growth factor receptor-1c. J Neurosci Res. 2003;72:436-443
152. Ohmachi S, Watanabe Y, Mikami T et al. FGF-20, a novel neurotrophic factor, preferentially expressed in the substantia nigra pars compacta of rat brain. Biochem Biophys Res Commun. 2000;277:355-360
153. van der Walt JM, Noureddine MA, Kittappa R et al. Fibroblast growth factor 20 polymorphisms and haplotypes strongly influence risk of Parkinson disease. Am J Hum Genet. 2004;74:1121-1127
154. Clarimon J, Xiromerisiou G, Eerola J et al. Lack of evidence for a genetic association between FGF20 and Parkinson's disease in Finnish and Greek patients. BMC Neurol. 2005;5:11
155. Satake W, Mizuta I, Suzuki S et al. Fibroblast growth factor 20 gene and Parkinson's disease in the Japanese population. Neuroreport. 2007;18:937-940
156. Gasser T, Muller-Myhsok B, Wszolek ZK et al. A susceptibility locus for Parkinson's disease maps to chromosome 2p13. Nat Genet. 1998;18:262-265
157. Hicks AA, Petursson H, Jonsson T et al. A susceptibility gene for late-onset idiopathic Parkinson's disease. Ann Neurol. 2002;52:549-555
158. Pankratz N, Nichols WC, Uniacke SK et al. Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations. Am J Hum Genet. 2002;71:124-135
159. Bras J, Simon-Sanchez J, Federoff M et al. Lack of replication of association between GIGYF2 variants and Parkinson disease. Hum Mol Genet. 2009;18:341-346
160. Lautier C, Goldwurm S, Durr A et al. Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease. Am J Hum Genet. 2008;82:822-833

161. Meeus B, Nuytemans K, Crosiers D et al. GIGYF2 has no major role in Parkinson genetic etiology in a Belgian population. Neurobiol Aging. 2009;23:23
162. Nichols WC, Kissell DK, Pankratz N et al. Variation in GIGYF2 is not associated with Parkinson disease. Neurology. 2009;11:11
163. Chartier-Harlin MC, Kachergus J, Roumier C et al. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. Lancet. 2004;364:1167-1169
164. Gwinn-Hardy K, Chen JY, Liu HC et al. Spinocerebellar ataxia type 2 with parkinsonism in ethnic Chinese. Neurology. 2000;55:800-805
165. Hardy J, Singleton A, Gwinn-Hardy K. Ethnic differences and disease phenotypes. Science. 2003;300:739-740
166. Rosenberg RN, Fowler HL. Autosomal dominant motor system disease of the Portuguese: a review. Neurology. 1981;31:1124-1126
167. Subramony SH, Hernandez D, Adam A et al. Ethnic differences in the expression of neurodegenerative disease: Machado-Joseph disease in Africans and Caucasians. Mov Disord. 2002;17:1068-1071
168. Tuite PJ, Rogaeva EA, St George-Hyslop PH, Lang AE. Dopa-responsive parkinsonism phenotype of Machado-Joseph disease: confirmation of 14q CAG expansion. Ann Neurol. 1995;38:684-687
169. Neudorfer O, Giladi N, Elstein D et al. Occurrence of Parkinson's syndrome in type I Gaucher disease. Qjm. 1996;89:691-694
170. Golbe LI, Lazzarini AM, Spychala JR et al. The tau A0 allele in Parkinson's disease. Mov Disord. 2001;16:442-447
171. Maraganore DM, Hernandez DG, Singleton AB et al. Case-Control study of the extended tau gene haplotype in Parkinson's disease. Ann Neurol. 2001;50:658-661
172. Gilman S, Low P, Quinn N et al. Consensus statement on the diagnosis of multiple system atrophy. American Autonomic Society and American Academy of Neurology. Clin Auton Res. 1998;8:359-362
173. Schrag A, Wenning GK, Quinn N, Ben-Shlomo Y. Survival in multiple system atrophy. Mov Disord. 2008;23:294-296
174. Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990. Neurology. 1997;49:1284-1288
175. Schrag A, Ben-Shlomo Y, Quinn NP. Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. Lancet. 1999;354:1771-1775
176. Tison F, Yekhlef F, Chrysostome V, Sourgen C. Prevalence of multiple system atrophy. Lancet. 2000;355:495-496
177. Gilman S, Wenning GK, Low PA et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology. 2008;71:670-676
178. Nee LE, Gomez MR, Dambrosia J et al. Environmental-occupational risk factors and familial associations in multiple system atrophy: a preliminary investigation. Clin Auton Res. 1991;1:9-13
179. Wenning GK, Wagner S, Daniel S, Quinn NP. Multiple system atrophy: sporadic or familial? Lancet. 1993;342:681
180. Gwinn-Hardy K, Mehta ND, Farrer M et al. Distinctive neuropathology revealed by alpha-synuclein antibodies in hereditary parkinsonism and dementia linked to chromosome 4p. Acta Neuropathol. 2000;99:663-672
181. Hara K, Momose Y, Tokiguchi S et al. Multiplex families with multiple system atrophy. Arch Neurol. 2007;64:545-551

182. Wullner U, Abele M, Schmitz-Huebsch T et al. Probable multiple system atrophy in a German family. J Neurol Neurosurg Psychiatry. 2004;75:924-925
183. Wullner U, Schmitt I, Kammal M et al. Definite multiple system atrophy in a German family. J Neurol Neurosurg Psychiatry. 2009;80:449-450
184. Litvan I, Goetz CG, Jankovic J et al. What is the accuracy of the clinical diagnosis of multiple system atrophy? A clinicopathologic study. Arch Neurol. 1997;54:937-944
185. Osaki Y, Wenning GK, Daniel SE et al. Do published criteria improve clinical diagnostic accuracy in multiple system atrophy? Neurology. 2002;59:1486-1491
186. Furiya Y, Hirano M, Kurumatani N et al. Alpha-1-antichymotrypsin gene polymorphism and susceptibility to multiple system atrophy (MSA). Brain Res Mol Brain Res. 2005;138:178-181
187. Kim HS, Lee MS. Frequencies of single nucleotide polymorphism in alcohol dehydrogenase7 gene in patients with multiple system atrophy and controls. Mov Disord. 2003;18:1065-1067
188. Morris HR, Vaughan JR, Datta SR et al. Multiple system atrophy/progressive supranuclear palsy: alpha-Synuclein, synphilin, tau, and APOE. Neurology. 2000;55:1918-1920
189. Cairns NJ, Atkinson PF, Kovacs T et al. Apolipoprotein E e4 allele frequency in patients with multiple system atrophy. Neurosci Lett. 1997;221:161-164
190. Soma H, Yabe I, Takei A et al. Associations between multiple system atrophy and polymorphisms of SLC1A4, SQSTM1, and EIF4EBP1 genes. Mov Disord. 2008;23:1161-1167
191. Bandmann O, Sweeney MG, Daniel SE et al. Multiple-system atrophy is genetically distinct from identified inherited causes of spinocerebellar degeneration. Neurology. 1997;49:1598-1604
192. Nishimura M, Kuno S, Kaji R, Kawakami H. Brain-derived neurotrophic factor gene polymorphisms in Japanese patients with sporadic Alzheimer's disease, Parkinson's disease, and multiple system atrophy. Mov Disord. 2005;20:1031-1033
193. Nicholl DJ, Bennett P, Hiller L et al. A study of five candidate genes in Parkinson's disease and related neurodegenerative disorders. European Study Group on Atypical Parkinsonism. Neurology. 1999;53:1415-1421
194. Plante-Bordeneuve V, Bandmann O, Wenning G et al. CYP2D6-debrisoquine hydroxylase gene polymorphism in multiple system atrophy. Mov Disord. 1995;10:277-278
195. Cho S, Kim CH, Cubells JF et al. Variations in the dopamine beta-hydroxylase gene are not associated with the autonomic disorders, pure autonomic failure, or multiple system atrophy. Am J Med Genet A. 2003;120:234-236
196. Lim SY, Wadia P, Wenning GK, Lang AE. Clinically probable multiple system atrophy with predominant parkinsonism associated with myotonic dystrophy type 2. Mov Disord. 2009;13:13
197. Tan EK, Zhao Y, Puong KY et al. Fragile X premutation alleles in SCA, ET, and parkinsonism in an Asian cohort. Neurology. 2004;63:362-363
198. Garland EM, Vnencak-Jones CL, Biaggioni I et al. Fragile X gene premutation in multiple system atrophy. J Neurol Sci. 2004;227:115-118
199. Yabe I, Soma H, Takei A et al. No association between FMR1 premutations and multiple system atrophy. J Neurol. 2004;251:1411-1412
200. Kamm C, Healy DG, Quinn NP et al. The fragile X tremor ataxia syndrome in the differential diagnosis of multiple system atrophy: data from the EMSA Study Group. Brain. 2005;128:1855-1860

201. Bannister R, Mowbray J, Sidgwick A. Genetic control of progressive autonomic failure: evidence for an association with an HLA antigen. Lancet. 1983;1:1017
202. Nee LE, Brown RT, Polinsky RJ. HLA in autonomic failure. Arch Neurol. 1989;46:758-759
203. Combarros O, Infante J, Llorca J, Berciano J. Interleukin-1A (-889) genetic polymorphism increases the risk of multiple system atrophy. Mov Disord. 2003;18:1385-1386
204. Nishimura M, Kawakami H, Komure O et al. Contribution of the interleukin-1beta gene polymorphism in multiple system atrophy. Mov Disord. 2002;17:808-811
205. Nishimura M, Kuno S, Kaji R, Kawakami H. Influence of a tumor necrosis factor gene polymorphism in Japanese patients with multiple system atrophy. Neurosci Lett. 2005;374:218-221
206. Infante J, Llorca J, Berciano J, Combarros O. Interleukin-8, intercellular adhesion molecule-1 and tumour necrosis factor-alpha gene polymorphisms and the risk for multiple system atrophy. J Neurol Sci. 2005;228:11-13
207. Ozelius LJ, Foroud T, May S et al. G2019S mutation in the leucine-rich repeat kinase 2 gene is not associated with multiple system atrophy. Mov Disord. 2007;22:546-549
208. Tan EK, Skipper L, Chua E et al. Analysis of 14 LRRK2 mutations in Parkinson's plus syndromes and late-onset Parkinson's disease. Mov Disord. 2006;21:997-1001
209. Zhang J, Montine TJ, Smith MA et al. The mitochondrial common deletion in Parkinson's disease and related movement disorders. Parkinsonism Relat Disord. 2002;8:165-170
210. Benussi L, Ghidoni R, Pegoiani E et al. Progranulin Leu271LeufsX10 is one of the most common FTLD and CBS associated mutations worldwide. Neurobiol Dis. 2009;33:379-385
211. Ozawa T, Takano H, Onodera O et al. No mutation in the entire coding region of the alpha-synuclein gene in pathologically confirmed cases of multiple system atrophy. Neurosci Lett. 1999;270:110-112
212. Chen R, Forno L, DiMonte D et al. Mutation Screening in the a-Synuclein Gene in MSA. Parkinsonism Relat Disord. 1999;5:S28
213. Lincoln SJ, Ross OA, Milkovic NM et al. Quantitative PCR-based screening of alpha-synuclein multiplication in multiple system atrophy. Parkinsonism Relat Disord. 2007;13:340-342
214. Ozawa T, Okuizumi K, Ikeuchi T et al. Analysis of the expression level of alpha-synuclein mRNA using postmortem brain samples from pathologically confirmed cases of multiple system atrophy. Acta Neuropathol. 2001;102:188-190
215. Vogt IR, Lees AJ, Evert BO et al. Transcriptional changes in multiple system atrophy and Parkinson's disease putamen. Exp Neurol. 2006;199:465-478
216. Ozawa T, Healy DG, Abou-Sleiman PM et al. The alpha-synuclein gene in multiple system atrophy. J Neurol Neurosurg Psychiatry. 2006;77:464-467
217. Healy DG, Abou-Sleiman PM, Quinn N et al. UCHL-1 gene in multiple system atrophy: a haplotype tagging approach. Mov Disord. 2005;20:1338-1343
218. Jakobsson M, Scholz SW, Scheet P et al. Genotype, haplotype and copy-number variation in worldwide human populations. Nature. 2008;451:998-1003
219. Steemers FJ, Gunderson KL. Whole genome genotyping technologies on the BeadArray platform. Biotechnol J. 2007;2:41-49
220. Rosenberg NA. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet. 2006;70:841-847
221. Rosenberg NA, Pritchard JK, Weber JL et al. Genetic structure of human populations. Science. 2002;298:2381-2385

222. Conrad DF, Jakobsson M, Coop G et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat Genet. 2006;38:1251-1260
223. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 2006;78:629-644
224. Wang K, Li M, Hadley D et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17:1665-1674
225. Simon-Sanchez J, Scholz S, Fung HC et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet. 2007;16:1-14
226. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes. 2004;4:137-138
227. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23:1801-1806
228. Ramachandran S, Deshpande O, Roseman CC et al. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci U S A. 2005;102:15942-15947
229. McVean GA. A genealogical interpretation of linkage disequilibrium. Genetics. 2002;162:987-991
230. Reich DE, Cargill M, Bolk S et al. Linkage disequilibrium in the human genome. Nature. 2001;411:199-204
231. Bayless TM, Rosensweig NS. A racial difference in incidence of lactase deficiency. A survey of milk intolerance and lactase deficiency in healthy adult males. Jama. 1966;197:968-972
232. Wang Y, Harvey CB, Pratt WS et al. The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element. Hum Mol Genet. 1995;4:657-662
233. Zhang J, Feuk L, Duggan GE et al. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet Genome Res. 2006;115:205-214
234. Wang S, Lewis CM, Jakobsson M et al. Genetic variation and population structure in native Americans. PLoS Genet. 2007;3:e185
235. Li M, Li C, Guan W. Evaluation of coverage variation of SNP chips for genome-wide association studies. Eur J Hum Genet. 2008;16:635-643
236. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006;38:659-662
237. Bamshad M, Wooding S, Salisbury BA, Stephens JC. Deconstructing the relationship between genetics and race. Nat Rev Genet. 2004;5:598-609
238. Bastos-Rodrigues L, Pimenta JR, Pena SD. The genetic structure of human populations studied through short insertion-deletion polymorphisms. Ann Hum Genet. 2006;70:658-665
239. Rosenberg NA, Mahajan S, Ramachandran S et al. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. 2005;1:e70
240. Bowcock AM, Kidd JR, Mountain JL et al. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci U S A. 1991;88:839-843

241. Jorde LB, Watkins WS, Bamshad MJ et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet. 2000;66:979-988
242. Shriver MD, Kennedy GC, Parra EJ et al. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics. 2004;1:274-286
243. Watkins WS, Rogers AR, Ostler CT et al. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res. 2003;13:1607-1618
244. Lewin R. The Origin of Modern Humans. Scientific American Library, W.H. Freeman and Company, NY. 1993
245. Harvey CB, Hollox EJ, Poulter M et al. Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. Ann Hum Genet. 1998;62:215-223
246. Harvey CB, Pratt WS, Islam I et al. DNA polymorphisms in the lactase gene. Linkage disequilibrium across the 70-kb region. Eur J Hum Genet. 1995;3:27-41
247. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008;455:237-241
248. Aitman TJ, Dong R, Vyse TJ et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature. 2006;439:851-855
249. Diskin SJ, Hou C, Glessner JT et al. Copy number variation at 1q21.1 associated with neuroblastoma. Nature. 2009;459:987-991
250. Fanciulli M, Norsworthy PJ, Petretto E et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet. 2007;39:721-723
251. Hollox EJ, Huffmeier U, Zeeuwen PL et al. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet. 2008;40:23-25
252. Sebat J, Lakshmi B, Malhotra D et al. Strong association of de novo copy number mutations with autism. Science. 2007;316:445-449
253. Stefansson H, Rujescu D, Cichon S et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;455:232-236
254. Walsh T, McClellan JM, McCarthy SE et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008;320:539-543
255. Camargos S, Scholz S, Simon-Sanchez J et al. DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA. Lancet Neurol. 2008;7:207-215
256. Dogu O, Johnson J, Hernandez D et al. A consanguineous Turkish family with early-onset Parkinson's disease and an exon 4 parkin deletion. Mov Disord. 2004;19:812-816
257. Johnson J, Hague SM, Hanson M et al. SNCA multiplication is not a common cause of Parkinson disease or dementia with Lewy bodies. Neurology. 2004;63:554-556
258. Purcell S, Neale B, Todd-Brown K et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-575
259. Patel CV, Handy I, Goldsmith T, Patel RC. PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR. J Biol Chem. 2000;275:37993-37998
260. D'Acquisto F, Ghosh S. PACT and PKR: turning on NF-kappa B in the absence of virus. Sci STKE. 2001;2001:RE1
261. Lee Y, Hur I, Park SY et al. The role of PACT in the RNA silencing pathway. Embo J. 2006;25:522-532

262. Rowe TM, Rizzi M, Hirose K et al. A role of the double-stranded RNA-binding protein PACT in mouse ear development and hearing. Proc Natl Acad Sci U S A. 2006;103:5823-5828
263. Seibler P, Djarmati A, Langpap B et al. A heterozygous frameshift mutation in PRKRA (DYT16) associated with generalised dystonia in a German patient. Lancet Neurol. 2008;7:380-381
264. Fung HC, Scholz S, Matarin M et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol. 2006;5:911-916
265. Maraganore DM, de Andrade M, Lesnick TG et al. High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet. 2005;77:685-693
266. Satake W, Nakabayashi Y, Mizuta I et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet. 2009;41:1261-1262
267. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry. 1992;55:181-184
268. Hyten DL, Song Q, Choi IY et al. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet. 2008;116:945-952
269. Steemers FJ, Gunderson KL. Illumina, Inc. Pharmacogenomics. 2005;6:777-782
270. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263-265
271. Gabriel SB, Schaffner SF, Nguyen H et al. The structure of haplotype blocks in the human genome. Science. 2002;296:2225-2229
272. Simon-Sanchez J, Schulte C, Bras JM et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet. 2009;41:1308-1312
273. Cronin KD, Ge D, Manninger P et al. Expansion of the Parkinson Disease-Associated SNCA-Rep1 Allele Up-Regulates Human {alpha}-Synuclein in Transgenic Mouse Brain. Hum Mol Genet. 2009;4:4
274. Pankratz N, Wilk JB, Latourelle JC et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet. 2008
275. Tobin JE, Latourelle JC, Lew MF et al. Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study. Neurology. 2008;71:28-34
276. Vandrovcova J, Pittman AM, Malzer E et al. Association of MAPT haplotype-tagging SNPs with sporadic Parkinson's disease. Neurobiol Aging. 2007
277. Myers AJ, Gibbs JR, Webster JA et al. A survey of genetic human cortical gene expression. Nat Genet. 2007;39:1494-1499
278. Arima K, Hirai S, Sunohara N et al. Cellular co-localization of phosphorylated tau- and NACP/alpha-synuclein-epitopes in lewy bodies in sporadic Parkinson's disease and in dementia with Lewy bodies. Brain Res. 1999;843:53-61
279. Duda JE, Giasson BI, Mabon ME et al. Concurrence of alpha-synuclein and tau brain pathology in the Contursi kindred. Acta Neuropathol. 2002;104:7-11
280. Giasson BI, Forman MS, Higuchi M et al. Initiation and synergistic fibrillization of tau and alpha-synuclein. Science. 2003;300:636-640
281. Goris A, Williams-Gray CH, Clark GR et al. Tau and alpha-synuclein in susceptibility to, and dementia in, Parkinson's disease. Ann Neurol. 2007;62:145-153
282. Ishizawa T, Mattila P, Davies P et al. Colocalization of tau and alpha-synuclein epitopes in Lewy bodies. J Neuropathol Exp Neurol. 2003;62:389-397

283. Mamah CE, Lesnick TG, Lincoln SJ et al. Interaction of alpha-synuclein and tau genotypes in Parkinson's disease. Ann Neurol. 2005;57:439-443
284. Clarimon J, Scholz S, Fung HC et al. Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. Am J Hum Genet. 2006;78:1082-1084; author reply 1092-1084
285. Goris A, Williams-Gray CH, Foltynie T et al. No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am J Hum Genet. 2006;78:1088-1090; author reply 1092-1084
286. Li Y, Rowland C, Schrodi S et al. A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. Am J Hum Genet. 2006;78:1090-1092; author reply 1092-1094
287. Spillantini MG, Crowther RA, Jakes R et al. Filamentous alpha-synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. Neurosci Lett. 1998;251:205-208
288. Wakabayashi K, Yoshimoto M, Tsuji S, Takahashi H. Alpha-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. Neurosci Lett. 1998;249:180-182
289. Scholz SW, Houlden H, Schulte C et al. SNCA variants are associated with increased risk for multiple system atrophy. Ann Neurol. 2009;65:610-614
290. Langerveld AJ, Mihalko D, DeLong C et al. Gene expression changes in postmortem tissue from the rostral pons of multiple system atrophy patients. Mov Disord. 2007;22:766-777
291. Singleton A, Myers A, Hardy J. The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases. Hum Mol Genet. 2004;13:123-126
292. Human Gene Mutation Database (HGMD) Professional 2009.2. (Accessed July 30, 2009, at http://www.hgmd.cf.ac.uk/ac/index.php).
293. Klein C, Lohmann-Hedrich K, Rogaeva E et al. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. Lancet Neurol. 2007;6:652-662
294. Farrer M, Chan P, Chen R et al. Lewy bodies and parkinsonism in families with parkin mutations. Ann Neurol. 2001;50:293-300
295. Foroud T, Uniacke SK, Liu L et al. Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. Neurology. 2003;60:796-801
296. Klein C, Pramstaller PP, Kis B et al. Parkin deletions in a family with adult-onset, tremor-dominant parkinsonism: expanding the phenotype. Ann Neurol. 2000;48:65-71
297. Sun M, Latourelle JC, Wooten GF et al. Influence of heterozygosity for parkin mutation on onset age in familial Parkinson disease: the GenePD study. Arch Neurol. 2006;63:826-832
298. Brooks J, Ding J, Simon-Sanchez J et al. Parkin and PINK1 mutations in early-onset Parkinson's disease: comprehensive screening in publicly available cases and control. J Med Genet. 2009;46:375-381
299. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402-408
300. Wang C, Lu R, Ouyang X et al. Drosophila overexpressing parkin R275W mutant exhibits dopaminergic neuron degeneration and mitochondrial abnormalities. J Neurosci. 2007;27:8563-8570
301. Hertz JM, Ostergaard K, Juncker I et al. Low frequency of Parkin, Tyrosine Hydroxylase, and GTP Cyclohydrolase I gene mutations in a Danish population of early-onset Parkinson's Disease. Eur J Neurol. 2006;13:385-390

302. Bertoli-Avella AM, Giroud-Benitez JL, Akyol A et al. Novel parkin mutations detected in patients with early-onset Parkinson's disease. Mov Disord. 2005;20:424-431
303. Kay DM, Moran D, Moses L et al. Heterozygous parkin point mutations are as common in control subjects as in Parkinson's patients. Ann Neurol. 2007;61:47-54
304. Hedrich K, Kann M, Lanthaler AJ et al. The importance of gene dosage studies: mutational analysis of the parkin gene in early-onset parkinsonism. Hum Mol Genet. 2001;10:1649-1656
305. Oliveira SA, Scott WK, Martin ER et al. Parkin mutations and susceptibility alleles in late-onset Parkinson's disease. Ann Neurol. 2003;53:624-629
306. Clark LN, Afridi S, Karlins E et al. Case-control study of the parkin gene in early-onset Parkinson disease. Arch Neurol. 2006;63:548-552
307. Satoh J, Kuroda Y. Association of codon 167 Ser/Asn heterozygosity in the parkin gene with sporadic Parkinson's disease. Neuroreport. 1999;10:2735-2739
308. Mellick GD, Siebert GA, Funayama M et al. Screening PARK genes for mutations in early-onset Parkinson's disease patients from Queensland, Australia. Parkinsonism Relat Disord. 2008
309. Valente EM, Salvi S, Ialongo T et al. PINK1 mutations are associated with sporadic early-onset parkinsonism. Ann Neurol. 2004;56:336-341
310. Klein C, Djarmati A, Hedrich K et al. PINK1, Parkin, and DJ-1 mutations in Italian patients with early-onset parkinsonism. Eur J Hum Genet. 2005;13:1086-1093
311. Cherif W, Ben Turkia H, Tebib N et al. [Mutation spectrum of Gaucher disease in Tunisia: high frequency of N370S/Rec NciI compound heterozygous]. Arch Inst Pasteur Tunis. 2007;84:65-70
312. Clark LN, Ross BM, Wang Y et al. Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. Neurology. 2007;69:1270-1277
313. Fung HC, Evans J, Evans W et al. The architecture of the tau haplotype block in different ethnicities. Neurosci Lett. 2005;377:81-84
314. Binkofski F, Reetz K, Gaser C et al. Morphometric fingerprint of asymptomatic Parkin and PINK1 mutation carriers in the basal ganglia. Neurology. 2007;69:842-850
315. Fiorio M, Valente EM, Gambarin M et al. Subclinical sensory abnormalities in unaffected PINK1 heterozygotes. J Neurol. 2008
316. Hagenah JM, Becker B, Bruggemann N et al. Transcranial sonography findings in a large family with homozygous and heterozygous PINK1 mutations. J Neurol Neurosurg Psychiatry. 2008;79:1071-1074
317. Hagenah JM, Konig IR, Becker B et al. Substantia nigra hyperechogenicity correlates with clinical status and number of Parkin mutated alleles. J Neurol. 2007;254:1407-1413
318. Hilker R, Klein C, Hedrich K et al. The striatal dopaminergic deficit is dependent on the number of mutant alleles in a family with mutations in the parkin gene: evidence for enzymatic parkin function in humans. Neurosci Lett. 2002;323:50-54
319. Khan NL, Scherfler C, Graham E et al. Dopaminergic dysfunction in unrelated, asymptomatic carriers of a single parkin mutation. Neurology. 2005;64:134-136
320. Lincoln SJ, Maraganore DM, Lesnick TG et al. Parkin variants in North American Parkinson's disease: cases and controls. Mov Disord. 2003;18:1306-1311
321. Marongiu R, Ferraris A, Ialongo T et al. PINK1 heterozygous rare variants: prevalence, significance and phenotypic spectrum. Hum Mutat. 2008;29:565
322. Gandhi S, Muqit MM, Stanyer L et al. PINK1 protein in normal human brain and Parkinson's disease. Brain. 2006;129:1720-1731

323. Huang Y, Song YJ, Murphy K et al. LRRK2 and parkin immunoreactivity in multiple system atrophy inclusions. Acta Neuropathol. 2008;116:639-646
324. Murakami T, Moriwaki Y, Kawarabayashi T et al. PINK1, a gene product of PARK6, accumulates in alpha-synucleinopathy brains. J Neurol Neurosurg Psychiatry. 2007;78:653-654
325. Shimura H, Hattori N, Kubo S et al. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat Genet. 2000;25:302-305
326. Brooks JA, Houlden H, Melchers A et al. Mutational Analysis of Parkin and PINK1 in Multiple System Atrophy. Neurobiol Aging. 2010; *in press*
327. Conrad FD, Pinto D, Redon R et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010; *in press*

6 Appendix

Table S1. PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
OSBPL6			
exon 1-2	325	AGTGTCTCATTGTCTGTCTG	ATCAAGTTTCCTCTCAAATG
exon 3	300	GATTGGGATGGTCTTAAGTG	ACATCAATGAACGTACCTACAG
exon 4	260	GACTGTGTTGTGTATGATTTC	TGTAAACGTTATTCACAGTTC
exon 5	287	CTTAAGAATCTGCTTTAAGTG	TGCCATATTCTTAGTTTCC
exon 6	216	AAGCAACATTAGACCCAC	AGGCTTGTACAATTACAGTG
exon 7	309	TGTAGTATTCCACTTCCCATAG	ATAACTCTCCTGCCTACCAC
exon 8	395	GAGCACATCTATCATTTGAAGG	ATTACACACGTGAGCCACAG
exon 9	309	TTCCTCTCTTTGTTAATACG	CTCATAATTGACCTCTTCG
exon 10	270	CATAATTTGCATTGACCC	CCTTATATATTCATAGAACAACC
exon 10a	224	TTATCTTTCATCCTCCCATC	ACAGGGGAGTCATCCTTG
exon 11	263	TAACTTGCTGTGCTCTTG	AATCTAACGCCCATAAAC
exon 12	398	CAAGTGTGTTATCAGAGAGG	ATCCAACCATTTAATCTTTC
exon 13	347	CCTATTTGTGAGTGAAAGTC	CTGCAAACTGTACTATCCC
exon 14	315	GAGCTAAGAACATCACAGAC	ATTAACTGTGAAAGAACTGG
exon 15	328	TTAAAGAGTAGAAATCTGGC	AACAAATGAAAGTAATCACC
exon 16	278	CTAGGACCTCTGCTTCTC	AATGCCTGTTCATTAAGTC
exon 17	406	GAAAGTTATTTCCCTTCTTG	CATAATGAATTCTGGCTTAG
exon 18	300	ATCTAGAGGGATGAGGAAC	AACAGATTGTTATCGAAATG
exon 19	217	TTCAGTAAGAGAAATGGATG	AATACAGTGTCCCTTTGG
exon 20	279	TTCCAAAGTGTTTCATAAAG	AAAGAGAAAGAAACCCTTAC
exon 21	367	AAATGGGAGGAGAATGAG	ATACAGGACAGATTCACTTG
exon 22	315	AGGGACTTCAAGGCTAAG	ATGACCTGATGTTCTCTTC
exon 23	297	GGGTAAGTCAGCTTCTTC	AAGTGATCTTCCCTCCTC
exon 24	322	GACTCCGTCCGTTACTGTG	TAACTGGAGTTTCTCATGGC
exon 25	318	ACTAACTGGCTTACAAACACTG	GACCCAGAAACCACACATAG
PRKRA			
exon 1	312	CCTCGCTGGAGCAACGCAAG	GGCACGGCTTTACCCAGAATG
exon 2	516	TCTAAAGACCTCGCTCAC	TGAGAGGTCTCAGTTTCAG
exon 3	268	TGACTTTGTTTTGTGTATTG	AACTGTTCACTTTGTTGC
exon 4	266	GAATGACAAGAGCAAAGAC	ATTAATTCCTTGTGTTAGCC
exon 5	290	GAAATGGGATCAAAATTAAG	AATATTTGAAAACATTACGAC
exon 6	347	AAACAAAGTTATCAGGTCAG	AATCACAACTCTGAAGTAGC
exon 7	525	AATGTTGTCTTGTTTAAATTG	TACTATCCACAAGAATGGG
exon 8	408	GGTGTAGTATACCATGGAG	GAGTGTTGATGGAATCTATG
DFNB59			
exon 1	211	GTGCTATAGCCTGCCATTCC	AGGTAAGCCCAAATAATATCCC
exon 2	196	TGCCTTTCTCTAACATTTGGG	AAAATACAACCTCAATGTTTTAAGC
exon 3	142	GGATTGCCTTGATTTACTATTAGGTG	CAACTGCAGCTCTTTCATCAG
exon 4	118	TGGACCAATTGGATCTCTGC	AACGGAGGGTCTTTCTCAGC
exon 5	99	GTGGGCCAGAGACCTATTTG	CCACTAGTCATGAAATAATTTGGC
exon 6	293	GCCAAATTATTTCATGACTAGTGG	TTTAATTTAATTCATCAAAGCAAACC

table continued on page 189

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
FKBP7			
exon 1	409	CCCTTTGTGACGCAGGAC	CAACCAATGAGGCTTCCG
exon 2	544	TGGGAATGATGGTTAGAGAG	GCCTCAAGTGATTCTACCAC
exon 3			
exon 4	419	TAATATGGTTATGGCAGTTC	ATTGAACAGGATGACTCAC
exon 5	354	TAATCTTGCACATGAGTTTC	AAATAGCAAATACAACTTGG
PLEKHA3			
exon 1	181	CTGCGCCTACCCCATCAC	GAAGAGGAGCCCCGACC
exon 2	243	AATCACAGTTTTCTATGTGG	TGAAGAACCCTTGACTATC
exon 3	349	ATGAAGCATCTGATACTGTG	CCTTGGGATAGATACTAAAAG
exon 4	275	TACATTGCTCAAAGAATATG	TCATCTTATGGAATGACTG
exon 5	289	TGTATATGATTCCATGTAGTTG	AAATGAAACGAAAACCAC
exon 6	176	TTAGAGCTGCATGATTCTAC	CAAAGAACTTCCCTAGATGC
exon 7	255	TAAAATTCCAGTTGAGTGTG	AAATTACAGAGCCCCAC
exon 8	289	GAGGACTGGAGTTAATATTGAC	ATAAAGTGCAGAGTCCTAAGG
TTN			
exon 2	322	GCCCACAGATGACCTATGGAGC	AAGCAGGGCTTAAACTTGGCGTC
exon 3	430	AGTGGCATTCCATGCTGGAAAGC	AGAGCTGCATCTACCTCTGCCC
exon 4	536	GTGAGCTGGGTTCAAGGACATTC	GCATTCTTCCCAGGGCTGTGAG
exon 5-6	659	CCATTGGGATATCCAGAAAATC	GGACACTGAAGAAGCGAACC
exon 7-8	1,034	CTCTGCCTTAAAGCACTTCCAGC	GCCAAGCTCAGTGGATGGTGG
exon 9	332	GGATTTCACCATGTTGGCGAGGCTGG	AACAACCATGTGGCCCAAGGAAC
exon 10	326	TTAGTGACATCTGCAGAAAAGGGA	GACTATAAGCTACCTGCAGCTGGCTG
exon 11	353	ATCGTTCACCACTAAGCATCC	GAAGAGGTGGAAGTGAAGAAGTG
exon 12-13	917	TTGCCTCCTTTCTTTTTGACTTACGT	AAGGAATTTCACATGATATGTGG
exon 14	561	ATCTGCCTGTCCTGATGCC	GTAAAGTGAGGGCTAGCTGTGC
exon 15	307	CCTCACTGCTGTGGTCATCTG	AAACACACGCACACACACATC
exon 16-17	676	GCCTGTCCTGATGCCCAGGAG	CCAGAGATGCTCTGTTTCACAGG
exon 18-19-20	1,284	GGCTCAGACTCAGTCCAAGAGGGC	AACTTATGCATTTCAACTGCCAC
exon 21	1,073	GTGGCAGTTGAAATGCATAAGTT	GGCTCCACAATGAAAGAGCCTATGC
exon 22	398	CCTGAATGCATAGTGGTATGACTG	AACTTGGACCTTCTAATAGCTGTC
exon 23-24	711	AAGTCCTCTAGATTCTTCTCCCAG	ACCTGATTCTGCTCACTGGC
exon 25-26	835	GTCTTCTAGATTCTTCTCCCAGC	GTTGCCAATGGTAAGTTTCTGTGCC
exon 27	946	GGTGCCTTCTAATACGTGCAAGAAG	GGGATATTTATGAGGCACAATGATG
exon 28_1	506	GAGAATGGCTCATGAAGGTGC	TTTCATGGGTAATTCTTTCAGC
exon 28_2	781	TGAACCAAGGCCTGAGTTTC	TTCTCAACCTGAGGAATGGG
exon 29	282	CCAAGCAATTGATCACTTTCAC	CTTCAACAATAAGTTTAGCAGTCG
exon 30-31	546	GATGGACCTGTCTTTGTATTCAA	TTCTGGCTATGAAACTTATAGTCAA
exon 32-32	819	CAAGACATCATCTCTATCCTAAG	ATCATACACACACAAGATGAATG
exon 34	863	CGTGTCTCTGTCTATAGTGAGTC	ACAGCATAAGCAGAGGCATG
exon 35	461	CATGTCCAGTTGGATCCCAG	TCAACTTACTCTCCACGTGCAG
exon 36	271	AACCCTATCACTGAATTGCTCC	TCCAATTGCCCGACCACAG
exon 37	868	CAGAGGTGCTGCATAACTGT	GGTAAAATCAAAGAGCACTTCAG
exon 38	496	CTGTAATCTCCTCAACAGACG	GATATCCCATCGTATGGCTATC
exon 39-40	621	CTTTAAATTGGTGGATTGTACAG	CTAAGCCAAGTAAGAATGAGCTG
exon 41	463	TCACATGCCATTCATTGAC	CTGGAACCACATGAATACCATAG
exon 42-43	819	GGATGTATAACTAGCCAGCAAC	GTGATGGAGGAGAAGCTGAC

table continued on page 190

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
TTN			
exon 44	483	CTATCATTACATTTCCTTCTTGAAG	CAACAATGATTTACATGAACTCTTG
exon 45	661	ACAAGGACAAGGCAGTAGGAG	GGACCTAGCGGGAGTTATTG
exon 46	687	CTAACATGGGCAAGTGACAAAG	TGAGGATGAAATGAAGCAAGTC
exon 47	332	GGGGAATGTAGTTTTCTGTTGT	CCAAGCATGCGACATAGTAATAT
exon 48_1	677	GGATTTGTGAGATCTGTAGCATG	GAGAGCCAGTAAACCTCAGGTC
exon 48_2	788	GCAATTAATAGTGAAGGGTCTGC	GCCATAATTTCTTCCAGCTGTCC
exon 48_3	643	AATCCACATGCCCTTTCTAATC	GTGGGATTGGCATGTCATTG
exon 48_4	782	ACCACATGCCAGCGAATC	CTTGAGGCATTGCTTTAGGTTC
exon 48_5	606	TGGAACCTAAAGCAATGCC	CTCTAGAAGGTATGCAACGCAC
exon 48_6	787	CAGCTGGTCCCTTTACTCCC	TCCTGTTGTTCCCTTTCTTGTG
exon 48_7	721	GCAAAGAGAATCAAGATCAAGC	GGTTGGAAGTAGGGCACATG
exon 48_8	659	AAAGAAGTCCGACATGTACAGG	TTAGTTTCAGGAACCTCACGC
exon 48_9	694	AGGTTCCTGAAACTAAAGACACC	TGCTCATTTGGTGTACCGTC
exon 48_10	263	TGCTCATTTGGTGTACCGTC	GAGAGATACTCCACACCTCCAGG
exon 48_11	706	CTCCTCAAGAAATTGTCCTGG	GGAGAGCCACGAACTAAGCAC
exon 48_12	555	GTGTAGGAGAGGCTATCTGTCG	GAATTCAAATATGGTGGACCTG
exon 49_1	971	GCAGAAGAAGGCCATCAACTGGT	CTGCGGGACCCTTTAAGGGTGTC
exon 49_2	1,031	TCAGAGGGAAGACAGTGGCCTCT	GCTCTTCTTTGAGCAGTACCTGC
exon 49_3	787	TAACCCTAATGGCTCTGGAAC	CACTCTTTGCTCTCTGTTGGTG
exon 49_4	757	ATCTTCCCAAGTGCCATGTC	GCTCTTGGGTGATGTTTACAGC
exon 49_5	756	TGACAACGTGGTGATGCC	CCCTCAGCTTCCTGTATCTTTAC
exon 49_6	816	GAGGGTCCTAGAATTCAGCAAG	GATAGCATAGCACCCAACAGG
exon 50	503	AAACTGCTGTAATCCATCTCACCA	GGTAATACAAGCATTTCCCACACA
exon 51	765	TTATGGCTCATGTCTATCTAAAGACT	ACTGACTGAATTGTTTGCCAACAGT
exon 52-53	908	GGAACCTATTAAATGCACCACAGCT	ATAGGACAAGAGAAGAAAGTACCAGT
exon 54-55	865	CACTATGAAATAGTAGTATTGGGACA	AGGTCAATATAGAAGAGTGCTCAGT
exon 56	511	GCTCCTGTGAGACTACATTCACTG	AGGGTTGATCTAAGGGAAGATGAC
exon 57-58	933	TTGTGCAAGGTTGGCTGGAGACT	CATGACATCGTACTCCTTTAAGACT
exon 59-60	909	CCACTGCAGTAGCATTGTAATAGTC	CGGTGATTGTTGCAGGTTCTGTAG
exon 61	648	CAAAGATGTTCTGCATTACTCTCAGT	TTATGGGATGTGACCCAGCCACT
exon 62	706	CAGAGAGGATTGGGCATTTCCATG	TAGAGGTGTCATCCTTCCAAACTGA
exon 63	636	AATGCAGTGGGCATCTGACAGTC	ATTCCTCAGCACTAAGACTGGACT
exon 64	564	GTCTGGTCTCTACATCTCCTCACT	CTCCACAGGCTTCAGCTCTCTGA
exon 65	579	TTATGTGTGTGAAGCTCGAAATGAC	CTTAATGAATGATGGTGGTTCTGTGA
exon 66	926	CAGCTGCTCATGCAGTACTAGAGT	TGAGTTACATTGGTAACTCCACTAG
exon 67-68	986	GTCAGGAGTTGAGAGGTGTTATGC	ACTCTTGGAAAGAGGCAATGGCTGA
exon 69	593	CAGCCATTGCCTCTTTCCAAGAGTG	GAGTCAGGAAAGAGGAGAGTATCAG
exon 70	694	CTGTGTCTTTGGTTCTCTGTATTCAG	GTTGAGGTAATCAGAATTATGACACTG
exon 71-72	963	CCATGAGAGTGTTAGTGCTGGACT	GATCAACTATGTCTTTCCTCTAGAC
exon 73-74	844	TCACTGGGAGAAACAATCTTAGTGA	TTTGTCAACATAGATGTGCACCTGA
exon 75-76	893	AAATTGCAACTCCATCTGAAGC	AAGGAAGAAAGGCTCACAGTCAG
exon 77-78	981	GGAGCTCTGGTATCTACATTAGG	CATCTATAGATACAGGCTTGATGTC
exon 79	564	CACAGCAAGAGGTTTGTTCACACG	ATGTTGCCTCCAACACTAATACAC
exon 80	431	TAAATTTATGGGATCTTCAGCTAC	CTTTGCTAAGAGCCCAAATCAGAG
exon 81	456	GTCATATGATCACACCAGTCACTC	GGCCACACAAGTTATTAGTTAGGC
exon 82	483	GCCCCTTCCTCCTATCTGTTC	TCAAACTCAAGAGTGCAATCACTG

table continued on page 191

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
TTN			
exon 83	519	CCTCAAATATTCCTCTTATCCAGT	TCACGTCTTTCAGCTTGCGGATG
exon 84	604	CGTCCATGTTTCTGGTGAGTA	GATGGTGGTTCTAGATATTGC
exon 85	476	GTCTTCATGTTTAGTCTTAGCT	CTATAAGGAGAAAACATGTGGGTA
exon 86	510	TCGTTACAAATGATGCTGGCAGT	GTAGTCAGCAAGTCAGTCATGCC
exon 87	487	GTCTTAGGTGTGTTGACAGCTAC	CAGGTGAGAAGGCATCCACAAC
exon 88	569	GTATTCATGCAAGGCAGACAACAG	GCAGAAACTAAATGGCACCAGC
exon 89	470	GCTGGTGCCATTTAGTTTCTGC	GACCCTGGTCCTAAACATAGCT
exon 90	700	TTAACTCTTCTTCCTTGAACTGCAT	CCTTTCAGTGTCTCTATAGGATG
exon 91	556	GTCCTCTTAACTTCCTGTTTGACT	ATAGGCTGCTAGTGATAAGACTG
exon 92	602	GATGAGATATTCAACTTTCCTACT	TTCTACAATTGTTGCAGGCTCTG
exon 93	537	TTCGCCACGCTATCCGTTCTCG	GATTCAGAAGATGTCGGACTCATG
exon 94	479	AGCCAAATGGATCTAGGGATAGTC	TTGTGAATTGTTATGACAAACATGC
exon 95	554	TCCTTTGACTGTGAGAGGTCAGT	TTCTCTATGCTGTAATCATTCAGC
exon 96	471	GTTTGTCAGACTGCTCTTCTAGAT	CAGCCAAGAGAGATAATCAATCAG
exon 97	511	CATACTAGGTTTGTCATGATTGCA	CTGTAAATAGGATTACTTAAGCAG
exon 98	475	GCATATATACGTGCGTTTGTAT	CTGGACCATGTCAGTTTACTG
exon 99	519	GTGGATTCTTAAACTTTCATTTCT	GTCCTAAATTGCATTCATCATC
exon 100	730	GTGATTCTCTAAATGGCCATGAGT	ACACCTCCTTCGGGAAGCCACAT
exon 101	842	TTCTGTCCTACTACAGCAACAG	AGAAGATAAAGCAGCTTCTCAGAC
exon 102-3	977	TCTGAGAAGCTGCTTTATCTTCTG	CAGTCATCCCTCTCTATGCCTG
exon 104	514	TTACAATGGTCCTAATCCATCTGATC	GCTGGTTAGGTCATATTAAATGAC
exon 105-6	710	GTCATTTAATATGACCTAACCAGC	TTCAGAGTCACATCCTGAAGATG
exon 107	496	ACATCGCTAGCGCTAAACTAACTG	GACCACTTTGGTGTGTGTCTACAG
exon 108	678	ATTTCACAGAGGAATAAGTTGTAG	GCATTCTAGATAGAATGAGAGCAT
exon 109-11	951	GCTCATGCTCTCATTCTATCTAG	GACTCTGCTGCAGTTTGGTCGCT
exon 112	422	CGACCAAACTGCAGCAGAGTCAG	CACTAGAGGGCCAATGCGTTGTA
exon 113	457	CAGTGAAGATGTAAGATCACAGTAC	GACTGAGTCAAAGACGAGGGCG
exon 114	381	ATGAGTAGTAGTCATAGCTCCTAC	AGACCATTGCACAGAGGCAGCAT
exon 115	550	AGTGAAGCAGTTGGATGGATAGA	CTGCAAATCAGGTTCATAGCATTG
exon 116	198	AACAACCTAAGAGAACTTACC	TATACATTGCTGCCAAAC
exon 117-18	437	AGAGAGAGAGAATTTTAATAGG	ATGTACACATGTCCATACAC
exon 119-20	718	GATGAAGGCTGCATGCATTGTTTG	AGTGGCTATAGGTATGCACACTG
exon 121	249	TCTTACAACAACATATGCC	TATAGATGGTTTACAGATGC
exon 122	266	TTCTCACACTGTTAGTGAAG	GCATAGTACATATGAAGATCG
exon 124-27	1,086	GCTGATGCATAGATAATCCATAACA	CTGCAAGTGAGACAATGGATCAG
exon 128	487	TTTGCTAGTCAAAGTGGTGTACTG	CCAAGACTGAGAATGATACTTGAG
exon 129	509	TCATTCGAAGGAAGTGGATGAGTC	AATGAGACTGGTAGATGTTCACTG
exon 130-31	767	TCCCATACTCTTCATTAACGCATG	TGCCTTACATATGAAGTGACACAG
exon 132	991	TGAACGTATTTACTGAGTATAGTC	AAAGGCAGCCATAAAGTAGTAGC
exon 135	492	GTCAAGTGTTTATGTTTGGGCTAC	AAAGCAAGCATCTAGGCAGATGAG
exon 136	582	TAATCAGGCACTGTTATGGTTGTC	CATATTCTTCTTCCCGTTGTACTG
exon 137	808	GCCTTCTAGTTCCCAATCTGAC	TTAACCATGTACTAGAATGATCAC
exon 138-40	942	AGCTTGTTCTGCTTGGATCTCTG	CATCCAACATTCTGCTGACAACTA
exon 141-42	887	TGAAGCTCAGAATCCTTTCCACAG	TTCTTCTTGGGAACAGGAACAG
exon 143-44	882	GATACCAGTCAAAGGTAGATGATGC	GGACTGGTGATAAAGACAAGACAC
exon 145-47	945	GATAGAATATCTATGCTGCTTCAG	GAAAGGCAAATATTCTGTGATCAC

table continued on page 192

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
TTN			
exon 148	614	CATTAATCTGTAACAGAGCCATGGT	GGTCAATCACAGAGGGTAAAGGAT
exon 149	514	CATGACAGACCTGTCTTGAGCGA	GTCCTGTGTGGATAGAACCACAG
exon 151	213	TATGTTGGAAACTGGAAAC	GACAATAATAAGACAAGGATG
exon 152	231	TTTTCAGATATTATCACTGTAGTC	TGAGAAAGATTTGGAACAC
exon 153	445	CCTTTACACCTCTAGGGTTCCTA	ATTCATACGTAAACACATGCACACA
exon 154-56	962	GATGTCCCACAGTACAGAGTGAC	ATGGGGTTTCTAATCTTCCAAACTG
exon 157-58	1,133	GTACTAACGTTAGTAGTAGTGCTCC	GTCATTCATAGCCATCTTGTGGCAT
exon 161-63	617	CCACTAATTTGCCTCTATGTGC	CTTCAGTTGGAGGATGTTCTGG
exon 164-65	313	TGTCATGTCTTTATGCCTTTGC	CTCTTGAGTGACTCACTGGGAC
exon 166	110	GCACCAGAAGCTATGAAAGAAG	CCACATTTGTTCAGAGGTAACG
exon 167-69	727	GAGCTCTGTGAGGTTCTGGAG	GGTGGTCCTTTCTATCGCC
exon 170-72	344	CTTGTTACCACTGATCTCAGCTC	CGGATAACCTCTTTGGAAGC
exon 173-74	484	GGCTTCAGATGATGTATCTCTTC	CCACCAAGATATTTTGGATAGCGA
exon 175	92	AGTGCCTGAAGTGCTGCCAC	CTTTAGGAGGTGGAGCTTCTG
exon 176-77	361	GCTTGGGTAAATGTGGAGTCG	CTGACATGTACCTGTAACTGCG
exon 180	82	GTGCCTGAAGTCCCACAAG	CTGTAACAGGTGGAACTTCTGG
exon 181	221	CAGAAGATCCAATTCCTC	ATATTTTCCAGAGCACAAG
exon 184	482	CAATCGCTATCCAAAATATCTTGG	CAGTGATTGTGAGGGGTACAGAC
exon 190	221	CAGAAGATCCAATTCCTC	ATATTTTCCAGAGCACAAG
exon 191-93	906	GGTTCTGTTTATATGATGGATTACG	GTACAACTTCTTTAGGAGCTTCAG
exon 200-1	633	TAATTCTTTTGCAGTTGAAG	AAGTTATGAAGACCATTAGG
exon 203	125	CTTTCAAGTGCCTGAGGCTC	CAATATCAAACACAGCACCATG
exon 204-7	1,025	GACACCCTCATGGTGCTGTGTTTG	CATGTACCTTTTGCTGGTGGGAC
exon 208-11	1,686	AGCCCCCAGCTTCTAAAAGGG	CAAACATATAATACAACACAACAC
exon 214-15	1,360	GCAAATGAAGAAAGCTGAGAAG	GACAAGCCACAGTTGACATGAG
exon 217	654	CATGTAACTTGTACTGTTTCTCTCAT	CAGATATCCCGGATGACGCTATC
exon 218	400	CATGGTTCGAATGCTGCTTGTATG	CAACACTCATGAGCCAGGACCAT
exon 219-20	1,463	GGAGCAACATCCTCTACATGCAT	ATCACAGATGAAATATGAAGCCATG
exon 221-23	1,285	GAAGTGATTTAGAAGGAGAGAGCT	CCTCCGAAGTTGAAATGATACCTA
exon 224	386	TAGCTTGAGCTCTTCAGTCTCAG	CCAGAAAACATGCTAGTTTTAGTAG
exon 225-26	1,033	TATCCAGCAGGCAGCCAAGTATG	GTTATTCCCCTCTTAGGTACAAGAT
exon 227-28	786	ATCTTGTACCTAAGAGGGGAATAAC	CCCGAATCTAGGATATAGATCCTG
exon 229-30	748	ATGACAGGGAATTCTCTGCTTACT	GGGGTTGTTTTGGTAACACTGTGA
exon 231-32	810	ACTCAGGCAGATTAGAATGGAATC	GCATGACTGAACTAATAAACTGCAG
exon 233-34	622	CTGTTGAGGGTGAGTTGCTCAAC	AAGTGGACAGTGGATGAAGTCAG
exon 235-36	841	CCTGATACACCCCTGCATTGATC	TAAGGAGTTGGGCTGCTTTCATGC
exon 237-38	672	CTCAGAACTCTGTGAGTTGTGTC	GCATCCACTCTGACTTTCACCTG
exon 239	486	CTCTGTGATGCTGCTAGTTGGCT	AGTGGCAAATACAAGAGAGCCAG
exon 240	404	GCCAGTCCTGTGGCTTCTTGAAG	CAGCAGAAATCAGGCTAAAGGCG
exon 241	426	ACTAACCTGCACATTGTGCACATG	TTGAGAGTTGGCATTGTTCATGAG
exon 242	678	TTATCAAGCTAGGTATCCATCACT	CAGGTTAGGTCAGTGTCAATACTAG
exon 243	346	AGTGATTTCAAAGTCAGTTGTAGTC	GTGGTGACCAGAGAAGTTGTGAC
exon 244-45	894	CACACTATGAAGTCTTGTTGGTTGC	CTCCAGATTCAAGTGTCTTATCATC
exon 246-47	994	TGCTGGCAACTACAACTGTCGAC	ATTCACCTTGCACCAGAATGTGAC
exon 248-49	941	CAAGCCAACTATAATGTGTCTTTGAC	ATAATCTTGTGGAGGCCTGATGATC
exon 250	570	AGAAGACAGGAAGTCTCGTGCTAG	CTCAAGACACTTACTTTGTAGTGAG

table continued on page 193

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
TTN			
exon 251-53	1,370	TTCTCTGCTATAAAGTTACTGTAGC	CACTAAGTAAAGTAGGCACAGTCTG
exon 254-56	999	CAGACTGTGCCTACTTTACTTAGTG	CACTCACGAGTCACTGTTAATAGTC
exon 257	462	GACTATTAACAGTGACTCGTGAGTG	GGTACTTAGCAGGTACCAAGTCAG
exon 258-59	740	CTGCAGTTTGTATCCCTGGTCATG	CCCCTGAGCAGGCAGTACATTCAG
exon 260-61	892	CTGCTGAATGTACTGCCTGCTCAG	GGTTCCATGTTAGAAGACATGACTC
exon 262-63	1,128	GTTGATAGTAAGAGAAGTGACACTG	TTAACTGCCACACACTTTAGACTG
exon 264-66	1,461	CCAGTATATTATGATAATGACCACTG	CCTCTGTATTGGTCTTCCACCATC
exon 267-69	1,079	GCGACTACACCGAAATTGAGGACT	AGTCTCCAGGTCAATTGTTGGCTT
exon 270	872	ATTGGTGTCAGCGAGCCATCTGA	TGTCAGTGGTATCTACCTCAAGTC
exon 271-72	1,065	TTAGTAGTTGGGAGCCAAGGATGC	CCAGAAATACTTGCATCAAGTGCTA
exon 273-74	1,535	TAATCCTACAGCGGAGGCAATGAC	CTAGCATACTGTTGCTGGTAACATC
exon 275-76	1,028	GAGTACCTCTTCCGTGTAAGAGCT	CAGTTCCACAGCTGGAGGCTCTG
exon 277-78	1,138	ACTGTGGCTGAACCACAAGGTAC	GAATTCTCTTTAACTACTACGCTAG
exon 279-80	1,151	AAGCTTCACAGACTTACTTTCCTAG	CGGTCTGGAAAGGAATCAACAGA
exon 281	522	CACGTTGAAGTATATGGTAAGTGAC	GATGTATTGTACCTGATAGTGGATC
exon 282	779	GATCCACTATCAGGTACAATACATCT	CTTCCGATAGTCTATGTGAACATG
exon 283-85	927	GACTCCAGCACTTCATCAGCTTCA	GCCTGTATGAGATCGTTTACACTC
exon 286-88	1,229	GGAGTATTTATACTACAACTGACTG	ATGTCTACATTCAAGCCATAGTAGC
exon 289-91	1,314	CTGAGCTACAAAGTCTGCCTACAGT	GGCTGCAAAGAGCCAGTATACGT
exon 292	556	AGACCCAGCGACTGCTAGAGATC	GGAATAACTTGGCAGGAGTTATGC
exon 293-94	870	AGTTAAATAGGCTGCTGGTGATGC	ATGGATGCATGTCTTCTTGTTAGC
exon 295	414	TCCAGATGTTCCAGGTCCC	TCAAGTGAATGAAATGTACGGC
exon 296-97	1,239	AAGTGCTGGCATTACCACTCCTAG	GGAGGCTTCCAAGAGATAACCATG
exon 298-99	1,004	GTGGAGAACAGTACAGGCTCTAG	GTGGCTGCCAAGTTAGATCGACT
exon 300-1	1,047	CCTGATCTTCTAGAAGGATGTCAG	ATTTCACTGACTTCCAGATCTCTAG
exon 302-3	1,028	GTGACTCCAGTTGGTAGCAAGCT	ATTTCACATCTAGCTCCACGGATG
exon 304_1	1,198	GTAGATGTTGACAAGACTGAAGTC	GCTTAGTCCTGCAAGATTTTCAGC
exon 304_2	856	CTGGCTACTATATGGAACGTCGA	GAAATTGTACAGTTCTCTTTGGTTAC
exon 304_3	1,420	CATGGCAAGTATATCATCTCAGCT	CACTCTTGTTGACAGGGACCCATC
exon 304_4	743	CAGTCATAGCCAAAACCAAGTATG	CGGTGCACCACCATCATAGACTG
exon 305	528	GGATGACCTATATTTGATTAACTGAG	TCTCCGGAGGTTCTGCAAATGAC
exon 306-7	860	ATACTAGGTAAGTCCTTCCCTTAGC	ATGGTATTAGTATCTTGACCTTGCTA
exon 308	557	GTACTGAACTACAGGTTCTATGTAC	TTTGAAAGCCACTGTTCTCTACTGT
exon 309-10	853	CAAGCATTAATTTGCACCAGTATAGC	GAGATGAAGACAAGGAAGATGTCAG
exon 311	526	CTCCAAAGATGGTTGCGCAGT	TACTACTCTCTGTGTCTTGGAGTC
exon 312	610	CTTCCTGTTGAGCTTGTTACTTAGC	TTATATAGTTGCCACAGTTGTGTATC
exon 313	480	CCATTCTTGAAAGAAGCAGCATATG	GAATGTCTTCTCCCACATTATTCTGT
exon 314-15	1,087	CCTATGCACAGAATTATTCTTATGAC	GACTCACAGAGCTGCGAGTTGTATC
exon 316	570	GGACCCTCAGTGTAAGTGTTCATG	TAAAGTACCCACCCAGCTCTCCTA
exon 317	600	CCTGGTTAAGCCTAAAGCATAGTG	CCAGGAGTATCTGGATAAATAGTAG
exon 318-19	967	TTGTTGTGAAAGTGCTTGGTAAGTC	TTAAGTCAAGATCAGGAGCCACTG
exon 320-21	998	CACTCTTGTGTCCCACTCACTAGC	CAGATAATGTTATCGATAAGCACATG
exon 322-23	823	GTATAGATACTAGCAAGAATGACTAC	CAAGGACAATTGTTGGTGCCTCTG
exon 324	881	GCAGGTGCTATCAGTGCTCCATC	GTTCTTAGTGACATTTGATACCTCTG
exon 325	539	AGACAGATTTGGTATGTAGAGCATG	TTTGAAGGTGGTCCTGGAGGATC
exon 326_1	1,079	GAGGCTTCAGATTCTGTTCTGATG	CGGTGCTCTTAGTTATGTCCATGA

table continued on page 194

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
TTN			
exon 326_2	1,145	CTACATTGTAGAGAAACGTGAAGC	GTGGCTCATGCCAGCTAATTGTCA
exon 326_3	1,013	CCAGGCATCACATCAGCATGCAT	AGTGGAGGGAACCATGATAGTAC
exon 326_4	1,111	AGATTCAACAAGAAGGGATAGTGG	CACCTCAGGTGCAGGCCTTCC
exon 326_5	1,049	CTGGATGCTGACCTGCGCAAAGT	CAGGCGTCCTTTATTGTCAGTAGTG
exon 326_6	1,293	AGACAACTGAGTGTGCCAGTGAT	AACCTTGAATGATTCACCAGCATG
exon 326_7	966	TAACAGCTAGAGATGAGGTAGATC	CTGAAGATCTGGTTATGTCTATGAC
exon 326_8	959	GCTGCTGGACTTAGTGAACCAAG	CCAGGAGGTTGTGGCACTTCTG
exon 326_9	829	AGAGTCACAGCTGAGAATGAGTA	GTGTAGTTCTAGCAACAGTAGCAG
exon 326_10	897	GCTGCCAAATCACCAACTACATTG	CATGGACATCAGCCTCAAGTCTG
exon 326_11	857	CTGACAGTACTGGACCAATAACTG	GACCCTGAATTCATACTCATGATC
exon 326_12	913	CCATGACCGTCTGTTGGAACCGT	GGCTCCCATACCAGGAAGGCAG
exon 326_13	940	ACAGCAGTGGATCAAAGTCTGCT	ACTGGGCCAACTGGAGGTCCAG
exon 326_14	923	CAGCTACCTCAACTGTTTTGCAC	GTCTTCTACAAGTCCACTCACTG
exon 326_15	921	ACCTGCCTATGATGGTGGTAGC	TCGGTACCTCCGTCGTCTACTG
exon 326_16	934	GAATCTGGGCCTGTTACGGCCTG	TTGGTGAGTCAAGAACTCTGACG
exon 326_17	1,087	GTCGTTCTGCGTGCAAGTGCTAC	GAACTGTTATGGATCCAGCACTG
exon 326_18	964	GACTCTTAAAGAGACAACTAGAGTC	ACGCTGATCTTCAGTAAGTTCAG
exon 326_19	962	TGGAGCTAAGATCACAGGATACAT	TTACACGCACCCATCTTAGGCTA
exon 326_20	1,223	CTATGACACTTTGCTGGTCAAGAC	AAAGCATTTAAGATGAAATAACGACTA
exon 327	633	TGTGGGAGCCTTGGTTGGTTCAG	CCAGCTCTAACATATACAGTGTGAC
exon 328	830	TCCTGTTCTTATTAAGGAGCAACTAG	GATGGTGGTCCAGGTGTTGCTAC
exon 329-30	975	TGTGACTGTCAAACTACCATACAGT	TTTCCTCAGACTCCCAGCTAGTC
exon 331-32	991	TTAAGTCCCTGCACAACTCACAGTC	AAGCATATGCACAGGTTAGCGTAG
exon 333-34	922	TGTTCTTGGAAAGAATGCGAACATC	CAATCTCTGGTGCCTCTGTAGAC
exon 335_1	739	ACCTATAGAAATGAATGAACCTGTAC	TGGTACTTCAGCAGCCTTCACTG
exon 335_2	639	GGAGAAGACTCCATTCTTCTTCAG	CACACTGCATTATCAAGAATAACAGT
exon 335_3	809	ATTAAGAATGCCAAGAAGGAGCATG	CAAGAAGATATGTAAGAAGGTGATGC
exon 336	582	GACCAATATGAGTTTGGGAGATTGA	AGAAAGACATAATTCATCCACTTAGC
exon 337-38	935	TGGCCTTGTGATATGGCACTACTG	GTTATGTTGCTAGGTTCTGGAATGC
exon 339_1	546	CAGGCCACTTGATTCTGATCCAGT	ATTTCCATGCCACCATCAAACACTG
exon 339_2	625	CTGCAAGTGGCATCTCAAGACTCA	ACTTTCACGGTGAATGTGATTGACT
exon 339_3	685	GATATCCATACAACAGATTCCTTCAG	CTCTGGGTTCACTATAGCCAGCAT
exon 339_4	876	CACACCAAACAGCTAACTTTCACAGT	TGGCAGACACATGGTAGACCTCAG
exon 340	585	CTAGAATCAGCACCAATAATTGCTGA	TACTGTTGATCTTGTGACATCTGTCA
exon 341	550	AAGCTTCAAGACCTATAATGGCTCA	GGCTTTGGGTGGAGCTGTCAGT
exon 342-43	972	CTGTCACTTGCCGAGATGAATACG	CAGTCCCAACTTCTTCAGGTATGC
exon 344	591	CAATTGTAGCCAGAAACTCATTCAGT	ATGATCATGCAAATAAGCATGTGCT
exon 345-46	870	AAGAGCAATAATGCCTTTGTGAGTC	CATAGTTCCTTTCTAGAGAGGTGATC
exon 347	847	GACCTATGATTGATCACCTCTCTAG	AGAAAGATCGATTGTTGGCAACACT
exon 348-49	962	TGTACAAGACCTCAGAGGTATGTAC	GTTTAGAAACCTTAGAAAGACCACAG
exon 350	850	AGGAGGACTGGCAATGTTATAGTG	AAACGGACACAAGGGAACTTTACG
exon 351	577	GTTAACAGTCCGTTTCAACAACTCA	TCTGATGACGCCACCTTGCCTTAC
exon 352	809	GAAATGCTTGGTGAGACATATGATC	AAGCACACATGTATTAGAATACAGTC
exon 353	662	CCCTTGAAATCTGAGGAACCAGTC	AGGTTTCTCCCACTGTAGAGTGAC
exon 354	587	TCTGAACCAGTTGTTTGCAAAGATC	CTTGTCATATTTCCTGTGTGTGTAC
exon 355	936	CTACATAAAATGTGACGATGTGCTCA	AGTTCTTCAATAGAGCTTCGATCAC

table continued on page 195

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
TTN			
exon 356	558	GGAACAGTTGATGCCATCCTTGATG	ACTTGTACTTTATCTGAATCTTTGCTA
exon 357	1,013	AGCAAAGATTCAGATAAAGTACAAGTG	TGTATCTTAGCTGGAACTGTATCAG
exon 358_1	838	ATCATCAAGTTGAGGCTAAATGAGT	ACTCACCACGCCCAAGATCTTCAG
exon 358_2	855	GAAGATAAGACCAGAGCTATGAAC	GAGAGCCTCCGATGCTGTCATG
exon 358_3	882	CCAACAGATCATTGAGAATATCATG	CCCTTGTCTGACTCAAATGTGTAC
exon 358_4	869	GACAGCTTATGTAGGTGAAAATGTC	GGTCTTCTTCTATGGTAGTCTGC
exon 358_5	834	CGGCTGTAAGCACCAAGACTGT	GTTCGCTTTTGTATTCTGAGAGATG
exon 358_6	814	GCTAGTACAAGTTATGCAGAACTG	CACTGTCATCAGTATGACAGTCCAG
exon 358_7	857	GCTCGCACAGGGTACCATGTG	GACTCTTGGTGGTGATGTCACAG
exon 358_8	957	GAGATCTCTTCAGTCCAGGCTTC	CCTCCTTGTGTAATGGCCTGTAG
exon 359	498	TGCAGTTAAGGCTACTGGAGAAC	ATGTGTGTTTCTGCTTTGGTGTAC
exon 360-61	1,206	CCTACATTGCATGAAGAAGTTGACT	GAATTCACATCAGTTGGCTGTCC
exon 362-63	890	CTGTAAGTCAGCGTCATATGTCATC	CAGTTCAGAAAGATTAGTCCGTGTG
FLJ39502			
exon 1	379	TTCAGAAAATGTAGTTTTAGG	GATTAATAGACAGCTGAACC
exon 2	294	AAAGCTGTATCTCTTGGG	CTCATGTTAGCCTGTGTC
exon 3	263	CAAATTTGCTGGAGATG	CAATTAATCAGGTAAACTGC
exon 4	394	AACAAATGAAAGAGTGATTC	TACTAGAAGACACAATTTGG
exon 5	352	ATGTGCTATAGAAATTAGGC	CATATACTAAGCAATCTGCC
exon 6	290	AGGTAAATAGCCTACTCGTC	ATTTAGACACATGCTTTAGG
exon 7	281	TCAGATGCTTTGATTTATTG	CAATAATTGTGTAGTACATGC
exon 8	335	TCAGTATTTAATCACTAGCATC	GCTTTAAAGTTGCAAAATG
exon 9	315	CTTCACTGGTCAAAATTAAG	GGCTCAGTATAGATTTGTTC
exon 10	251	TACCTGGACTTTTCTTCTTG	CCAGAAGACATTTGATTAAC
exon 11	257	CCAAGTAGTTAGTGTATGGC	CAATGGACAATGACTTTAAC
exon 12_1	553	TTTTACTGTAATTGAAACTGG	ATGTTCAGCACTGATTCTG
exon 12_2	571	AACATTTTCAAGTCATTTTG	CTAAACCTTTCAGTTAGTGC
SESTD1			
exon 1	194	TGTAGTTGACATTTGACTTTC	ATGCATCTGTCATACTTAAC
exon 2	354	TTACAATCTATAAGTAGCCATTAG	AAGCAGATAAGATAGCAAAG
exon 3	225	ATTATGAAGAACGAAATCTG	ATGCATTTGCCATTTTC
exon 4	297	AGCATAATCTTGACTGTCTG	GCTTAGTAACCAATGCTG
exon 5	315	TTTCGTGGATTTTAATAGTG	TATGCTATAATGTGGTATGC
exon 6	292	AAACATATCCTAGCTGGC	TTCTTTTAGCTTGCATTAAC
exon 7	225	ATACTTTTGTGCCCCTG	ATTTAATGAACGAATCAATC
exon 8	352	ACTAGAATTGGGAATATTGAG	GTCACATAGTAGAAATTATAAGG
exon 9	255	AATAAGTTAGAATGAAAGTTTG	TATGCTACTAATTACCCACG
exon 10	352	TCATGTATTACTTAGTTTATCTGG	ACGTGTAGGTAAGTGTTCAG
exon 11	254	AACACTTTAGATGCTGATTAC	AAGTTGACTTAATTGTCTAACC
exon 12	340	ATATCTGATGAGACGATTTG	AGTTAAGAACGTTTTGTCTG
exon 13	255	CATGTACCAAATGATGATG	TTTTAGGACCATGAAAAG
exon 14	252	CATTTTGATACAATAAGTAAAGAG	ATTATAATCATGCAGAAAGC
exon 15	464	GTGCCACTAGATTTATTCAG	TCTAAACAAACCGTAACAAC
exon 16	246	GTGATCCTCTGAAACTGC	TGATTTCACTTTAAGACCAC
exon 17	259	CACTGCAGTTAATAGATTTG	GTGGCTAATGCTGTAGTATG

table continued on page 196

Table S1 (continued). PCR and Sequencing Primers

Exon(s)	Product length (bp)	Forward primer	Reverse primer
LOC728984			
exon 1	236	ACCCGCTGGAAAATAAGGAC	CCGCCTCTGCTCCCAAC
exon 2_1	547	CTCCGGCGCTTTGTCTG	CTGTACCGCCGTCACCC
exon 2_2	487	GGGAACGCTAGCGAGGTG	GGTGTCAAGTAACGGACCCTC
LOC64476			
exon 1_1	509	ATAGGGGAGACTGAGCTAC	TTTAGCCCGTCTGTCAC
exon 1_2	498	GTGAGGGTCCAGTTGTAG	ACTTGCTGTTTCTCCATC
exon 1_3	497	AGCTTCCTGACTTCTGAG	AAAGTAGTTTTCCAAAGTCC
ZNF533			
exon 1	206	CAGGCAGTACATGCCACAC	CACAAATGTAATGAAATTGATGGAG
exon 2	238	TTATTTGCTCACACTGTTG	AGATCAATCAGACTTTTAGC
exon 3	316	ATCCACTGTGGTGTTTG	AGAGTAACATGCAACAACTG
exon 4	377	TCTATGGTGATTTATCTTTG	TCTGATGGGAAAAGTTG
exon 5	313	TCAAGGATAAGAAAACAAAG	GAAGATACACAGTCAAGTGG
exon 6	264	CGTTTTACATGGAATAAGAC	GAAACCTTTCATGACATTAG
exon 7	353	TACTTATGAAAAGATGCCTG	GTGGCTTTCTTTCTCAAC
LOC729001			
exon 1	631	AGAAAGAATCTGATTGAACC	ATCCTAAACTCTCTACCTGC
exon 2	582	GTGAGGGTCCAGTTGTAG	TCCTTCCCTCTTATTTTC

Table S2. PCR Cycling Program

	1 cycle	4 cycles			6 cycles			8 cycles		
Phase	Den.	Den.	Ann.	Ext.	Den.	Ann.	Ext.	Den.	Ann.	Ext.
Temp.°C	94°C	94°C	70°C	72°C	94°C	68°C	72°C	94°C	66°C	72°C
Temp. Inc.°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C
Duration	4min	30sec	30sec	45sec	30sec	30sec	45sec	30sec	30sec	45sec

	10 cycles			12 cycles			14 cycles		
Phase	Den.	Ann.	Ext.	Den.	Ann.	Ext.	Den.	Ann.	Ext.
Temp.°C	94°C	64°C	72°C	94°C	62°C	72°C	94°C	60°C	72°C
Temp. Inc.°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C	0°C
Duration	30sec	30sec	45sec	30sec	30sec	45sec	30sec	30sec	45sec

	16 cycles			1 cycle	1 cycle
Phase	Den.	Ann.	Ext.	Ext.	Cool
Temp.°C	94°C	58°C	72°C	72°C	4°C
Temp. Inc.°C	0°C	0°C	0°C	0°C	0°C
Duration	30sec	30sec	45sec	10min	hold

Den., denaturation; Ann., annealing; Ext., extension

Table S3. Sequencing Program

	30 cycles			1 cycle
Phase	Den.	Ann.	Ext.	Cool
Temp. °C	96°C	50°C	60°C	4°C
Temp. Inc. °C	0°C	0°C	0°C	0°C
Duration	30sec	15sec	4min	hold

Den., denaturation; Ann., annealing; Ext., extension

Table S4. SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs2736990	4	90,897,564	*SNCA*	5.69E-09	Yes	OK	3.97E-09
rs3857059	4	90,894,261	*SNCA*	3.60E-08	Yes	OK	5.45E-07
rs415430	17	42,214,305	*WNT3*	4.50E-08	Yes	NC	-
rs11931074	4	90,858,538	*SNCA*	4.78E-08	Yes	OK	6.59E-07
rs199533	17	42,184,098	*NSF*	5.05E-08	Yes	OK	7.59E-07
rs356220	4	90,860,363	*SNCA*	6.99E-08	Yes	Bad cluster	-
rs169201	17	42,145,386	*NSF*	1.25E-07	Yes	OK	5.90E-07
rs393152	17	41,074,926	*C17orf69*	1.42E-07	Yes	OK	9.31E-09
rs12185268	17	41,279,463	*IMP5*	1.90E-07	Yes	MAF < 0.05	-
rs1981997	17	41,412,603	*MAPT*	2.02E-07	Yes	OK	5.36E-07
rs2532274	17	41,602,941	*KIAA1267*	2.22E-07	Yes	OK	3.89E-05
rs2532269	17	41,605,885	*KIAA1267*	2.70E-07	Yes	OK	2.40E-06
rs8070723	17	41,436,901	*MAPT*	3.36E-07	Yes	OK	5.55E-07
rs17563986	17	41,347,100	*MAPT*	3.44E-07	Yes	OK	1.89E-07
rs2668692	17	41,648,797	*LOC644246*	3.94E-07	Yes	OK	2.81E-07
rs11648673	16	317,795	*AXIN1*	4.77E-07	Yes	OK	0.4128
rs12373139	17	41,279,910	*IMP5*	4.91E-07	Yes	OK	6.51E-07
rs239748	23	18,793,279	*LOC441484*	1.17E-06	Yes	Bad cluster	-
rs12431733	14	53,360,580	*BMP4*	1.52E-06	Yes	OK	0.04338
rs11591754	10	35,247,159	*LOC646213*	1.68E-06	Yes	OK	0.307
rs7013027	8	2,911,376	*CSMD1*	1.85E-06	Yes	OK	0.06337
rs11012	17	40,869,224	*LOC201175*	2.85E-06	Yes	OK	6.08E-05
rs7004938	8	140,328,407	*COL22A1*	2.97E-06	Yes	NC	-
rs10857899	1	111,929,241	*LOC643329*	3.06E-06	Yes	OK	0.968
rs6542651	2	3,737,705	*LOC728597*	3.34E-06	Yes	OK	0.04105
rs2285459	16	30,402,913	*ITGAL*	3.38E-06	Yes	Bad cluster	-
rs13027881	2	6,473,384	*LOC391349*	3.76E-06	Yes	OK	0.01554
rs2492448	10	35,235,412	*LOC646213*	3.84E-06	Yes	OK	0.7996
rs4957473	5	39,378,920	*C9*	4.24E-06	Yes	OK	0.002219
rs3775439	4	90,928,764	*SNCA*	4.43E-06	Yes	OK	0.006902
rs6734894	2	6,507,379	*LOC391349*	4.63E-06	Yes	OK	0.01455
rs4556079	8	140,302,117	*COL22A1*	4.79E-06	Yes	OK	0.5891
rs2896905	12	38,779,683	*SLC2A13*	5.03E-06	Yes	OK	0.3969
rs11781101	8	140,316,844	*COL22A1*	5.31E-06	Yes	OK	0.869
rs11783351	8	140,328,721	*COL22A1*	5.51E-06	Yes	OK	0.983
rs817097	17	64,889,657	*MAP2K6*	6.22E-06	Yes	OK	0.9806
rs11644916	16	299,568	*AXIN1*	6.24E-06	Yes	OK	0.09621
rs2856336	12	11,847,265	*ETV6*	7.69E-06	Yes	OK	0.2275

table continued on page 198

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs764660	2	165,921,543	*TTC21B*	7.83E-06	Yes	NC	-
rs11244079	9	135,174,347	*LOC653163*	8.66E-06	Yes	Bad cluster	-
rs2733333	15	55,223,218	*TCF12*	9.31E-06	Yes	OK	0.5624
rs183211	17	42,143,493	*NSF*	1.05E-05	Yes	OK	4.66E-06
rs11878803	19	57,193,133	*ZNF615*	1.07E-05	Yes	OK	0.9152
rs7176873	15	55,136,454	*TCF12*	1.15E-05	Yes	OK	0.7076
rs17654531	1	111,601,208	*LOC728204*	1.16E-05	Yes	OK	0.9846
rs7923172	10	35,349,373	*CUL2*	1.43E-05	Yes	OK	0.1616
rs2116658	2	165,874,918	*SCN2A2*	1.48E-05	Yes	OK	0.08783
rs4934704	10	35,372,170	*CUL2*	1.48E-05	Yes	OK	0.1611
rs163321	5	178,679,053	*ADAMTS2*	1.49E-05	Yes	OK	0.8043
rs869714	1	165,392,874	*LOC391130*	1.54E-05	Yes	OK	0.1256
rs1865648	15	45,665,913	*SEMA6D*	1.57E-05	Yes	OK	0.07442
rs7175191	15	54,973,689	*LOC145783*	1.64E-05	Yes	OK	0.9741
rs10827492	10	35,469,831	*CREM*	1.69E-05	Yes	OK	0.1752
rs10505762	12	11,881,877	*ETV6*	1.72E-05	Yes	OK	0.9894
rs13139027	4	4,977,067	*LDHAL1*	1.75E-05	Yes	OK	0.8203
rs2491015	10	70,436,819	*KIAA1279*	1.76E-05	Yes	OK	0.4443
rs4934540	10	35,514,705	*CREM*	1.83E-05	Yes	OK	0.1584
rs10784359	12	38,732,017	*SLC2A13*	2.01E-05	Yes	Bad cluster	-
rs10437024	1	82,673,852	*LPHN2*	2.02E-05	Yes	OK	0.435
rs2224437	14	28,060,178	*RPL26P3*	2.09E-05	Yes	OK	0.5943
rs7099036	10	35,389,580	*CUL2*	2.13E-05	Yes	OK	0.1556
rs11593858	10	35,421,379	*LOC729811*	2.14E-05	Yes	OK	0.1709
rs12644119	4	90,822,442	*SNCA*	2.15E-05	Yes	OK	0.00167
rs13025009	2	165,878,319	*SCN2A2*	2.16E-05	Yes	OK	0.01459
rs3792738	5	76,283,540	*CRHBP*	2.19E-05	Yes	OK	0.03942
rs7651825	3	14,787,122	*C3orf20*	2.20E-05	Yes	OK	0.3565
rs7920095	10	35,343,340	*CUL2*	2.20E-05	Yes	OK	0.3113
rs1635291	17	41,107,696	*C17orf69*	2.27E-05	Yes	Bad cluster	-
rs11595898	10	35,366,269	*CUL2*	2.31E-05	Yes	OK	0.1411
rs12768019	10	35,429,849	*LOC729811*	2.31E-05	Yes	OK	0.1886
rs17115100	10	104,581,383	*CYP17A1*	2.46E-05	Yes	OK	0.000864
rs558076	17	64,847,377	*ABCA5*	2.53E-05	Yes	OK	0.8766
rs4247113	17	228,978	*RPH3AL*	2.62E-05	Yes	OK	0.4698
rs894278	4	90,953,558	*SNCA*	2.64E-05	Yes	OK	0.000343
rs7527143	1	165,393,978	*LOC391130*	2.65E-05	Yes	OK	0.1569
rs4934719	10	35,416,775	*LOC729811*	2.68E-05	Yes	lMISS > 0.1	-
rs16977243	15	55,181,297	*TCF12*	2.74E-05	Yes	OK	0.1249
rs7844468	8	2,916,843	*CSMD1*	2.74E-05	Yes	OK	0.3636
rs2378554	9	82,805,138	*TLE1*	2.93E-05	Yes	OK	0.398
rs7988279	13	78,528,580	*LOC390415*	3.05E-05	Yes	lMISS > 0.1	-
rs1793949	12	46,657,862	*COL2A1*	3.14E-05	Yes	OK	0.1007
rs4603502	15	45,661,946	*SEMA6D*	3.15E-05	Yes	OK	0.05544
rs4934724	10	35,433,675	*LOC729811*	3.21E-05	Yes	OK	0.1511
rs1686103	2	36,208,125	*CRIM1*	3.26E-05	Yes	OK	0.2357
rs1926554	10	35,384,975	*CUL2*	3.32E-05	Yes	OK	0.1329

table continued on page 199

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs6992732	8	2,911,645	*CSMD1*	3.32E-05	Yes	OK	0.381
rs12834774	23	139,529,169	*LOC286411*	3.38E-05	Yes	HWE < 10^{-7}	-
rs2042079	2	16,969,333	*FAM49A*	3.40E-05	Yes	OK	0.0125
rs859522	7	38,936,691	*VPS41*	3.41E-05	Yes	OK	0.7848
rs7215239	17	41,123,556	*C17orf69*	3.43E-05	Yes	OK	2.88E-06
rs2708909	7	48,018,204	*SUNC1*	3.44E-05	Yes	OK	0.3968
rs2666128	2	36,199,103	*CRIM1*	3.46E-05	Yes	OK	0.9083
rs1395993	3	21,344,590	*VENTXP7*	3.49E-05	Yes	OK	0.003222
rs10849446	12	6,349,553	*SCNN1A*	3.66E-05	Yes	OK	0.001208
rs3824754	10	104,604,340	*C10orf32*	3.92E-05	Yes	OK	0.00864
rs11191425	10	104,615,960	*C10orf32*	3.96E-05	Yes	OK	0.000948
rs12261843	10	35,594,060	*C10orf9*	4.05E-05	Yes	OK	0.333
rs11564162	12	38,729,159	*SLC2A13*	4.08E-05	Yes	OK	0.6368
rs4563067	17	69,700,141	*RPL38*	4.11E-05	Yes	OK	0.8312
rs6794137	3	134,371,136	*TMEM108*	4.27E-05	Yes	OK	0.0296
rs9458499	6	162,633,417	*PARK2*	4.27E-05	Yes	OK	0.6411
rs9574355	13	78,574,175	*LOC390415*	4.28E-05	Yes	OK	0.1002
rs6044224	20	16,606,300	*RPL7AL3*	4.37E-05	Yes	OK	0.8112
rs11778693	8	22,518,797	*KIAA1967*	4.45E-05	Yes	Bad cluster	-
rs2001893	10	35,340,629	*CUL2*	4.48E-05	Yes	HWE < 10^{-7}	-
rs207481	21	24,133,087	*TUBAP*	4.56E-05	Yes	OK	0.9715
rs7707022	5	37,802,142	*WDR70*	4.58E-05	ND	-	-
rs11183395	12	44,890,335	*SLC38A1*	4.59E-05	Yes	OK	0.5671
rs935378	2	46,962,655	*MCFD2*	4.63E-05	Yes	OK	0.2252
rs9918939	9	82,836,491	*TLE1*	4.63E-05	Yes	OK	0.3737
rs1005511	11	57,123,232	*SERPING1*	4.68E-05	Yes	OK	0.5326
rs12413409	10	104,709,086	*CNNM2*	4.69E-05	Yes	OK	0.002757
rs17071181	18	59,307,479	*SERPINB5*	4.86E-05	Yes	OK	0.02886
rs12324092	15	54,990,533	*LOC145783*	4.97E-05	ND	-	-
rs13085998	3	143,746,502	*ATR*	5.02E-05	Yes	OK	0.9789
rs12411886	10	104,675,289	*CNNM2*	5.05E-05	Yes	OK	0.002692
rs2227928	3	143,764,302	*ATR*	5.11E-05	Yes	OK	0.5652
rs10246477	7	83,103,936	*SEMA3E*	5.25E-05	Yes	OK	0.7576
rs12578421	12	11,846,605	*ETV6*	5.26E-05	Yes	OK	0.1197
rs1563863	8	2,917,883	*CSMD1*	5.39E-05	Yes	OK	0.7274
rs1024889	3	70,546,020	*LOC654340*	5.46E-05	Yes	OK	0.8825
rs10894203	11	129,833,092	*ADAMTS15*	5.46E-05	Yes	OK	0.422
rs3936503	10	35,589,263	*C10orf9*	5.50E-05	Yes	OK	0.2224
rs8080993	17	13,650,646	*LOC644361*	5.53E-05	Yes	OK	0.7936
rs9525776	13	42,928,966	*LOC647049*	5.53E-05	Yes	OK	0.844
rs7454430	6	123,331,736	*RLBP1L2*	5.54E-05	Yes	OK	0.8676
rs9924026	16	48,147,935	*ZNF423*	5.94E-05	Yes	OK	0.08927
rs4843467	16	85,441,321	*LOC729979*	6.01E-05	Yes	OK	0.5675
rs1934828	13	75,431,142	*FLJ35379*	6.08E-05	Yes	NC	-
rs11642990	16	85,467,695	*LOC729979*	6.25E-05	Yes	OK	0.1937
rs1449587	13	48,311,469	*LOC338099*	6.26E-05	Yes	OK	0.2692
rs6599389	4	929,113	*MGC4618*	6.28E-05	Yes	NC	-

table continued on page 200

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs265120	1	215,659,568	*GPATCH2*	6.38E-05	Yes	OK	0.1926
rs265120	1	215,659,568	*GPATCH2*	6.38E-05	Yes	OK	0.1926
rs1498046	4	126,152,922	*LOC729377*	6.39E-05	Yes	OK	0.7799
rs1396003	3	21,331,410	*VENTXP7*	6.48E-05	Yes	OK	0.001428
rs17690703	17	41,281,077	*IMP5*	6.55E-05	Yes	OK	1.60E-07
rs3740484	10	102,737,353	*MRPL43*	6.57E-05	Yes	OK	0.9038
rs7183808	15	96,641,692	*FLJ39743*	6.61E-05	Yes	OK	0.01689
rs636508	9	83,582,101	*TLE1*	6.65E-05	Yes	OK	0.09658
rs6812193	4	77,418,010	*STBD1*	6.67E-05	Yes	OK	0.001925
rs9530494	13	75,434,275	*FLJ35379*	6.82E-05	Yes	OK	0.1611
rs6959225	7	8,499,029	*NXPH1*	6.84E-05	Yes	NC	-
rs8111509	19	57,003,307	*FPRL2*	6.84E-05	Yes	OK	0.3644
rs7077361	10	15,601,549	*ITGA8*	6.88E-05	Yes	OK	0.01262
rs4409766	10	104,606,653	*C10orf32*	6.92E-05	Yes	OK	0.002806
rs9839984	3	162,251,606	*PPM1L*	6.92E-05	Yes	OK	0.9053
rs2240914	9	131,938,127	*GPR107*	6.98E-05	Yes	OK	0.9165
rs4661747	1	16,612,540	*SPATA21*	7.02E-05	Yes	OK	0.8089
rs4584384	1	152,762,321	*TDRD10*	7.09E-05	Yes	OK	0.3791
rs807302	6	119,968,450	*LOC728727*	7.25E-05	Yes	OK	0.1395
rs595046	21	43,626,445	*FLJ41733*	7.27E-05	Yes	OK	0.2841
rs2686831	7	47,959,511	*PKD1L1*	7.31E-05	Yes	Bad cluster	-
rs4242434	8	22,557,775	*BIN3*	7.31E-05	Yes	OK	0.04666
rs6481928	10	35,276,450	*LOC646218*	7.31E-05	Yes	OK	0.4771
rs1580254	3	162,249,972	*PPM1L*	7.35E-05	Yes	OK	0.9136
rs6582668	12	37,052,871	*ALG10B*	7.41E-05	Yes	OK	0.2451
rs1526123	17	41,139,123	*C17orf69*	7.52E-05	Yes	OK	5.76E-05
rs207521	21	24,103,584	*TUBAP*	7.53E-05	Yes	OK	0.8755
rs12425761	12	114,119,258	*TBX3*	7.75E-05	Yes	OK	0.03189
rs6780193	3	71,988,566	*PROK2*	7.93E-05	Yes	OK	0.6032
rs16944593	12	113,641,262	*TBX3*	7.95E-05	Yes	OK	0.1255
rs11973020	7	25,549,191	*LOC646588*	7.98E-05	Yes	OK	0.8942
rs417968	17	41,084,159	*C17orf69*	8.03E-05	Yes	OK	1.95E-06
rs7436941	4	159,654,523	*RXFP1*	8.13E-05	Yes	OK	0.1329
rs7911697	10	14,525,211	*FAM107B*	8.19E-05	Yes	OK	0.7798
rs12255903	10	35,269,643	*LOC646218*	8.26E-05	Yes	OK	0.4782
rs7903802	10	12,997,566	*CCDC3*	8.38E-05	Yes	IMISS > 0.1	-
rs9285433	6	119,979,004	*LOC728727*	8.40E-05	Yes	OK	0.1392
rs699038	12	25,050,907	*LOC645177*	8.53E-05	Yes	OK	0.1829
rs7897198	10	35,277,737	*LOC646218*	8.55E-05	Yes	OK	0.5479
rs6446700	4	4,962,543	*LDHAL1*	8.56E-05	Yes	OK	0.8794
rs2515501	8	6,400,033	*ANGPT2*	8.85E-05	Yes	OK	0.8016
rs7112698	11	129,835,261	*ADAMTS15*	8.96E-05	ND	-	-
rs928939	12	11,882,934	*ETV6*	9.02E-05	Yes	OK	0.7836
rs2387807	12	36,733,128	*LOC727847*	9.12E-05	Yes	OK	0.4389
rs2794256	6	119,985,100	*LOC728727*	9.17E-05	Yes	OK	0.2503
rs935920	2	35,985,642	*MRPL50P1*	9.24E-05	Yes	OK	0.03587
rs9480154	6	150,652,308	*RNU4P1*	9.25E-05	Yes	OK	0.2809

table continued on page 201

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs7559362	2	196,027,545	*LOC391470*	9.26E-05	Yes	OK	0.6706
rs2470179	15	49,471,703	*GLDN*	9.27E-05	ND	-	-
rs12777747	10	123,989,646	*TACC2*	9.30E-05	Yes	OK	0.1074
rs1224671	15	45,682,835	*SEMA6D*	9.32E-05	Yes	OK	0.0268
rs2708851	7	48,052,327	*LOC136288*	9.33E-05	Yes	OK	0.437
rs8014371	14	32,579,884	*NPAS3*	9.34E-05	Yes	OK	0.3856
rs1605527	3	21,351,855	*VENTXP7*	9.41E-05	Yes	OK	0.408
rs6440096	3	143,814,050	*PLS1*	9.52E-05	Yes	OK	0.1154
rs7485262	12	37,088,013	*ALG10B*	9.54E-05	Yes	OK	0.2864
rs13264187	8	22,562,043	*BIN3*	9.55E-05	Yes	OK	0.04817
rs560271	17	64,870,692	*ABCA5*	9.60E-05	ND	-	-
rs748088	21	38,320,720	*DSCR4*	9.66E-05	Yes	OK	0.01838
rs6596287	5	135,507,247	*SMAD5*	9.72E-05	Yes	OK	0.6976
rs9544996	13	78,590,677	*RBM26*	9.82E-05	Yes	OK	0.1583
rs7655536	4	77,395,792	*SCARB2*	9.86E-05	Yes	IMISS > 0.1	-
rs662616	13	78,593,828	*RBM26*	0.000101	Yes	OK	0.3867
rs359079	10	132,035,527	*LOC728497*	0.000102	Yes	OK	0.9392
rs10122587	9	2,681,951	*KCNV2*	0.000102	Yes	OK	0.253
rs999826	5	67,311,112	*LOC643631*	0.000103	Yes	OK	0.4743
rs11100188	4	159,655,208	*RXFP1*	0.000103	Yes	OK	0.1072
rs7225002	17	41,544,850	*KIAA1267*	0.000103	Yes	OK	0.007533
rs2168330	8	52,111,821	*SNTG1*	0.000105	Yes	IMISS > 0.1	-
rs11136568	8	2,922,585	*CSMD1*	0.000105	Yes	OK	0.351
rs1542384	4	159,664,766	*RXFP1*	0.000107	Yes	OK	0.5712
rs2000731	18	46,150,625	*C18orf24*	0.000107	Yes	OK	0.9792
rs305163	2	36,221,313	*CRIM1*	0.000107	Yes	NC	-
rs3759407	12	46,481,336	*HDAC7A*	0.000107	Yes	OK	0.3208
rs12460684	19	1,624,793	*TCF3*	0.000108	Yes	NC	-
rs12412945	10	54,468,113	*MBL2*	0.000109	Yes	OK	0.532
rs10064163	5	112,007,605	*APC*	0.000111	Yes	OK	0.3395
rs7224296	17	42,155,230	*LOC644315*	0.000113	Yes	OK	4.82E-05
rs1036745	13	78,499,702	*LOC390415*	0.000113	Yes	OK	0.6067
rs11128994	3	21,195,185	*VENTXP7*	0.000115	Yes	OK	0.02904
rs836109	11	34,526,784	*LOC729710*	0.000116	Yes	OK	0.9343
rs11747238	5	4,831,369	*LOC340094*	0.000116	Yes	OK	0.2399
rs4789632	17	69,702,146	*RPL38*	0.000116	Yes	OK	0.1311
rs4823506	22	46,643,740	*RP11-191L9*	0.000116	Yes	OK	0.531
rs6481654	10	30,401,130	*KIAA1462*	0.000116	Yes	OK	0.2113
rs4142010	1	39,938,607	*HPCAL4*	0.000116	Yes	OK	0.4934
rs706858	6	120,079,876	*LOC728727*	0.000117	Yes	OK	0.4436
rs11217299	11	98,250,101	*CNTN5*	0.000117	Yes	OK	0.6855
rs11185726	9	136,240,925	*RXRA*	0.000118	Yes	OK	0.2468
rs10044636	5	150,714,711	*SLC36A2*	0.00012	Yes	OK	0.1681
rs12361904	11	127,208,320	*LOC387820*	0.00012	Yes	OK	0.6577
rs4508240	12	37,186,069	*CPNE8*	0.000121	Yes	OK	0.3038
rs10515822	5	160,512,672	*GABRB2*	0.000122	Yes	OK	0.3855
rs2686830	7	47,957,563	*PKD1L1*	0.000123	Yes	OK	0.4921

table continued on page 202

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filter	*p* value[c] (Stage II)
rs11708730	3	72,008,678	*PROK2*	0.000123	Yes	OK	0.9601
rs974627	12	37,205,791	*CPNE8*	0.000124	Yes	OK	0.26
rs727549	7	83,065,910	*SEMA3E*	0.000125	Yes	OK	0.342
rs37391	5	123,553,864	*ZNF608*	0.000127	Yes	OK	0.7778
rs10839984	11	8,137,470	*RIC3*	0.000127	Yes	OK	0.9743
rs1639304	7	83,017,558	*SEMA3E*	0.000127	Yes	OK	0.148
rs1693389	7	83,014,198	*SEMA3E*	0.000127	Yes	OK	0.1697
rs6532197	4	91,016,324	*MMRN1*	0.000128	Yes	OK	0.00245
rs10136071	14	32,580,195	*NPAS3*	0.000132	Yes	OK	0.1549
rs2949065	2	36,044,594	*MRPL50P1*	0.000132	Yes	OK	0.04779
rs934397	2	237,371,678	*CXCR7*	0.000132	Yes	OK	0.07947
rs4478801	1	152,731,196	*SHE*	0.000133	Yes	OK	0.4124
rs9924308	16	30,062,241	*MAPK3*	0.000134	Yes	OK	0.464
rs972427	3	21,339,149	*VENTXP7*	0.000135	Yes	OK	0.00123
rs2681051	7	11,615,690	*KIAA0960*	0.000135	Yes	OK	0.0779
rs8078967	17	41,363,929	*MAPT*	0.000136	Yes	OK	0.9231
rs7900480	10	35,378,705	*CUL2*	0.000137	Yes	OK	0.2513
rs8034843	15	55,368,601	*TCF12*	0.000137	Yes	OK	0.9363
rs12118128	1	221,904,876	*LOC388743*	0.000137	Yes	OK	0.2898
rs643786	19	55,547,761	*NAPSA*	0.000138	Yes	OK	0.5972
rs922687	15	71,635,861	*NPTN*	0.000139	Yes	OK	0.01475
rs11082819	18	46,144,485	*C18orf24*	0.000141	Yes	OK	0.8669
rs4697508	4	24,576,450	*DKFZp761B107*	0.000142	Yes	OK	0.1544
rs11595185	10	25,231,376	*PRTFDC1*	0.000143	Yes	OK	0.04961
rs11107270	12	92,834,257	*CRADD*	0.000144	Yes	MAF < 0.05	-
rs1430961	4	90,771,943	*SNCA*	0.000145	Yes	OK	0.01876
rs1353615	7	83,013,723	*SEMA3E*	0.000147	Yes	OK	0.1337
rs1950712	14	68,550,317	*RPS29P1*	0.000147	Yes	OK	0.4661
rs11852946	15	36,975,713	*FLJ35695*	0.000148	ND	-	-
rs6473485	8	51,979,283	*SNTG1*	0.000150	Yes	OK	0.891
rs11656130	17	64,912,041	*MAP2K6*	0.00015	Yes	OK	0.7426
rs1035833	2	230,417,442	*TRIP12*	0.000152	Yes	OK	0.8748
rs4789636	17	69,710,434	*RPL38*	0.000152	Yes	OK	0.4873
rs7589111	2	131,887,358	*LOC389043*	0.000153	Yes	Bad cluster	-
rs16925839	10	70,277,867	*STOX1*	0.000153	Yes	OK	0.3681
rs12714369	2	3,041,245	*LOC729897*	0.000156	Yes	OK	0.01134
rs4742236	9	676,753	*ANKRD15*	0.000156	Yes	OK	0.1345
rs11222109	11	129,841,446	*ADAMTS15*	0.000157	Yes	OK	0.9962
rs3823906	7	83,067,501	*SEMA3E*	0.000159	Yes	OK	0.2898
rs12624568	20	2,187,669	*TGM3*	0.000160	ND	-	-
rs7251728	19	62,492,200	*ZNF460*	0.000160	Yes	NC	-
rs6044218	20	16,605,508	*RPL7AL3*	0.000160	Yes	OK	0.7729
rs11602361	11	122,068,548	*STS-1*	0.000164	Yes	OK	0.5877
rs1782322	10	70,156,331	*CCAR1*	0.000165	Yes	OK	0.7161
rs700802	9	78,424,532	*KIAA0367*	0.000166	Yes	OK	0.4141
rs1147255	12	125,968,187	*LOC121296*	0.000166	Yes	OK	0.4072
rs12817211	12	48,865,665	*LIMA1*	0.000166	Yes	OK	0.1543
rs13076171	3	71,988,841	*PROK2*	0.000167	Yes	OK	0.6182

table continued on page 203

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs2686821	7	47,941,607	*PKD1L1*	0.000167	Yes	OK	0.4518
rs333917	17	64,863,054	*ABCA5*	0.000167	Yes	OK	0.6572
rs11226125	11	103,418,178	*PDGFD*	0.000168	Yes	OK	0.02021
rs10826765	10	30,447,990	*KIAA1462*	0.000170	Yes	OK	0.5874
rs12130076	1	238,984,728	*LOC645939*	0.000171	Yes	OK	0.4978
rs6079575	20	14,755,599	*C20orf133*	0.000171	Yes	MAF < 0.05	-
rs1438339	3	29,457,050	*RBMS3*	0.000173	Yes	OK	0.4274
rs7164298	15	55,370,593	*TCF12*	0.000173	Yes	OK	0.4743
rs10826764	10	30,447,878	*KIAA1462*	0.000174	Yes	OK	0.7499
rs7678831	4	159,660,989	*RXFP1*	0.000174	Yes	OK	0.05006
rs2317515	8	137,044,531	*KHDRBS3*	0.000175	Yes	OK	0.5258
rs9957722	18	46,100,998	*CXXC1*	0.000176	Yes	OK	0.9822
rs681210	21	43,603,771	*FLJ41733*	0.000177	Yes	OK	0.8667
rs3924079	5	117,871,768	*DTWD2*	0.000180	Yes	HWE < 10^{-7}	-
rs9988732	10	30,438,875	*KIAA1462*	0.000180	Yes	OK	0.1922
rs699052	12	25,077,644	*LOC645177*	0.000181	Yes	OK	0.3761
rs4505549	2	195,996,381	*LOC391470*	0.000182	Yes	OK	0.7072
rs1443562	8	92,997,929	*MRPS16P1*	0.000186	Yes	OK	0.8276
rs997120	22	31,438,536	*SYN3*	0.000187	Yes	OK	0.2715
rs7813504	8	88,868,935	*SOX5P*	0.000187	Yes	OK	0.2821
rs4534200	9	25,642,508	*TUSC1*	0.000187	Yes	OK	0.2083
rs787274	9	114,590,184	*SNX30*	0.000187	Yes	Bad cluster	-
rs1481088	15	71,629,314	*NPTN*	0.000189	Yes	OK	0.03829
rs6851219	4	77,398,854	*SCARB2*	0.000190	Yes	OK	0.005657
rs2141202	23	144,519,417	*LOC347422*	0.000191	Yes	HWE < 10^{-7}	-
rs4784045	16	57,094,125	*NDRG4*	0.000191	Yes	OK	0.5436
rs3764941	5	135,497,426	*SMAD5*	0.000191	Yes	OK	0.7109
rs4965303	15	98,641,293	*ADAMTS17*	0.000192	Yes	OK	0.7292
rs2243988	21	35,124,310	*RUNX1*	0.000192	Yes	OK	0.0488
rs13273663	8	51,712,711	*SNTG1*	0.000193	ND	-	-
rs999867	10	104,494,554	*SFXN2*	0.000194	Yes	OK	0.000559
rs11705555	22	26,536,912	*MN1*	0.000194	Yes	OK	0.9235
rs9619283	22	31,363,048	*SYN3*	0.000195	Yes	OK	0.2266
rs1882832	14	74,870,997	*FOS*	0.000195	Yes	OK	0.06192
rs2736050	8	15,718,351	*LOC137012*	0.000196	Yes	OK	0.7151
rs4775330	15	59,022,174	*RORA*	0.000196	Yes	OK	0.7178
rs12172730	1	104,913,058	*LOC642337*	0.000198	Yes	OK	0.02137
rs9288651	2	230,439,641	*TRIP12*	0.000198	Yes	OK	0.9841
rs1108089	1	221,907,077	*LOC388743*	0.000198	Bad cluster	-	-
rs1533969	2	36,036,959	*MRPL50P1*	0.000198	Yes	OK	0.124
rs2710697	12	53,684,257	*NEUROD4*	0.000199	Yes	OK	0.7104
rs7622285	3	146,367,022	*GM2AP*	0.000201	Yes	OK	0.5467
rs1863985	16	57,181,595	*CNOT1*	0.000201	Yes	OK	0.06933
rs10908183	11	68,962,197	*LOC390218*	0.000201	Yes	OK	0.6039
rs1662694	17	64,869,168	*ABCA5*	0.000207	Yes	OK	0.7883
rs10854627	22	31,433,918	*SYN3*	0.000207	Yes	IMISS > 0.1	-
rs2102480	3	87,657,428	*LOC643766*	0.000207	ND	-	-
rs13081379	3	193,918,614	*FGF12*	0.000208	Yes	OK	0.6943

table continued on page 204

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs7767977	6	5,530,769	*FARS2*	0.0002075	ND	-	-
rs2009817	8	15,721,394	*LOC137012*	0.0002089	Yes	OK	0.9717
rs9374809	6	119,977,195	*LOC728727*	0.0002103	Yes	OK	0.2666
rs10504370	8	65,150,536	*LOC729415*	0.0002138	Yes	OK	0.911
rs4340037	11	8,084,734	*TUB*	0.0002142	Yes	OK	0.8093
rs6800573	3	143,828,507	*PLS1*	0.0002171	Yes	OK	0.1516
rs4669060	2	6,429,026	*LOC391349*	0.0002179	Yes	NC	-
rs903056	3	14,785,294	*C3orf20*	0.000218	Yes	Bad cluster	-
rs880183	19	2,000,314	*MKNK2*	0.0002187	Yes	OK	0.2389
rs1491923	12	38,877,384	*LRRK2*	0.0002197	Yes	OK	0.007736
rs2487707	10	70,435,098	*KIAA1279*	0.0002243	Yes	OK	0.459
rs264122	5	129,675,680	*CHSY-2*	0.0002248	Yes	OK	0.6788
rs1223271	20	13,244,912	*C20orf82*	0.0002276	Yes	OK	0.03009
rs11248060	4	954,359	*DGKQ*	0.0002278	Yes	OK	0.0464
rs6800015	3	180,262,240	*ZMAT3*	0.0002301	Yes	OK	0.5391
rs17275640	1	176,155,379	*SEC16B*	0.0002313	Yes	Bad cluster	-
rs2169166	8	105,999,385	*LOC644103*	0.0002325	Yes	OK	0.8604
rs7666265	4	77,395,305	*SCARB2*	0.0002325	Yes	OK	0.03444
rs2586469	17	45,682,246	*LOC729160*	0.0002351	Yes	OK	0.214
rs6903627	6	120,169,406	*LOC728727*	0.000236	Yes	OK	0.8461
rs4517741	15	90,711,310	*ST8SIA2*	0.0002373	Yes	OK	0.6458
rs7920137	10	73,806,703	*CBARA1*	0.0002383	Yes	OK	0.1865
rs10765137	10	129,278,636	*DOCK1*	0.0002387	Yes	OK	0.8456
rs9377145	6	148,961,594	*SASH1*	0.000241	Yes	OK	0.544
rs2962101	5	35,171,498	*PRLR*	0.0002432	Yes	OK	0.201
rs11766212	7	47,910,478	*PKD1L1*	0.0002434	Yes	IMISS > 0.1	-
rs9299039	9	680,460	*ANKRD15*	0.0002436	Yes	OK	0.192
rs473532	2	165,021,032	*GRB14*	0.0002438	Yes	OK	0.7161
rs1369642	15	45,692,185	*SEMA6D*	0.0002442	Yes	OK	0.02119
rs2512139	11	117,320,774	*TMPRSS13*	0.0002451	Yes	OK	0.2448
rs13437473	6	72,711,578	*RIMS1*	0.0002457	Yes	OK	0.9729
rs356229	4	90,825,620	*SNCA*	0.000246	Yes	OK	1.25E-06
rs7671488	4	122,618,788	*LOC729112*	0.0002491	Yes	OK	0.2968
rs398293	9	118,023,823	*PAPPA*	0.0002494	Yes	OK	0.5005
rs9846960	3	143,790,760	*ATR*	0.0002503	Yes	OK	0.7395
rs10520381	4	178,769,035	*LOC285500*	0.0002518	Yes	OK	0.7613
rs2051569	22	31,355,180	*SYN3*	0.0002518	Yes	OK	0.1105
rs9593152	13	75,495,388	*FLJ35379*	0.0002521	Yes	OK	0.3368
rs2616510	8	89,087,241	*MMP16*	0.0002531	Yes	OK	0.1443
rs7625872	3	45,132,572	*CDCP1*	0.0002542	Yes	OK	0.6413
rs4769388	13	24,393,134	*CENPJ*	0.0002548	Yes	OK	0.9777
rs4505777	4	164,419,696	*LOC133332*	0.0002595	Yes	OK	0.772
rs6427069	1	165,391,458	*LOC391130*	0.0002597	Yes	OK	0.8003
rs207465	21	24,120,442	*TUBAP*	0.0002599	Yes	IMISS > 0.1	-
rs16959883	15	45,698,555	*SEMA6D*	0.0002604	Yes	OK	0.0149
rs1881925	3	110,444,079	*DPPA2*	0.0002624	Yes	OK	0.01282
rs1484127	8	51,888,207	*SNTG1*	0.000263	Yes	OK	0.9963
rs13192471	6	32,779,081	*HLA-DQB1*	0.000265	Yes	OK	0.8693

table continued on page 205

Table S4 (continued): SNPs Selected for the Replication Stage

SNP ID	Chr.	Position[a]	Gene	*p* value[b] (Stage I)	Designable for Stage II	QC filtering	*p* value[c] (Stage II)
rs1613367	11	104,416,241	*COPl*	0.00026	Yes	IMISS > 0.1	-
rs26286	5	14,219,402	*TRIO*	0.00027	Yes	OK	0.1182
rs11136092	8	22,522,433	*KIAA1967*	0.00027	Yes	OK	0.08695
rs996243	6	137,895,153	*LOC391040*	0.00027	Yes	OK	0.2965
rs7094852	10	129,223,387	*DOCK1*	0.00027	Yes	OK	0.1899
rs2150279	14	53,343,338	*BMP4*	0.00027	Yes	OK	0.08
rs1868108	3	162,073,563	*PPM1L*	0.00027	Yes	NC	-
rs7854502	9	686,683	*ANKRD15*	0.00027	Yes	OK	0.1234
rs19334	8	9,047,316	*PPP1R3B*	0.00027	Yes	OK	0.1008
rs2046065	12	65,458,480	*GRIP1*	0.00027	Yes	OK	0.01911
rs925030	8	25,298,518	*DOCK5*	0.00027	Yes	OK	0.1013
rs9842991	3	70,554,058	*LOC654340*	0.00028	Yes	OK	0.8779
rs9876540	3	45,137,320	*CDCP1*	0.00028	Yes	OK	0.7709
rs6480643	10	74,204,180	*CCDC109A*	0.00028	Bad cluster	-	-
rs9489765	6	120,038,756	*LOC728727*	0.00028	Yes	OK	0.4417
rs10996742	10	67,294,261	*CTNNA3*	0.00028	Yes	OK	0.4747
rs2896159	7	114,989,770	*TFEC*	0.00028	Yes	OK	0.1442
rs10996743	10	67,294,317	*CTNNA3*	0.00028	Yes	OK	0.5027
rs1453815	4	24,566,751	*DKFZp761*	0.00028	Yes	OK	0.1274
rs11096577	2	19,114,014	*FLJ41481*	0.00028	Yes	OK	0.6959
rs7024926	9	82,766,092	*TLE1*	0.00029	Yes	OK	0.2269
rs2878172	14	54,443,420	*GCH1*	0.00029	Yes	OK	0.3818
rs10510622	3	29,670,592	*RBMS3*	0.00029	Yes	OK	0.6978
rs4422788	8	51,843,584	*SNTG1*	0.00029	Yes	OK	0.9555
rs6050867	20	25,678,968	*LOC728882*	0.00029	Yes	OK	0.5659
rs187286	1	215,659,907	*GPATCH2*	0.00029	Yes	OK	0.2446
rs1027493	1	165,399,506	*LOC391130*	0.00029	Yes	OK	0.02147

This table has been published elsewhere (see reference [272]). Excluded SNPs are highlighted in grey.

[a] SNP positions are shown according to NCBI genome build 36.3.

[b] *p* values are shown according to the trend test model in the screening stage

[c] *p* values are shown according to the trend test model in the replication stage

Chr., chromosome; HWE, Hardy-Weinberg equilibrium; IMISS > 0.1, individual SNP call rate >90%; MAF, minor allele frequency; NC, nonconcordant; ND, not designable

Table S5. Comparison with Previously Reported Candidate Loci from Three GWA Studies in PD

SNP ID	Chr.	Position[a]	Candidate gene	*p* value in this study[b] (trend model)	OR (95% CI)
Maraganore et al.					
rs7702187	5	9,385,281	*SEMA5A*	not in filtered dataset	-
rs10200894	2	228,525,376	-	not in filtered dataset	-
rs2313982	4	139,007,510	-	0.52	1.0 (0.8 - 1.1)
rs17329669	7	36,818,454	-	not in filtered dataset	-
rs7723605	5	5,407,615	-	not in filtered dataset	-
ss46548856	10	58,986,929	-	not in filtered dataset	-
rs16851009	2	166,338,953	*GALNT3*	not in filtered dataset	-
rs2245218	1	14,012,413	*PRDM2*	not in filtered dataset	-

table continued on page 206

Table S5 (continued). Comparison with Previously Reported Candidate Loci from Three GWA Studies in PD

SNP ID	Chr.	Position[a]	Candidate gene	*p* value in this study[b] (trend model)	OR (95% CI)
Maraganore et al.					
rs7878232	X	150,597,031	*PASD1*	not in filtered dataset	-
rs1509269	4	138,973,174	-	not in filtered dataset	-
rs11737074	4	125,300,823	-	not in filtered dataset	-
Fung et al.					
rs10501570	11	84,095,494	*DLG2*	not in filtered dataset	-
rs281357	17	19,683,106	*ULK2*	not in filtered dataset	-
rs2242330	4	68,129,844	*BRDG1*	not in filtered dataset	-
rs1480597	10	44,481,115	-	not in filtered dataset	-
rs6826751	4	68,116,450	*BRDG1*	not in filtered dataset	-
rs4888984	16	78,066,835	-	not in filtered dataset	-
rs4862792	4	188,438,344	-	0.10	0.9 (0.8 - 1.0)
rs3775866	4	68,126,775	*BRDG1*	0.08	1.1 (1.0 - 1.2)
rs2235617	20	47,988,384	*ZNF313*	not in filtered dataset	-
rs988421	1	72,322,424	*NEGR1*	0.22	1.1 (0.9 - 1.1)
rs7097094	10	44,530,696	-	not in filtered dataset	-
rs999473	10	44,502,322	-	0.74	1.0 (0.9 - 1.1)
rs1912373	11	56,240,441	-	not in filtered dataset	-
rs1887279	1	182,176,783	*GLT25D2*	0.02	0.9 (0.8 - 1.0)
rs2986574	1	182,173,237	*GLT25D2*	not in filtered dataset	-
rs11090762	22	46,133,989	-	not in filtered dataset	-
rs6125829	20	48,002,336	*ZNF313*	not in filtered dataset	-
rs7796855	7	49,627,992	-	0.25	1.0 (1.0 - 1.1)
rs355477	4	68,079,120	*BRDG1*	not in filtered dataset	-
rs3010040	1	182,174,845	*GLT25D2*	not in filtered dataset	-
rs2296713	1	182,176,340	*GLT25D2*	not in filtered dataset	-
rs355461	4	68,063,319	*BRDG1*	not in filtered dataset	-
rs355506	4	68,068,677	*BRDG1*	not in filtered dataset	-
rs355464	4	68,061,719	*BRDG1*	not in filtered dataset	-
rs1497430	4	68,040,409	*BRDG1*	not in filtered dataset	-
rs11946612	4	68,018,566	*BRDG1*	not in filtered dataset	-
Pankratz et al.					
rs1564282	4	842,313	*GAK/DGKQ*	1.44E-03	1.2 (1.1 - 1.4)
rs11248051	4	848,332	*GAK/DGKQ*	not in filtered dataset	-
rs11248060	4	954,359	*GAK/DGKQ*	2.28E-04	1.2 (1.1 - 1.4)
rs4811072	20	48,519,524	*COX6CP2/PTPN1*	not in filtered dataset	-
rs1997791	20	48,529,835	*COX6CP2/PTPN1*	not in filtered dataset	-
rs2654735	4	112,618,062	*LOC729075*	0.67	1.0 (0.9 - 1.1)
rs1806506	4	112,686,700	*LOC729075*	0.88	1.0 (0.9 - 1.1)
rs11729080	4	112,723,321	*LOC729075*	0.67	1.0 (0.9 - 1.1)
rs4736788	8	40,947,586	*LOC727725/ZMAT4*	not in filtered dataset	-
rs10094981	8	40,950,451	*LOC727725/ZMAT4*	not in filtered dataset	-
rs898528	17	74,678,398	*HRNBP3*	not in filtered dataset	-
rs12871648	13	113,018,663	*LAMP1*	not in filtered dataset	-
rs4670322	2	33,309,246	*LTBP1*	0.39	1.0 (0.9 - 1.1)
rs11592212	10	110,407,383	-	not in filtered dataset	-
rs4106153	4	90,463,499	*SNCA/GPRIN3/MMRN1*	0.53	1.0 (0.9 - 1.1)
rs1504489	4	90,477,611	*SNCA/GPRIN3/MMRN1*	6.60E-03	1.1 (1.0 - 1.2)

table continued on page 207

Table S5 (continued). Comparison with Previously Reported Candidate Loci from Three GWA Studies in PD

SNP ID	Chr.	Position[a]	Candidate gene	*p* value in this study[b] (trend model)	OR (95% CI)
Pankratz et al.					
rs356229	4	90,825,620	*SNCA/GPRIN3/MMRN1*	2.46E-04	1.2 (1.1 - 1.3)
rs356188	4	90,910,560	*SNCA/GPRIN3/MMRN1*	2.07E-03	0.8 (0.8 - 0.9)
rs3775478	4	91,061,863	*SNCA/GPRIN3/MMRN1*	6.07E-04	1.3 (1.1 - 1.5)
rs4431442	6	100,320,236	*PRDM13/MCHR2*	0.26	1.0 (0.9 - 1.0)
rs10937194	3	186,201,412	*VPS8*	0.36	1.0 (0.9 - 1.0)
rs4901519	14	54,088,930	*CGRRF1/SAMD4A*	not in filtered dataset	-
rs11012	17	40,869,224	*C17orf69/PLEKHM1/MAPT*	not in filtered dataset	-
rs1724425	17	41,137,530	*C17orf69/PLEKHM1/MAPT*	not in filtered dataset	-
rs12638253	3	158,108,785	*LEKR1*	0.84	1.0 (0.9 - 1.1)
rs9655034	7	39,258,636	*POU6F2*	0.14	1.1 (1.0 - 1.2)
rs1197313	3	134,583,142	*TMEM108*	0.04	0.9 (0.8 - 1.0)
rs7312607	12	95,350,301	*LOC728328/PCTK2*	not in filtered dataset	-
rs9859577	3	193,571,219	*FGF12*	not in filtered dataset	-
rs2108521	12	124,901,417	*LOC652429/TMEM132B*	not in filtered dataset	-

[a] SNP position is shown according to NCBI Build 36.3.

[b] *p* values for the screening stage data are shown.

Chr., chromosome; OR, odds ratio; CI, confidence interval

Table S6. Primer Sequences and Fragment Lengths

SNP	Product length (bp)	Forward primer	Reverse primer
rs2856336	198	GGTAGACAGCTGCCACATCAAG	CAAAGCAAAGAGCACCAAGC
rs11931074	286	TATTCCGCCCATCCTGTG	ACCCATTTGGCTATTGTTATTG
rs9480154	168	ACTGGAGAGCTGAGGGTGAG	TGAAGTTAACCTGGCAATCACTAC
rs2515501	319	ATAGTCGGATGGCACACCTG	GTTCCATGTGGCCTTCTCTTC
rs3857059	424	TTGGTATCGTCATGGAATTTG	ACTGAGGCGATGTTTCAAGTAG
rs2794256	321	TTACAAATCGCATGAAGGAGAC	CCTGCCTGTGCAATCCTTTG
rs10515822	181	GGAGTGATGAGGTGAATAGCAC	AGTGCCCAGAGTCCCTAAAG
rs13139027	321	TACATGGAGCACACCTCAGAGC	AGTGAGTTCTCTGGAGCTGGTC
rs2042079	298	GTGAGCTAGAAGACACTATAGGGC	CAGGATAAGCATGATGTTCTCC
rs2896159	376	GGCATCACTCTCTTATCTCTGTTC	TCCCATAATGCCACAAAGATAG

This table has been published elsewhere (see reference [289]).

Table S7. PCR and Sequencing Primers for *PRKN* and *PINK1*

Exon	Product length (bp)	Forward primer	Reverse primer
PRKN			
exon 1	112	GCGCGGCTGGCGCCGCTGCGCGCA	GCGGCGCAGAGAGGCTGTAC
exon 2	308	ATGTTGCTATCACCATTTAAGGG	AGATTGGCAGCGCAGGCGGCATG
exon 3	516	TCAGTGTTGTTTGTCTACCGTG	ACTAAATATGCACCCGGTGAG
exon 4	261	ACAAGCTTTTAAAGAGTTTCTTGT	AGGCAATGTGTTAGTACACA
exon 5	326	ACCTAGCACATCCCTTGAAAG	GGCAAACAGTGAAGATGTCATC
exon 6	378	CTGGGAAAGGTTTGATGCTG	AAGGCTCGTGTGGCAGAAC
exon 7	239	TGCCTTTCCACACTGACAGGTACT	TCTGTTCTTCATTAGCATTAGAGA
exon 8	541	CCAGGGTCACATGCAACTG	GCTTGAGTCTTGAGATGAAATGC
exon 9	278	GGGTGAAATTTGCAGTCAGT	AATATAATCCCAGCCCATGTGCA
exon 10	507	GACCCATCATCTCTGTAGCCC	CCACTGGGAAGTCTACACTGTG
exon 11	303	ACAGGGAACATAAACTCTGATCC	CAACACACCAGGCACCTTCAGA
exon 12	255	GTTTGGGAATGCGTGTTTT	AGAATTAGAAAATGAAGGTAGACA
PINK1			
exon 1	730	GGGAAAGTCACTGCTAGAGGC	AATAGATGAGCTTTAAGGACCCTC
exon 2	526	TTTATTGATCTGGTCGACGTG	ACCTTTCCTGTGGATAATCTGTC
exon 3	400	TATCTCGAAGGTCAGAGCCAATTC	TACCTGCAATCTTTCCCAAGTG
exon 4	400	TGTGGCCTTAGGTTATTCTTTCC	TCCCTTTGCATGGCTTTC
exon 5	299	ACGTATTGGGAGTCGTCGAT	CCTGAAGAGTCAGTCCTAAATGC
exon 6	301	GCTATGTCTTGCTGGTGGCTTTA	CAAGGCATCGAGTCTCCTGC
exon 7	420	TTAGCCCATGGATCAGGTG	TCTGTCACTGTGGCTCTGGC
exon 8	501	GGAAGAATTGGGTTGGGACC	CGAGGCCTTTTCCGGCTA

This table has been published elsewhere (see reference [298]).

Table S8. 60 Touchdown 50 PCR Cycling Program

	1 cycle	8 cycles			20 cycles			16 cycles		
Phase	Den.	Den.	Ann.	Ext.	Den.	Ann.	Ext.	Den.	Ann.	Ext.
Temp.°C	94°C	94°C	60°C	72°C	94°C	60°C	72°C	94°C	50°C	72°C
Temp. Inc.°C	0°C	0°C	0°C	0°C	0°C	-0.5°C	0°C	0°C	0°C	0°C
Duration	4min	30sec	30sec	30sec	30sec	30sec	30sec	30sec	20sec	30sec

	1 cycle
Phase	Ext.
Temp.°C	72°C
Temp. Inc.°C	0°C
Duration	5min

Den., denaturation
Ann., annealing
Ext., extension

Table S9. *PRKN* Primers Used for Gene Dosage Experiments

Exon	Product length (bp)	Forward primer	Reverse primer
exon 1	62	CCACCTACCCAGTGACCATGA	CGGCGCAGAGAGGCTGTA
exon 2	68	CCCAGTGGAGGTCGATTCTG	CCCCCTGTCGCTTAGCAA
exon 3	74	TTTACCTTGCTCCCAAACAGAATT	TCTCCACGGTCTCTGCACAA
exon 4	82	TTCTTCTCCAGCAGGTAGATCAATC	TTTTCCCGGCTGCACTCTT
exon 5	77	CCGGATGAGTGGTGAATGC	AGAGGAATGAATGTGACCAGGTACT
exon 6	76	GCACACCCCACCTCTGACA	TGCAAGTGATGTTCCGACTATTTG
exon 7	64	CCGCCACGTGATTTGCTTA	CTGCCGATCATTGAGTCTTGTC
exon 8	61	GCAGCCTTTGAGATGCTCACT	AGAGCTCCATCACTTCAGGATTCT
exon 9	64	GGACACACTCCTCTGCACCAT	CAATCTGCTTTTTGGGTTTTGC
exon 11	56	GCTCGGCGGCTCTTTCA	ACGCCTTTCCTCTTTGTTTCC
exon 12	64	CGAACCCACCACACCTTTGT	TGCGGACACTTCATGTGCAT
β-globin	68	TGGGCAACCCTAAGGTGAAG	GTGAGCCAGGCCATCACTAAA

This table has been published elsewhere (see reference [298]).

Table S10. *PRKN* Gene Dosage Probes

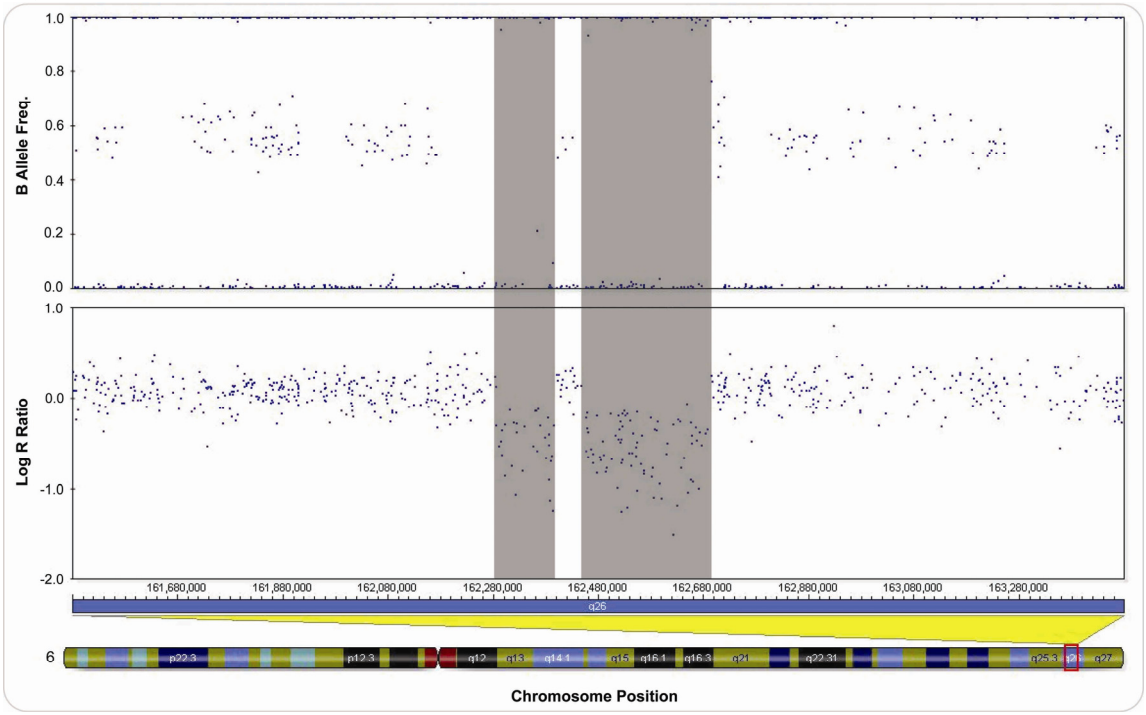
Exon	Label	Sequence
exon 1	VIC	TACGTGGGTACCTGCC
exon 2	VIC	CCAGCATCTTCCAGCTCAAGGAGGTG
exon 3	VIC	TGACCTGGATCAGCAGAGCATTGTTCA
exon 4	VIC	TTTTATGTGTATTGCAAAGGCCCCTGTCA
exon 5	VIC	CCACACTGCCCTGGGACTAGTGCA
exon 6	VIC	AAACATCAGTAGCTTTGCACCTGATCGCA
exon 7	VIC	CTGTTTCCACTTATACTGTG
exon 8	VIC	ACCTGCTCTTCTCC
exon 9	VIC	CTGCTGGTACCGGTTG
exon 11	VIC	CGACTCTGTAGGCCTG
exon 12	VIC	TTCTGCCCCCAACAGGAGGCTG
β-globin	FAM	CTCATGGCAAGAAAGTGCTCGGTGC

This table has been published elsewhere (see reference [298]).

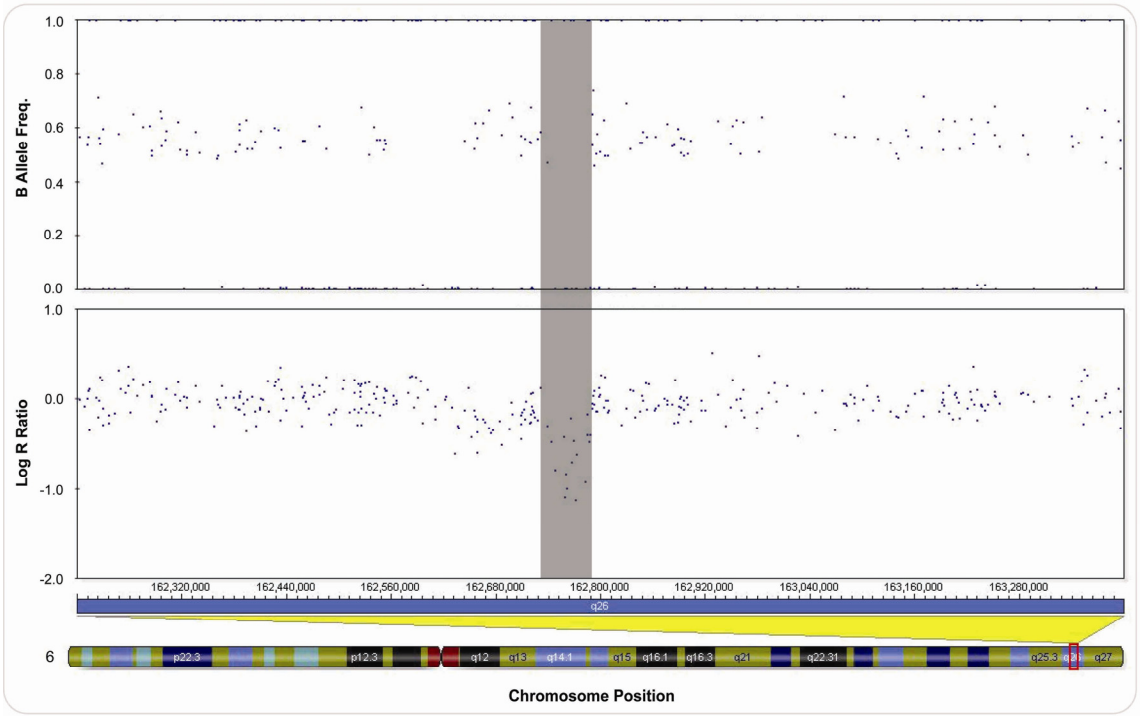
Table S11. Cycling Conditions for Gene Dosage Experiments

	1 cycle	1 cycle	40 cycles	
Temp. °C	50°C	95°C	95°C	60°C
Temp. Inc. °C	0°C	0°C	0°C	0°C
Duration	2min	10min	15sec	1min

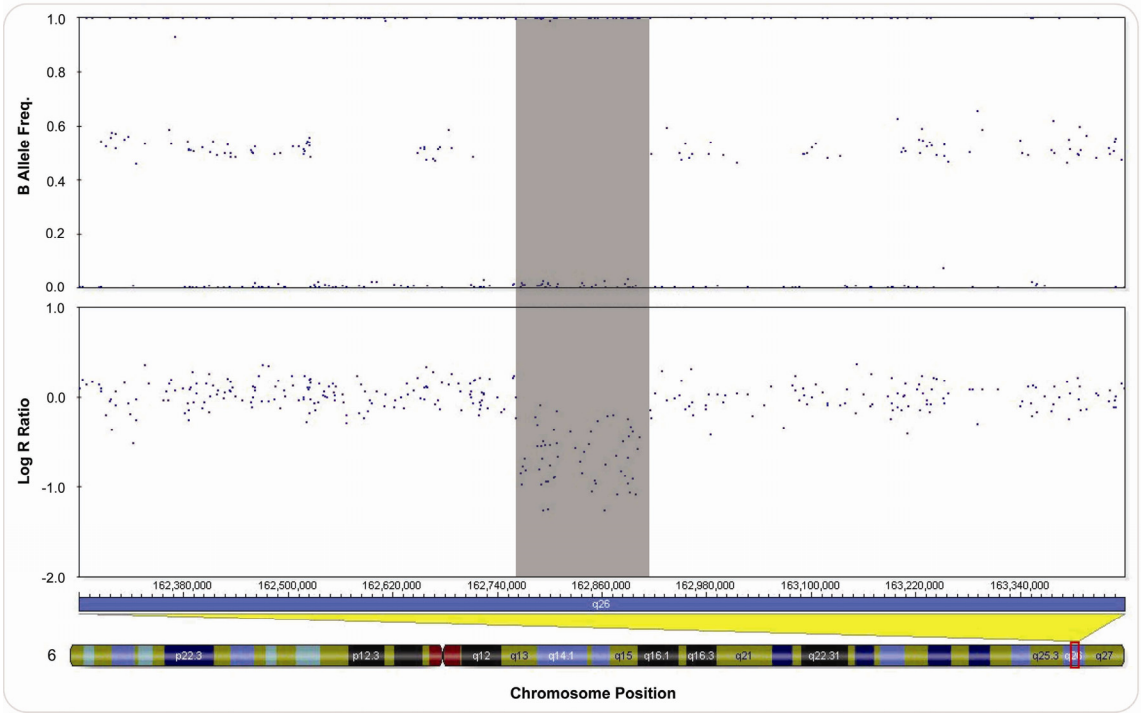
Figure S1. Heterozygous Deletions at the *PRKN* Locus in Four PD Patients



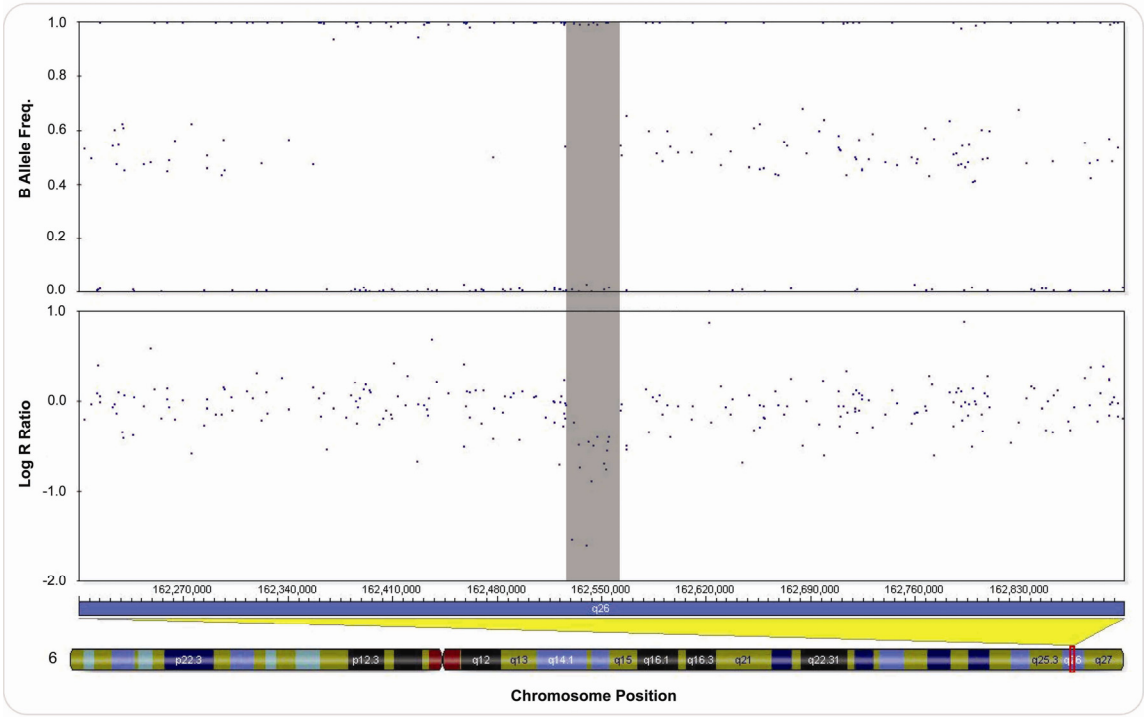
Shown are two heterozygous deletions (highlighted in grey) in sample ND00153 embracing exons 5-6 in *PRKN*.



Shown is a heterozygous deletion (highlighted in grey) in sample ND01119 involving exon 2 of *PRKN*.

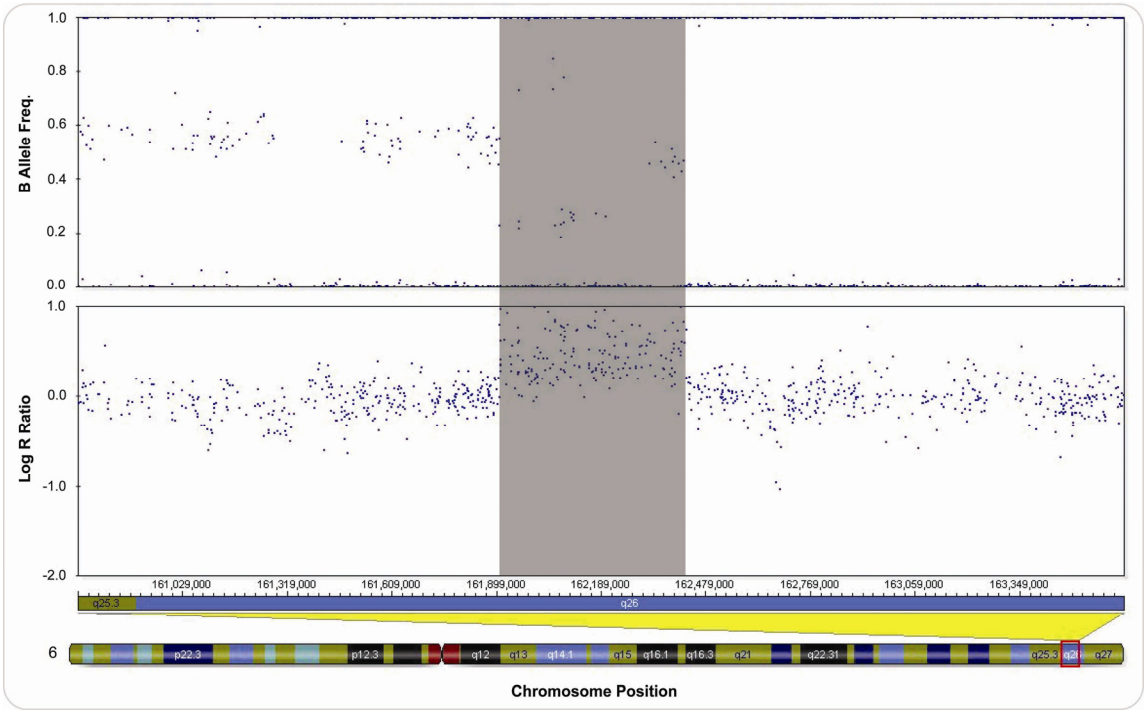


Shown is a heterozygous deletion (highlighted in grey) in sample ND04581 involving exon 2 of *PRKN*.

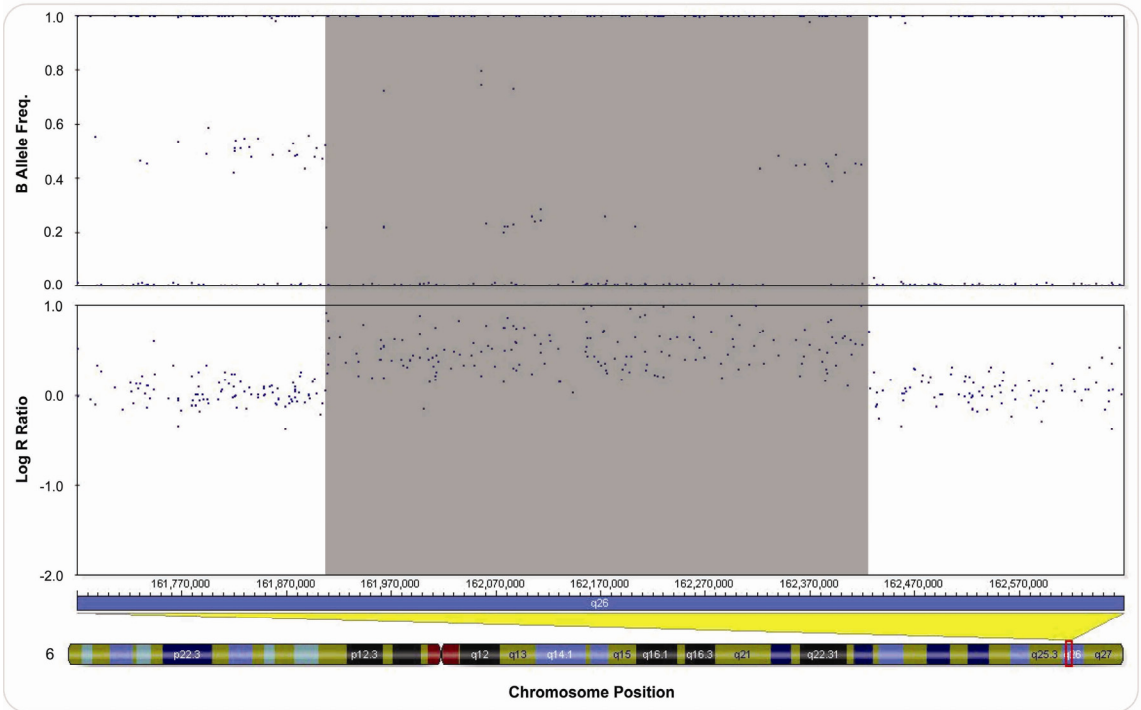


Shown is a heterozygous deletion (highlighted in grey) in sample ND05921 involving exon 4 of *PRKN*.

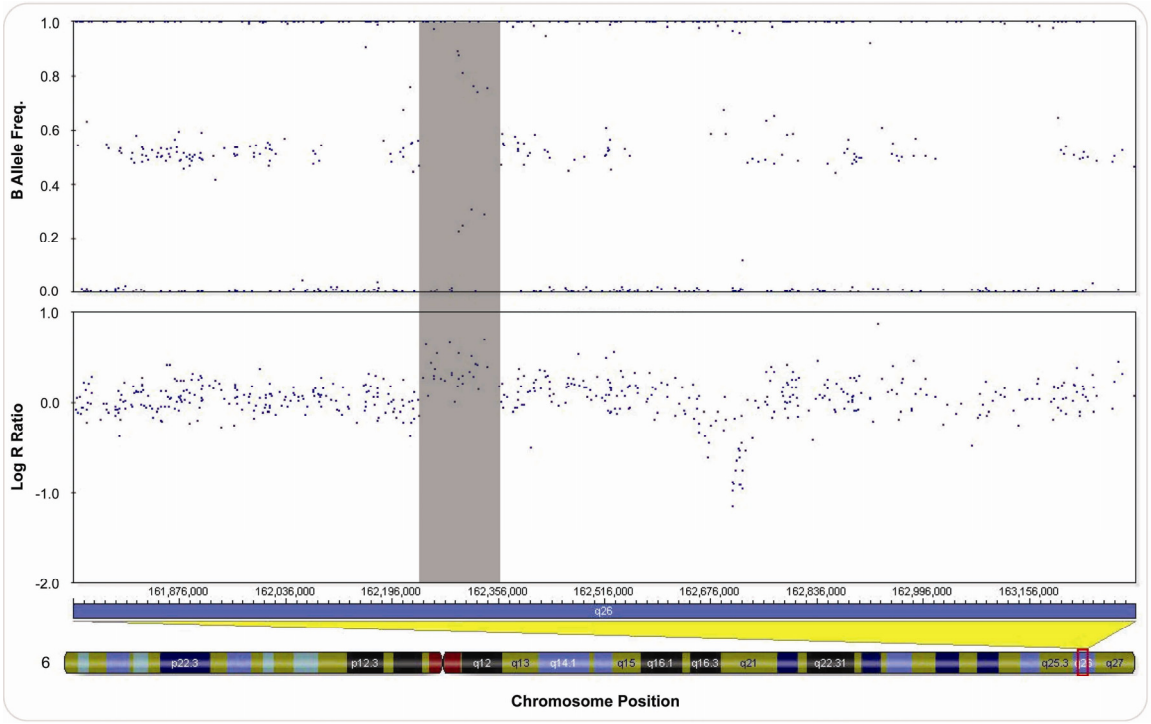
Figure S2. Heterozygous Duplications at the *PRKN* Locus in Four PD Patients



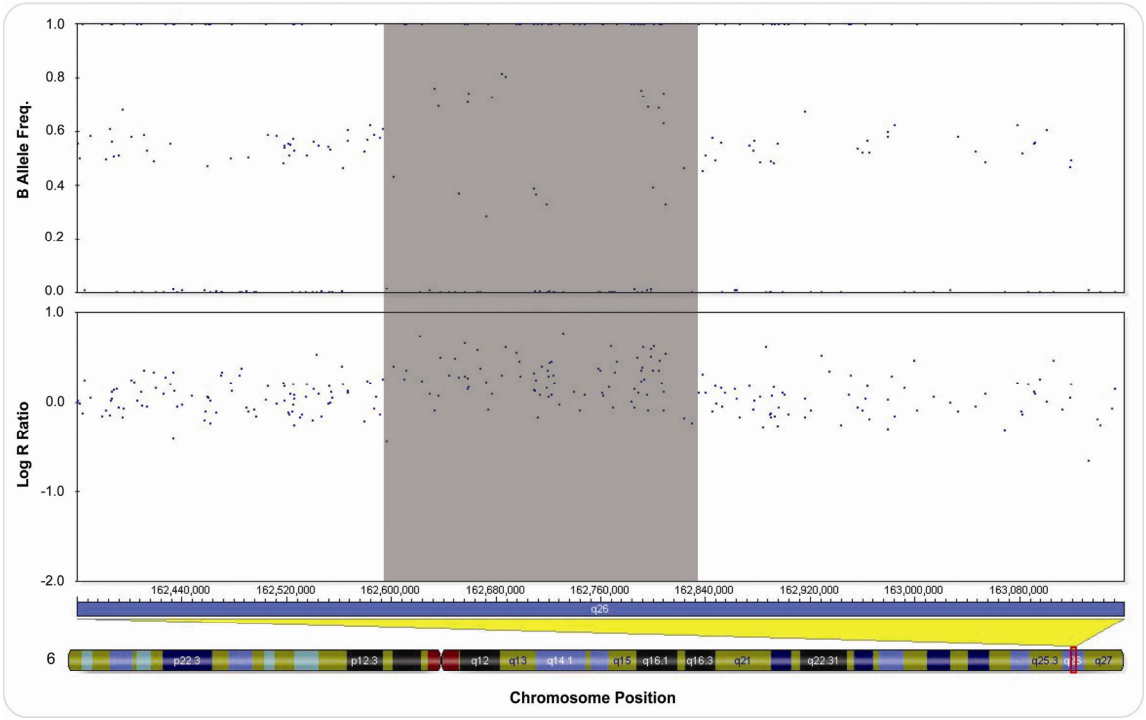
Shown is a heterozygous duplication (highlighted in grey) in sample ND00187 involving exons 5-9 of *PRKN*.



Shown is a heterozygous duplication (highlighted in grey) in sample ND00429 involving exons 5-9 of *PRKN*. Individual ND00429 is a sibling of ND00187 (shown above).

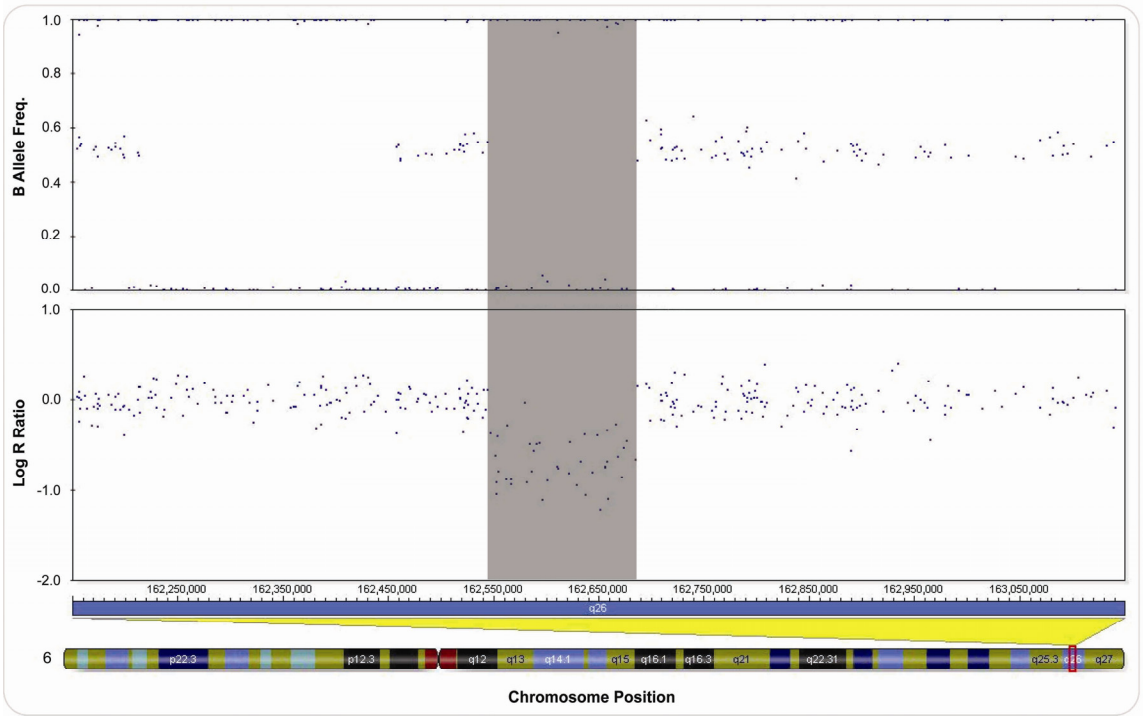


Shown is a heterozygous duplication (highlighted in grey) in sample ND02789 involving exon 6 of *PRKN*.

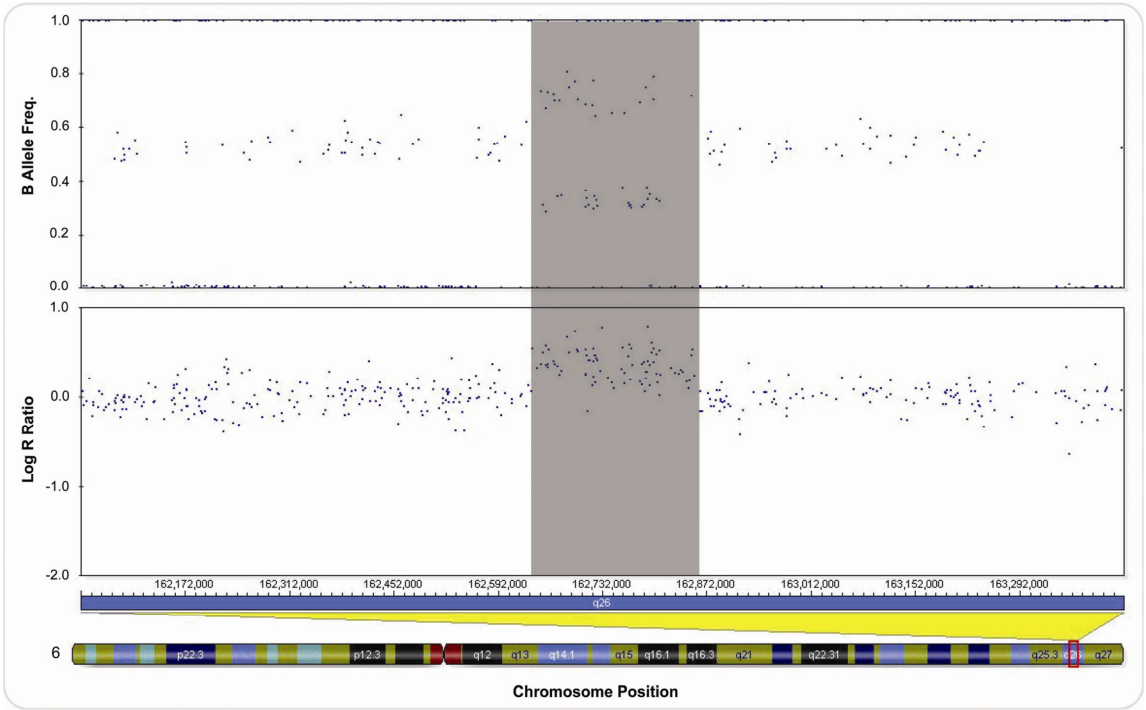


Shown is a heterozygous duplication (highlighted in grey) in sample ND06330 involving exons 2-3 of *PRKN*.

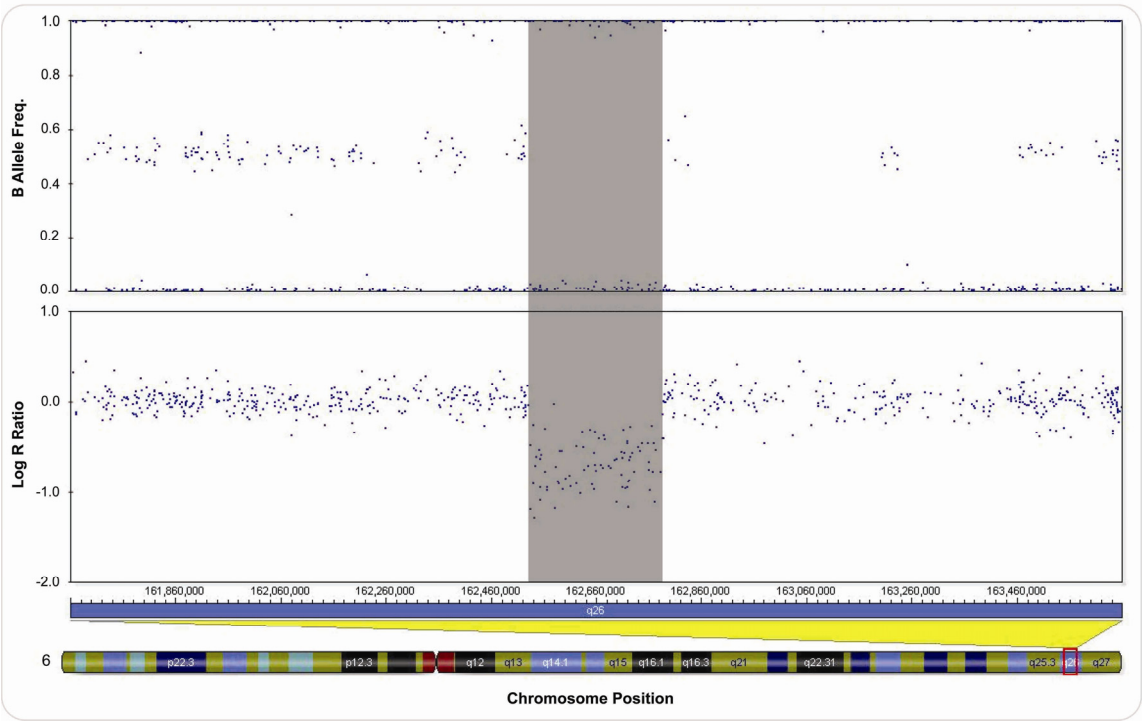
Figure S3. Copy Number Variants at the *PRKN* Locus in Three Normal Controls



Shown is a heterozygous deletion (highlighted in grey) in sample ND03967 involving exon 3 of *PRKN*.



Shown is a heterozygous duplication (highlighted in grey) in sample ND04990 involving exon 2 of *PRKN*.



Shown is a heterozygous deletion (highlighted in grey) in sample ND08538 involving exons 2-4 of *PRKN*.