

VNIVERSITAT Đ VALÈNCIA

Facultad de Biología

Departamento de Biología Celular y Parasitología

**“GENETIC BASIS OF PARKINSON’S DISEASE  
AND RELATED DISORDERS”**

Doctoral thesis

Javier Simón Sánchez

Valencia, 2008





VNIVERSITAT Đ VALÈNCIA

Facultad de Biología

Departamento de Biología Celular y Parasitología

**“GENETIC BASIS OF PARKINSON’S DISEASE  
AND RELATED DISORDERS”**

Doctoral thesis

Javier Simón Sánchez

Valencia, 2008







MINISTERIO DE  
CIENCIA E  
INNOVACIÓN



CONSEJO SUPERIOR DE  
INVESTIGACIONES  
CIENTÍFICAS

D. Jordi Pérez Tur, Doctor en Biología y Científico Titular del Instituto de Biomedicina de Valencia del Consejo Superior de Investigaciones Científicas (CSIC),

**Certifica:**

que **Javier Simón Sánchez**, licenciado en Ciencias Biológicas por la Universidad Autónoma de Madrid, ha realizado bajo su dirección parte del trabajo titulado **“Genetic basis of Parkinson’s disease and related disorders”**, para la obtención del grado de Doctor en Biología.

Para que conste a todos los efectos, firma la presente certificación en:

Valencia, a 15 de Octubre de 2008

Dr. Jordi Pérez Tur



DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Laboratory of Neurogenetics  
National Institute on Aging  
Porter Neuroscience Research Center,  
35 Lincoln Drive  
Bethesda, Maryland 20892

Monday, September 22, 2008

To the Committee,

**RE: Mr Javier Simon-Sanchez, Doctoral Thesis Work**

With this letter I certify that Mr. Javier Simon-Sanchez with a BSc degree on Biochemistry and Molecular Biology at the Autonomous University of Madrid, has performed work within the Laboratory of Neurogenetics, of which I am Chief, toward the degree of doctor in Biology.

If I can be of any further help please do not hesitate to contact me.

Yours truly,

A handwritten signature in blue ink, appearing to read "Andrew Singleton", with a long horizontal stroke extending to the right.

Andrew Singleton PhD,  
Chief, Laboratory of Neurogenetics

Andrew Singleton, Laboratory of Neurogenetics, National Institute on Aging.  
Secretary, Joan Ward; T 1 (301) 451 6081; F 1 (301) 451 5466; E wardjoan@mail.nih.gov

## FUNDING SOURCES

This work was supported by the following funding sources:

**1. Caracterización genómica y proteómica de la enfermedad de Parkinson.**

- Funding source: Ministerio de Ciencia y Tecnología. Acción estratégica en Genómica y Proteómica (GEN2001-4851-C06-01).

**2. Identificación de dianas terapéuticas relacionadas con la enfermedad de Parkinson.**

- Funding source: Fundación Ramón Areces. (Ref. DI-TER Parkinson)

**3. Intramural research programs** of the National Institute on Aging and the National Institute on Neurological Disorders and Stroke, National Institutes of Health, US Department of Health and Human services.

Javier Simón-Sánchez was recipient of the following fellowships:

**1. Predoctoral fellowship** for the development of doctoral thesis with interests for the industrial sector (I3P-BPD2003-1). February 2004 through October 2005.

**2. National Institutes of Health fellowship** for foreign visiting scientists. October 2005 through October 2008.



## ACKNOWLEDGMENTS

In first place, I would like to express my deepest gratitude to my dearest parents for their continuous patience and support during these five years, and for making this work possible. I would also like to give my warmest thanks to my brothers Nacho and Yoryi and to all of my granpas, for their continuous support and interest on what I was doing while I was away.

In the second place I would like to express my deepest thanks to my supervisors, Drs. Jordi Pérez Tur and Andrew B. Singleton for their marvellous guidance in this thesis and their smart ideas throughout the realization of the work presented herein. I also want to express my gratitude to Adriano Jiménez Escrig for giving me the opportunity to start doing research in his laboratory at the *Hospital Ramon y Cajal de Madrid*, and for introducing me to Drs. Andrew Singleton and John Hardy in summer 2002. On this regard I would like to thank them both for receiving me as a summer student in their lab at the National Institutes of Health (NIH) that same year.

I would also like to thank Anetxu, Salo, JM, JV, Silvia, Anuska and Vicente (from the *Unitat de Genètica Molecular*) who helped me out so many times at the beginning of my thesis. I must express my deepest gratitude to Vicente for printing and binding this thesis and helping me out with all the university paperwork while I was abroad.

I can not forget to thank all the people at the *Instituto de Biomedicina de Valencia* for making my stay in Valencia so much easier.

I also want to thank all the staff at the Laboratory of Neurogenetics at the NIH with especial attention to Mark, Dena, Raph, Sonja, José, Rita, Coro, Mar, Mike and Jordi for their help, friendship and support during the three years I worked at the NIH.

Finally I want to thank all my friends from all around the world with special mention to, Belén for her wholehearted support during my stay in Valencia, *Los praderianos*, my university classmates in Madrid (especially to the *DYC* lovers), the Valencian crew, all those I met during my stay in Washington DC and especially to Maca from Santiago de Chile. Without you this work would have simply not been fulfilled.

Thank you all.

# **GENETIC BASIS OF PARKINSON'S DISEASE AND RELATED DISORDERS**

A thesis presented by

Javier Simón Sánchez

To

Departamento de Biología celular y Parasitología

Facultad de Medicina

Universidad de Valencia

Under the supervision of Drs. Jordi Pérez Tur and Andrew B. Singleton

October 2008

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

. “Sequencing analysis of *OMI/HTRA2* shows previously reported pathogenic mutations in neurologically normal controls”. **Javier Simón-Sánchez** and Andrew B. Singleton. *Human Molecular Genetics*. 2008 March 25.

. “DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA”. Sarah Camargos<sup>#</sup>, Sonja Scholz<sup>#</sup>, **Javier Simón-Sánchez**<sup>#</sup>, Coro Paisán-Ruiz<sup>#</sup>, Patrick Lewis, Dena Hernandez, Jinhui Ding, J Raphael Gibbs, Mark R Cookson, Jose Bras, Rita Guerreiro, Catarina Resende Oliveira, Andrew Lees, John Hardy, Francisco Cardoso, Andrew B Singleton. *Lancet Neurology*. 2008 Mar; 7 (3): 207-215.

. “Genome-wide SNP assay reveals mutations underlying Parkinson’s disease”. **Javier Simón-Sánchez**<sup>#</sup>, Sonja Scholz<sup>#</sup>, Maria del Mar Matarin<sup>#</sup>, Hon-Chung Fung<sup>#</sup>, Dena Hernandez, J Raphael Gibbs, Angela Britton, John Hardy, Andrew Singleton. *Human Mutation*. 2008 Feb; 29 (2): 315-322.

. “A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release”. Mar Matarin<sup>#</sup>, W. Mark Brown<sup>#</sup>, Sonja Scholz<sup>#</sup>, **Javier Simón-Sánchez**<sup>#</sup>, Hon-Chung Fung<sup>#</sup>, Dena Hernandez, Wavrant De Vrieze, Cynthia Crews, Angela Britton, J Raphael Gibbs, Carl Langefeld PhD, Fabienne, Thomas G. Brott, Robert D. Brown Jr, Bradford B. Worrall, Michael Frankel, Scott Silliman, L. Douglas Case, Andrew Singleton, John A Hardy, Stephen S Rich, James F. Meschia. *Lancet Neurology*. 2007 May; 6(5): 414-420.

. “Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data”. Fung HC<sup>#</sup>, Scholz S<sup>#</sup>, Matarin M<sup>#</sup>, **Simón-Sánchez J**<sup>#</sup>, 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. *Lancet Neurology*. 2006 Nov; 5 (11): 911-916.

. “Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals”. **Simón-Sánchez J**<sup>#</sup>, Scholz S<sup>#</sup>, Fung HC<sup>#</sup>, Matarin M<sup>#</sup>, Hernandez D, Gibbs JR, Britton A, de Vrieze FW, Peckham E, Gwinn-Hardy K, Crawley A, Keen JC, Nash J, Borgaonkar D, Hardy J, Singleton A. *Human Molecular Genetics*. 2007 Jan 1; 6(1):1-14.

. “Parkinson’s disease due to the R1441G mutation in Dardarin: A founder effect in the Basques” **Simón-Sánchez J**<sup>#</sup>, Marti-Masso JF<sup>#</sup>, Sanchez-Mut JV, Paisan-Ruiz C, Martinez-Gil A, Ruiz-Martinez J, Saenz A, Singleton AB, Lopez de Munain A, Perez-Tur J. *Movement Disorders*. 2006 Nov; 21 (11): 1954-1959.

. “LRRK2 is expressed in areas affected by Parkinson's disease in the adult mouse brain”. **Javier Simón-Sánchez**, Vicente Herranz-Pérez, Francisco Olucha-Bordonau, Jordi Pérez-Tur. *European Journal of Neuroscience*. 2006 Feb; 23 (3): 659-666.

. “Analysis of *SCA-2* and *SCA-3* repeats in Parkinsonism: Evidence of *SCA-2* expansion in a family with autosomal dominant Parkinson's disease”. **Simón-Sánchez J**, Hanson M, Singleton A, Hernandez D, McInerney A, Nussbaum R, Werner J, Gallardo M, Weiser R, Gwinn-Hardy K, Singleton AB, Clarimon J. *Neurosci Letters*. 2005 Jul 1-8; 382 (1-2): 191-194. Epub 2005 Apr 1.

. “Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease”. Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, **Simón J**, van der Brug M, de Munain AL, Aparicio S, Gil AM, Khan N, Johnson J, Martinez JR, Nicholl D, Carrera IM, Pena AS, de Silva R, Lees A, Marti-Masso JF, Perez-Tur J, Wood NW, Singleton AB. *Neuron*. 2004 Nov 18;44(4):595-600.

# These authors equally contributed to this work.

## ABBREVIATION LIST

AAO	Age at onset	GL	Granular layer
AD	Alzheimer's disease	GP	Globus pallidum
Add	Additive	GWAS	Genome-wide association study
ALS	Amyotrophic Lateral Sclerosis	Haplo	Haplotype
APT	Anterior pretecal nucleus	HapMap	Haplotype Map
Arc	Nucleus arcuatus	HL	Hindlimb cortex
ARMS	Amplification refractory mutation system	HWE	Hardy-Weinberg equilibrium
AR-JP	Autosomal recessive juvenile parkinsonism	IBD	Identity by descent
ASPE	allele-specific primer extension	IBS	Identity by state
ASt	Amygdalostriatal transition area	Icj	Islands of Calleja
BLA	Basolateral amygdala	IL	Infralimbic medial prefrontal cortex
BMA	Basomedial amygdala	InG	Intermediate grey layer of the superior colliculus
Bp	Base pair	IP	Interpeduncular nucleus
CA1	Field 1 of the Ammon's horn	JPT	Japanese from Tokyo
CA2	Field 2 of the Ammon's horn	KRS	Kufor-Rakeb Syndrome
CA3	Field 3 of the Ammon's horn	LA	Lateral amygdala
CDEs	Clinical data elements	LB	Lewy Body
CeA	Central amygdala	LCL	Lymphoblastoid cell line
CEPH	Centre d'Etude du Polymorphisme Humain)	LD	Linkage disequilibrium
CEU	CEPH from Utah	lfp	Longitudinal fasciculus of the pons
Cg	Cingular cortex	LHA	Lateral hypothalamic area
CGEMS	Cancer genetic markers of susceptibility	LHb	Lateral habenula
CHB	Chinese from Beijing	LL	Lateral lemniscus
Chr	Chromosome	LN	Lewy Neurite
Ci	Capsula interna	LOH	Loss of heterozygosity
CI	Confidence interval	LOT	Nucleus of the lateral olfactory tract
CL	Clumsiness	LPV	lop p value
CNV	Copy number variation	LRS	Likelihood ratio statistic
CoA	Cortical amygdala	MAF	Minor allele frequency
DF	Degree of freedom	MDS	Multidimensional scaling
DG	Dentate gyrus	MeA	Medial amygdala
DHA	Dorsal hypothalamic area	mer	Base pair
DLB	Dementia with Lewy bodies	MGB	Medial geniculate body
dmVMH	Dorsomedial division of the ventromedial hypothalamic nucleus	MHb	Medial habenula
Dom	Dominant	MHC	Major Histocompatibility Complex
EBV	Epstein-Barr virus	MJD	Machado-Joseph disease
EW	Edinger Westfal nucleus	ML	Molecular layer
F	Inbreeding coefficient	MPP+	1-methylphenylpyridinium
Fr	Frontal cortex	MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
FTDP17	Frontotemporal dementia with parkinsonism linked to chromosome 17	MS	Multiple sclerosis

### Abbreviation list (continued)

MSA	Multiple system atrophy	RPO	Nucleus reticularis pontis oralis
NaN	Not a number	RT	Reticular thalamic nucleus
NCBI	National Center for Biotechnology Information	S	Subculiculum
nsSNP	non-synonymous SNP	SBE	Single-base extension
OMIM	Online Mendelian Inheritance in Man	SCA	Spinocerebellar ataxia
OR	Odds ratio	SNC	Substantia nigra pars compacta
ot	Optic tract	SNP	Single nucleotide polymorphism
PAG	Periaqueductal grey	SNR	Substantia nigra pars reticulata
Par	Parietal cortex	Te	Temporal cortex
PCR	Polymerase chain reaction	tSNP	tag SNP
PD	Parkinson's disease	TT	Tenia tecta
PFA	Perifornical area	TuMC	Tuberal magnocellular division of the lateral hypothalamic area
Pir	Piriform cortex	UFMG	Universidade Federal de Minas Gerais
PnN	Pontine nuclei	UPDRS	Unified Parkinson's disease rating scale
PSP	Progressive supranuclear palsy	UPS	Ubiquitin-proteasome system
Pu	Putamen	UT	Unilateral Tremor
PV	p value	VEnd	Ventral endopiriform division of the claustrum
PVP	Paraventricular posterior thalamic nucleus	vIVMH	Ventralateral division of the ventromedial hypothalamic nucleus
Q-Q	Quantile-Quantile	VP	Ventral pallidum
Rec	Recessive	VPL	Ventral posterolateral thalamic
RefSeq	Reference sequence	WM	White matter
RLS	Restless syndrome	XDP	X-linked recessive dystonia
RN	Red nucleus	YOPD	young onset Parkinson's disease
Roc	Ras on complex proteins (domain)	YRI	Yoruban from Ibadan, Nigeria
ROS	Reactive oxygen species	ZI	Zona incerta

# TABLE OF CONTENTS

<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>1.1. BACKGROUND .....</b>	<b>1</b>
<b>1.2. EPIDEMIOLOGY .....</b>	<b>2</b>
1.2.1. <i>Prevalence.....</i>	<i>3</i>
1.2.2. <i>Incidence.....</i>	<i>5</i>
1.2.3. <i>Parkinson’s disease epidemiology in Spain.....</i>	<i>7</i>
<b>1.3. CLINICAL FEATURES.....</b>	<b>9</b>
<b>1.4. PATHOLOGY.....</b>	<b>12</b>
1.4.1. <i>Lewy bodies.....</i>	<i>13</i>
1.4.2. <i>Lewy neurites.....</i>	<i>14</i>
1.4.3. <i>Pale bodies.....</i>	<i>15</i>
<b>1.5. MECHANISMS OF NEURODEGENERATION.....</b>	<b>19</b>
1.5.1. <i>Protein degradation and mishandling.....</i>	<i>20</i>
1.5.2. <i>Mitochondrial dysfunction and oxidative stress.....</i>	<i>21</i>
<b>1.6. ETIOLOGY OF PD .....</b>	<b>23</b>
1.6.1. <i>Environmental versus genetic factors.....</i>	<i>23</i>
1.6.2. <i>Mendelian variants of PD.....</i>	<i>26</i>
1.6.3. <i>Other genetic factors related to PD.....</i>	<i>44</i>
1.6.4. <i>Genome-wide association studies (GWAS) and PD.....</i>	<i>58</i>
<b>2. OBJECTIVES.....</b>	<b>62</b>
<b>3. METHODS.....</b>	<b>63</b>
<b>3.1. GENETIC ANALYSIS OF <i>PARK8</i>-LINKED PARKINSON’S DISEASE PATIENTS. ....</b>	<b>63</b>
3.1.1. <i>Basque late-onset PD families .....</i>	<i>63</i>
3.1.2. <i>Minimal haplotype shared by UGM03, UGM04, UGM05 and UGM06.....</i>	<i>66</i>
3.1.3. <i>Analysis of p.R1441G mutation in a Basque PD cohort.....</i>	<i>71</i>
3.1.4. <i>Haplotype analysis of Basque p.R1441G carriers.....</i>	<i>74</i>
3.1.5. <i>Haplotype analysis of p.R1441G carriers from Asturias.....</i>	<i>79</i>
<b>3.2. <i>IN SITU</i> HYBRIDIZATION OF <i>LRRK2</i> mRNA IN ADULT MOUSE BRAIN.....</b>	<b>80</b>
3.2.1. <i>Probe synthesis .....</i>	<i>80</i>
3.2.2. <i>In situ hybridization .....</i>	<i>88</i>
3.2.3. <i>Relative regional expression .....</i>	<i>90</i>
3.2.4. <i>Image handling.....</i>	<i>91</i>
3.2.5. <i>Quality-control procedures .....</i>	<i>91</i>
3.2.6. <i>Nomenclature .....</i>	<i>91</i>
<b>3.3. SCREENING FOR MUTATIONS IN <i>OMI/HTRA2</i>.....</b>	<b>91</b>
3.3.1. <i>Subject collection.....</i>	<i>91</i>
3.3.2. <i>Sequence analysis of <i>OMI/HTRA2</i>.....</i>	<i>93</i>
3.3.3. <i>Statistical analysis.....</i>	<i>95</i>
<b>3.4. ANALYSIS OF <i>SCA-2</i> AND <i>SCA-3</i> GENES IN PARKINSONISM .....</b>	<b>97</b>
3.4.1. <i>Subject collection.....</i>	<i>97</i>
3.4.2. <i>Genomic DNA extraction.....</i>	<i>100</i>
3.4.3. <i>Genotyping.....</i>	<i>100</i>
<b>3.5. WHOLE-GENOME SNP GENOTYPING IN PARKINSON’S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 1. ....</b>	<b>102</b>
3.5.1. <i>Subject collection.....</i>	<i>102</i>
3.5.1.1. <i>Neurologically normal control samples. ....</i>	<i>102</i>
3.5.1.2. <i>Parkinson’s disease samples.....</i>	<i>103</i>
3.5.2. <i>Genotyping.....</i>	<i>104</i>

3.5.2.1.	Platform used.....	104
3.5.2.2.	Infinium Workflow.....	106
3.5.3.	<i>Quality-Control procedures</i> .....	114
3.5.3.1.	Low quality genotyping.....	114
3.5.3.2.	Population substructure.....	115
3.5.4.	<i>Statistical analysis</i> .....	115
3.5.5.	<i>Beyond association</i> .....	116
3.5.6.	<i>Confirmation of structural alterations by Real-time PCR</i> .....	120
3.6.	<b>WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 2</b> .....	123
3.6.1.	<i>Subject collection</i> .....	123
3.6.1.1.	Late-onset PD samples.....	123
3.6.1.2.	YOPD samples.....	124
3.6.1.3.	Coriell Institute neurologically normal control samples.....	125
3.6.1.4.	CGEMS initiative control samples.....	125
3.6.2.	<i>Genotyping</i> .....	126
3.6.3.	<i>Quality-control procedures</i> .....	127
3.6.3.1.	Low quality genotyping.....	127
3.6.3.2.	Gender problems.....	128
3.6.3.3.	Population substructure.....	128
3.6.3.4.	Non-reported relatedness.....	129
3.6.4.	<i>Statistical analysis</i> .....	130
3.6.5.	<i>Combined analysis</i> .....	131
3.6.5.1.	German cohort.....	132
3.6.5.2.	Japanese cohort.....	132
3.7.	<b>AUTOZYGOSITY MAPPING IN BRAZILIAN FAMILIES</b> .....	133
3.7.1.	<i>Subject collection</i> .....	133
3.7.1.1.	Family DYT16-1.....	135
3.7.1.2.	Family DYT16-2.....	137
3.7.1.3.	Individual 2035-61.....	140
3.7.2.	<i>Exclusion of known genes</i> .....	141
3.7.3.	<i>Genome-wide SNP genotyping</i> .....	145
3.7.4.	<i>Autozygosity mapping</i> .....	145
3.7.5.	<i>Sequencing of genes within critical intervals</i> .....	146
3.7.6.	<i>Assay for the c.665C&gt;T (P222L) mutation</i> .....	146
4.	<b>RESULTS</b> .....	149
4.1.	<b>GENETIC ANALYSIS OF PARK8-LINKED PARKINSON'S DISEASE PATIENTS.</b> .....	149
4.1.1.	<i>Minimal haplotype shared by UGM03, UGM04, UGM05 and UGM06</i> .....	149
4.1.2.	<i>A founder effect of p.R1441G in Basque population</i> .....	150
4.2.	<b>IN SITU HYBRIDIZATION OF LRRK2 mRNA IN ADULT MOUSE BRAIN</b> .....	155
4.2.1.	<i>Telencephalon</i> .....	156
4.2.2.	<i>Diencephalon</i> .....	160
4.2.3.	<i>Brainstem</i> .....	160
4.2.4.	<i>Cerebellum</i> .....	161
4.3.	<b>SEQUENCING ANALYSIS OF OMI/HTRA</b> .....	164
4.4.	<b>ANALYSIS OF SCA-2 AND SCA-3 REPEATS IN PARKINSONISM</b> .....	170
4.5.	<b>WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 1</b> .....	171
4.5.1.	<i>Whole-genome association</i> .....	171
4.5.2.	<i>Beyond association</i> .....	178
4.5.2.1.	Extended homozygosity in control population.....	178
4.5.2.2.	Copy number variation in control population.....	182
4.5.2.3.	Copy number variation in PD population.....	191
4.6.	<b>WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 2</b> .....	196
4.6.1.	<i>Quality-control approaches</i> .....	196
4.6.2.	<i>Statistical analysis</i> .....	199

4.6.3.	<i>Combined association analysis.</i>	202
4.6.3.1.	German cohort.....	202
4.6.3.2.	Japanese cohort.....	205
4.7.	<b>AUTOZYGOSITY MAPPING IN BRAZILIAN FAMILIES</b>	207
4.7.1.	<i>Exclusion of known genes.</i> .....	207
4.7.2.	<i>Autozygosity mapping.</i> .....	208
<b>5.</b>	<b>DISCUSSION</b> .....	<b>213</b>
5.1.	<b>GENETIC ANALYSIS OF PARK8-LINKED PARKINSON'S DISEASE PATIENTS.</b>	<b>213</b>
5.2.	<b><i>LRKK2</i> IS EXPRESSED IN BRAIN AREAS AFFECTED IN PD IN THE ADULT MOUSE BRAIN</b> .....	<b>217</b>
5.3.	<b><i>OMI/HTR42</i> IS NOT ASSOCIATED WITH PD IN A NORTH AMERICAN POPULATION.</b> .....	<b>221</b>
5.4.	<b>ANALYSIS OF SCA-2 AND SCA-3 GENES IN PARKINSONISM</b> .....	<b>225</b>
5.5.	<b>WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 1</b> .....	<b>226</b>
5.5.1.	<i>Whole-genome association</i> .....	226
5.5.2.	<i>Extended homozygosity in control population</i> .....	232
5.5.3.	<i>Copy number variation in neurologically normal controls</i> .....	233
5.5.4.	<i>Copy number variation in Parkinson's disease patients</i> .....	237
5.6.	<b>WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 2.</b> .....	<b>239</b>
5.7.	<b>AUTOZYGOSITY MAPPING IN BRAZILIAN FAMILIES</b> .....	<b>243</b>
<b>6.</b>	<b>SUMMARY</b> .....	<b>250</b>
<b>7.</b>	<b>CONCLUSIONS</b> .....	<b>255</b>
<b>8.</b>	<b>BIBLIOGRAPHY</b> .....	<b>258</b>
<b>9.</b>	<b>SUPPLEMENTARY MATERIAL</b> .....	<b>CD</b>

# 1. INTRODUCTION

## 1.1. BACKGROUND

Parkinson's disease (PD) was first described by James Parkinson in 1817 (Parkinson, 1817). It is the second most common neurodegenerative disorder affecting 1-2% of the population over the age of 65 years. Although 5-10% of PD cases belong to a familial form of the disease, the remaining 90-95% of are apparently sporadic.

The neuropathological features of idiopathic PD are the appearance of intracellular protein aggregates known as Lewy bodies (LBs) (Lewy, 1912) and the degeneration of neurons in the *pars compacta* tier of the *substantia nigra* (SNc) with consequent loss of dopaminergic innervation to the caudate and putamen (nigro-striatal pathway). Other areas of the brain can also be affected in idiopathic PD including *locus coeruleus* or the dorsal motor nucleus of the vagus and the limbic system. Whether the LBs existence is triggering this neuronal degeneration or the neuronal dysfunction is promoting the appearance of these aggregates is still unknown. The phenotype of idiopathic PD is characterized by resting tremor, rigidity, postural instability and bradykinesia, which at first respond to L-DOPA treatment.

Additional non-motor features like postural abnormalities, dysautonomia, dystonic cramps, anxiety, depression or dementia are characteristic. Most of these symptoms tend to get worse over time.

The term parkinsonism refers to a syndrome including a combination of all these clinical features as well as many other diseases which most likely result from loss or dysfunction

of the dopaminergic neurons in the SN. Although the most common cause of this syndrome is PD, other etiologies can lead to its appearance: Dementia with Lewy bodies (DLB; OMIM #127750), progressive supranuclear palsy (PSP; OMIM #601104), frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP17; OMIM #600274), autosomal recessive form of Segawa's syndrome (OMIM #605407), X-linked recessive Filipino type of dystonia (XDP; OMIM #314250), multiple system atrophy (MSA) or drug-induced parkinsonism (Shy and Drager 1960; Steele, Richardson *et al.* 1964; Lee, Pascasio *et al.* 1976; Langston, Ballard *et al.* 1983; Foster, Wilhelmsen *et al.* 1997; Hansen LA 1997; Segawa 1999; Wenning and Geser 2003).

## 1.2. EPIDEMIOLOGY

Data on the prevalence and incidence of PD are of interest for several reasons. Epidemiological studies can provide insights into suspected risk factors, protective factors and primary causes of the disease; these data can also provide critical information to inform public health planning. This latter point has become of special interest in the last years, as longevity of the population increases and the incidence of PD rises with age. Since PD was first described, a number of epidemiological studies have been performed. However, since these studies were methodologically limited, being mostly small case-control or register-based studies, it was not until the end of the twentieth century that large studies reached a stage with sufficient number of PD patients to examine incidence and potential risk factors. A limitation of this work is that a reliable and easily applicable

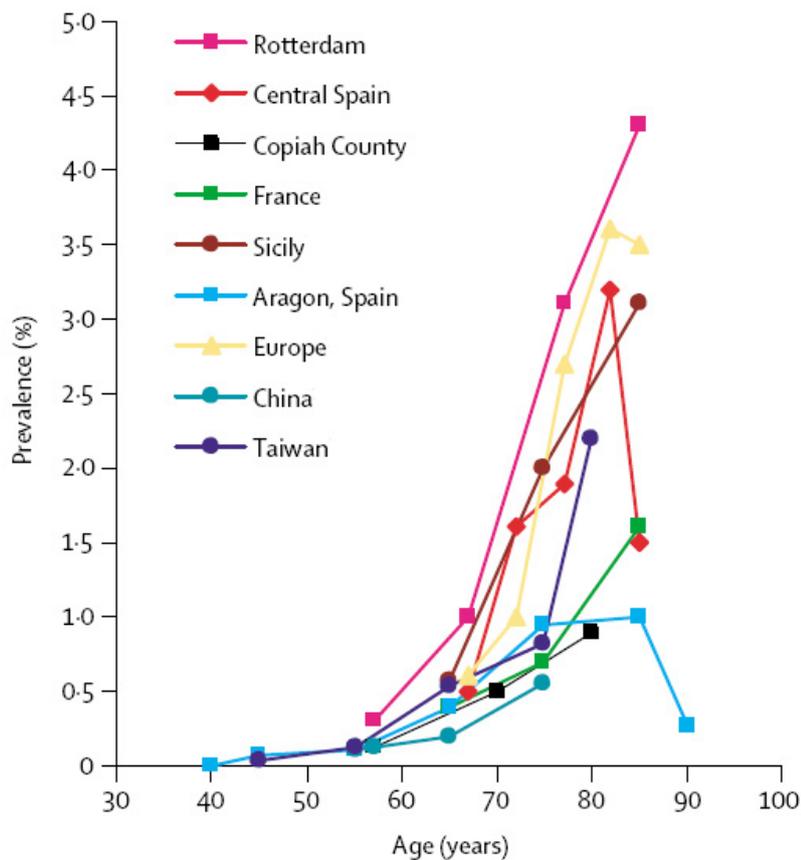
diagnostic test for PD is not yet available. Sophisticated imaging techniques such as CT or PET may be helpful to diagnose PD, but their usefulness for population epidemiological studies is still limited. Thus, a diagnosis of PD for these kind of studies is primarily based on clinical criteria. As will be argued latter, these criteria are still controversial and a long-term follow-up of patients is needed to improve accuracy of diagnoses. Definitive diagnosis requires post-mortem confirmation. Thus, estimates of the prevalence and incidence of PD may vary according to the methodology utilized to make the diagnoses. Hence, the stricter the diagnostic criteria used, the lower the estimates of prevalence and incidence are.

### **1.2.1. Prevalence**

Prevalence denotes the total number of cases of disease on a given time. The prevalence of PD in industrialized countries is generally estimated at 0.3% of the general population and about 1-2% in people over 60 years of age. Age, however, is clearly the greatest factor in PD expression. A European study of 15,000 participants over 65 reported that the prevalence of PD from 65 to 89 years old increases steadily from 0.6 to 3.5% (De Rijk, Launer *et al.*, 2000). De Lau and colleagues reviewed age-specific prevalence rates obtained from studies across different populations (figure 1). This review clearly shows that PD is an age-related disease, being rare before 50 in all reviewed populations (De Lau and Breteler, 2006). As the elderly population increases in size, the prevalence of PD increases as well, making prevalence studies difficult to perform.

From an etiological point of view, PD may result from different environmental exposition and genetic background that are both due to intercultural differences. Some studies have

addressed the issue of ethnicity in relation to the occurrence of PD, obtaining conflicting results. Thus, some groups support the idea that prevalence of PD varies by ethnicity, being less common in black and Asian populations than in white and hispanic groups (Mayeux, Marder *et al.*, 1995; Morens, Davis *et al.*, 1996; Van Den Eeden, Tanner *et al.*, 2003). These reported differences may result from difference in response rates, survival, and case-ascertainment rather than from real differences in PD across ethnic groups. It is also possible that a delay in diagnosis due to limited access to appropriate health services among some groups could have resulted in the observed discordant rates of disease.



**Figure 1: Population-based prevalence studies of Parkinson's disease** (Figure modified from De Lau and Breteler, 2006).

Many studies have reported higher prevalence of PD in men than in women (Mayeux, Marder *et al.*, 1995; Morens, Davis *et al.*, 1996; Baldereschi, Di Carlo *et al.*, 2000; Van Den Eeden, Tanner *et al.*, 2003). This difference may be due to the hypothetical neuroprotective effect of estrogen that has been shown to have numerous effects on dopamine synthesis, transmission and release (Saunders-Pullman, 2003). Another possible explanation for this difference is that PD increases with age and women constitute a greater proportion of aged population than men, although many studies attempt to adjust for this potential confound.

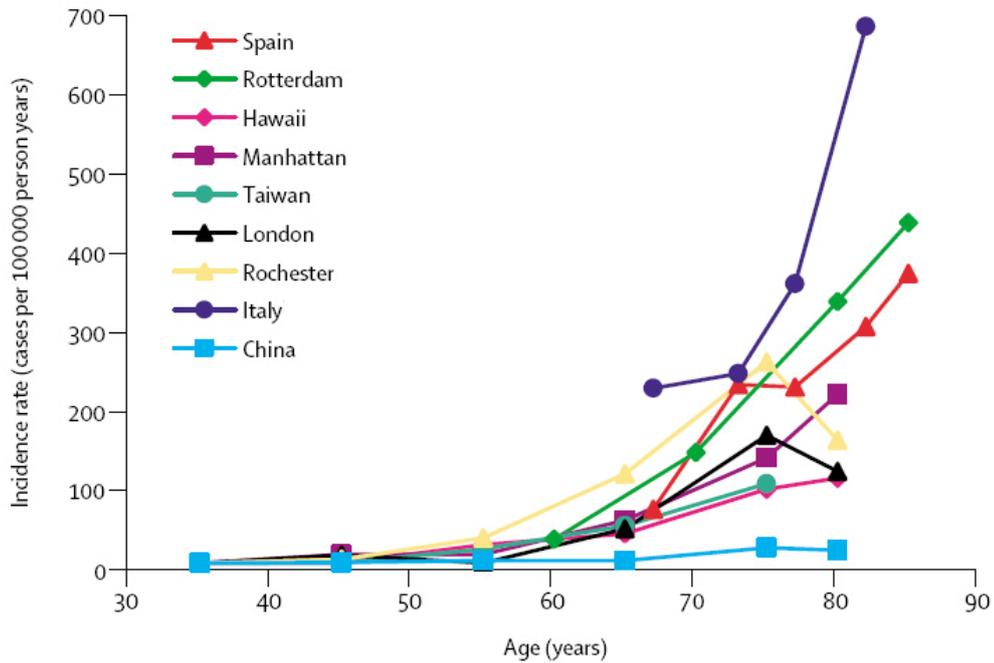
### **1.2.2. Incidence**

The term incidence refers to the number of new cases of a disease per unit of time, usually 1 year. Incidence rates are theoretically not affected by differences in survival of patients and therefore better measures of the risk of disease than prevalence estimates. However, population-based incidence estimates are much harder to obtain than record-based data, as they require large cohorts and long follow-up periods. Also, if the follow-up of a cohort is incomplete, substantial misclassification may occur, which will typically lead to underestimate disease risk. The incidence of PD has been difficult to ascertain with a high degree of certainty, but more accurate estimates have become available with the advent of better diagnostic guidelines as those mented previously.

The correlation of PD and aging is even more striking from studies of disease incidence. A meta-analysis of 25 incidence studies (Twelves, Perkins *et al.*, 2003) estimated an overall population incidence of 16-19 per 100,000 per year. However, De Lau and collaborators in 2004 (De Lau, Giesbergen *et al.*, 2004) reported that incident rates

increase after the age of 50, especially in men, and rapidly increase over the age of 75. They stated that risk of developing any parkinsonian features between the ages of 55 and 85 years is 8.5% for men and 7.7% for women. Between the ages of 75 and 85, the incident rate is 6.8 for men and 4.5 for women per 1,000 people per year. Over the age of 85, it increases to 12.1 and 10.2 for men and women respectively.

Similar incidence rates have been published by different groups worldwide, establishing that onset of PD rarely occurs before 50 with a sharp increase after age of 60. (Rajput, Offord *et al.*, 1984; Wang, Shi *et al.*, 1991; Mayeux, Marder *et al.*, 1995; Morens, Davis *et al.*, 1996; Baldereschi, Di Carlo *et al.*, 2000; Macdonald, Cockerell *et al.*, 2000; Chen, Chang *et al.*, 2001; Benito-Leon, Bermejo-Pareja *et al.*, 2004; De Lau, Giesbergen *et al.*, 2004). Interestingly, some studies observed a decline in incidence in the highest age groups (Morens, Davis *et al.*, 1996; Bower, Maraganore *et al.*, 1999), but this is likely an artifact caused by increased diagnostic uncertainty due to comorbid disorders and selective loss to follow-up (figure 2).



**Figure 2: Prospective population-based incidence studies of Parkinson's disease** (Figure modified from De Lau and Breteler, 2006).

### 1.2.3. Parkinson's disease epidemiology in Spain

In 2005, Von Campenhausen and colleagues (Von Campenhausen, Bornschein *et al.*, 2005) provided an overview on the prevalence and incidence of PD in certain European countries by systematically searching for epidemiological studies performed in these countries. As for Spain, nine accessible studies were identified (figure 3). In the first three, a three-phase door-to-door study was performed in Bidasoa region (Bergareche, De La Puente *et al.*, 2004), central Spain (Benito-Leon, Bermejo-Pareja *et al.*, 2003) and Cantalejo region (Claveria, Duarte *et al.*, 2002). In these reports, individuals were first screened by interviewers and those who screened positive were examined by a neurologist. With this methodology, prevalence estimates were 1.5% (Bergareche *et al.*),

1.5% (Benito-León *et al*) and 1.3% (Claveria *et al*) in persons older than 65 years. Benito-León and coworkers (Benito-Leon, Bermejo-Pareja *et al.*, 2004), also performed a study on PD incidence by following 5,160 parkinsonism-free subjects for a period of 3 years, obtaining an annual incidence rate of 186.8 standardized to the European population.

Martinez-Suarez and Blázquez-Menes (Martinez-Suarez and Blazquez-Menes, 2000) performed a study in Asturias to estimate the prevalence of PD in this Spanish region. Records of prescriptions were obtained from the pharmacy sub-directorate of Health and consumption Ministry, which covers 100% of the population in Asturias and estimated a crude prevalence of PD of 0.2%.

Similar results were obtained by Errea and colleagues in lower Aragon (Errea, Ara *et al.*, 1999), by collecting data from hospital record, medical files from specialist in the region, and retirement homes. A 0.2% prevalence of PD was estimated. Viñés and collaborators (Vines, Larumbe *et al.*, 1999) estimated PD incidence in Navarra on 8.2/100,000 between 1994 and 1995 by collecting data of diagnosed cases from family doctors, physicians, private neurologists and National Health Service.

Finally, Criado-Álvarez and Artazcoz Sanz (Artazcoz Sanz and Vines Rueda, 1995; Criado-Alvarez, Romo-Barrientos *et al.*, 1998), by obtaining information from the Pharmaceutical Service of Health and Consume Ministry of Spain, estimated that PD prevalence was 2.7% and 0.16% in Castilla la Mancha and Navarra respectively.



**Figure 3: Parkinson’s disease epidemiology in Spain.** Map of Spain showing prevalence (blue) and incidence (red) data obtained by studies performed by Bergareche, De La Puente *et al.* (1); Benito-León, Bermejo-Pareja *et al.* (2); Benito-León, Bermejo-Pareja *et al.* (3); Claveria, Duarte *et al.* (4); Martínez-Suarez and Blázquez-Menes (5); Errea, Ara *et al.* (6); Viñés, Larumbe *et al.* (7); Criado-Álvarez, Romo-Barrientos *et al.* (8); and Artazcoz Sanz and Viñés Rueda (9).

### 1.3. CLINICAL FEATURES

Parkinson’s disease is a clinical syndrome whose phenotype is characterized by the cardinal features akinesia/bradykinesia, muscular rigidity and resting tremor. Not all cardinal features need to be present, but at least two should be for a diagnosis of PD, with at least one of them being resting tremor or akinesia/bradykinesia. Among these cardinal symptoms, akinesia (meaning absence of movement) is the one contributing more to the disability caused by PD. The terms bradykinesia/hypokinesia (meaning slowness and

decreased amplitude of movement) are more appropriate, as only a minority of patients suffer from absolute loss of movement. This symptom affects all voluntary and involuntary movements. Thus, automatic and habitual movements such as swinging the arms while walking, eyes blinking, swallowing saliva or making gestures while talking are partially lost. These, together with monotonous speech, impair the patient's ability to communicate. Although slowness is a general feature of parkinsonian movements there are instances in which repetitive movements are preformed at abnormally high frequencies with very small amplitude. Parkinsonian subjects usually have problems with both executing and programming a certain movement. Thus, patients with advanced disease status may suffer what is known as "freezing", a situation in which they are completely unable to start walking. Interestingly, this impairment is more pronounced when patients have to self-initiate movements rather than when they respond to an external stimuli.

Another of the cardinal clinical features of PD is muscular rigidity. This is defined as a constant increased resistance of a joint to passive movement, resulting in stiffness and a reduced ability to relax limb muscles. When this feature is present along with resting tremor, a characteristic type of resistance is observed that has been given the name of "cogwheel-like" rigidity.

The third cardinal feature, resting tremor, is the most conspicuous of parkinsonism and was the reason for the original name, shaking palsy, given by James Parkinson in 1817 (Parkinson, 1817). Resting tremor mostly occurs in the upper limbs but also in the legs and head. Since this tremor is absent during voluntary actions, this feature is not

necessarily impairing. PD patients can also suffer from other kinds of tremor such as postural and action tremor, re-emergent tremor and orthostatic tremor.

Additional features of PD that aid in the differential diagnosis are an asymmetric onset of symptoms (only in one side of the body), resting tremor (although this may be absent in PD it is almost always absent in diseases such as PSP, FTDP-17, etc...) and a substantial response to L-dopa therapy.

Resting tremor is often the first symptom recognized in PD, but the illness sometimes starts with bradykinesia and resting tremor may be absent throughout the course of the disease. The early symptoms of PD (resting tremor, bradykinesia and rigidity) are usually improved with L-dopa therapy or other dopamine agonists. As time goes on, symptoms that do not respond to L-dopa, such as flexed posture, freezing and loss of postural reflexes, develop. Moreover, the bradykinesia that first responded to L-dopa in the earlier stages increases and stops responding to the therapy. Abnormal involuntary movements (dyskinesia) are common in PD patients primarily as a consequence of the prolonged dopaminergic therapy.

Besides the features discussed so far, non-motor symptoms have been described to be a frequent feature of PD and, in many cases, these reduce quality of life to a greater extent than the motor symptoms. These include autonomic, sensory (pain), cognitive and neuropsychiatric symptoms. These include symptoms like orthostatic hypotension, bladder disturbances, obstipation (a severe constipation caused by intestinal obstruction), fatigue, anxiety, sleep disturbances, decreased motivation and apathy, depression, slowness in thinking (bradyphrenia) and cognitive decline.

#### 1.4. PATHOLOGY

Motor symptoms of PD are primarily related to extensive loss of the dopaminergic neurons of the SNC, causing a decrease in the dopaminergic innervation in the caudate and putamen. There appears to be substantial redundancy in these systems as frank clinical manifestation of PD only presents after the pathology has reached an advanced stage in which approximately 50% of the dopaminergic cells of the SNC are lost with a consequent depletion of about 80% of striatal dopamine. Dopaminergic cell loss also affects other neurotransmitter systems. Peripherally there is decreased tyrosine hydroxylase (TH) activity and decreased dopamine in the adrenal medulla, possibly explaining the autonomic manifestations in PD patients. Centrally, neurotransmitter level alterations are detected in the nucleus basalis of Meynert, *locus coeruleus* and Raphe nuclei, as well as changes in the glutamatergic and GABAergic pathways. Thus, PD is not only a loss-of-dopamine-related pathology, but a multisystem neurodegenerative disease in which diverse neural pathways and several neurotransmitter systems are involved.

The pathological hallmark lesions of PD are the Lewy body (LBs), which are neuronal eosinophilic cytoplasmic inclusions containing aggregated ubiquitinated proteins; spindle-like or thread-like branching Lewy neurites (LNs) in surviving neurons; and distinct neuronal inclusions called pale bodies.

#### 1.4.1. Lewy bodies

LBs were first described by Frederic H. Lewy in the nucleus basalis of Meynert and the dorsal vagal nucleus in patients with PD (Lewy, 1912) and were named Lewy bodies in his honor by Tretiakoff who confirmed their presence in the *substantia nigra* (Tretiakoff, 1919). There are two types of LBs: the brainstem (classical) type and the cortical type. Brainstem-type LBs are easily seen by light microscopic examination of hematoxylin and eosin (HE)-stained sections as extraordinarily large spherical eosinophilic intracytoplasmatic inclusions that displace other cell components, possessing a dense core and a peripheral halo (figure 4A). Cortical LBs are also eosinophilic, but irregular in shape, being poorly defined structures often without a conspicuous halo or central core (figure 4B).

Ultrastructurally, both brainstem-type and cortical LBs are composed of filamentous structures (Duffy, 1965; Kosaka, 1978).

LBs are typically found in dopaminergic neurons of the SNC, but are also seen in all other populations of neurons that selectively degenerate in the PD brain. Thus, they can be found at the hypothalamus, nucleus basalis of Meynert, *locus ceruleus*, dorsal raphe nucleus, dorsal vagal nucleus, intermediolateral nucleus of the spinal cord, and sacral autonomic nucleus (Jager and Bethlem, 1960; Ohama and Ikuta, 1976; Oyanagi, Wakabayashi *et al.*, 1990; Kakita, Takahashi *et al.*, 1994; Wakabayashi and Takahashi, 1997a). LBs are also seen in the neurons of the amygdaloid nucleus and cerebral cortex, particularly in deep layers (V and VI) of the limbic system (Braak, Braak *et al.*, 1994a; Wakabayashi, Hansen *et al.*, 1995). Similar inclusions are also found in the peripheral autonomic nervous system, including the sympathetic ganglia, enteric nervous system of

the alimentary tract, cardiac and pelvic plexuses, adrenal medulla, salivary gland and skin (Den and Bethlem, 1960; Okazaki, Lipkin *et al.*, 1961; Ohama and Ikuta, 1976; Kosaka, 1978; Wakabayashi, Takahashi *et al.*, 1988; Oyanagi, Wakabayashi *et al.*, 1990; Takeda, Yamazaki *et al.*, 1993; Braak, Braak *et al.*, 1994a; Kakita, Takahashi *et al.*, 1994; Wakabayashi, Hansen *et al.*, 1995; Wakabayashi and Takahashi, 1997a; 1997b; Iwanaga, Wakabayashi *et al.*, 1999). The widespread distribution of LB pathology may correspond to a variety of motor and non-motor symptoms of PD.

Despite their size it is not clear whether these LBs are causative of PD pathology or a consequence of it.

Their exact composition is unknown, but more than 70 different molecules have been identified in LBs (Wakabayashi, Tanji *et al.*, 2007). These molecules are structural elements of the LB fibril,  $\alpha$ -synuclein-binding proteins, synphilin-1-binding proteins, components of the Ubiquitin-Proteasome System, proteins implicated in cellular response, proteins associated with phosphorylation and signal transduction, cytoskeletal proteins, cell cycle proteins, cytosolic proteins that passively diffuse into LBs, etc. Of these components,  $\alpha$ -synuclein (Spillantini, Schmidt *et al.*, 1997), dj-1 (Bandopadhyay, Kingsbury *et al.*, 2004), Irfk2 (Greggio, Jain *et al.*, 2006; Zhu, Babar *et al.*, 2006), parkin (Schlossmacher, Frosch *et al.*, 2002) and pink1 (Gandhi, Muqit *et al.*, 2006) are PD-linked gene products (figure 4C, D).

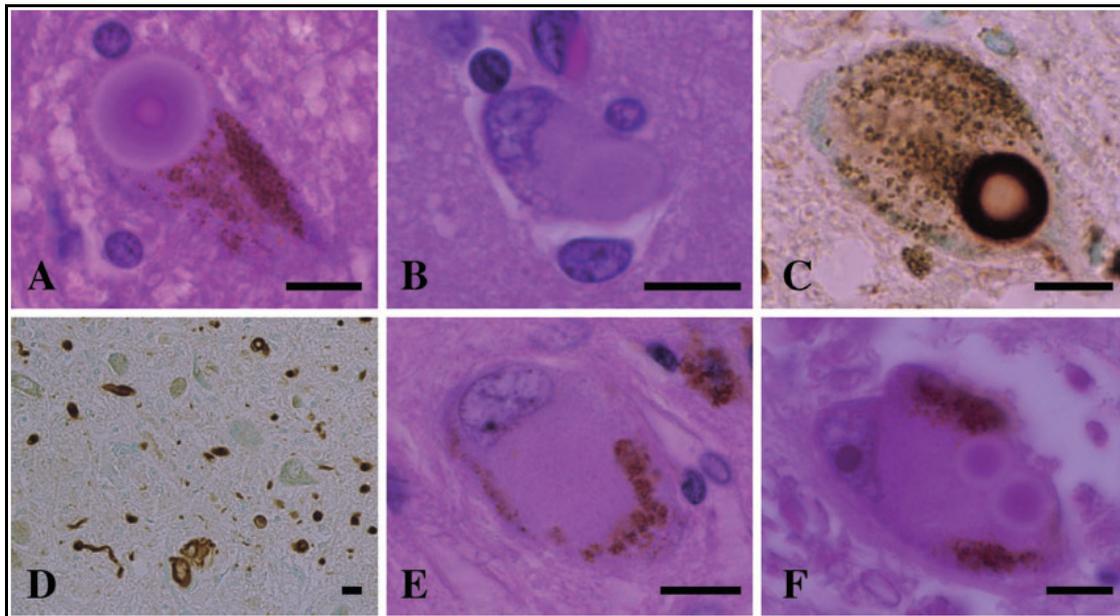
#### **1.4.2. Lewy neurites**

Lewy neurites are degenerating neural processes found in affected brain regions both in PD and in diffuse Lewy body disease. The term was first coined by Braak and colleagues

in 1994 (Braak, Braak *et al.*, 1994b). They share some histological properties with LBs, including ubiquitin reactivity and may result from the same pathological process as LBs. In many affected brain regions LNs are more numerous than LBs and this abundance may facilitate studies into LN pathogenesis.

#### **1.4.3. Pale bodies**

Besides, in the SN and *locus ceruleus*, distinct neuronal inclusions called pale bodies are seen in the cytoplasm of pigmented neurons, showing well-defined, less eosinophilic areas without halos (figure 4E) (Pappolla, Shank *et al.*, 1988; Dale, Probst *et al.*, 1992). These bodies contain sparse granular and vesicular substructures and filaments identical to those seen in LBs (Takahashi, Ohama *et al.*, 1994). In these nuclei, pale bodies are frequently found in close association with LBs (figure 4F) (Gibb, Scott *et al.*, 1991; Takahashi, Ohama *et al.*, 1994). Pale bodies are weakly positive for ubiquitin and are intensely immunolabeled with anti- $\alpha$ -synuclein (Wakabayashi, Hayashi *et al.*, 1998; Kuusisto, Parkkinen *et al.*, 2003).



**Figure 4: Pathological hallmark lesions of Parkinson's disease.** Lewy bodies (LBs) and Lewy neurites in patients with Parkinson's disease (A,C,E,F), dementia with LBs (B) and incidental LB disease (D). **A:** Concentric LB in a pigmented neuron in the substantia nigra. HE. **B:** Cortical LB in the temporal cortex. HE. **C:** Typical LB in a pigmented neuron in the SN labeled with anti- $\alpha$ -synuclein. Note that the central core is not stained. **D:** Many Lewy neurites in the dorsal vagal nucleus labeled with anti- $\alpha$ -synuclein. **E:** Pale body in a pigmented neuron in the substantia nigra. HE. **F:** Co-occurrence of a pale body and two LBs in a pigmented neuron in the substantia nigra. HE. A-F: bar, 10  $\mu$ m. (Figure modified from Wakabayashi, Tanji *et al.*, 2007).

Some authors (Braak, Del Tredici *et al.*, 2003; Braak, Ghebremedhin *et al.*, 2004) devised a staging system of LB pathology in PD with six consecutive clinical stages (1-3 being premotor and 4-6 motor stages) in which different brain areas are gradually affected marked by the continual development of LBs and LNs (figure 5): During the course of PD, neuronal damage in the brain is not random. The reason for the preferential vulnerability of some neuronal types and the resistance of others is still not properly understood. Thus, this damage occurs almost completely around areas related to motor functions, particularly the superordinate centers of the somatomotor, visceromotor, and limbic systems. All vulnerable cells in PD are projecting neurons with axons disproportionately long and thin in relation to the size of their soma, which is poorly

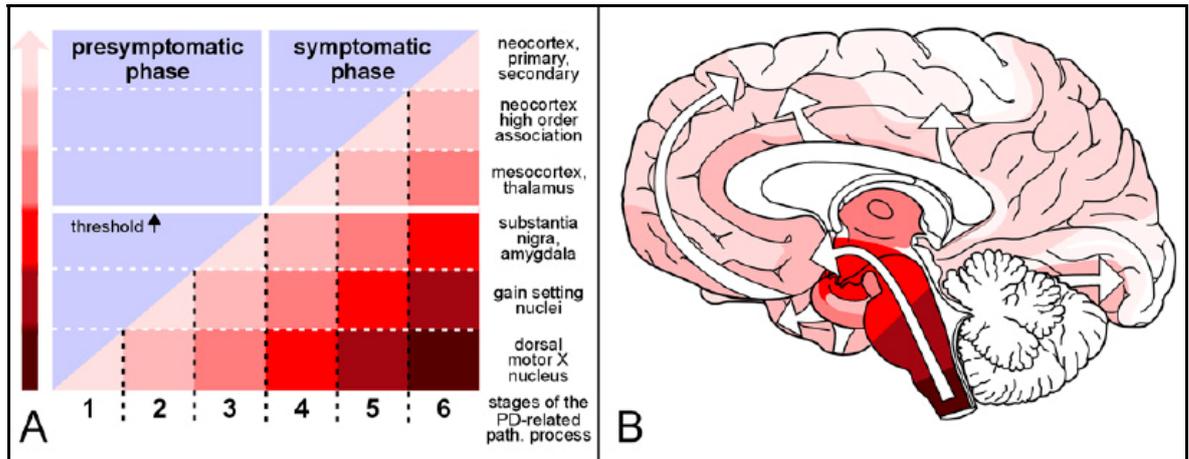
myelinated or unmyelinated. These intraneural lesions evolve sequentially in a predictable manner.

At first (stage 1) no more than a few isolated LNs appear at the dorsal motor nucleus and in the adjoining intermediate reticular zone. The large projection cells of this nucleus generate long unmyelinated preganglionic fibers that connect the central nervous system with the postganglionic nerve cells of the enteric nervous system. A few spindle-shaped LNs also appear in these axons. LNs can be found in select neuronal types within the enteric nervous system in vasoactive intestinal polypeptide neurons of the Auerbach plexus.

The pathology of the dorsal motor nucleus worsens in stage 2 and the following stages and spreads to the lower Raphe nuclei and magnocellular portions of the reticular formation, in particular the gigantocellular reticular nucleus. In addition, first LNs are seen in the coeruleus-subcoeruleus complex. All these nuclei function together as components of the “gain setting” system whose descending tracts comprise a pain control system. In addition, they work as a motor control system by regulating the sensitivity and excitability of medullary and spinal pre motor and motor neurons.

In stage 3, damage reaches the pontine tegmentum and enters the basal portions of the midbrain and forebrain. In this stage, the first LNs appear in the SNC, followed by the appearance of granular aggregates, pale bodies and LBs within the melanized projecting neurons. Macroscopic changes can not be distinguished in this stage. At the same time, the damage reaches the central subnucleus of the amygdala and extends to the basolateral nuclei. A distinctive network of LNs gradually fills the central subnucleus and makes it distinguishable from surrounding structures. Additional regions involved in this stage are the cholinergic tegmental pedunculopontine nucleus, the oral raphe nuclei, the

cholinergic magnocellular nuclei of the basal forebrain and the hypothalamic tuberomammillary nucleus. These three stages may be correlated with soft non-motor clinical findings such as those described before. In subsequent stages, the pathological alterations take an ascending course with increasing involvement of the cerebral cortex. Thus, in stage 4 a thick network of LNs appears in a specific portion of the cerebral cortex between the halo and the neocortex. Many of the projecting neurons display LBs. The Amygdala and Ammon's horn also become affected and display a plexus of LNs. In the final stages (4-5) the neurodegenerative process reaches its greatest extent. SNC appears almost devoid of melanin-containing neurons and is pale upon microscopic inspection. Starting at the temporal mesocortex, inclusion bodies gradually appear in the entire neocortex. In these latest motor phases, a broad cell loss is detected accompanied by the presence of LBs in the surviving cells not only in the SNC, but also in the *locus coeruleus*, nucleus basalis of Meynert, brainstem cranial motor nuclei and even peripheral divisions of the autonomic nervous system. Severe damage to the autonomic, limbic and somatomotor systems that began in the presymptomatic or premotor phase can become compounded by functional deficits on the part of the cerebral cortex.



**Figure 5: Staging system of Lewy body pathology in Parkinson’s disease. A, B** PD presymptomatic and symptomatic phases. **A:** The presymptomatic phase is marked by the appearance of Lewy neurites/bodies in the brains of asymptomatic persons. In the symptomatic phase, the individual neuropathological threshold is exceeded (black arrow). The increasing slope and intensity of the colored areas below the diagonal indicate the growing severity of the pathology in vulnerable brain regions (right). The severity of the pathology is indicated by darker degrees of shading in the colored arrow left. **B:** Diagram showing the ascending pathological process (white arrows). The shading intensity of the colored areas corresponds to that in A. (Figure modified from Braak, Ghebremedhin *et al.*, 2004.

### 1.5. MECHANISMS OF NEURODEGENERATION

There is extensive neural loss in PD in the ventrolateral and caudal portions of the SNC, differing from the pattern observed in normal aging brains (primarily located in the dorsomedial SNC cells). The molecular mechanisms underlying neurodegeneration in these cells in PD patients are still unknown. However, there is a large body of evidence suggesting that oxidative stress, mitochondrial dysfunction, and aberrant proteolytic degradation (proteotoxicity) may be relevant to the pathogenesis.

### 1.5.1. Protein degradation and mishandling

Protein aggregation and abnormal processing of proteins have been reported to be involved in the pathology of several neurodegenerative diseases like Huntington's (OMIM #143100) and Alzheimer's disease (AD, OMIM #104300). In PD  $\alpha$ -synuclein is present in LBs (Spillantini, Schmidt *et al.*, 1997) and mutations in the gene encoding  $\alpha$ -synuclein (*SNCA*) cause a form of familial PD (Polymeropoulos, Lavedan *et al.*, 1997). In its native state,  $\alpha$ -synuclein is a soluble and unfolded protein. Owing to a central hydrophobic region, it tends to aggregate and form soluble protofibrils first and insoluble fibrils last. These soluble protofibrils, are thought to be able to form pores in the membranes, collapsing the ion gradients (Volles, Lee *et al.*, 2001). The insoluble fibrils created later, are the main constituent of LBs. There are various factors promoting the aggregation of  $\alpha$ -synuclein: mutations (Conway, Harper *et al.*, 1998; Narhi, Wood *et al.*, 1999; Conway, Lee *et al.*, 2000; Tabrizi, Orth *et al.*, 2000), post-translational modifications (phosphorylation, nitration and glycosylation) (Giasson, Duda *et al.*, 2000; Shimura, Schlossmacher *et al.*, 2001; Fujiwara, Hasegawa *et al.*, 2002; Anderson, Walker *et al.*, 2006), exposure to certain chemicals (like MPTP or paraquat) and dopamine itself by forming a dopamine- $\alpha$ -synuclein adduct (Conway, Rochet *et al.*, 2001).

Abnormal and misfolded proteins are normally targeted via ubiquitination to the proteasome, where they are degraded in an ATP-dependent manner. Dysfunction of the Ubiquitin Proteasome System (UPS) may thus, lead to the accumulation of aggregated proteins. The first evidence of a role of the UPS in neurodegeneration emerged after the identification of mutation in *PARK2* in autosomal recessive PD patients (Kitada, Asakawa *et al.*, 1998). This gene encodes an E3 ligase protein (parkin) that catalyzes the

addition of ubiquitin molecules to target proteins for their degradation in the proteasome. Patients with mutations in this gene have a notable absence of LBs when a homozygous deletion is present. When compound heterozygous deletions are noted, LBs have been rarely found. This suggests that parkin has a significant role in LBs formation, but that neurodegeneration leading to PD can occur without LBs.

Further support for the role of the UPS in neurodegeneration was provided when mutations in *UCHL-1* (a gene encoding ubiquitin C-terminal hydrolase L1) were identified in an apparent autosomal dominant PD family (Leroy, Boyer *et al.*, 1998). The protein encoded by this gene has an important role on recycling polyubiquitin chains back to monomeric ubiquitin. However, the evidence for this mutation as a cause of PD is weak.

### **1.5.2. Mitochondrial dysfunction and oxidative stress**

Mitochondrial involvement in the pathogenesis of PD began to be studied with the observation that exposure to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), an herbicide and contaminant of illicit street drugs, caused symptoms matching those of PD in human substance abusers and in animal models (Burns, Chiueh *et al.*, 1983; Langston, Ballard *et al.*, 1983). MPTP is metabolized to 1-methylphenylpyridinium (MPP<sup>+</sup>) by the monoamine oxidase B (MAO-B) in glial cells. MPP<sup>+</sup> is subsequently selectively taken up by dopaminergic terminals and concentrated in neuronal mitochondria, where it inhibits complex I of the respiratory chain leading to an inhibition of ATP synthesis and the generation of free radicals (Volles, Lee *et al.*, 2001). Additional support for mitochondrial dysfunction in PD pathogenesis comes from evidence that the insecticides

such as rotenone and paraquat, induce a conformational change in  $\alpha$ -synuclein and significantly accelerate the rate of formation of fibrils *in vitro* (Uversky, Li *et al.*, 2001). A parkinsonian-like syndrome in animal models with protein deposits that resemble LBs was seen when the animals were exposed to widely-used herbicide paraquat (McCormack, Thiruchelvam *et al.*, 2002). These chemicals also inhibit mitochondrial complex I, and dopamine neurons seem to be most severely affected (Nicklas, Vyas *et al.*, 1985; Ramsay and Singer, 1986).

As briefly noted above, inhibition of complex I can lead to the depletion of ATP levels, hence impairing all ATP-dependent cellular processes, and augmenting the production and accumulation reactive oxygen species (ROS) that cause oxidative stress. There is evidence of oxidative stress in postmortem PD brains, showing elevated levels of lipid peroxidation markers and nitrated proteins within the SNC and LBs (Ilic, Jovanovic *et al.*, 1999; Agil, Duran *et al.*, 2006).

The preferential loss of nigral neurons in PD has been attributed to the highly oxidative intracellular environment within dopaminergic neurons. Metabolism of dopamine either by monoamine oxidase or by auto-oxidation leads to the production of free radicals and other ROS that, along with those created from the respiratory chain failure, can lead to both modifications of proteins and depletion of intracellular glutathione (GSH), a potent molecular antioxidant and an essential cofactor for the glutathione peroxidase family of enzymes. Reduced levels of glutathione and oxidized glutathione are indeed the earlier markers of nigral cell loss in PD brains. Interestingly, ROS species target the respiratory chain itself, resulting in further damage to the cell.

Identification of mutations in genes encoding the mitochondrial protein pink1, and a protein directly related to mitochondrial function, dj-1, in familial autosomal recessive PD (Bonifati, Rizzu *et al.*, 2003; Valente, Abou-Sleiman *et al.*, 2004), further support a primary role of oxidative stress in PD.

## 1.6. ETIOLOGY OF PD

### 1.6.1. Environmental *versus* genetic factors

Whether genetic factors are related to the appearance of PD has been an area of much discussion. The idea that there was little or no genetic contribution to the etiology of the disease was supported by a series of twin studies in the 1980s that failed to demonstrate significantly higher concordance of PD in monozygotic versus dizygotic twins. (Duvoisin, Eldridge *et al.*, 1981; Ward, Duvoisin *et al.*, 1983; Eldridge and Ince, 1984; Ward, Duvoisin *et al.*, 1984). This, along with striking results from epidemiological studies linking PD to environmental causes such as viral infection or neurotoxins, strengthened the idea that genetic factors contributed little to the etiology of PD. Part of this latter evidence was driven by the 1918's pandemic influenza associated with post-encephalic PD, pointing to viral infection as a major cause for PD. This association seemed so strong that it led some authors to establish that PD would disappear as a clinical entity (Poskanzer and Schwab, 1961; Lipczynska-Lojkowska, 1979).

Other facts supporting the theory that environmental factors are critical in the etiology of PD, was the observation that drug users in the 1980s exposed to 1-methyl-4-phenyl-

1,2,3,6-tetrahydropyridine (MPTP) developed parkinsonian-like features (Davis, Williams *et al.*, 1979; Langston, Ballard *et al.*, 1983). This work led to intense scrutiny of neurotoxins, including pesticides, as possible causes of PD. Some epidemiological studies have found that pesticide exposure was associated with a high risk of PD (Semchuk, Love *et al.*, 1992; Seidler, Hellenbrand *et al.*, 1996; Liou, Tsai *et al.*, 1997; Gorell, Johnson *et al.*, 1998; Uversky, Li *et al.*, 2001; Petrovitch, Ross *et al.*, 2002). However, these reports have not been consistent and structured prospective epidemiological studies are underway to establish the role of such factors.

Since all these reports were published, data on the familial aggregation of PD is still conflicting. In 1990 Johnson *et al.* (Johnson, Hodge *et al.*, 1990), reexamined twin studies performed in the early 1980s (Duvoisin, Eldridge *et al.*, 1981; Ward, Duvoisin *et al.*, 1983; Eldridge and Ince, 1984; Ward, Duvoisin *et al.*, 1984) and argued that since nearly 80% of the function of dopaminergic neurons of the SN must be lost before symptoms occur, reliance on clinical diagnosis alone and lack of follow up, makes a determination of concordance extremely insensitive. In 1992 Burn and collaborators used positron emission tomography to measure <sup>18</sup>F-fluorodopa uptake (18F-dopa PET) into basal ganglia for the diagnosis of preclinical dysfunction of the SN (Burn, Mark *et al.*, 1992). Concordance rate amongst monozygotic versus dizygotic twins increased when using this methodology. Two consecutive studies performed by Piccini and collaborators further supported the idea of inheritance in PD. They first reported that unaffected family members of PD patients showed nigrostriatal dopaminergic dysfunction. Later they found an increased concordance rate in monozygotic compared to dizygotic twins (Piccini, Morrish *et al.*, 1997; Piccini, Burn *et al.*, 1999).

Later, by means of case-control studies, positive family history was found to be the greatest risk factor for PD next to age (Duvoisin and Johnson, 1992; De Michele, Filla *et al.*, 1995; Marder, Tang *et al.*, 1996). In these studies, 10-24% of PD patients were found to have a positive family history for the disease. The relative risk in first degree relatives of PD patients ranged from 4 to 10. Interestingly, the relative risk among relatives of a PD subject, seemed to increase with the number of affected family members (Lazzarini, Myers *et al.*, 1994).

An important step forward in favor of a genetic contribution to the etiology of PD was taken with the identification of several families in which PD was inherited in an autosomal dominant manner (Waters and Miller, 1994; Payami, Bernard *et al.*, 1995; Plante-Bordeneuve, Taussig *et al.*, 1995; Wszolek, Pfeiffer *et al.*, 1995; Golbe, Di Iorio *et al.*, 1996; Maraganore, Schaid *et al.*, 1996; Morrison, Godwin-Austen *et al.*, 1996). The largest of these families (Golbe, Di Iorio *et al.*, 1990) was a kindred with autosomal dominant Lewy body parkinsonism, originally from Contursi, Italy. As discussed below, further genetic analysis identified mutations in *SNCA* as causative of the disease in this family (Polymeropoulos, Lavedan *et al.*, 1997). After this finding, family studies and whole-genome linkage analysis have allowed the identification of twelve more *loci* related to the appearance of different variants of the disease, eight of which have been identified to be following classical Mendelian inheritance patterns (table 1). Although these cases account only for 5% of the total PD cases, they have generated considerable interest because understanding their pathology may help to understand the molecular basis of sporadic PD.

### 1.6.2. Mendelian variants of PD

To date there is strong evidence linking six genes to Mendelian forms of PD with both, autosomal dominant (*SNCA*, *PARK1*, OMIM #168601 and *PARK4*, OMIM #605543; *LRRK2*, *PARK8* OMIM #607060) and autosomal recessive PD (*PRKN*, *PARK2* OMIM #602544; *PINK1*, *PARK6* OMIM #605909; *DJ-1*, *PARK7* OMIM #606324; *ATP13A2*, *PARK9* OMIM #606693). All these *loci* are referred as *PARK loci* because patients with mutations in these genes have clinical features overlapping with those of idiopathic PD (table 1).

**Table 1:** Genetic causes of Parkinson's disease. AD, autosomal dominant; AR, autosomal recessive; J, juvenile; YOPD, early-onset PD.

<i>PARK loci</i>	Origin	Gene	Chromosome	Form of PD	Age at onset
<i>PARK1</i>	Italy	<i>SNCA</i>	4q21	AD	40s
<i>PARK2</i>	Japan	<i>PRKN</i>	5q25.2-q27	AR	20-40
<i>PARK3</i>	Europe	Unknown	2p13	AD	60s
<i>PARK4</i>	Iowa	<i>SNCA</i>	4q21	AD	30s
<i>PARK6</i>	Italy	<i>PINK1</i>	1p35-p36	AR	30-40
<i>PARK7</i>	Europe	<i>DJ-1</i>	1p36	AR	30-40
<i>PARK8</i>	Japan	<i>LRRK2</i>	12q12	AD	~ 60
<i>PARK9</i>	Jordan, Italy and Brazil	<i>ATP13A2</i>	1p36	AR, Kufor-Rakeb	20-40

. ***SNCA*, *PARK1* and *PARK4*:** As previously noted, Golbe and colleagues (Golbe, Di Iorio *et al.*, 1990) , identified a kindred with autosomal dominant Lewy body parkinsonism, originally from Contursi (Italy) that emigrated to New Jersey/New York in 1905. This was the largest PD family characterized thus far, and consisted of at least 60 affected members in four generations. Apart from a relative paucity of tremor, younger onset (~ 46 years) and long disease course, there were not major clinical features that differed from idiopathic PD. Linkage analysis of this family performed by Polymeropoulos

and colleagues in 1996 (Polymeropoulos, Higgins *et al.*, 1996), mapped the disease to chromosome 4q21-q23. As this region was the first to be related to PD, it was given the designation of *PARK1*. Among the candidate genes in that region, *SNCA* (Gene: *SNCA*, Protein:  $\alpha$ -synuclein) was one of the best candidates. This 140 aminoacid protein was a neuron-specific presynaptic membrane protein originally identified because it encodes the NACP peptide (non-Amyloid- $\beta$  component of Alzheimer's disease amyloid plaques) precursor. Sequencing analysis of this gene, identified a heterozygous G>A transition at position 209 of exon 4 (A53T) in this family. A group of three apparently unrelated Greek families were later identified as carrying the same mutation. Since it was found in three unrelated PD kindreds and it was absent in 314 chromosomes from control European samples (200 from an area near Contursi) it was defined as causative of the disease (Polymeropoulos, Lavedan *et al.*, 1997). Later, two other point mutations in this gene (A30P and E46K) were identified to be responsible for autosomal dominant PD in a German (Kruger, Kuhn *et al.*, 1998) and a Spanish kindred (Zarranz, Alegre *et al.*, 2004).  $\alpha$ -synuclein is expressed in the central nervous system and is the major component of LBs and Lewy neurites (Spillantini, Schmidt *et al.*, 1997). Although it has an unknown function it has been suggested to be related to synaptic plasticity, neuronal differentiation, axonal and synaptic vesicles transport and molecular chaperone activity. It is thought that *SNCA* mutations may promote PD pathogenesis by accelerating LBs formation through the  $\alpha$ -synuclein fibril formation. Thus, it has been shown *in vitro* that the A53T and A30P mutations accelerate  $\alpha$ -synuclein fibril formations. However, the fibrilization caused by A30P mutation is slower (Conway, Harper *et al.*, 1998). Although little work

on the E46K mutation has been performed to date, the pathology in these cases suggests that E46K mutation would have an effect on fibril formation.

Singleton and colleagues in 2003 (Singleton, Farrer *et al.*, 2003), gave an exciting breakthrough on the understanding of the pathology of PD with the discovery of a genomic triplication of the wild type form of *SNCA* in an family with autosomal dominant early-onset parkinsonism with dementia, known as the Iowa kindred (Waters and Miller, 1994). This *locus* is usually referred as *PARK4* because the disease was initially linked to another region on chromosome 4p (Farrer, Gwinn-Hardy *et al.*, 1999). The study by Singleton and colleagues (Singleton, Farrer *et al.*, 2003) not only suggested that increased expression in wild-type  $\alpha$ -synuclein might enhance the risk of developing PD but also that the actual level of wild type protein might determine disease onset. It has been shown that *SNCA* triplication, which occurs in the Iowa family, causes a doubling in the amount of  $\alpha$ -synuclein in blood (Miller, Hague *et al.*, 2004).

These results were consistent with haplotype data suggesting that genetic variability in the *SNCA* promoter contributes to the risk of developing PD. Thus, several groups (Kruger, Vieira-Saecker *et al.*, 1999; Tan, Matsuura *et al.*, 2000; Farrer, Maraganore *et al.*, 2001) reported a significant association of a dinucleotide polymorphism (NACP-rep1) in the promoter region of *SNCA* with PD compared to healthy controls in a white population from Europe and America. Similar results were obtained in Asian populations (Mizuta, Nishimura *et al.*, 2002; Tan, Tan *et al.*, 2003; Tan, Chai *et al.*, 2004). These publications, along with those showing that these alleles increase  $\alpha$ -synuclein expression (Chiba-Falek and Nussbaum, 2001; Chiba-Falek, Touchman *et al.*, 2003; Chiba-Falek, Kowalak *et al.*, 2005), support evidence that mutant  $\alpha$ -synuclein behaves differently from the wild-type

protein in a quantitative rather than qualitative manner. While other groups did not find such an association in different cohorts (Parsian, Racette *et al.*, 1998; Izumi, Morino *et al.*, 2001; Khan, Graham *et al.*, 2001; Holzmann, Kruger *et al.*, 2003; Spadafora, Annesi *et al.*, 2003), meta analysis at all studies showed a robust association (Mellick, Maraganore *et al.*, 2005).

The earlier triplication report (Singleton, Farrer *et al.*, 2003) promoted several groups to screen for similar mutations in other PD pedigrees. In 2004 Farrer and colleagues described *SNCA* multiplication in one out of forty two familial Swedish-American probands with early-onset parkinson they assayed (Farrer, Kachergus *et al.*, 2004). That same year, duplications in *SNCA* were described in families with clinical phenotypes resembling that of idiopathic PD (Chartier-Harlin, Kachergus *et al.*, 2004; Ibanez, Bonnet *et al.*, 2004). The results from these reports lead to the conclusion that the severity of the disease is affected directly by the dosage effect according to the gene copy number. Therefore, *SNCA* duplication appears to cause a disease similar to idiopathic PD, whereas *SNCA* triplication causes a disorder that ranges clinically and neuronpathologically from PD to diffuse LB disease with a earlier age at onset (Singleton and Gwinn-Hardy, 2004). Efforts to identify *SNCA* copy number changes in normal sporadic PD cases have largely failed, concluding that *SNCA* multiplications are not a common cause of PD (Johnson, Hague *et al.*, 2004).

Although *in vivo* data corroborate the hypothesis that elevated levels of  $\alpha$ -synuclein might be deleterious to dopaminergic neurons (Masliah, Rockenstein *et al.*, 2000; Lo Bianco, Ridet *et al.*, 2002), the underlying mechanism by which this occurs remains to be identified. Of potential relevance to this matter is the work of Cuervo *et al.* (Cuervo,

Stefanis *et al.*, 2004) showing that certain species of  $\alpha$ -synuclein impair degradation machinery in the cell, resulting in a toxic gain-of-function effect caused by the accumulation of damaged and misfolded proteins.

. ***PRKN, PARK2***: Mutations in *PRKN* are the most common cause of Autosomal Recessive Juvenile Parkinsonism (AR-JP). AR-JP as a clinically defined entity has been studied for decades, primarily in Japan (Yamamura, Sobue *et al.*, 1973; Yamamura, Arihiro *et al.*, 1993; Takahashi, Ohama *et al.*, 1994; Ishikawa and Tsuji, 1996). It is defined as a kind of parkinsonism with age at onset before 40 years, foot dystonia at onset, hyperreflexia of the lower limbs, diurnal fluctuations, slow progression of the disease, good levodopa response and early onset levodopa-induced dyskinesias. Although according to this description it is clinically distinguishable from idiopathic PD, some molecular-genetics-based studies suggested that these features are not *locus*-specific but more frequent in AR-JP than in idiopathic PD (Klein, Pramstaller *et al.*, 2000; Lucking, Durr *et al.*, 2000).

Since impairment of the function of mitochondrial respiratory chains had been reported in PD (Mizuno, Ohta *et al.*, 1989; Schapira, Cooper *et al.*, 1989; Hattori, Tanaka *et al.*, 1991), a causal relationship between the inhibition of respiratory chain functions and the release of superoxide anions within mitochondria was reported (Hasegawa, Takeshige *et al.*, 1990; Guidot, Mccord *et al.*, 1993), Mn-superoxide dismutase is inducible by these released anions (Hassan, 1988), and increased activity of this enzyme had been reported in SN of PD patients (Saggu, Cooksey *et al.*, 1989); thus a possible role of this protein in PD was suspected. By linkage analysis of a diallelic polymorphism within the Mn-

superoxide dismutase gene (*SOD2*), Matsumine and colleagues, found a family with AR-JP showing perfect segregation of the disease with the *SOD2 locus* (Matsumine, Saito *et al.*, 1997). By extending the linkage analysis to thirteen families with AR-JP, they discovered strong evidence for the localization of the AR-JP gene at chromosome 6q25.2-27, including the *SOD2 locus*. However, sequence analysis of the *SOD2* coding region did not reveal any causative mutation. These results were confirmed a year later in several families with distinct ethnic backgrounds (Jones, Yamamura *et al.*, 1998). Further analyses, with additional markers in this region, allowed them to construct a detailed genetic map of the linked region which localized the AR-JP *locus* to a 6,9 cM region within this chromosome band, excluding *SOD2*. Since a high proportion of families with different ethnicity were found to have linkage to this AR-JP *locus*, it was established as a common *locus* for early-onset Parkinsonism (Tassin, Durr *et al.*, 1998).

In 1998, Kitiada and collaborators (Kitada, Asakawa *et al.*, 1998) investigated a small chromosomal deletion in a Japanese AR-JP patient, previously described by Matsumine's group (Matsumine, Yamamura *et al.*, 1998). By positional cloning within this deletion, they isolated a 2,960 bp cDNA with a 1,395-base-pair open reading frame, encoding for a 465 aminoacid protein with moderate similarity to ubiquitin at the amino terminus and a RING-finger motif at the carboxy terminus. The gene spanned more than 500 kb and had 12 exons, five of which were deleted in that particular patient. The gene was named *PARK2/PRKN* and the encoded protein parkin.

Parkin is an E3 Ubiquitin-Protein ligase (Shimura, Hattori *et al.*, 2000; Zhang, Gao *et al.*, 2000) with a ubiquitin-like domain (32% identity to Ubiquitin) at its N-terminus end, and two RING finger motifs separated by a cysteine-rich region known as IBR (in Between

Ring) at its C-terminus end. E3 ligases are part of the cellular machinery that target proteins with ubiquitin molecules for their degradation in the proteasome.

As extensively reviewed (Giasson and Lee, 2001; Ciechanover and Brundin, 2003), this ubiquitin–proteasome system (UPS) plays a major role in many vital cellular processes, and its dysfunction has been implicated in the pathogenesis of neurodegenerative disorders. Ubiquitin is a 76 aminoacid protein that is covalently linked to the lysine side chain of a substrate. This is called ubiquitination and occurs through three sequential steps catalyzed by ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin-ligase (E3) enzymes. By repeating this process, additional ubiquitin molecules are linked to a previously linked ubiquitin resulting in the formation of a polyubiquitin chain. Polyubiquitinated proteins are recognized by the 26S proteasome, a protein complex with proteolytic activity, for their degradation into peptide fragments. The 19S cap subunit (located on both ends of a central 20S subunit) recognizes the polyubiquitin chain, removes it and cleaves it into single ubiquitin monomers with deubiquitinating enzymes such as ubiquitin C-terminal hydrolase L1 (UCH-L1). Afterwards it unfolds the targeted protein and feeds it into the inner chamber of the 20S subunit. There, the target protein is proteolytically cleaved into small peptides that are released to the cytosol. It is easy to hypothesize how a failure in parkin (E3 ubiquitin-protein ligase), may avoid some proteins to be properly degraded causing dysfunction and eventually the death of certain neurons. Several groups have focused their research on identifying different parkin substrates. Thus CDCrel-1, Synphilin-1, glycosilated  $\alpha$ -synuclein, Pael-Receptor, a-Tubulin/b-Tubulin, Synaptogamin XI, p38, WSEPT5\_v2/CDCrel-2 and Cyclin E has been proposed as putative parkin substrates (Zhang, Gao *et al.*, 2000; Chung, Zhang *et*

*al.*, 2001; Imai, Soda *et al.*, 2001; Shimura, Schlossmacher *et al.*, 2001; Choi, Snyder *et al.*, 2003; Corti, Hampe *et al.*, 2003; Huynh, Scoles *et al.*, 2003; Ren, Zhao *et al.*, 2003; Staropoli, Mcdermott *et al.*, 2003).

Since the first mutation was identified in *PRKN* (Kitada, Asakawa *et al.*, 1998) a multitude of mutations have been described: deletions of single or multiple exons, duplications or triplications of exons, frameshift mutations, point mutations (both missense and nonsense mutations) and splice site (intronic) mutations (Abbas, Lucking *et al.*, 1999; Lucking, Durr *et al.*, 2000; West, Periquet *et al.*, 2002; Mata, Alvarez *et al.*, 2005). A comprehensive review of 42 articles (those published between 1998 and 2003) on the identification of parkin mutations revealed a total of 95 different mutations including 40 exon rearrangements (26 deletions and 14 multiplications), 43 single base pair substitutions and 12 small deletions or insertions of one or several base pairs. The most common mutations were deletions of exon 4, exon 3, exons 3 and 4, a point mutation in exon 7 (c.924C>T), and a single base pair deletion in exon 2 (255/256delA). These five alterations account for 35% of all *PRKN* mutations described thus far (Hedrich, Eskelson *et al.*, 2004).

Whether heterozygous *PARK2* mutations are involved in the development of late-onset autosomal dominant PD is still a field of much discussion. Thus, some groups have reported heterozygous Late-onset-PD-cases-mutation carriers in families with parkin mutation. A good example is a Brazilian kindred reported by Khan and colleagues (Khan, Horta *et al.*, 2005). In this family, an affected family member is identified to carry a heterozygous deletion in exon 4 of *PARK2*. Interestingly, another sibling who possessed an identical *PARK2* haplotype by descent was unaffected. The authors argued that this

supported the idea that genomic variation elsewhere or environmental influence, might be interacting with the single deletion to cause disease, although clearly it could also suggest no causal relationship of disease. Besides this report, several have been published showing *PARK2* mutations in late-onset PD cohorts. Klein and collaborators described two affected female individuals carrying a heterozygous deletion in heterozygous state in late onset PD patients, concluding that this gene may be important in the etiology of the more frequent late-onset typical Parkinson's disease (Klein, Pramstaller *et al.*, 2000). Oliveira *et al* showed that heterozygous mutations, especially in exon 7, trended to a later age at onset than compound heterozygotes and homozygotes, concluding that heterozygous mutations in *PARK2* act as susceptibility alleles for late-onset PD (Oliveira, Scott *et al.*, 2003) . Similar results were later obtained by other groups (Foroud, Uniacke *et al.*, 2003; Schlitter, Kurz *et al.*, 2006).

Conversely, several groups concluded after their experiments that *PARK2* is not involved in the pathogenesis of late-onset PD, either familial or sporadic. Oliveri and collaborators (Oliveri, Zappia *et al.*, 2001) failed to find any *PARK2* mutation in 23 patients from 18 families with autosomal recessive late-onset PD and 95 patients with sporadic late-onset PD. However, since they did not perform dosage analysis, their results are incomplete. Beside this “incomplete” report, some groups failed to find any association between heterozygous mutations in *PARK2* and late-onset PD (Lincoln, Maraganore *et al.*, 2003; Munhoz, Sa *et al.*, 2004; Poorkaj, Moses *et al.*, 2005; Kay, Moran *et al.*, 2007).

. ***PARK3***: In 1998 Gasser and colleagues mapped PD to a 10.3 cM region within chromosome 2p13 in 6 families with autosomal dominant PD that clinically resembled

typical sporadic PD (Gasser, Muller-Myhsok *et al.*, 1998). The mean age of onset ranged from 54 to 63 years and LBs were present. No further neurological deficit was observed except for dementia, which was prominent in several affected individuals. Results from F-Dopa-PET and J-IPT-SPECT scanning further supported the diagnosis of typical PD.

Further analysis of genetic markers within this 10.3 cM region showed that two of families supporting linkage to this *locus*, shared a common haplotype segregating with the disease. This haploype covered an interval of 3.2 cM and was given the name of *PARK3*. Although no genealogical link was found between these two families, the ancestors of both of them originate from neighboring regions of southern Denmark and northern Germany. This finding may indicate the existence of a founder mutation which might also be present in other PD cases from the same area. In 2001, in order to narrow down the shared haplotype in these two families, West and colleagues (West, Zimprich *et al.*, 2001), genotyped 17 microsatellite markers on chromosome 2p13 in all available family members. They managed to refine the disease *locus* to a 2.5 Mb region within chromosome 2p13. Sequencing of all 14 genes contained in that area failed to find any causative mutation. In 1999 Klein and collaborators (Klein, Vieregge *et al.*, 1999), evaluated patient and control population from northern Germany for those markers in chromosome 2p13. They found that more controls than PD patients shared the presumed full or partial haplotype with affected members of the previously described two families. Therefore, the putative founder mutation on chromosome 2p13 appears to be a rare cause of sporadic or familial PD in patients from southern Denmark/northern Germany. In 2002 DeStefano and colleagues (Destefano, Lew *et al.*, 2002), performed a genomewide scan using 103 multiplex families and identified four regions with evidence of linkage to PD.

One of these regions was on chromosome 2, coinciding with the location of *PARK3 locus*. In addition to evidence of linkage, they also observed association between a specific allele of a marker within *PARK3 locus* and the age at onset of PD. Given this association they focused on this marker when typing additional individuals and observed both an increase in the lod score value and a decrease in the P value. These results suggest that this marker is in linkage disequilibrium with a putative causative mutation or a factor influencing age of onset in these families. In 2003 Karamohamed and colleagues (Karamohamed, Destefano *et al.*, 2003), defined a smaller haplotype containing three SNPs associated with onset age containing the sepiapterin reductase gene (*SPR*). Pankratz *et al.* in 2004 (Pankratz, Uniacke *et al.*, 2004), performed genome-wide linkage analysis over families with two or more members having a reported age at PD onset and identified several chromosomal regions linked to the age at onset, including one overlapping with *PARK3 locus*. Although the gene containing mutations that cause this variant of PD has not yet been described, this *locus* seems to be related to the age at onset. Until mutations are identified, this should be considered as a putative *PARK locus*.

. ***PINK1, PARK6***: Valente and collaborators on 2001 (Valente, Bentivoglio *et al.*, 2001), identified a family (the Marsala kindred) for which linkage to *PARK2* was excluded. After performing homozygosity mapping, they identified a second *locus* for AR-JP (*PARK6*) on the short arm of chromosome 1 (1p35-p36). They mapped a 12.5 cM region containing a large number of both genes and predicted transcripts.

In order to assess the role of *PARK6* in parkin-negative AR-JP in Europe, the same group performed linkage analysis in 28 European AR-JP families in which *PARK2* mutations

had been excluded (Valente, Brancati *et al.*, 2002). Multipoint analysis in 8 families showing evidence of linkage to *PARK6* and the subsequent haplotype analysis, refined the *PARK6 locus* to a 9 cM region on chromosome 1p35-p36. Hatano and colleagues in 2004 (Hatano, Sato *et al.*, 2004), found five Asian families with linkage to *PARK6*, indicating that this form of parkinsonism is not limited to European countries.

That same year, Valente and collaborators identified a consanguineous Spanish family with linkage to *PARK6 locus*. This family presented a recombination event that narrowed the candidate area to 3.7 cM. Further SNPs mapping and new short tandem repeat markers, allowed them to define a 2.8 Mb region of homozygosity containing 40 genes. Sequence analysis of candidate genes in affected members from each of the described consanguineous families, led to the identification of two homozygous mutations in the PTEN-induced putative kinase 1 gene (*PINK1*): G390D in the Spanish family and W4370PA in both Italian families (Valente, Abou-Sleiman *et al.*, 2004). *PINK1* contains eight exons spanning 1.8 kilobases and encodes a 581 aminoacid protein. The transcript is ubiquitously expressed and is predicted to encode a 34 aminoacid mitochondrial targeting motif and a highly conserved protein kinase domain that shows a high degree of homology to the serine/threonine kinases of the Ca<sup>2+</sup>/calmodulin family. The mutations they described occurred in the serine/threonine kinase domain and thus, putatively affected the kinase activity or substrate recognition. They hypothesized that pink1 might phosphorylate mitochondrial proteins in response to cellular stress, protecting against mitochondrial dysfunction.

Six novel pathogenic mutations in exons 3, 4, 5 and 6, in six unrelated families from different countries in Asia where described that same year by Hatano and colleagues

(Hatano, Li *et al.*, 2004). These results indicate that pathogenic mutations in *PINK1* are not limited to Europeans.

To further evaluate the pathogenic role of *PINK1* in early-onset PD and to find a the correlation between genotype and phenotype, Valente and collaborators (Valente, Salvi *et al.*, 2004), performed *PINK1* mutation analysis in 100 AR-JP. Two of them carried two mutations in *PINK1*, one homozygous and the other a compound heterozygous carrier. Interestingly, five sporadic cases carried heterozygous missense changes in *PINK1*. This supports the idea of heterozygous *PINK1* mutations being a risk factor for the appearance of sporadic PD.

Further mutation screenings in different populations, identified a three-nucleotide insertion in *PINK1* (Klein, Djarmati *et al.*, 2005) and a homozygous four-base insertion (Rohe, Montagna *et al.*, 2004). Several mutations has been described since (Bonifati, Rohe *et al.*, 2005; Klein, Djarmati *et al.*, 2005; Hedrich, Hagenah *et al.*, 2006; Ibanez, Lesage *et al.*, 2006; Leutenegger, Salih *et al.*, 2006; Tan, Yew *et al.*, 2006). *PINK1* mutations account for approximately 1 to 7% of AR-JP in Caucasians (Healy, Abou-Sleiman *et al.*, 2004a; Rohe, Montagna *et al.*, 2004; Valente, Salvi *et al.*, 2004), about 8.9% in Japanese AR-JP families (Li, Tomiyama *et al.*, 2005), and 2 to 3% in sporadic and familial PD of Chinese origin (Tan, Yew *et al.*, 2005; Tan, Yew *et al.*, 2006).

The clinical phenotype in families with homozygous *PINK1* mutations are similar to those in *PARK2*-positive cases with an early onset of disease (later than *PARK2*-positive cases), excellent and sustained response to L-dopa, slow progression, frequent L-dopa-induced motor fluctuations and dyskinesias. However, distinguishing features at onset such as dystonia, sleep benefit, or psychiatric disturbances, are absent.

. ***DJ-1, PARK7***: Duijn and colleagues, linked PD in a consanguineous family to chromosome 1p36 (Van Duijn, Dekker *et al.*, 2001). After saturating the region with additional markers and performing multipoint linkage analysis, they mapped the disease in this family to a 16 cM region in chromosome 1p36, 25 cM away from *PARK6 locus*. This *locus* was given the name of *PARK7*.

In four families with early-onset parkinsonism from Italy, Uruguay and the Netherlands, Bonifati and colleagues found two with linkage to *PARK7* (Bonifati, Breedveld *et al.*, 2002). Fine mapping reduced the critical region to a 5.6 Mb region, containing 90 genes. Sequencing of obvious candidate genes did not reveal any mutation in either of the two families analyzed. By means of systematic screening of transcripts in the region, they found a 14,082 bp homozygous genomic deletion in *DJ-1* segregating with the disease on the Dutch family (Bonifati, Rizzu *et al.*, 2003). Sequencing of the *DJ-1* cDNA revealed a homozygous point mutation in the Italian family. That same year, Hague and colleagues (Hague, Rogaeva *et al.*, 2003) found a frameshift and a splice site mutation in a 24 year old hispanic woman with PD.

Abou-Sleiman *et al* (Abou-Sleiman, Healy *et al.*, 2003) studied the prevalence of *DJ-1* mutations (including the 14,082 bp deletion previously reported) in a large cohort of PD patients. They found a homozygous missense mutation and a compound heterozygous mutation in two early onset PD patients. In addition, five other heterozygous variants were identified in patients with a later age of onset PD. This suggests either that these *DJ-1* variants are non-causal or that, as previously discussed, mutations in AR-JP-related genes in a heterozygous state, confer susceptibility for later onset PD. The reported 14 Kb deletion was not found in any of the samples assayed.

Since the proportion of AR-JP families with *DJ-1* mutations was still unknown, Ibanez *et al.* (Ibanez, De Michele *et al.*, 2003) evaluated its frequency in a large set of consanguineous and non-consanguineous families with AR-JP, not linked to *PARK2*. No mutations were found in either of the families assayed. Hedrich's group (Hedrich, Djarmati *et al.*, 2004) also failed on finding any homozygous *DJ-1* mutations in the cohort they studied. However, two carriers of single heterozygous mutations, including a heterozygous deletion of exons 5 to 7 and an 11 base pair deletion were found. Since 17 cases of these same cohort harbored mutations in *PARK2*, the authors concluded that mutations in *DJ-1* are less frequent than mutations in *PARK2* in AR-JP.

In aggregate several groups concluded that *DJ-1* mutations are a very rare cause of early onset PD. (Hague, Rogaeva *et al.*, 2003; Clark, Afridi *et al.*, 2004; Healy, Abou-Sleiman *et al.*, 2004c; Lockhart, Bounds *et al.*, 2004; Lockhart, Lincoln *et al.*, 2004; Tan, Tan *et al.*, 2004).

*DJ-1* encodes a 189 aminoacid highly conserved protein that belongs to the DJ-1/ThiJ/PfpI superfamily. Crystal structure analyses revealed that, at least *in vitro*, DJ-1's native structure is a homodimer (Moore, Zhang *et al.*, 2003). Its distribution is mainly cytoplasmatic with a slight proportion localized in the mitochondria (Zhang, Shimoji *et al.*, 2005). It appears to have several cell functions, such as RNA stabilization and protection against oxidative stress. DJ-1 can also directly act against Reactive Species of Oxygen (ROS) such as hydrogen peroxide, which he can eliminate by auto-oxidation means (Taira, Saito *et al.*, 2004). However its function remains unclear and further studies are needed. The clinical symptoms in these patients seem to be similar to those found in parkin families with a later age at onset and a slower progression of the disease.

. ***LRKK2, PARK8***: In 1997, a family with autosomal dominant PD, was reported in Sagamihara City, Japan. Clinical features of this family did not differ from common PD, however, the loss of melanin-containing cells was mild to moderate, the number of neurons of the *locus ceruleus* was maintained and LBs were not detected (Hasegawa and Kowa, 1997). Five years later, Funayama and colleagues (Funayama, Hasegawa *et al.*, 2002) performed genome-wide linkage analysis over this family and mapped the disease to a 13.6 cM interval of chromosome 12p11.23–q13.11. This region (*PARK8*) contained 116 genes or predicted transcripts. Later, Paisan-Ruiz and coworkers (Paisan-Ruiz, Saenz *et al.*, 2005), identified 4 families from the Basque region in northern Spain with symptoms clinically compatible with idiopathic autosomal dominant PD. After confirming linkage to *PARK8 locus*, they narrowed this susceptibility *locus* to a 9.35 cM region containing 75 genes and predicted transcripts. Sequence analysis of genes and predicted open frames in that region showed that all four Basque families carried a very rare allele within one of the candidate genes (*KIF21A*), indicating that these families were ancestrally related. Genotyping of several SNPs flanking this allele, allowed them to define a minimum haplotype shared by all the affected members in all four families spanning 2.6 Mb (11genes). By means of direct sequencing, they found a segregating mutation (R1441G) within a putative kinase domain containing transcript, DKFZp434H2111. Moreover, the change Y1699C was found within the same predicted transcript in an English kindred analyzed. At that same time Zimprich and colleagues identified four missense mutations and a putative splice site mutation in two large German-Canadian and Nebraskan families with autosomal dominant late-onset PD

previously linked to *PARK8 locus* by Zimprich and coworkers in 2004 (Zimprich, Biskup *et al.*, 2004; Zimprich, Muller-Myhsok *et al.*, 2004).

The protein encoded by this predicted transcript belongs to the ROCO complex protein family (Bosgraaf and Van Haastert, 2003). It contains an armadillo and ankyrin repeat region, a leucine-rich repeat domain, a kinase domain, a ROC (Ras On Complex) GTPase domain, a COR domain (C-terminal Of ROC) and a WD40 domain. The gene was given the name of *LRRK2* and the encoded protein dardarin, from the Basque word meaning tremor. Several mutations have been reported to date. By means of segregation analysis, six mutations have been proven to be pathogenic: I1122V, R1441C, R1441G, Y1699C, G2019S, and I2020T (Mata, Wedemeyer *et al.*, 2006). One of them, G2019S (Kachergus, Mata *et al.*, 2005), is the most prevalent and has been found at frequencies of more than 2% in general North American clinical population and English PD Brain Bank specimens. Early studies have quoted the prevalence of G2019S to be about 5 to 6% for familial and 1 to 2% for apparently sporadic cases of PD (Di Fonzo, Rohe *et al.*, 2005; Gilks, Abou-Sleiman *et al.*, 2005; Skipper, Shen *et al.*, 2005; Tan, Shen *et al.*, 2005). However subsequent studies have shown that this frequency depends on the ethnicity of the cohorts tested being at remarkably high frequencies in clinic-based Portuguese, Ashkenazi Jewish, and North African Arabian patients, even in the absence of a clear family history of disease (Bras, Guerreiro *et al.*, 2005; Lesage, Durr *et al.*, 2006; Ozelius, Senthil *et al.*, 2006; Ferreira, Guedes *et al.*, 2007b). Interestingly, a neurologically healthy octogenarian with the *LRRK2* G2019S mutation has been reported, suggesting that environmental epigenetic and other genetic factors may be implicated in expression of the disease phenotype (Kay, Kramer *et al.*, 2005).

. ***ATP13A2, PARK9***: In 1994 Al-Din and co-workers (Najim Al-Din, Wriekat *et al.*, 1994), reported an unusual neurological syndrome in an Arab family with five affected siblings. The affected subjects had clinical signs and symptoms of severe parkinsonism but the age at onset was early (11-16 years) and the disease was rapidly progressive. Additional features of these affected individuals were spasticity resulting from corticospinal tract degeneration, supranuclear up gaze paresis, and the development of dementia. They also showed significant atrophy of the globus pallidus and the pyramids, as well as generalized brain atrophy in later stages. Therapy with levodopa resulted in significant improvement in extrapyramidal dysfunction. Since there were multiple affected sibilings born from phenotypically normal consanguineous parents, an autosomal recessive inheritance was suggested. This syndrome was given the name of Kufor-Rakeb Syndrome (KRS), since this was the name of the village were the individuals were originally from. By means of autozygosity mapping, a common region of homozygosity in all affected subjects on chromosome 1p was identified (Hampshire, Roberts *et al.*, 2001). The disease was fine mapped to a 9 cM region on chromosome 1p36. In 2006, two different compound heterozygous mutations were found in *ATP13A2* within a Chilean family with autosomal-recessive PD that resembled the original KRS. A 22 bp duplication leading to a frameshift that causes a stop codon, was found in all affected members of the original Kufor-Rakeb syndrome family (Ramirez, Heimbach *et al.*, 2006).

### 1.6.3. Other genetic factors related to PD

In addition to these genes, candidate sequencing analysis and whole-genome linkage approaches have suggested a relationship between other genes and PD. We briefly mention some of these here:

- ***UCH-L1*, *PARK5***: Ubiquitin Carboxy-Terminal Hydrolase L1 (UCH-L1) is one of the most abundant proteins in the brain and it has been found in LBs (Lowe, Mcdermott *et al.*, 1990). It belongs to a family of deubiquitinating enzymes, and its role is to hydrolyze small C-terminal adducts of ubiquitin to generate ubiquitin monomers. Thus it is a perfect candidate gene to be implicated on the ethiology of PD. In 1998 Leroy and collaborators (Leroy, Boyer *et al.*, 1998), sequenced the coding region of *UCHL-1* in probands from 72 families with PD, and identified a missense mutation (I93M) in the fourth exon of *UCHL-1* in a German pedigree. They showed that this mutation caused a partial loss of the catalytic activity of the protease which could lead to aberrations in the proteolytic pathway and aggregation of proteins. Mutation analysis of the affected brother showed that he also carried this mutation. In both patients the phenotype was that of idiopathic PD with a slightly younger age at onset. The father of the affected sibling was not diagnosed with PD, indicating incomplete penetrance of this mutation in this family. Whether the I93M mutation is truly pathogenic is still an area of controversy. To date, no further mutations have been observed in follow-up studies (Harhangi, Farrer *et al.*, 1999; Lincoln, Vaughan *et al.*, 1999; Maraganore, Farrer *et al.*, 1999; Wintermeyer, Kruger *et al.*, 2000; Shi and Tao, 2003). Thus, on a pure genetic basis, the evidence of this mutation as a cause of PD is very weak.

In 1999 Lincoln and colleagues (Lincoln, Vaughan *et al.*, 1999), failed to find any mutations in *UCH-L1* in familial and sporadic PD. However they described a coding polymorphism (S18Y) within this gene. That same year, Maraganore and colleagues reported that S18Y is underrepresented in PD in European population, thus indicating a protective effect. Again, lot of controversy has been reported around this issue. Besides Maraganore's report several groups have supported the protective effect of S18Y (Wintermeyer, Kruger *et al.*, 2000; Satoh and Kuroda, 2001; Momose, Murata *et al.*, 2002; Facheris, Strain *et al.*, 2005; Tan, Puong *et al.*, 2006; Carmine Belin, Westerlund *et al.*, 2007). On the other hand, several studies failed on finding any association between S18Y and PD (Mellick and Silburn, 2000; Zhang, Hattori *et al.*, 2000; Levecque, Destee *et al.*, 2001; Savettieri, De Marco *et al.*, 2001; Wang, Zhao *et al.*, 2002; Elbaz, Levecque *et al.*, 2003). Interestingly, two of these studies (Wang, Zhao *et al.*, 2002; Elbaz, Levecque *et al.*, 2003) showed a protective effect when analysis was restricted to early onset PD patients.

Given this controversy, Maraganore *et al* performed a collaborative pooled analysis of all the association studies performed to date (the original plus 10 additional studies) to determine whether S18Y is actually associated with PD (Maraganore, Lesnick *et al.*, 2004). This meta-analysis confirmed an inverse association between S18Y variant and PD, particularly in younger subjects. However, although association findings were positive overall, they only considered one polymorphism. It is possible that this variant is only a marker for genetic association of PD with another *UCH-L1* variant or a variant within another gene.

Tan and colleagues (Tan, Puong *et al.*, 2006) believed that independent replication in Asian cohorts would further contribute to the debate of the protective effect of S18Y variant in PD. Thus, they performed a large case-control study of the *UCHL-1*, confirming the protective effect of S18Y against PD amongst the younger Chinese subjects.

. **NR4A2:** This gene encodes a member of the steroid-thyroid hormone-retinoid receptor superfamily called Nurr1. Encouraged by the fact that Nurr1 is critical for the midbrain dopaminergic cell development and survival, Xu and coworkers targeted it for a case-control association study they performed in 2002 (Xu, Liang *et al.*, 2002). As a result they found that the homozygous form of an insertion in intron 6 of this gene is significantly higher in both familial and sporadic PD compared to a control population. Although no statistically significant association was reported between the heterozygous form of this polymorphism and PD, Zheng and colleagues found it increased in both sporadic PD and diffuse Lewy Bodies disease (Zheng, Heydari *et al.*, 2003). That same year, two heterozygous mutations within the non-coding 5'-untranslated region of the gene was reported in 10 familial cases were reported. These mutations were verified to reduce the level of Nurr1 mRNA, implying a dominant negative effect of the mutated allele (Le, Xu *et al.*, 2003)

Although these results were encouraging, no further mutations have been found in several follow-up reports (Carmine, Buervenich *et al.*, 2003; Tan, Chung *et al.*, 2003; Wellenbrock, Hedrich *et al.*, 2003; Zimprich, Asmus *et al.*, 2003; Hering, Petrovic *et al.*, 2004; Ibanez, Lohmann *et al.*, 2004; Levecque, Destee *et al.*, 2004; Nichols, Uniacke *et*

*al.*, 2004), leading to the speculation by some authors that Nurr1 does not play a role in Parkinson's disease (Hering, Petrovic *et al.*, 2004).

. ***PARK10***: Hicks and collaborators (Hicks, Petursson *et al.*, 2002) using a genomewide scan in Icelandic patients with classic late-onset PD, found linkage to chromosome 1p32. The authors designated this *locus* *PARK10*. Later that same year, Li and collaborators (Li, Scott *et al.*, 2002) on the first genome screen for age at onset for both AD and PD, found linkage for PD on a region overlapping with *PARK10*. Although different traits were studied, these two studies provided strong evidence that there may be one or more genes of interest (*PARK10* or an age at onset gene) in this region. Interestingly, Li's group, has recently nominated *ELAVL4* to be affecting age at onset in PD (Noureddine, Qin *et al.*, 2005). The gene maps 2 Mb from the age at onset linkage peak on chromosome 1p32. Haugarvoll and colleagues (Haugarvoll, Toft *et al.*, 2007), suggested that the association of this gene, might be explained by a celtic founder effect. Thus, they used Norwegian and Irish case-control series, as well as a US series, to examine association between *ELAVL4* markers and age at onset or susceptibility to PD. Association was found between two markers and susceptibility to PD only in the Irish cases.

Oliveira *et al* (Oliveira, Li *et al.*, 2005) tested for association genes after a combined gene expression and linkage data, these included genes in *PARK10* (Li, Scott *et al.*, 2002). The eukaryotic translation initiation factor 2B3 gene (*EIF2B3*) showed the strongest association with age at onset of PD, and the HIV type 1 enhance-binding protein 3 gene (*HIVEP3*) was the only significant gene associated with the risk of developing PD.

Following this, Maraganore and colleagues (Maraganore, De Andrade *et al.*, 2005) in their two-tiered, whole-genome association study for PD, found that the two SNPs with the lowest P values tagged the *PARK10* late-onset PD susceptibility *locus*. These two SNPs were within *LOC200008* (now CUB Domain Containing Protein 2, *CDCP2*) which encodes a protein with inferred oxidoreductase activity and potential involvement in cholesterol biosynthesis and electron transport. They also found association between *HIVEP3* and PD.

Three recent follow-up studies failed to find any significant association between *CDCP2* and different PD cohorts. (Farrer, Haugarvoll *et al.*, 2006; Goris, Williams-Gray *et al.*, 2006; Li, Rowland *et al.*, 2006).

Considering that all the studies performed thus far had focused on unrelated cases-control samples, Li *et al.* (Li, Deng *et al.*, 2007) increased the density of SNPs within *HIVEP3* and *CDCP2* and expanded the study to a familial dataset. They concluded that *HIVEP3*, but not *CDCP2*, is a strong candidate gene associated with the risk of developing PD. As they did not find significant association with the age at onset in this cohort, they proposed the possibility that two different genes regulate age at onset and risk on developing PD in this region.

. ***PARK11***: Pankratz *et al* (Pankratz, Nichols *et al.*, 2002), in an attempt to identify genes contributing to PD, undertook a genome-wide linkage analysis in a large sample of American sibling pairs. Under two different models of PD affection status the highest LOD score values were obtained on chromosome X and chromosome 2. After including

additional individuals in this cohort, evidence of linkage to a 20 cM region of chromosome 2q36-37 was reported (Pankratz, Nichols *et al.*, 2003).

Prestel and colleagues (Prestel, Sharma *et al.*, 2005), failed to replicate Pankratz's results in an European population. In the whole-genome association study performed by Maraganore's group (Maraganore, De Andrade *et al.*, 2005), one of the SNPs with the lowest p values, tagged that region on chromosome 2q36-37.

Genome-wide linkage analyses such as those performed by Scott and Hicks did not find linkage at this *locus* (Scott, Nance *et al.*, 2001; Hicks, Petursson *et al.*, 2002).

. **PARK12:** As noted previously, Pankratz *et al.* in their two-model genome-wide linkage analysis (Pankratz, Nichols *et al.*, 2002) obtained the strongest evidence of linkage on chromosome X. These results were confirmed after they expanded the sample to 754 affected individuals, consisting of 425 sibling pairs from 362 families (Pankratz, Nichols *et al.*, 2003). Previous studies had noted linkage to this same *locus* (Scott, Nance *et al.*, 2001; Hicks, Petursson *et al.*, 2002).

An interesting candidate gene in this chromosomal region may be in the *locus* for sex-linked dystonia with parkinsonism (XDP), a disease identified at high incidence in Panay, Philippines. See below.

. **X-linked recessive dystonia with parkinsonism:** X linked recessive dystonia with parkinsonism (XDP) is a movement disorder unique to adult Filipino men whose ancestry can be traced to Panay Island, Philippines. It is characterized by severe, progressive torsion dystonia, with parkinsonian features occurring later in the disease course. The

syndrome was first noted in 1970 and referred to as “dystonia musculorum deformans”, Lee *et al* (Lee, Pascasio *et al.*, 1976), where the first on publishing an unusually high frequency of torsion dystonia in Panay, the sixth largest island in Philippines. Pedigree analysis of the 28 reported cases revealed six families with several members affected. No male to male transmission occurred, suggesting x-linked recessive transmission. In addition, it was noted that some patients had parkinsonian features and relatives with parkinsonism. Fahn *et al* confirmed the coexistence of dystonia and parkinsonism in XDP in these cases and called the disease “lubag” based on the local term used to describe the abnormal movemen (Fahn, 1988).

In 1990 Kupke *et al* (Kupke, Lee *et al.*, 1990) performed linkage analysis in seven Filipino families, using microsatellite markers within chromosome X. The maximum LOD score value was obtained for a marker located at Xq21.3. The *locus* was later refined to chromosome Xq12-13.1 from linkage analyses performed by different groups (Wilhelmsen, Weeks *et al.*, 1991; Kupke, Graeber *et al.*, 1992). Mapping of the disease to this *locus* (*DYT3*) was confirmed afterwards (Graeber, Kupke *et al.*, 1992; Muller, Haberhausen *et al.*, 1994).

Later, by means of a YAC contig construction, they disease was further refined to a 1.8 Mb region in chromosome Xq13.1 (Haberhausen, Schmitt *et al.*, 1995) and narrowed down to <350 kb by Nemeth and colleagues (Nemeth, Nolte *et al.*, 1999). Subsequently, Nolte and colleagues used PCR-based sequencing and screening analyses to report four SNPs and five disease-specific sequence changes in the “Multiple Transcript System” within 260 kb of the *DYT3* region (Nolte, Niemann *et al.*, 2003). Since PCR often fails to find large sequence variants such and brain specimens were not included in this analysis,

Makino and collaborators performed complete genomic DNA sequence of that region and detailed expression analysis of the gene in brain specimens obtained from patients with XDP. Their results suggest that SVA retrotransposon insertion into *TAF1* may cause XDP by altering the expression of taf1 isoforms (Makino, Kaji *et al.*, 2007).

. ***Omi/HtrA2, PARK13***: This gene was first linked to neurodegeneration when Gray and collaborators (Gray, Ward *et al.*, 2000) showed that Omi/HtrA2 interacts with presenilin-1, previously related to young-onset inherited forms of Alzheimer's disease (Sherrington, Rogaev *et al.*, 1995). Further support for a role of Omi/HtrA2 in neurodegeneration came when Park and coworkers demonstrated that Amyloid- $\beta$ , a key factor implicated in the pathogenesis of Alzheimer's disease, is directly and efficiently cleaved by the Omi/HtrA2 serine-protease activity both *in vitro* and *in vivo* (Park, Seong *et al.*, 2004). In addition Jones and colleagues (Jones, Datta *et al.*, 2003) found a missense mutation (p.S276C) in the protease domain of Omi/HtrA2 of the mouse mutant *mnd2* (motor neuron degeneration 2). Loss of neurons in the striatum resulting in a neurodegenerative disorder with parkinsonian features was described in *OMI/HTRA2*-knockout mice (Martins, Morrison *et al.*, 2004); these findings lead Strauss and colleagues to look for mutations in *OMI/HTRA2* in a large sample of 518 German PD (Strauss, Martins *et al.*, 2005). They identified two novel mutations in the *Omi/HtrA2* gene that resulted in defective activation of the protease activity of the protein codified by this gene *in vitro*. Moreover, these mutations induced mitochondrial dysfunction associated with altered mitochondrial morphology that made the cells more susceptible to stress-induced cell death than wild-type. These results provide a novel link between

mitochondrial dysfunction and neurodegeneration in PD. Interestingly, they also found this protein as a component of LBs in idiopathic PD patients.

• **Association with Gaucher's disease:** Gaucher disease (GD) is a lipid storage disorder due to the inherited deficiency of the lysosomal enzyme acid  $\beta$ -glucosidase due to mutations in the  $\beta$ -glucocerebrosidase gene (*GBA*). There are three subgroups within the GD phenotype: Type 1 (OMIM #230800), 2 (OMIM #230900) and 3 (OMIM #231000). GD type 1 is the most common and least severe variant and has not been associated with neurological manifestations. Type 2 is a rapidly progressive disorder with death before the first year of life. Type 3 has a varying degree of systemic involvement as well as at least one neurological manifestation. This is not a strict classification scheme and particular cases can cross boundaries. Several GD type 1 cases have been reported to develop parkinsonian symptoms (Vanbogaert L, 1939; Sack, 1980; Mckeran, Bradbury *et al.*, 1985; Sun, 1986; Turpin Jc and Boutry Jm, 1987; Cormand, Vilageliu *et al.*, 1995; Tytki-Szymanska, Millat *et al.*, 1996). However, the first publication to clearly report the coexistence of PD with type 1 GD in six cases, was that by Neudorfer in 1996 (Neudorfer, Giladi *et al.*, 1996). These subjects suffered from relatively mild GD symptoms but with early onset and L-dopa refractory parkinsonian manifestations including tremor, bradykinesia, rigidity, and often cognitive decline. Analysis of the *GBA* gene in 57 autopsy-confirmed PD cases with broad age of onset identified 8 mutation carriers (Lwin, Orvisky *et al.*, 2004). In 2005 association between *GBA* mutation and PD was reported, suggesting that heterozygous mutations in *GBA* might be a very rare cause of PD (Sato, Morgan *et al.*, 2005).

The way mutations in this gene might lead to the development of PD is not known. However, a hypothesis would be that misfolding of aberrant glucocerebrosidase may affect lysosomal targeting and proteasome function, predisposing to synucleinopathies, leading to improper degradation of  $\alpha$ -synuclein and contribute to aggregate formation (Lwin, Orvisky *et al.*, 2004), has also been reported.

. **Relationship between SCA2/SCA3 and PD:** Autosomal dominant spinocerebellar ataxias (SCAs) were first described at the end of the 19<sup>th</sup> century. (Brown, 1892; Marie, 1893; Babinski, 1899). The phenotype was investigated in detail by Schut (Schut, 1950) and subclassifications were attempted by Greenfield (Greenfield, 1954) and Harding (Harding, 1983).

Orozco *et al.* described a large cohort from the Holguin province in Cuba with autosomal dominant ataxia (Orozco, Estrada *et al.*, 1989). Linkage to *SCA1 locus* was excluded in this population, providing the first evidence of *locus* heterogeneity for this disorder (Auburger, Diaz *et al.*, 1990). Three years later the disease-causing *locus* was mapped to a 30 cM region in chromosome 12q23-q24.1 (Gispert, Twells *et al.*, 1993). This region was narrowed down 2 years later thanks to the linkage results obtained by Hernandez and collaborators over 11 families from the same geographical region (Hernandez, Magarino *et al.*, 1995). Moreover, their haplotype analysis results were consistent with the hypothesis of a founder effect of *SCA2* mutations in Holguin province in Cuba.

Pulst *et al.* (Pulst, Nechiporuk *et al.*, 1993) identified an Italian pedigree with linkage to chromosome 12q. In contrast with the Cuban pedigree, this one showed a remarkable degree of anticipation of age at onset. Given that (CAG)<sub>n</sub> triplet expansion had been

identified in *ATXN-1* (Orr, Chung *et al.*, 1993), they proposed that something similar could be underlying the disease in this pedigree. This was confirmed when a 150 Kda protein in a patient linked to *SCA2*, using an antibody that identifies only extended polyglutamine tracts, was detected (Trottier, Lutz *et al.*, 1995). They latter found (CAG)<sub>n</sub> repeats in a novel gene (*ATXN-2*) encoding ataxin 2 (Pulst, Nechiporuk *et al.*, 1996).

Given that anticipation had been described on familial parkinsonism (Payami, Bernard *et al.*, 1995) and that probands with parkinsonism had been reported in families linked to *SCA2* (Gwinn-Hardy, Chen *et al.*, 2000; Shan, Soong *et al.*, 2001), Payami and collaborators decided to look for *ATXN-2* mutations in patients with parkinsonism in which family history had been reported (Payami, Nutt *et al.*, 2003). They found two probands with expanded CAG alleles of *ATXN-2*.

On the other hand, Kawaguchi and collaborators, found 68-69 CAG expansions within *ATXN-3* in patients diagnosed with Machado-Joseph disease (MJD) (Kawaguchi, Okamoto *et al.*, 1994). In 2001, encouraged by the fact that parkinsonism had previuously been reported as the predominant clinical feature of some variants of MJD patients (Tuite, Rogaeva *et al.*, 1995), Gwinn-Hardy and colleagues successfully screened for (CAG)<sub>n</sub> trinucleotide repeats in a large black family with apparent autosomal dominant PD (Gwinn-Hardy, Singleton *et al.*, 2001).

These results together, supported the hypothesis that some forms of parkinsonism with anticipation, may be caused by trinucleotide expansions. Several studies have further supported the hypothesis that *ATXN-2* and *ATXN-3* expansion mutations may be a cause of disease in patients with a history of parkinsonism. (Shan, Soong *et al.*, 2001; Furtado, Farrer *et al.*, 2002; Ragothaman, Sarangmath *et al.*, 2004; Wilkins, Brown *et al.*, 2004;

Simon-Sanchez, Hanson *et al.*, 2005; Lim, Zhao *et al.*, 2006; Modoni, Contarino *et al.*, 2007).

. **Microtubule-Associated Protein Tau:** One of the regions giving the highest lod score in the genomic screen for PD Scott and collaborators performed (Scott, Nance *et al.*, 2001), was on chromosome 17q21. This region contains the gene encoding Microtubule-Associated Protein Tau (*MAPT*). Tau is a neuronal microtubule-associated protein that plays a central role in microtubule assembly and maintenance. Pathological accumulations of tau, known as neurofibrillary tangles, characterize a number of tauopathic conditions, including AD, PSP, corticobasal degeneration, prion diseases, and amyotrophic lateral sclerosis/parkinsonism-dementia complex (Conrad, Andreadis *et al.*, 1997; Bennett, Bonifati *et al.*, 1998; Baker, Litvan *et al.*, 1999; Bonifati, Joosse *et al.*, 1999).

Mutations in *MAPT* have been found to cause familial variants of frontotemporal dementia (Clark, Poorkaj *et al.*, 1998; Hutton, Lendon *et al.*, 1998; Poorkaj, Bird *et al.*, 1998; Spillantini, Murrell *et al.*, 1998). Besides, there is a strong association between common variants in *MAPT* and cortical basal degeneration, Pick's disease, PSP and AD. (Conrad, Andreadis *et al.*, 1997; Conrad, Vianna *et al.*, 2002; Zhukareva, Mann *et al.*, 2002).

Pastor and colleagues first studied *MAPT* in a cohort of PD patients and controls (Pastor, Ezquerra *et al.*, 2000). They reported association of a dinucleotide polymorphism (A0) with PD compared with controls. These results were confirmed in a meta-analysis

performed by Golbe and coworkers that included all the studies published thus far (Golbe, Lazzarini *et al.*, 2001).

Martin and colleagues in 2001 (Martin, Scott *et al.*, 2001) underwent a positional association study utilizing many of the families they used in their former PD whole genome linkage screen (Scott, Nance *et al.*, 2001). They found that the H1 haplotype, previously reported to be related to PSP (Baker, Litvan *et al.*, 1999), was associated with PD. They also found allele A0 to be significantly overrepresented, and A3 to be underrepresented in PD patients. Kwok and colleagues (Kwok, Teber *et al.*, 2004) identified a series of biallelic variants in the promoter region of the *MAPT* gene, which were in complete linkage disequilibrium and are part of haplotypes H1 and H2. They showed that the H1 promoter haplotype, which is the strongest at initiating transcription, is significantly overrepresented in idiopathic PD patients compared with normal controls, suggesting that increased *MAPT* expression confers susceptibility for PD. This suggested that tau protein levels are an important factor in the cause of idiopathic PD. They also looked for recombination events between the promoter haplotypes and the coding region ones. Their data strongly suggested that the promoter haplotypes form part of an extended haplotype with the entire *MAPT* gene. Since then, those polymorphisms defining H1 haplotype have been investigated for an association with PD with conflicting results. Two meta-analyses support an association between haplotype H1 and PD (Healy, Abou-Sleiman *et al.*, 2004b; Zhang, Song *et al.*, 2005).

Besides association studies, other studies support the involvement of tau in PD pathogenesis. Hence, there is evidence that tau can co-aggregate with  $\alpha$ -synuclein in LBs (Ishizawa, Mattila *et al.*, 2003) and that these two proteins interact to promote and

propagate the polymerization of each other into fibrils (Ishizawa, Mattila *et al.*, 2003). In addition, tau aggregation has been seen in AR-JP caused by parkin gene mutations (Mori, Kondo *et al.*, 1998).

. **Apolipoprotein E:** Apolipoprotein E is a polymorphic protein, highly expressed in the brain (*APOE*, chromosome 19q13.2). There are three major isoforms:  $\epsilon 2$ ,  $\epsilon 3$  and  $\epsilon 4$ . Since the  $\epsilon 4$  isoform is a very well known genetic risk factor for familial and sporadic AD (Corder, Saunders *et al.*, 1993; Saunders, Schmechel *et al.*, 1993), several studies have been published about a possible association between *APOE* and PD. Most have reported no association between  $\epsilon 4$  and the development of PD itself or dementia in PD patients (Arai, Muramatsu *et al.*, 1994; Marder, Maestre *et al.*, 1994; Rubinsztein, Hanlon *et al.*, 1994; Ibarreta, Gomez-Isla *et al.*, 1995; Poduslo, Riggs *et al.*, 1995; Helisalmi, Linnaranta *et al.*, 1996; Whitehead, Bertrand *et al.*, 1996; Bon, Jansen Steur *et al.*, 1999; Kruger, Vieira-Saecker *et al.*, 1999; Li, Hauser *et al.*, 2004). However, in a meta-analysis performed by Huang and colleagues (Huang, Chen *et al.*, 2004), an association between  $\epsilon 2$  carriers and sporadic PD was found. This had been previously prompted by Harhangi and collaborators (Harhangi, De Rijk *et al.*, 2000).

The effect of *APOE* variants on age of onset of PD has also been investigated. Thus, Zarepari and colleagues (Zarepari, Camicioli *et al.*, 2002) presented data that suggested that patients with  $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$  genotypes had an age at onset much earlier than those with  $\epsilon 3/\epsilon 3$  genotype. Moreover, they saw a trend of earlier onset of PD in  $\epsilon 2/\epsilon 3$  patients than in  $\epsilon 3/\epsilon 3$ . Interestingly, Maraganore *et al.* (Maraganore, Farrer *et al.*, 2000) had previously described an earlier age at onset of those PD patients with the  $\epsilon 2/\epsilon 3$  *APOE*

genotype. However, this is still a controversial issue (De La Fuente-Fernandez, Sellers *et al.*, 1998; Inzelberg, Schechtman *et al.*, 2002; Parsian, Racette *et al.*, 2002; Li, Hauser *et al.*, 2004; Pankratz, Byder *et al.*, 2006). *APOE* has also been related to other features related with PD like appearance of drug-induced hallucination or psychosis along the course of PD (De La Fuente-Fernandez, Nunez *et al.*, 1999; Goetz, Burke *et al.*, 2001; Feldman, Chapman *et al.*, 2006).

#### **1.6.4. Genome-wide association studies (GWAS) and PD.**

Although six monogenic forms of PD have been identified, ~ 95% of the PD cases are apparently sporadic variant in which both genes and environmental risk factors are thought to play an important role. In this variant of the disease, genetic variability, unlike that causing familial disease, refers to common variants more or less extended in normal population and only exerting adverse or protective effects when appearing in combination with other genetic or environmental factors. This grouping into two apparently separate variants does not imply that these two types of the disease should be considered as different entities. In fact, because of the clinical and pathological similarities existing between familial and non-monogenic PD, knowledge of the genes causing the familial disease could bring new ideas for the research of sporadic forms of the disease.

Before the advent of GWAS the search for genetic variants that alter risk for disease was dominated by candidate gene association studies, usually focused on genetic variants that altered the coding sequence of a gene. Generally, positive associations would be reported and negative associations, unless a refutation of a previous report would go largely unpublished. Genuine success required sufficient understanding of the disease process to

allow selection of the correct gene, the right variants to type within the gene and most importantly the presence of variability within the gene that altered function/expression. If we assume there are ~30,000 genes and a couple of million common variants, the odds are clearly stacked against being successful. GWAS attempts to solve this inherent bias by offering association across the entire genome, including both regions containing genes and those between genes. This approach has been made possible by three primary developments. The first, the inception and execution of the International Haplotype Map Project (<http://www.hapmap.org>), a study designed to catalog and correlate genetic variability in the genome in a population specific manner (Frazer, Ballinger *et al.*, 2007). This project allows the use of SNP genotypes to infer genotype information about neighboring SNPs; thus a large proportion of the information content in several million SNP variants can be inferred using a smaller set of several hundred thousand informative tagging SNPs. The second is the development of reliable high-throughput genotyping methodologies, which allow cost-effective typing of millions of SNP loci in parallel. The third is the existence of repositories harboring DNA from well-characterized cases and controls. This is especially relevant in case of heterogeneous disorders such as Parkinson's or Alzheimer's disease, in which many clinical phenotypes have been described. A good example of such a repository is the NINDS-funded Coriell Institute for Medical Research, which banks DNA from subjects with a variety of diseases as well as controls (<http://ccr.coriell.org/>). The intersection of these three developments allows the application of genome wide SNP association studies to understand the genetic basis of complex disease.

Given the technical demands, it was not until 2002 that the first whole-genome association study was published (Ozaki, Ohnishi *et al.*, 2002). In this study, by means of genotyping 92,788 gene-based SNPs, they identified a candidate *locus* on chromosome 6p21 associated with susceptibility to myocardial infarction. Three years later Maraganore and colleagues performed the first whole-genome association study for PD (Maraganore, De Andrade *et al.*, 2005). On tier 1 of their study, they genotyped 198,345 uniformly spaced SNPs in 443 sibling pairs discordant for PD. Those SNPs that presented a P value below 0.1 (1,793) were genotyped in 332 case-control unrelated pairs (tier 2a). This way they found 11 SNPs associated with PD. The lowest combined P value was found on a SNP located on a gene encoding a protein with an important role on neurogenesis and neural apoptosis, *SEMA5A*. A second SNP was located at the *PARK11 locus*. They also performed a tier 2b in which they genotyped SNPs with borderline P values but located in biologically interesting targets. This way, the SNP with lowest P value was found at the *PARK10 locus*.

In genetic research, almost as important as an initial description of an association between a certain trait and a location in the genome related to that trait, is the replication of that particular finding. Although the results obtained by Maraganore's group appeared to be exciting, several groups have failed to confirm these findings in different cohorts (Clarimon, Scholz *et al.*, 2006; Elbaz, Nelson *et al.*, 2006; Farrer, Haugarvoll *et al.*, 2006; Goris, Williams-Gray *et al.*, 2006; Li, Rowland *et al.*, 2006). The results from these studies, along with the fact that *SNCA* (the most common causes of PD) was not identified in Maraganore's experiments (Maraganore, De Andrade *et al.*, 2005), were both disappointing and discouraging. However, since sporadic PD is likely caused by an

interaction of genetic and environmental risk factors, the chances of discovering a causative gene, depends on its relative effect on disease risk (penetrance) and the frequency of the disease-associated allele. Thus, a larger number of both PD cases and control subjects would allow minimizing the bias created by this phenomenon.

An interesting point extracted from Perez-Tur comment in 2006 (Perez-Tur, 2006), is that related to data sharing. Since we are about to witness an explosion in the number of genome-wide association studies aimed at uncovering genetic variability related to common diseases, implementation of an open access policy regarding the availability of data to all researchers, is important. This way, other groups are allowed to interpret data produced in a different way, possibly selecting groups of patients on the basis of certain clinical characteristics and selecting genotypes within specific genes in order to perform independent association studies.

## 2. OBJECTIVES

As can be observed from this introduction, PD is a very heterogeneous disorder in which genetic factors play an important role. In the last 12 years, six distinct genetic *loci* have unequivocally been linked with familial forms of PD, and several others related to both sporadic and familial forms of the disease.

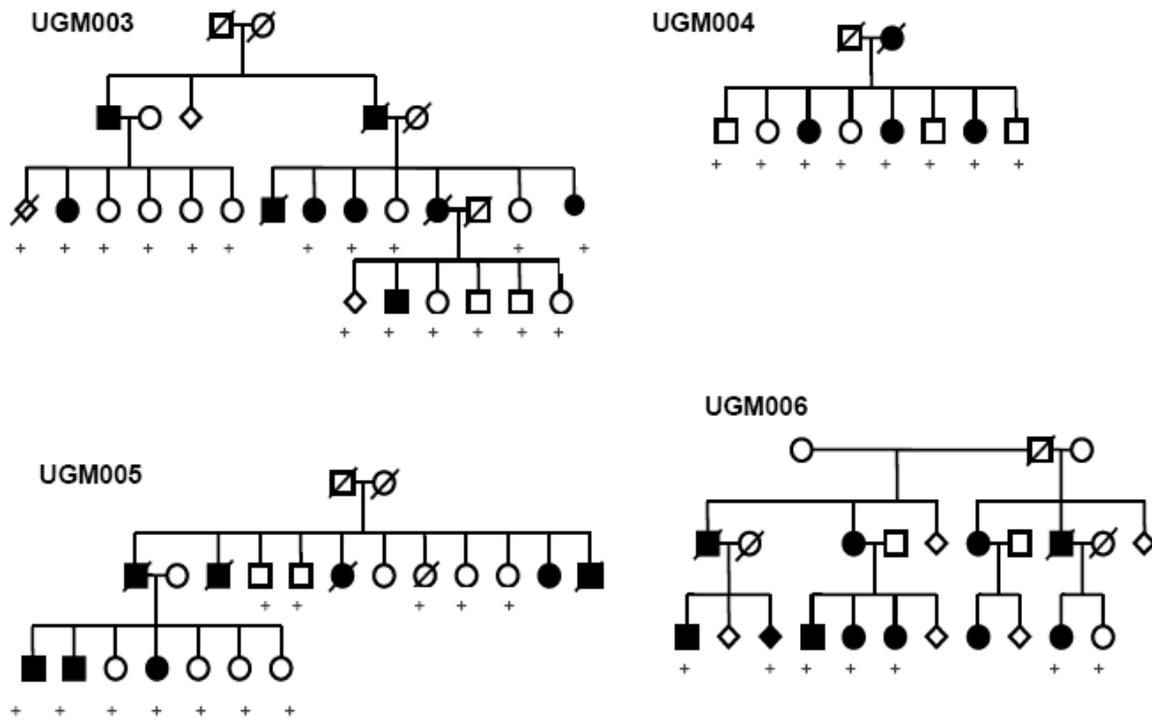
In the current study **we aim to identify the responsible genetic factors representing a risk of PD in both familial and sporadic forms of parkinsonism**. To achieve this purpose different strategies have been utilized including family-based and population approaches in both single genomic locations and genome-wide.

### 3. METHODS

#### 3.1. GENETIC ANALYSIS OF *PARK8*-LINKED PARKINSON'S DISEASE PATIENTS.

##### 3.1.1. Basque late-onset PD families

DNA samples were collected from PD patients with a positive family history of late onset PD. All patients were examined by Drs. Jose Félix Martí-Massó and Dr. Angel Martinez Gil. A total of 4 families were identified, of which 3 presented PD with a mode of inheritance clearly resembling apparent autosomal dominant disease (UGM03, UGM05 and UGM06). Limited genealogical information in the fourth family (UGM04), did not allow assignment of any particular mode of inheritance. DNA samples from a total of 75 individuals (15 affected and 60 unaffected) were collected, all belonging to one of these four families (figure 6). Age at onset in all four families ranged from 50 to 80 with an average of 61. Disease duration is known for only 2 of the four families, UGM03 and UGM05, with mean disease duration of 18 and 10 years respectively. All affected members respond well to L-Dopa (table 2).



**Figure 6: Families UGM003, UGM004, UGM005 and UGM006.** Affected members are shaded, and those whose DNA was available are denoted by a plus sign.

**Table 2:** Summary of clinical features of affected members of autosomal dominant PD from families UGM3, UGM4, UGM5 .and UGM6.

Kindred	Person	Gender	Age at onset	Initial Symtom	Response to L-Dopa	Duration (years)	Disease complications
UGM03	III:1	F	62	Leg tremor, clumsiness	Good response	18	Dyskinesias, falls
	III:10	F	66	Unilateral tremor	Good response	22	Dyskinesias and motor oscillations
	III:12	F	80	Unilateral tremor, severe gait disorder	Good response	9	Dyskinesias, motor oscillations and severe gait disorder
	III:16	F	69	Unilateral hand tremor	Good response	7	Dyskinesias and motor oscillations
	III:18	F	69	Gait disorder clumsiness	Good response	10	Dyskinesias and motor oscillations
	IV:7	F	51	Unilateral tremor	Good response	17	Dyskinesias and motor oscillations
	IV:27	M	51	Unilateral tremor	Good response	5	-
UGM04	II:3	F	59	Unilateral tremor	Good response	5	-
	II:5	F	67	Unilateral leg tremor	Good response	4	-
	II:7	F	63	Unilateral tremor	Good response	5	-
UGM05	II:14	M	57	Clumsiness, unilateral tremor	Good response	2	-
	III:1	M	58	Global clumsiness, unilateral tremor, foot dystonia	Good response	7	-
	III:2	M	58	Unilateral clumsiness, unilateral tremor, foot dystonia	Good response	2	-
	III:4	F	60	Unilateral tremor	Good response	11	Severe gait disorder
UGM06	III:1	M	65	Unilateral leg tremor	Good response	17	Dyskinesias and motor oscillations
	III:2	M	60	Unilateral leg tremor	Good response	6	-
	III:5	M	62	Clumsiness, unilateral tremor	Good response	14	Dyskinesias
	III:6	F	60	Unilateral leg tremor	Good response	12	Dyskinesias, motor oscillations, gait ignition failure
	III:7	F	60	Unilateral leg tremor	Good response	7	-

### 3.1.2. Minimal haplotype shared by UGM03, UGM04, UGM05 and UGM06.

After linkage analysis showing linkage to *PARK8* was performed in these four families (Paisan-Ruiz, Saenz *et al.*, 2005) and a large kindred from the United Kingdom (Khan, Jain *et al.*, 2005), identification of a rare polymorphism within *KIF21A* in three of the four Basque families suggested that they were ancestrally related and might share a minimal interkindred disease haplotype. In order to define this common minimum haplotype, 23 SNPs were selected from the region with strongest linkage to PD in these families (D12S331 to D12S1668). These SNPs were selected from dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) choosing those most likely to be informative. In order to maximize our ability to obtain phase, all available family members were genotyped.

These SNPs were genotyped by means of PCR amplification and direct sequencing (primer sequences shown in table 3). To ensure maximum accuracy on these SNPs genotypes, these assays were performed twice by two different researchers (Coro Paisán-Ruiz and Javier Simón Sánchez). Amplification of each SNP was carried out with Eppendorf's Taq DNA polymerase using a Mastercycler thermal cycler (<http://www.eppendorf.com>) as shown below:

Taq DNA Polymerase:..... 0.15 µl  
 Primer F (20 µM):..... 1.5 µl  
 Primer R (20 µM): .....1.5 µl  
 dNTP mix (10µM).....1.5 µl  
 10x PCR Buffer: .....5 µl  
 H<sub>2</sub>O: .....40.85 µl  
 DNA (20 ng/µl): ..... 1 µl

**Table 3:** Forward and reverse primer sequences used for genotyping 23 SNPs in the region flanked by D12S331 and D12S1668.

SNP ID	Forward primer (5'-3')	Reverse primer (5'-3')
rs1523118	TAAGTATCTGTATGTGAAAGC	TGTGTACAGGATTCTGCTTG
rs11169992	AAAATATCTAGTTCTTTTCC	CATTATAACTATGCCTTGCA
rs10876410	CTTAGATATGAGCAGGAACAAC	TCACCTCCAGTCAACTATCAG
rs10876646	GACAGAGATTCAAACCCAAG	GAAATGCTTATCATAGGATT
rs10876876	CATTGTGAAAATTAAGTGAG	TTGCATGCTGAGAAATTTCT
rs11171789	TATACTAGGACTGCTGATTT	AACCACCTAGAAAGAACTCT
rs10876886	GGAAATAAGCACTAGTCAAGC	TGTTGAGAGTACTGTTCCCTG
rs11172282	TGACAATAGATTAGAATGCAG	TTTACTCATACTAAGGCTG
rs11172541	ATATTGGGTCACAAGTACTG	TTCAAGACAACATGCTAAAT
rs10877201	GCATATTTGAGTAACTGAGGC	TGCACTCGAGAATTATAGCC
rs4548690	GTCTAATTGCTCTTGCCTTG	GTGAACTTGAAAGACGGATATG
rs7294916	CACAAGAAGGCTTTAGCTG	GGGATAGCAGACGTAATTG
rs4423249	ACATTACTCTGGTGAGAGCC	GTAATTTGTCATGGCAGCA
rs515205	AGTTTAAAGGGATGGACTAG	CAAAGGGAGGCTGAATTTAT
rs515291	CATGCATTATGATACTATTACC	GTAAGCCTAGCCTAGGTGAC
rs937110	GAAGCCTCTTGAGTCTATC	GTCTTTGCTTCCTTAATATTG
rs7399259	GTAGAGACGGGGTTTCACCA	AATCATGGCAGAAGGCAAAG
rs4768224	GATACCAGCATTGTGAACATAG	CATTTTCCTCAACCATGATG
rs1820544	CAAGAAATGGCTGTAATCAAG	GAAGAGAAAAACAACAGCATG
rs10784800	CCAATCATCTCCCACCAG	GAAGTGTAAAGAGGCTATGTG
rs10879192	GCTTAACCCATGTCATTGTC	ATTTATTAGCTCTGTGCACAAG
rs10747736	TGAGGGTAACTTGTTCACAGC	ATGCTTTGGAGGGTCATTTG
rs12423567	CCTCAGTTTGGTCTGTTGAAAC	GTGTTTAAAGGTGGCTGTGGTC
rs10784616	GTATTGCCTCTCCTGTTTATGG	GGCAAGAATGTTCTGTTGGTAC

Thermocycling programs used were 60TD50 for rs10876410, rs11169992, rs10876886, rs10877201, rs4548690, rs11171789, rs7294916, rs4426249, rs9371100, rs7399259, rs10876876, rs1820544, rs10879192 and rs12423567 (table 4); 57TD52 for rs1523118, rs10876646, rs11172282, rs11172541, rs515291 and rs4768224 (table 5); 62x35 for rs10784800 and rs10784616 (table 6); and 55x35 for rs10747736 (table 7).

**Table 4:** Thermocycling conditions corresponding to program 60TD50.

	Stage 1	Stage 2			Stage 3			Stage 4			Stage 5	
	1 Cycle	8 Cycles			20 Cycles			12 Cycles			1 Cycle	
Temperature	94°C	94°C	60°C	72°C	94°C	60°C	72°C	94°C	50°C	72°C	72°C	°C
Time	5 min	30 seg	30 seg	45 seg	30 seg	30 seg	45 seg	30 seg	30 seg	45 seg	5 min	HOLD
Temperature increase	0	0	0	0	0	- 0.5°C	0	0	0	0	0	0

**Table 5:** Thermocycling conditions corresponding to program 57TD52.

	Stage 1	Stage 2			Stage 3			Stage 4			Stage 5	
	1 Cycle	15 Cycles			10 Cycles			14 Cycles			1 Cycle	
Temperature	94°C	94°C	57°C	72°C	94°C	57°C	72°C	94°C	52°C	72°C	72°C	°C
Time	5 min	30 seg	30 seg	45 seg	30 seg	30 seg	45 seg	30 seg	30 seg	45 seg	5 min	HOLD
Temperature increase	0	0	0	0	0	- 0.5°C	0	0	0	0	0	0

**Table 6:** Thermocycling conditions corresponding to program 62x35.

	Stage 1	Stage 2			Stage 3	
	1 Cycle	35 Cycles			1 Cycle	
Temperature	94°C	94°C	62°C	72°C	72°C	°C
Time	5 min	30 seg	30 seg	45 seg	5 min	HOLD
Temperature increase	0	0	0	0	0	0

**Table 7:** Thermocycling conditions corresponding to program 55x35.

	Stage 1	Stage 2			Stage 3	
	1 Cycle	35 Cycles			1 Cycle	
Temperature	94°C	94°C	55°C	72°C	72°C	°C
Time	5 min	30 seg	30 seg	45 seg	5 min	HOLD
Temperature increase	0	0	0	0	0	0

PCR products were verified in a 2% agarose-TBE gel with ethidium-bromide, and visualized with ultraviolet transillumination. Before loading each PCR product on a gel, 5 µl were mixed with 6X Orange G Loading Dye Solution (10 mM Tris-HCl (pH 7.6), 0.15% orange G, 60% glycerol, 60mM EDTA).

After verification, PCR products were purified with MultiScreen 96-Well Filter Plates (Millipore, <http://www.millipore.com>) following the protocol below:

- . Add 80 µl of di H<sub>2</sub>O to 15 µl of PCR product and transfer to a MultiScreen 96-Well Filter Plate.
- . Apply vacuum at 20 inches Hg until wells are dry.
- . Add 20 µl of distilled, deionized H<sub>2</sub>O (Molecular Grade Water, Mediatech, <http://www.cellgro.com>) to each well containing sample
- . Vortex at 1000 rpm for 10-15 minutes.
- . Collect samples.

Each purified product was sequenced using thermocycling conditions shown in table 8, using the same forward or reverse primers as in the PCR reaction and with Applied Biosystems BigDye terminator v3.1 sequencing chemistry as below:

BigDye v3.1:..... 0.5 µl  
 5x Sequencing buffer:.....2 µl  
 Primer F or R (5µM) .....0.5 µl  
 H<sub>2</sub>O: .....5 µl  
 Purified PCR product: .....1 µl

**Table 8:** Thermocycling conditions used for sequencing with Applied Biosystems BigDye terminator v3.1.

	Stage 1			Stage 2
	25 Cycles			1 Cycle
Temperature	96°C	50°C	60°C	4°C
Time	30 seg	15 seg	4 min	HOLD
Temperature increase	0	0	0	0

Sequencing products were then purified with MultiScreen 96-Well Filter Plates (Millipore, <http://www.millipore.com>) following the protocol below:

- Add 20 µl of Molecular Grade Water to 10 µl of sequencing product and transfer to a MultiScreen 96-Well Filter Plate.
- Apply vacuum at 20 inches Hg until wells are dry.

- . Add 20 µl of di H<sub>2</sub>O to each well containing sample
- . Apply vacuum at 20 inches Hg until wells are dry.
- . Add 20 µl of Molecular Grade Water to each well containing sample
- . Vortex at 1000 rpm for 10-15 minutes.
- . Collect samples.

Purified sequencing reactions were then electrophoresed on an ABIPrism 3100 Genetic Analyzer (Applied Biosystems, <http://www.appliedbiosystems.com>) and analyzed with Staden sequence analysis package (<http://staden.sourceforge.net>).

### **3.1.3. Analysis of p.R1441G mutation in a Basque PD cohort.**

Given that p.R1441G was not only found in those families presented in section 3.1.1, but also in apparently sporadic cases with Basque ancestry (Paisan-Ruiz, Jain *et al.*, 2004), we proceeded to determine its frequency in a larger population from this same region in Spain. Thus, blood from 238 individuals was drawn from the clinical practice of three public hospitals of San Sebastián. Patients were diagnosed according to the Gelb criteria for PD (Gelb, Oliver *et al.*, 1999). All study participants provided their informed consent, and the study was approved by the Ethics Committee of the Hospital Donostia (San Sebastián). Study participants were classified according to either the presence or absence of family history of PD (61 familial patients; 177 sporadic patients). Family history was considered positive when at least one first-degree relative of the patient was reported to suffer from PD. The information about familial history was obtained from the patients or their proxies; thus, it is possible that there may be under-reporting of positive family

history. Unfortunately, it was not possible to analyze the family members of the probands; therefore, no estimates of penetrance are provided.

Since 30% of Guipuzcoa inhabitants (730,000 total inhabitants) do not have Basque origin, the ethnicity of patients was assessed during the personal interview. Spanish handling of surnames makes it reasonably easy to know immediate ancestry of a patient by simply examining his or her surnames. Upon this basis, we classified patients as being of Basque descent when at least one of the patient’s surnames was of Basque origin and as non-Basques when none of their surnames could be considered as ethnically Basque (table 9).

**Table 9:** Clinical and demographic characteristics of the analyzed population.

	Whole sample	Basque origin	Non Basque	Positive familial history		Sporadic	
				Basque	Non Basque	Basque	Non Basque
N	238			41	20	117	60
Mean age at onset	64.6±10.2	65.4±10.2	63.0±10.3	65.1±10.9	62.3±9.4	65.4±10.0	63.3±10.6
Sex ratio (% men)	57.7%	0.544	0.588	0.634	0.75	0.513	0.533
Years of evolution	8.1±6.4	8.0±6.5	8.4±6.0	6.6±4.9	9.1±7.7	8.4±7.0	8.1±5.4
Main presenting symptom	UT	UT	UT	UT	UT	UT	UT
	60.0%	53.6%	66.7%	52.6%	72.7%	63.1%	62.7%

UT: Unilateral tremor.

The presence of the p.R1441G mutation was determined in all cases by the use of allele-specific PCR (ARMS) (Newton, Graham *et al.*, 1989). This technique allows amplification and discrimination of specific alleles or DNA sequence variants at the same *locus*. Specificity is achieved by designing PCR primers so that they partially overlap the site of sequence difference between the amplified alleles. For each change, we designed a common primer, distal to the tested variant, and two allele-specific primers whose 3’ end was specific of the alleles of that *locus*. For a given variant, each sample is PCR-amplified in two different PCR reactions containing one of the allele specific primers and

the common primer. Absence/presence of the expected band in an agarose-TBE gel indicates absence/presence of that particular allele in that sample's genotype. Thus, if only one of the reactions produces a visible band in an agarose-TBE gel, that sample is considered as homozygous for the allele recognized by the specific-allele primer used in that reaction. Otherwise, if amplification is achieved after both separate reactions, a heterozygous genotype is suspected.

In this case, this technique was carried out using common primer: hiLRRK2-X31-ARMS-F (5'-CCTTGTGATTGAATCACCAC-3') and an allele-specific primer: heLRRK2-X31-ARMS-R-C (5'-GTGTCTTTCCTCAGGCTC-3'), that detects the wild-type allele, or heLRRK2-X31-ARMS-R-G (5'-GTGTCTTTCCTCAGGCTG-3') to detect the mutant allele. PCR amplification was carried out with Qiagen's Taq polymerase chemistry ([www1.qiagen.com](http://www1.qiagen.com)), using the following protocol:

Taq DNA Polymerase:.....	0.08 µl
hiLRRK2-X31-ARMS-F (20 µM): .....	0.6 µl
heLRRK2-X31-ARMS-R (20 µM): .....	0.6 µl
dNTP mix (10 mM): .....	0.6 µl
10x PCR Buffer:.....	1.5 µl
H <sub>2</sub> O: .....	11.92 µl
DNA (20 ng/µl): .....	0.3 µl

The thermal cycling program used was a 60x27 program, which consists of the steps displayed in table 9.

**Table 9:** Thermocycling conditions corresponding to program 62x27.

	Stage 1	Stage 2			Stage 3	
	1 Cycle	27 Cycles			1 Cycle	
Temperature	94°C	94°C	60°C	72°C	72°C	°C
Time	5 min	30 seg	30 seg	45 seg	5 min	HOLD
Temperature increase	0	0	0	0	0	0

After resolving the PCR product on a 2% agarose-TBE gel, a band of 571 bp indicates the presence of the allele corresponding to the allele-specific primer used in that reaction. Every allele-specific PCR was performed along with both a negative and a positive control whose genotype had previously been verified by means of direct sequencing. Positive results were also verified by direct sequencing.

**3.1.4. Haplotype analysis of Basque p.R1441G carriers.**

To explore the possibility that all mutation carriers inherited their disease allele from a common ancestor, the 11 SNPs that defined a common haplotype in our previous study in the four Basque families (Paisan-Ruiz, Jain *et al.*, 2004), were directly sequenced after PCR amplification in all carriers of p.R1441G (see section 3.1.2 for conditions).

In order to determine the haplotype frequency in general population, 80 healthy individuals from the Basque country and 82 non-Basque Spanish healthy individuals, were typed at each SNP. Haplotypes were determined with fast-PHASE (Scheet and Stephens, 2006) and HapAnalyzer (<http://hap.ngri.re.kr>). The standardized coefficients of disequilibrium  $r^2$  and  $D'$  between all combinations of SNP pairs was calculated using the

EMLD software, developed in the University of Texas by Drs. Huang and Anderson (<http://cge.mdanderson.org/qhuang/software/pub.htm>).

These Basque controls comprised 46.25% males with mean age at sampling of 68.66 years (range, 21-87 years) and 52.75% females with a mean age at sampling of 75 years (range, 44-89 years). The non-Basque Spanish controls comprised 63,64% males with mean age at sampling of 37 years (range, 23-67 years) and 36,36% females with mean age at sampling of 42 years (range, 24-67 years).

In order to determine the copy number of a chromosomal region of individual 01-407 between SNPs rs937110 and rs10747736, a semiquantitative PCR analysis was performed. For this purpose each SNP was amplified along with an endogenous control located on the first intron of the *MAPK8IP1* on chromosome 11p11.2, using fluorescent-labeled oligonucleotides MAPK8IP1-I1-F (5'-GCAGTGCCTCTTTAATAGGC-3') and MAPK8IP1-I1-R (5'- GTGCAATATCATCCTGGGGA-3'). Fluorescent labels were HEX for rs10747736 in its forward primer, TAMRA for rs937110 in its forward primer, and FAM for the endogenous control in its reverse primer (primer sequences for rs10747736 and rs937110 shown in table 3)

PCR products were electrophoresed on an ABIPrism 3100 Genetic Analyzer (Applied Biosystems) with an internal standard molecular weight (GeneScan 350-ROX; Applied Biosystems). Data was analyzed using GeneScan v3.6 (Applied Biosystems) to produce electrophoregrams showing the sizes of the peaks in base pairs and areas under the peaks representing the amount of PCR product present. To determine gene dosage for each amplicon the samples were first compared to the endogenous control on chromosome 11p11.2 to obtain a ratio. The resulting ration was then compared to the ratio obtained

from a control subject run in parallel. Thus, a ratio of 0.6 or less is indicative of heterozygous deletion, a ratio ranging from 0.8 to 1.2 is indicative of normal and ratios between 1.2 and 1.8 as heterozygous duplication.

The number of cycles of these PCR reactions was calculated previously so that when the reaction finishes, it is still on the log-linear range of the PCR (see section 3.7.7.2). This way, the amount of amplified product is directly proportional to the genomic DNA used as a template in the reaction. For this purpose, a control DNA was PCR-amplified in 11 different reactions for a certain number of cycles ranging from 20 to 40, with the thermocycling program shown in table 10. After electrophoresing the amplified products on a 10% polyacrylamide gel and resolving with silver staining (protocol below), PCR-reaction kinetics were calculated with Kodak digital Science 1D v.3.0.1 software (figure 7).

The protocol used for silver staining samples electrophoresed on a 10% polyacrilamide gel is:

- Cover the gel with fixer solution containing 7.5% of acetic acid, glacial in diH<sub>2</sub>O. This solution immobilizes the DNA molecules in the acrylamide gel matrix to avoid diffusion and subsequent image blurring.
- Rock the staining tray continuously on a platform rocker for 5-10 minutes.
- Following fixation, carefully decant the solution, taking care not to damage the gel or touch the gel surface.
- To wash the gel, cover the gel with diH<sub>2</sub>O

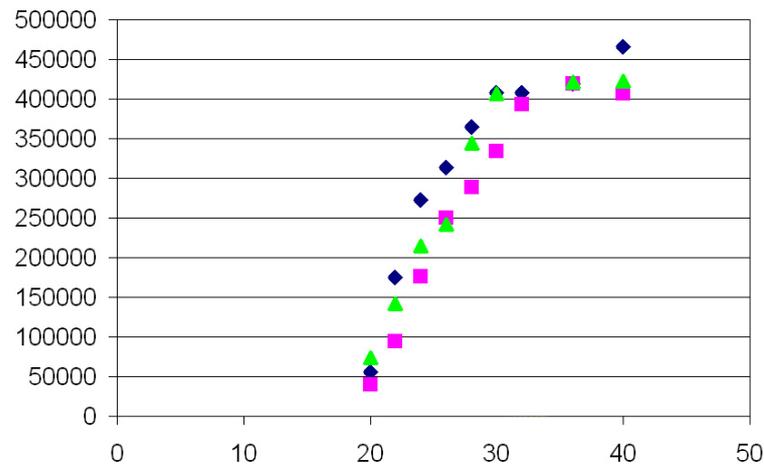
- Rock the staining tray continuously on a platform rocker for 2 min. This step removes acid and other trace substances that may interfere with staining, and provides a clear background to the final stain.
- Repeat the wash step twice.
- Cover the gel with fresh formaldehyde solution containing 15% formaldehyde in diH<sub>2</sub>O and gently rock for 5-10 minutes. This treatment is important for stain sensitivity and maximum image contrast.
- Carefully decant solution, taking care not to damage the gel or touch the gel surface.
- Cover the gel with fresh silver solution containing 0.1% AgNO<sub>3</sub> in diH<sub>2</sub>O and gently rock for 10-20. A careful examination of silver impregnation times showed that optimal staining was achieved after 20 minutes. However, as little as 10 minutes is sufficient for high-quality staining without significant loss of sensitivity. Impregnation times can be increased up to 60 minutes, but greater than 90 minutes can cause severe image loss.
- Following silver impregnation, carefully decant the solution, taking care not to damage the gel or touch the gel surface.
- Briefly rinse residual silver solution from the surface of the gel by covering it with diH<sub>2</sub>O for 5–10 s. Do not rinse the gel longer than 15 s, as this step removes silver from the gel.
- Cover gel with cold (4-10°C) developer solution containing 3% Na<sub>2</sub>CO<sub>3</sub>. Agitate the staining tray throughout image development so the developer solution is not stagnant. Image development begins as soon as the developer solution is added. Image development typically takes about 3 min depending on gel thickness, the reagents

used and the temperature of the reagents. Alternatively, you can add 600 ml of formaldehyde solution per 100 ml of final developer solution to improve image contrast.

- Decant the developer solution carefully, avoiding damage to the gel or touching the gel surface.
- Cover gel with cold (4°C) developer stop solution containing 7.5% of acetic acid, glacial in diH<sub>2</sub>O and incubate for 5-10 minutes.
- Decant the developer stop solution and rinse the gel with diH<sub>2</sub>O.
- Photograph the gel and dry overnight.

**Table 10:** Thermocycling conditions used to amplify chromosomal region between SNPs rs937110 and rs10747736.

	Stage 1	Stage 2		Stage 3	
	1 Cycle	20-40 Cycles		1 Cycle	
Temperature	95°C	94°C	60°C	72°C	4 °C
Time	15 min	30 seg	90 seg	60 seg	HOLD
Temperature increase	0	0	0	0	0



**Figure 7:** Band intensity (proportional to DNA concentration) *versus* number of PCR cycles of amplification of rs10747736 in three different control samples represented herein in three different colors.

### 3.1.5. Haplotype analysis of p.R1441G carriers from Asturias

Five p.R1441G carriers (PE008, PE155, PE040, PE066 and PE139) derived from a study performed by Mata and collaborators (Mata, Ross *et al.*, 2006) in a population of 225 patients from Asturias were typed for SNPs rs4768224, rs12423567, rs1820544 and rs10784616 (minimal common haplotype of p.R1441G carriers without taking into account individual 01-407. See section 4.1.1)

These 5 samples represent 2.7% of the original 225 samples selected by Mata and coworkers from three hospitals (Hospital Universitario Central de Asturias, Oviedo; Hospital Alvarez Buylla, Mieres; and Hospital de Cabueñes, Gijón) within Asturias. These 5 patients (2 females and 3 males) presented typical late-onset PD with at least two of the three cardinal signs tremor, rigidity and bradykinesia, and with a positive response to levodopa therapy when administered. Only one patient noted a family history of PD at

the time of diagnosis. The mean age at onset was 63.2 years (range, 46-74) and at the time of diagnosis they did not present any atypical disease signs. All patients gave their consent to participate in this study, which was approved by the Ethical Committee of all three hospitals.

### 3.2. *IN SITU* HYBRIDIZATION OF *LRRK2* mRNA IN ADULT MOUSE BRAIN

This work is based on five B2B6 adult male mice. All animal manipulations were approved by the Animal Care Ethics Committee of the C.S.I.C. and the School of Medicine of the University of Valencia. This technique can mainly be divided on two steps: Probe synthesis and *in situ* hybridization.

#### 3.2.1. Probe synthesis

*LRRK2* cDNA fragments from codon 1004 to codon 1172 (probe 1) and 2013 to 2171 (probe 2) were PCR-amplified from a mouse cDNA library (Invitrogen). Primers used for this stage along with size of the corresponding probe is shown in table 11.

**Table 11:** Primers and restriction sites used for cloning the different probes designed for the *in situ* hybridization.

Probe name	Forward primer (5'-3')	Reverse primer (5'-3')	5' Restriction site	3' Restriction site	Probe size (bp)
LRRK2-PROBE1	ATCATCTAGAGTTGCCTCAGTAGCCACC	ATCAGTCGACAAGGTAAGGCAGGCATTG	<i>Xba</i> I	<i>Sal</i> II	531
LRRK2-PROBE2	TACGTCTAGAGCGAAGATTGCGGACTAC	CACAGATATCCCAACCAGAGAGTCGCAC	<i>Xba</i> I	<i>Eco</i> RV	554

Non-complementary tails of the primers are shown in grey including the corresponding restriction site for enzyme.

The first 18 nucleotide of the primers used (3'-5') were sequence specific, the next 6 contain the restriction site in further steps, and the last 4 were randomly assigned and provided a physical site for the restriction enzyme (table 11). *Pfu turbo* polymerase from Stratagene (<http://www.stratagene.com>) was used for these amplifications following the protocol below:

dNTPs 2.5m mM:.....	4µl
Primer F (20ng/ul):.....	2µl
Primer R (20ng/ul):.....	2µl
Buffer 10X:.....	5µl
<i>Pfu turbo</i> polymerase:.....	1µl
cDNA:.....	1µl
H <sub>2</sub> O:.....	35µl

The PCR conditions used were 2 minutes at 95°C followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 58°C and 1 minute at 72°C. After these 35 cycles, a 10 minutes extension at 72°C was performed.

The amplified PCR products (probes 1 and 2) were resolved on a 1% agarose-TBE gel. The band corresponding to the expected fragment size was isolated and purified with Minelute gel extraction kit (Qiagen, <http://www.qiaquen.com>).

Minelute-purified products were next digested with *XbaI* and *Sall* (probe 1); or *XbaI* and *EcoRV* (probe 2) for 3 hours at 37°C as displayed in table 12. All restriction enzymes

used were from Fermentas (<http://www.fermentas.com>). The digested product was purified with minelute kit afterwards.

Meanwhile, 8 µg of pBluescript KSII<sup>-</sup> vector (Stratagene, <http://www.stratagene.com>) were also digested first with *XbaI* and then with either *SalI* or *EcoRV* for 3 hours at 37°C as displayed in table 13.

**Table 12:** Reaction mixture for the digestion of the Minilute-purified PCR products prior to ligation.

<b>Component</b>	<b>Probe 1</b>	<b>Probe 2</b>
<b>PCR Product</b>	9 µl	9 µl
<b>Y<sup>+</sup>Tango 2X</b>	4 µl	4 µl
<b><i>XbaI</i></b>	0.7 µl	0.7 µl
<b><i>Sal/EcoRV</i></b>	0.7 µl	0.7 µl
<b>H<sub>2</sub>O</b>	5.6 µl	5.6 µl

**Table 13:** Reaction mixtures for the digestion of pBluescript KSII<sup>-</sup> prior to ligation.

<b>First digestion</b>		<b>Second digestion</b>	
<b>Component</b>	<b>Amount</b>	<b>Component</b>	<b>Amount</b>
Y <sup>+</sup> Tango 2X	5 µl	O+/R	6 µl
XbaI	1.5 µl	SalI/EcoRV	1.5 µl
H <sub>2</sub> O	35.5 µl	H <sup>2</sup> O	4.5 µl
pBluescript KSII <sup>-</sup>	8 µl	Digested pBluescript KSII <sup>-</sup>	48 µl

QIAquick PCR Purification kit (Quiagen, <http://www.qiagen.com>) purification was performed after each digestion.

The entire digested vector was dephosphorylated for 1 hour at 37°C with Roche's Alkaline phosphatase (<http://www.roche.com>). For this purpose, 5 µl of alkaline phosphatase with to 6 µl of 10X buffer was utilized for each reaction. Purification with

QIAquick PCR Purification kit (Quiagen, <http://www.qiagen.com>) was performed afterwards.

After these digestions both PCR products (inserts) and pBluescript KSII<sup>-</sup> plasmid (vector), were ready for ligation (Roche's Rapid ligation kit). In order to estimate how much insert and vector we had to use for this step we ran 1 µl of each sample through a 0.8% agarose-TBS gel, along with REAL SCALA 1 DNA ladder (Durviz, <http://www.durviz.com>). This DNA ladder allows an estimation of the DNA concentration we ran in the gel. Thus, according to the formula [ $Insert\ ng = (3 * Vector\ ng * Insert\ bp) / Vector$ ] we concluded that the amount to be used of each product for the ligation reaction should be that shown in the table 14:

**Table 14:** Reaction mixtures for ligation of pBluescript KSII<sup>-</sup> with the corresponding digested PCR product.

<b>Component</b>	<b>Probe 1</b>	<b>Probe 2</b>
Digested pBluescript KSII <sup>-</sup>	6 µl	6 µl
Digested PCR Product	1.2 µl	1.1 µl
Buffer 2	2 µl	2 µl
Buffer 1	20 ul	20 ul
Ligase	1 ul	1 ul

After incubating for 10 minutes at room temperature, Roche's High Pure PCR Product Purification Kit (<http://www.roche.com>), was used to purify each reaction mix.

The samples were then transfected into DH5-α electrocompetent bacteria. For this purpose 40 µl of electrocompetent bacteria, 10 µl of ligation product, and 950 µl of SOC medium were exposed to 1700 V in a Eppendorf Electroporator 2510 (Eppendorf,

<http://www.ependorf.com>). After electroporation, bacteria were allowed to recover for 1 hour at 37°C.

SOC is a nutrient-rich medium usually utilized at the end of a transformation process. It is mainly composed of 1) tryptone and yeast extract which serve as rich nutrients to allow good growth of bacteria after transformation (Pronadisa, <http://www.condalab.com>); 2) sodium and potassium chloride to achieve optimal osmotic conditions; and 3) a magnesium source, needed in many enzyme reactions, such as those involved on DNA replication. The exact composition of every 100ml of this medium is as follows:

Tryptone:.....	2.0g
Yeast extract: .....	0.5g
NaCl (1M): .....	1ml
KCl (1M):.....	0.25ml
Filter-sterilized Mg <sup>2+</sup> stock (CL <sub>2</sub> Mg 1M + Mg SO <sub>4</sub> 1M): .....	1ml
Filter-sterilized glucose 2M: .....	1ml
H <sub>2</sub> O: .....	Up to 100ml

The completed medium had to be filtered through a 0.2 µm filter unit (Millipore, <http://www.mlilipore.com>). The pH was 7.0.

In order to check the integrity of our construct, 100 µl of the transfected cells were grown overnight at 37°C on an agar plate with ampicilin at a final concentration of 100 mg/ml. Two single colonies from each construct were grown in 200 µl of LB-ampicilin (10 g/l tryptone, 5g/l yeast extract, 5g/l NaCl and 100 µg/ml ampicilin) overnight at 37°C. After

extracting plasmid DNA from the bacterial cultures using QIAprep Spin Miniprep Kit (Qiagen, [www1.qiagen.com](http://www1.qiagen.com)), 5  $\mu$ l of the final product were sequenced with BigDye v3.1 chemistry as shown below. (Thermocycling conditions shown in table 8).

BigDye v3.1:..... 0.5  $\mu$ l  
 5x Sequencing buffer:.....2  $\mu$ l  
 T3/T7 Promoter primer (5ng/ $\mu$ l).....0.5  $\mu$ l  
 H<sub>2</sub>O: .....2  $\mu$ l  
 Construct: .....5  $\mu$ l

To further confirm that our inserts had successfully been cloned into pBluescript KSII<sup>+</sup>, each construct was digested with a different restriction enzyme. The digested product was then run in a 1% agarose-TBE gel, to see whether we could visualize bands with the expected size (table 15).

**Table 15:** Reaction mixture to confirm probe 1 and 2 construct integrity.

<b>Component</b>	<b>Probe 1 construct</b>	<b>Probe 2 construct</b>
B <sup>+</sup> 2X/H 2X	6 $\mu$ l	6 $\mu$ l
<i>DraI/DdeI</i>	1.2 $\mu$ l	1.1 $\mu$ l
H <sub>2</sub> O	2 $\mu$ l	2 $\mu$ l
Construct	20 $\mu$ l	20 $\mu$ l

The bands expected after digestion were of 19, 692, 1351 and 1405 bp (probe 1 construct), and 166, 409, 540, 821 and 1499 bp (probe 2 construct).

Both sequencing and restriction digestion, confirmed the integrity of the two constructs.

In order to linearize our constructs so that further reactions could be performed, 20 µg of each construct was digested with each of the enzymes we used prior to the ligation step. Thus, the probe 1 construct was digested with *XbaI* (for antisense probe synthesis) and *Sall* (for sense probe synthesis), and the probe 2 construct with *XbaI* (antisense probe synthesis) and *EcoRV* (sense probe synthesis) as follows (table 16).

**Table 16:** Reaction mixture used to linearize probe 1 and probe 2 constructs

<b>Probe 1 construct</b>		<b>Probe 2construct</b>	
<b>Component</b>	<b>Amount</b>	<b>Component</b>	<b>Amount</b>
Y <sup>+</sup> Tango 2X/ O <sup>+</sup> 2X	8 µl	Y <sup>+</sup> Tango 2X/ R <sup>+</sup> 2X	8 µl
<i>XbaI/Sall</i>	2 µl	<i>XbaI/EcoRV</i>	2 µl
H <sub>2</sub> O	60 µl	H <sub>2</sub> O	60 µl
Constuct (2000ng/µl)	10 µl	Constuct (2000ng/µl)	10 µl

After checking the results of these digestions in a 2% agarose-TBE gel, samples were purified by phenol purification following the protocol below:

- In order to remove proteins, add 1 volume of phenol: chloroform: iso-amyl alcohol (25: 24: 1).
- Vortex.
- Spin 5 minutes at 13000 rpm at 4°C.
- Transfer supernatant to a new tube avoiding aspiration of the interlayer or organic phase.
- In order to remove phenol, add 1 volume of chloroform: iso-amyl alcohol (24: 1).
- Vortex.
- Spin 5 minutes at 13000 rpm at 4°C.

- Transfer supernatant to a new tube avoiding aspiration of the interlayer or organic phase.
- In order to precipitate DNA, add 2.5 volumes of ethanol 100% at -20°C and 0.1 volumes of sodium acetate 3M.
- Vortex.
- Incubate 30 minutes at -80°C.
- Spin 30 minutes at 13000 rpm at 4°C.
- Carefully pour out supernatant without losing DNA pellet.
- Add 1 ml of ethanol 70% (do not vortex).
- Spin 15 minutes at 13000 rpm at 4°C.
- Carefully pour out supernatant without losing DNA pellet
- Air dry pellet at room temperature being careful of not overdrying it.
- Resuspend in 12  $\mu$ l of DEPC-treated H<sub>2</sub>O.

Once samples were purified, probes were synthesized by means of *in vitro* transcription with digoxigenin-labeled nucleotides using the PCR DIG Probe Synthesis Kit (Roche, <http://www.roche.com>) as shown in table 17.

**Table 16:** Reaction mixture used for *in vitro* transcription reactions.

Component	Probe 1		Probe 2	
	Antisense	Sense	Antisense	Sense
Construct concentration ng/ml	624	306	1806	930
Construct volume used	2.4 $\mu$ l	4.9 $\mu$ l	0.83 $\mu$ l	1.61 $\mu$ l
10X digoxigenin RNA labeling mix	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l
10X transcription buffer	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l
RNA polymerase	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l
RNAse inhibitor	0.5 $\mu$ l	0.5 $\mu$ l	0.5 $\mu$ l	0.5 $\mu$ l

Incubation for 2 hours at 37°C is required. The obtained product was now precipitated following the protocol below:

- Precipitate RNA with 2 µl of sodium acetate 3M and 40 µl of ethanol 100% at -20°C.
- Vortex.
- Incubate 30 minutes at -80°C.
- Centrifuge for 30 minutes at 1,3000 rpm and 4°C.
- Remove supernatant
- Air dry pellet at 4°C to avoid temperature rising
- Resuspend pellet on 12.5 µl of DEPC-treated water

1 µl of the purified RNA was then resolved on a 1% agarose-TBE gel and the rest was stored at -80°C with 50% formamide to avoid the formation of secondary structures.

### **3.2.2. *In situ* hybridization**

In order to avoid the action of ribonucleases, all material was washed with ethanol and heated for at least two hours at 200°C, prior to *in situ* hybridization. In addition, all solutions were treated with 1:10<sup>4</sup> diethyl-pyrocabonate (DEPC, Sigma-Aldrich, <http://www.sigmaaldrich.com>) and autoclaved. All steps are performed at 4°C if not otherwise specified.

Subcervical dislocation was performed on five B2B6 adult male mice. After decapitation, brains were quickly removed, washed on 0.1M PBS (0.137M NaCl, 2.68mM KCl, 6.48

mM Na<sub>2</sub>HPO<sub>4</sub>, 647mM KH<sub>2</sub>PO<sub>4</sub>) and immersed overnight in a fixative solution containing 4% paraformaldehyde in 0.1M PBS at 4°C. Brains were then washed for 30 minutes with 0.1M PBS and immersed in gelatine glutaraldehyde for 30-60 minutes, 50 µm thick coronal sections were prepared using a vibratome VT1000M (Leica, <http://www.histo-solutions.com>). Coronal sections were refixed in PBS-DEPC overnight. In order to remove the residual paraformaldehyde, sections were washed twice with PBST (PBS with 0.1% of Tween 20) before gradual dehydration and rehydration with 25%, 50%, 75%, 100% methanol in PBST, and vice-versa. All washes were 5 minutes long. Next, sections were washed with PBST and bleached with 1% H<sub>2</sub>O<sub>2</sub> in PBST for 10 minutes. Afterwards, sections were washed again with PBST, treated with 5µg/ml of proteinase K for 7 minutes and post-fixed with 4% paraformaldehyde-PBS for 20 minutes. Sections were next washed for at least 2 hours with a hybridization solution containing 50% dionized formamide, 5XSSC, 2% Blocking powder (Roche, <http://www.roche.com>), 0.1% Triton X-100, 1mg/ml yeast tRNA, 5mM EDTA and 50µg/ml heparin at 57°C. After removing this solution, fresh one was added containing the digoxigenin-labeled RNA probes at a final concentration of 1 µg/ml. The probes were left to hybridize overnight at 57°C. The next day, five high stringency washes were carried out for 30 minutes with CHAPS 0.1% and SSC 2X in DEPC-treated H<sub>2</sub>O at 57°C. After blocking the sections with 20% fetal calf serum in KTBT (50mM Tris-HCl pH7.5, 150 mM NaCl, 10mM KCl, 0.3% Triton X-100) for 3 hours, sections were incubated with the anti-DIG antibody (Roche, <http://www.roche.com>) overnight. We pre-absorbed the anti-DIG antibody with brain powder (see next paragraph). The next day 5 one-hour washes were performed with KTBT at room temperatures. An alkaline phosphatase-

mediated color reaction was subsequently carried out using 333 mg/ml of 4-nitroblue tetrazolium chloride (Boehringer Mannheim) and 165 mg/ml 5-bromo-4-chloro-3-indolyl phosphate (Boehringer Mannheim) in NTMT buffer (100mM Tris-HCl pH 9.5, 50mM MgCl<sub>2</sub>, 100mM NaCl, 0.1% Tween 20). When the desired level of signal was observed, the reaction was stopped by gently washing with KTBT for two hours with a final wash in PBS prior to mounting the sections.

Brain powder previously mentioned was obtained by homogenizing several frozen mouse brains in a minimal volume of PBS. Following centrifugation at 3000g for 30 minutes, the pellet was washed with ice-cold acetone, dried, and crushed to a fine powder

### **3.2.3. Relative regional expression**

To determine the level of the *LRRK2* expression in different areas of the brain we performed several experiments with different image development times (from 2 to 16 h). The level of expression was then correlated with the time needed to obtain clear staining in each of the areas shown in table 25. When clear staining was observed in a particular area during the first two hours of developing, it was assigned a +++ mark, those areas that were clearly stained after nine hours of chromogenic reaction were assigned as ++ whereas those stained after sixteen hours of staining were designated with +. Those which revealed no staining after sixteen hours were considered as having *LRRK2* expression too low to be detectable with this particular technique.

#### **3.2.4. Image handling**

Brightfield microscopic images were captured using a Nikon DMX-2000 camera (Nikon, <http://www.nikoninstruments.com>). The camera was attached to a Nikon Eclipse E600 microscope (Nikon, <http://www.nikoninstruments.com>) and connected to a computer. The captured images were automatically adjusted for brightness and contrast with Adobe Photoshop.

#### **3.2.5. Quality-control procedures**

We carried out control experiments for all the probes using the sense sequence, always at the same concentration as the respective antisense probe, to check possible non-specific hybridization.

#### **3.2.6. Nomenclature**

The nomenclature used in this report follows that of the atlas the adult mouse brain by Paxinos and Franklin (Paxinos G, 2001).

### **3.3. SCREENING FOR MUTATIONS IN *OMI/HTRA2***

#### **3.3.1. Subject collection**

All samples were taken directly from pre-compiled panels from the National Institute of Neurological Disorders and Stroke (NINDS) funded Neurogenetics repository hosted by

the Coriell Institute for research (<http://ccr.coriell.org>). All participants provided written informed consent.

Neurologically normal Caucasian control subjects were derived from 9 panels of DNA: *NDPT002*, *NDPT006*, *NDPT009*, *NDPT019*, *NDPT020*, *NDPT021*, *NDPT022*, *NDPT023* and *NDPT024* (see section 3.5.1.1). These contain DNA from 828 unrelated individuals from North America, 340 males and 488 females. Each panel contains 5 micrograms of DNA from 92 unrelated Caucasian individuals without history of Alzheimer's disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, brain aneurism, dementia, dystonia, or Parkinson's disease. None had any first-degree relative with a known primary neurological disorder and the mean age of participants was 74 (range, 55-88), 67 (range, 55-84), 68 (range, 55-84), 28 (range, 15-35), 40 (range, 36-48), 49 (range, 46-53), 67 (range, 56-91), 68 (range, 56-94) and 67 (range, 56-95) for panels *NDPT002*, *NDPT006*, *NDPT009*, *NDPT019*, *NDPT020*, *NDPT021*, *NDPT022*, *NDPT023* and *NDPT024* respectively. The mean age at onset of all controls was 58 years.

PD cases were taken from 7 panels of DNA: *NDPT001*, *NDPT005*, *NDPT014*, *NDPT015*, *NDPT016*, *NDPT017* and *NDPT018* (for more details see section 3.5.1.2).

These panels contain 5 micrograms of DNA from 644 unrelated Caucasian individuals from North America with PD, including 363 males and 281 females. The mean age at onset was 70 (range, 55-70), 67 (range, 55-81), 32 (range, 7-40), 43 (range, 40-47), 49 (range, 47-52), 66 (range, 56-84) and 66 (range, 56-87) respectively.

Here we have defined young-onset PD (YOPD) as onset of a parkinsonian syndrome at or before age 40 ( $\leq 40$ ); only those samples from *NDPT014* and 14 samples from *NDPT015*, were considered as YOPD. These samples (a total of 106 including 63 males and 43

females) had a mean age at onset of 31 (range 7-40), defined as when symptoms were first noted, including at least one of the following: resting tremor, rigidity, bradykinesia, gait disorder and postural instability.

The remaining samples comprise 538 samples, including 302 males and 236 females with late-onset PD. The mean age at onset is 59 years (range 41-87) and they all show at least one of the main clinical signs of PD such as resting tremor, rigidity, bradykinesia, gait disorder and postural instability at the disease onset.

All subjects (from both young- and late-onset panels) were questioned regarding family history of parkinsonism, dementia, tremor, gait disorders, and other neurological dysfunction. Subjects both with and without a reported family history of Parkinson's disease were included. None were included who had three or more relatives with parkinsonism, nor with clear Mendelian inheritance of PD. The mean age at onset of all PD cases was 55 years.

A more detailed description of both case and control samples, can be found at <http://ccr.coriell.org/Sections/Collections/NINDS/DNAPanels.aspx?PgId=195&coll=ND>.

### **3.3.2. Sequence analysis of *OMI/HTRA2***

The entire coding region of *OMI/HTRA2* (RefSeq NM\_013247) was sequenced in all samples. PCR was performed with thermocycling program 57TD52 (table 5) in a final volume of 15  $\mu$ l with Roche's FastStart PCR Master Mix, and forward and reverse primers listed on table 17.

Master Mix:..... 10 µl  
 Primer F (5µM):.....2 µl  
 Primer R (5µM):.....2 µl  
 DNA (10ng/µl): .....1 µl  
 DMSO: .....\*

\* 1µl of Dimethylsulfoxide (American Bioanalytical) was added for PCR reactions of exons 2 and 4.

PCR products were verified in a 2% ethidium-bromide-containing agarose-TBE gel and visualized via ultraviolet transillumination. Prior to loading each PCR product on a gel, it was mixed with 6X Orange G Loading Dye Solution (10 mM Tris-HCl (pH 7.6), 0.15% orange G, 60% glycerol, 60mM EDTA).

**Table 17:** Forward and reverse primer sequences used for amplification and direct sequencing of the coding region of *OMI/HTRA2*.

<b>Exon</b>	<b>Forward primer (5'-3')</b>	<b>Reverse primer (5'-3')</b>
Exon 1_1	GTCCTACTGTCCGCCTG	TTCCAGAGTTCTCTGAGGC
Exon 1_2	GGTTGTCTGTTGGGGTC	ACAGATAAGAGGAGGCTCTG
Exon 2	CTATATCGAGATCCTGGACC	GACCTGTCATCTGAGATGC
Exon 3	CTTATTTGCTCGCATCTTCAG	GCTCTATATAGGCTCATCCACC
Exon 4	AGAAAGTACCTACATCCTGG	CTTCCCCTATTTCCACC
Exon 5-6	CTATCTCTCAATATTCCAACCAG	ATTACACTGGTCACATTACATGT
Exon 7	AATGTGTTGATGAGAGACTTGAG	GGAGTACAAAGCCTAGTTCAAG
Exon 8	GATGTTTCAGCATGGTGTACTC	GTGATTCTTCAGAGCCCCAG

In order to remove unincorporated primers, dNTPs, DNA polymerase and salts used during PCR amplification that can interfere with downstream applications such as sequencing, the Agencourt AMPure PCR Purification system (Agencourt, <http://www.agencourt.com>) was utilized. This system utilizes Agencourt’s AMPure solid-phase paramagnetic bead technology for high-throughput amplification of PCR amplicons. Agencourt AMPure utilizes an optimized buffer to selectively bind PCR

amplicons 100bp and larger to paramagnetic beads. Excess oligos, nucleotides, salts, and enzymes can be removed using a simple washing procedure. The resulting purified PCR product is essentially free of contaminants and can be used in downstream applications such as sequencing.

Each purified product was sequenced using forward or reverse primers with Applied Biosystems BigDye terminator v3.1 sequencing chemistry as explained in section 3.1.2.

In order to efficiently purify the sequencing products to deliver superior quality sequencing data, Agencourt CleanSEQ technology was used.

Purified sequencing reactions were then electrophoresed on an ABI3730 XL genetic analyzer (Applied Biosystems, <http://www.appliedbiosystems.com>) and analyzed with Sequencher software v4.1.4 (Gene Codes, <http://www.genecodes.com>). All changes that derived from the wild type sequence were verified by PCR amplification of a fresh DNA aliquot and sequencing in both forward and reverse directions.

### 3.3.3. Statistical analysis.

Power calculations were performed using the program PS v2.1.30 (Dupont and Plummer, 1990). For variant p.A141S, based on the minor allele frequency we obtained after sequencing 828 control samples (0.02991), our series possesses sufficient power (98.5%) to detect a difference in allele frequency ( $p = 0.05$ ) with an effect of OR = 2.15. This is the OR that Strauss and coworkers obtained in their study for this same variant (Strauss, Martins *et al.*, 2005).

For p.G399S based on a minor allele frequency of 0.0036 (according to our control population), our study possesses enough power (80%) to detect a difference in allele frequency ( $p = 0.05$ ) with an effect of  $OR = 3.59$ .

Plink v0.99s (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (Purcell, Neale *et al.*, 2007) was used to perform Fisher's exact test of association between all identified variants and PD. Haploview v4.0RC2 (<http://broad.mit.edu/mpg/haploview/>) was used to verify the results obtained. (Barrett, Fry *et al.*, 2005).

### 3.4. ANALYSIS OF SCA-2 AND SCA-3 GENES IN PARKINSONISM

#### 3.4.1. Subject collection.

280 probands with PD with and without family history with different ethnic background (African American, Asian, Caucasian, Hispanic, Native American and Pacific Islandic) were selected for these experiments (table 18). Probands underwent a neurological history and physical examination and were diagnosed with parkinsonism by Drs. Katrina Gwinn-Hardy, Roberto Weiser, or Marisol Gallardo. Family history was based on a questionnaire completed by the proband prior to the clinic visit. This questionnaire elicits the medical history of first and second-degree relatives. Enrollment of other family members was carried out as follows: the proband was informed that those relatives wishing to participate should call the investigators in order to be considered for the study. If those individuals came forward and wished to enroll, informed consent was obtained and subsequently these additional family members were evaluated according to the above procedures. The diagnosis of Parkinson's disease was made based on Hughes' criteria (Hughes, Daniel *et al.*, 1992) for all 280 subjects.

In 114 cases with a positive family history of parkinsonism, additional diagnoses of dementia, dystonia, amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), restless legs syndrome (RLS), essential tremor and epilepsy were noted in other members of the kindred. Several of these diagnoses may have been present in the proband in addition to PD, these include: dementia, dystonia, essential tremor, and RLS.

Here we present a brief clinical description of members of the family 374, in which pathogenic *ATXN-2* expansions were identified (see results, section 4.4):

. **374-1:** This Caucasian woman, evaluated by us at age 59 years, was first diagnosed with Parkinson's disease at age 55 years. Her initial symptoms were shuffling gait and 4–6 Hertz resting tremor. On evaluation in our clinic, she had rare right lower extremity resting tremor while “on”, *en bloc* turns, small steps, and stooped posture. She had markedly decreased right arm swing when walking. There was right sided rigidity in both upper and lower extremities. She had no cerebellar findings. She responded well to carbidopa/levodopa therapy, which she stated made her feel “normal”. In addition to her symptoms of Parkinson's disease, she suffered from symptoms consistent with periodic leg movements of sleep, and restless legs syndrome, which responded to the dopamine replacement therapy. A sleep study has not been done.

. **374-5:** This 36-year old Caucasian woman's symptoms began at age 29 years, when she noted difficulties with “internal trembling”. Resting tremor in the right upper extremity became apparent to her at age 30 years, and involved the right lower extremity within the next several years. An extensive evaluation by a neurologist revealed no secondary cause for her symptoms. She was diagnosed with Parkinson's disease and responded well to carbidopa/levodopa therapy. Examination when “on” was remarkable for normal eye movements, mild dystonic posturing in the right foot, and no resting nor activation tremor. No retropulsion was noted on pull test. Gait was apparently normal. There were no cerebellar or other atypical signs.

. **374-9:** This patient is a 53-year old female sibling of the proband with essential tremor onset at age 49 years affecting both upper extremities. Additional diagnoses include vertigo consistent with peripheral vestibular disease, migraine and left hand paresthesias, possibly carpal tunnel syndrome.

**Table 18:** General Characteristics of parkinsonism patients.

	(n = 280)
Age, median, mean, range	61, 61±15, [22–91]
Age at onset median, mean, range	50, 51±15, [16–89]
Gender (n, freq)	
Male	152 (0.54)
Female	98 (0.35)
Unknown	30 (0.11)
Dopamine replacement therapy (n, freq)	
Responsive	149 (0.53)
Unresponsive	8 (0.03)
Unknown	123 (0.44)
Other features (n, freq)	
Dystonia	13 (0.05)
Dementia	16 (0.06)
Essential tremor	8 (0.03)
Family history	
Sporadic	166 (0.59)
First degree relative	57 (0.20)
Second degree relative 1	4 (0.05)
Extended (first and second degree)	40 (0.14)
Distant relative only	3 (0.01)
Mode of inheritance (n, freq)	
Autosomal dominant	66 (0.58)
Autosomal recessive	41 (0.36)
Unclear	7 (0.06)
Ethnic origin (n, freq)	
African American	21 (0.08)
Asian	10 (0.03)
Caucasian	165 (0.59)
Hispanic	81 (0.29)
Native American	2 (0.01)
Pacific Islander	1 (0.004)

### 3.4.2. Genomic DNA extraction.

Genomic DNA was isolated from whole blood using the Wizard Genomic DNA purification kit (Promega Corporation, <http://www.promega.com>) as per the manufacturer's instructions. Briefly, purification is based on a four-step process. The first step in the purification procedure lyses the cells and the nuclei. For isolation of DNA from white blood cells, this step involves lysis of the red blood cells with the Cell Lysis Solution, followed by lysis of the white blood cells and their nuclei in the Nuclei Lysis Solution. The cellular proteins are then removed by a salt precipitation step using the Protein Precipitation Solution, which precipitates the proteins but leaves the high molecular weight genomic DNA in solution. Finally, the genomic DNA is concentrated and desalted by isopropanol precipitation.

### 3.4.3. Genotyping

PCR amplification was performed in a 25  $\mu$ l reaction containing 25ng of genomic DNA, 10 pmol of both forward and reverse oligonucleotide custom primers (table 19), manufacturer's buffer, Q solution (Qiagen, <http://www.qiagen.com>), 200  $\mu$ mol/L deoxynucleotide triphosphates (Amersham Pharmacia Biotech Inc, <http://www.amersham.com>), and 0.2 U of TaqDNA polymerase (Qiagen, <http://www.qiagen.com>). For PCR amplification either the forward or the reverse primer utilized was labeled with 6-FAM (table 19). PCR reaction were run on a Touchdown Hybaid thermocycler (Hybaid, <http://www.thermo.com>) using the conditions in table 4. Five microliters of each reaction were then run on a 2% agarose-TBE gel and visualized via ultraviolet transillumination to confirm amplification and to exclude the possibility of

pathogenic extensions beyond the ladder size (500 bp) used on the automated sequencer. Prior to loading each PCR product on a gel, it was mixed with 6X Orange G Loading Dye Solution.

Two microliters of each PCR product along with 8 microliters of a solution 1:20 of GS500LIZ size standard (Applied Biosystems, <http://www.appliedbiosystems.com>) and doubly deionized formamide were analyzed on an ABIPrism 3100 Genetic Analyzer (Applied Biosystems, [www.appliedbiosystems.com](http://www.appliedbiosystems.com)). All samples were amplified and analyzed on at least two separate occasions. Product size was used to calculate triplet repeat sizes for all samples; this calculation was based on the product size relative to samples run in parallel with a known repeat number. To identify these samples with known repeat sizes for each repeat, three samples with homozygous repeat size were sequenced with BigDye terminator chemistry (Applied Biosystems, <http://www.appliedbiosystems.com>) using unlabeled primer pairs identical in sequence to those used for amplification (see section 3.1.2 for conditions). The observed number of repeats at each *locus* was then used to calculate repeat numbers for all other cases.

**Table 19:** Forward and reverse primer sequences used for amplification and direct sequencing of the region containing cag repeats within *ATX2* and *ATX3* genes.

<b>Gene/Exon</b>	<b>Forward primer (5'-3')</b>	<b>Reverse primer (5'-3')</b>	<b>Label</b>
<i>ATXN-2/Exon 1</i>	TTCCAGACCCTTCCATCC	ACGATGGATAACTAATTGGAGC	6-FAM
<i>ATXN-3/Exon 9</i>	CCAGTGACTACTTTGATTCG	TGCCTTTCACATGGATGTGAA	6-FAM

### **3.5. WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 1.**

#### **3.5.1. Subject collection.**

All samples utilized in this study were derived from the National Institute of Neurological Disorders and Stroke (NINDS) funded Neurogenetics repository hosted by the Coriell Institute for Medical research (<http://ccr.coriell.org>). The NINDS banks lymphoblast cell lines (LCLs) with DNA from subjects with cerebrovascular disease, epilepsy, motor neuron disease, Parkinsonism and Tourette syndrome as well as controls. Six precompiled panels each consisting of 92 cases or controls were selected for the analysis. The panels that contained samples from patients with Parkinson's disease were *NDPT001*, *NDPT005*, and *NDPT007*; these included DNA from 273 unique participants and three replicate samples. The panels that contained samples from neurologically normal controls were *NDPT002*, *NDPT006*, and *NDPT008*; these comprised DNA from 275 unique participants and one replicate sample.

DNA was extracted using a modification to the procedure developed by Miller *et al.* (Miller, Dykes *et al.*, 1988) as previously reported and transfected to Epstein-Barr virus immortalized LCLs. Epstein-Barr virus immortalization was performed as previously described (Miller, Shope *et al.*, 1972; Tumilowicz, Gallick *et al.*, 1984) and the average passage number for each line was 5 (range 5–7).

**3.5.1.1. Neurologically normal control samples:** Blood samples were drawn from neurologically normal, unrelated, white individuals at many different sites within the

USA. Each participant underwent a detailed medical history interview and asked specifically regarding the following disorders, which were not present: Alzheimer's, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, cerebrovascular disease, dementia, dystonia, Parkinson's, and schizophrenia. None had any first-degree relative with a known primary neurological disorder. The mean age of participants at sample collection was 68 years (range 55–88).

**3.5.1.2. Parkinson's disease samples:** Blood was obtained from unique and unrelated white individuals from the USA with idiopathic Parkinson's disease including 166 males and 110 females. The mean age at onset was 66 years, ranging from 55 to 84 years. Disease onset was defined as the time when symptom(s) of the disease were first noted, including at least one of the following: resting tremor, rigidity, bradykinesia, gait disorder and postural instability.

In order for subject inclusion, completion of the NINDS Repository Clinical Data Elements (<http://ccr.coriell.org/Sections/Collections/NINDS/CDE/pd.aspx?PgId=336>) were required. Furthermore, only subjects who met the UK Brain Bank Criteria idiopathic Parkinson's disease were included (Hughes, Daniel *et al.*, 1992). By those criteria, all subjects had bradykinesia, and at least one of the following: muscular rigidity, 4-6 Hz resting tremor and postural instability (not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction). None had exclusionary features. All had documentation of sustained, excellent response to dopamine replacement therapy.

All subjects were queried regarding family history of parkinsonism, dementia, tremor, gait disorders, and other neurological dysfunction. Subjects both with and without a

reported family history of Parkinson's disease were included on this panel. None were included who had three or more relatives with parkinsonism, nor with apparent Mendelian inheritance of PD.

### 3.5.2. Genotyping.

**3.5.2.1. Platform used:** The platform we used for this study was Illumina's Infinium genome-wide genotyping beadarray (Illumina, <http://www.illumina.com>). This technology is based on 3  $\mu\text{m}$  silica beads that self assemble by Van de Waals forces and hydrostatic interactions with the walls of micro-wells over a planar silica substrate in a slide format. When randomly assembled to this substrate, the beads have a uniform spacing of approximately 5.7  $\mu\text{m}$ . Each bead is covered with hundreds of thousands of copies of a specific 50 mer nucleotide that act as the capture sequences for the fragmented target DNA. These oligos are coupled to the beads through an amine linkage, creating a "bead type" for each bead-oligo combination. "Bead types" are pooled together to form assay panels up to 1,000,000 targets depending on the chip version.

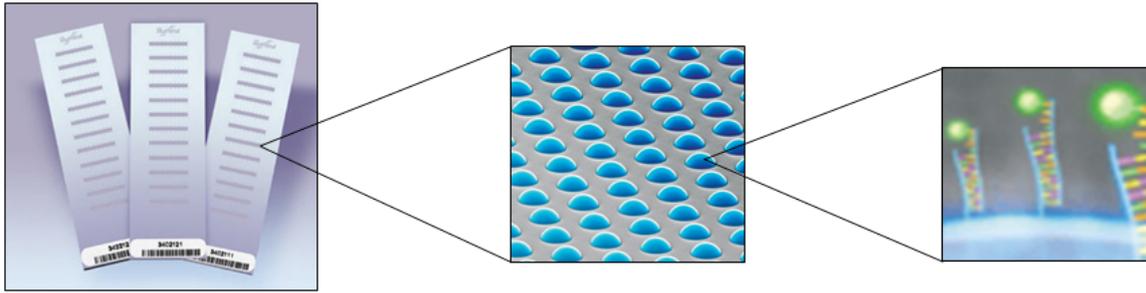
The concept of the Infinium whole-genome genotyping assays is based on direct hybridization of fragmented whole-genome amplified genomic DNA to these 50 mer *locus*-specific oligos. After hybridization, each SNP is "tagged" by an enzymatic-based extension assay, using labeled nucleotides. These "tags" are visualized by staining with an immunohistochemistry assay that increases the overall sensitivity of the assay (figure 8).

All samples in this study were assayed using Illumina's Infinium Human-1 (Infinium I) and Infinium HumanHap300 SNP chips (Infinium II).

Infinium Human-1 genotyping beadchip assays 109,365 gene-centric SNPs. Of those, 48% (49,183) are within 10Kb of an exon, 23% (24,292) within transcripts, 15% (16,086) in highly conserved regions and 14% (19,804) to achieve uniform spacing across genome. This product is suited for both whole-genome association studies and LOH/copy number analysis.

Infinium II HumanHap300 assays 317,511 haplotype tagging SNPs (tagSNPs) derived from Phase I HapMap Project. tagSNPs are *loci* that can serve as proxies for many other SNPs. The use of tagSNPs greatly improves the power of association studies, as the same information and power from a larger number of SNPs can be gathered by genotyping only a subset of *loci*, thus making possible to achieve more statistical power and genomic coverage using fewer SNPs and statistical tests, compared to other strategies using larger numbers of randomly chosen SNPs. This beadchip version contains a higher density of tag SNPs in regions of the genome that are within 10 kb of a gene or in evolutionary conserved regions. In addition, it contains about 7,300 nsSNPs (non-synonymous amino-acid changing SNPs), and a higher SNP density across the Major Histocompatibility Complex (MHC).

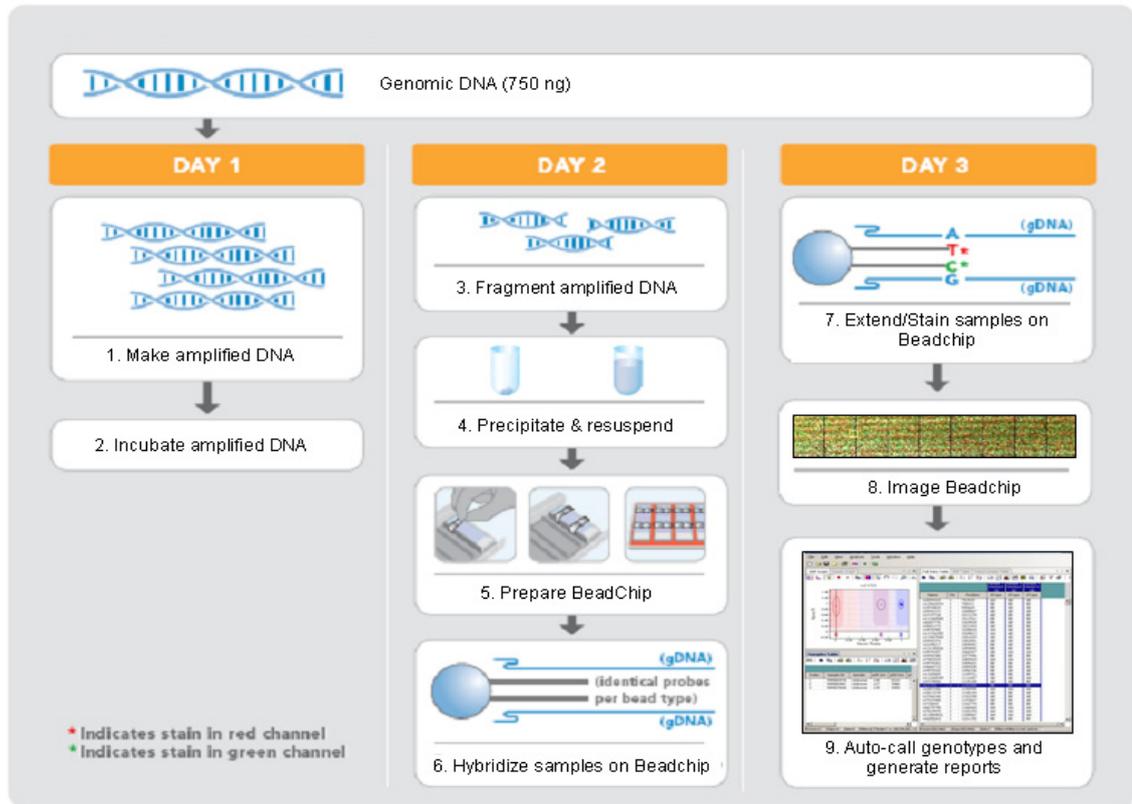
This product is ideally suited for both whole-genome association studies and LOH/copy number studies, and provides higher genomic coverage with fewer SNPs, compared to whole-genome genotyping strategies using randomly selected SNPs.



**Figure 8: Illumina's Infinium genome-wide genotyping beadarray.** Each bead is covered with thousands of copies of a specific 50 mer capture sequence that hybridize fragmented DNA, making a beadtype. Bead types are pooled together to form assay panels up to 1,000,000 targets depending on the chip version.

There are 18,073 SNPs in common between the Human-1 and HumanHap300 arrays; thus the assays combined provide data for 408,803 unique SNPs.

**3.5.2.2. Infinium Workflow:** The Infinium assay protocols, feature single-tube sample preparation without PCR or ligation steps, significantly reducing labor and potential sample handling errors. An enzymatic discrimination step provides high call rates and accuracy with unconstrained *locus* selection. The Infinium workflow can be divided on five main stages: 1) Sample preparation, 2) sample fragmentation and hybridization, 3) extension and staining, 4) scanning, and 5) data analysis (figure 9).



**Figure 9: Infinium II assay workflow.**

• **Sample preparation:** In stage 1, 750-1,000 ng of DNA is used as template for amplification over 20-24 hours at 37°C with WG# -AMM solution. This generates a sufficient quantity of each individual DNA sample to be assayed in further steps. Prior to 37°C incubation, DNA is denatured with NaOH 0.1N for ten minutes and neutralized by adding WG# -MP1 reagent. To facilitate plate-based amplification and isopropanol precipitation, each DNA sample is amplified in four separate wells within the same plate. After fragmentation, precipitation and resuspension in hybridization buffer, the four wells are recombined.

. **Sample fragmentation and hybridization:** After whole-genome amplification is performed in the first stage, each individual sample is enzymatically fragmented to an average size of around 300 bp by using an endpoint enzymatic reaction by the addition of WG# -FRG solution, vortexing at 1600 rpm for one minute and incubating at 37°C for one hour. This endpoint fragmentation process eliminates the risk of over-digestion.

Once the digestion is performed, WG# -PA1 reagent and isopropanol are added to precipitate each DNA sample. After 30 minutes of incubation at 4°C, and a 3,000 Xg centrifugation also at 4°C, excess of isopropanol and WG# -PA1 is removed, and pellets air-dried for one hour. Once the pellets are dry, the formamide-containing hybridization buffer (WG# -RA1) is added to each well, followed by a one-hour-incubation at 48°C and vortexing at 1800 rpm for one minute. After a 20 minute denaturation at 95°C, DNA is ready to be hybridized to the surface of the beadchip. Before performing this step, the beadchip has been prepared for hybridization by washing with 100% ethanol for 10 minutes and detergent WG# -PB1 for 5 minutes, followed by assembly in a Flow Through Chamber that allows solutions to be added to the surface of the beadchip. 100% formamide is dispensed into the Flow Through Chamber reservoir afterwards, to pre-wet the beadchip surfaces. Then WG# -RA1 is dispensed to equilibrate the Chamber and beadchip surfaces.

Now, utilizing Tecan's GenePaint automated slide processor (Tecan, <http://www.tecan.com>), DNA samples are loaded into the Flow Through Chambers at 22°C. These are then removed from the GenePaint processor and placed in a container humidified with WG# -PB2 for a minimum of 16 hours (24 hours maximum) at 48°C.

. **Extension and staining:** For the following steps Tecan's GenePaint automated slide processor is used. The GenePaint system employs a capillary gap flow-through chamber to enable reagent entrapment and exchange over the beadchip's active surface. Washing, blocking, extension, and signal amplification are all performed by simple reagent additions to the Flow Through Chamber. Although these processes can be performed manually, automated robotic processing assures maximum consistency from slide to slide.

Two different primer extension assays can be employed: allele-specific primer extension (ASPE) for Infinium I assays, and single-base extension (SBE) for Infinium II assays (figure 10).

The ASPE assay, a one color format, is specifically designed to allow the detection of all SNP classes by employing two identical bead types differing only at their 3' base of the attached oligo (allele-specific bead type). For a given SNP the capture sequence of the allele-specific bead type is the perfect match hybrid for allele A, and the other is the perfect match hybrid for allele B of a certain SNP. After hybridization, multiple biotin-labeled nucleotides are added only to those oligos perfectly matched to a captured DNA target.

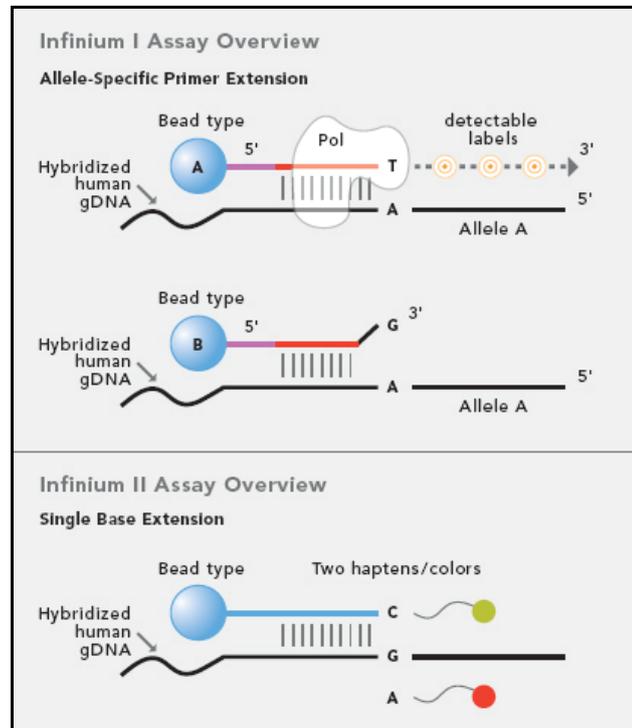
In order to perform this series of reactions, the beadchips are first treated with WG# -RA1 to wash away unhybridized and non-specific hybridized DNA sample. WG# -XB1 and WG# -XB2 reagents are used for preparing the beadchip for the extension reaction. WG# -EMM reagent, containing polymerase and biotin-labeled nucleotides, is dispensed to extend oligos correctly matched to DNA on the beadchip, incorporating biotin-labeled nucleotides at 37°C. NaOH-formamide treatment for two minutes removes the hybridized

DNA to reduce extraneous signal, leaving the extended oligos. After neutralization with WG# -XB3 reagent, the labeled extended oligos are stained in a multi-layer way to amplify and detect the incorporated label (WG# -LMM, WG# -ASM and WG# -XB3). Now the beadchips are removed from the Tecan's GenePaint system, the Flow Through Chambers, and washed during ten minutes in WG# -PB1 detergent. Once dried by gentle centrifuging, they are ready to be imaged.

The SBE assay format uses a single oligo per SNP with two color read-out. This characteristic reduces the required number of oligos needed by half compared to the ASPE assay. After target hybridization, those SNP *locus*-specific oligos perfectly matched to a captured DNA target are extended in the presence of labeled dideoxynucleotides. These dideoxynucleotides are biotin-labeled ddCTP and ddGTP, and 2,4-dinitrophenol (DNP)-labeled ddATP and ddUTP. These labeled dideoxynucleotides are incorporated by the polymerase and allow detection after a dual-color multi-layer immunohistochemical assay. The limitation of this assay is that only 83% of common bi-allelic polymorphisms can be detected on a single beadchip. A and T nucleotides are stained in one color and C and G in another. Thus, AT and GC polymorphisms can not be detected.

To perform this assay the beadchips are first washed with WG# -RA1. WG# -XC1 and WG# -XC2 are added later to condition the beadchips surface for the extension reaction. This SBE is performed by adding a polymerase and labeled dideoxynucleotide mix (WG# -TEM). NaOH-formamide wash is the performed to remove the hybridized DNA. After neutralization using WG# -XC3 reagent, the labeled extended primers undergo a multi-

layer staining process. Next the Flow Through Chambers are disassembled and beadchips washed with WG# -PB1 detergent and coated with WG# -XC4.



**Figure 10: Infinium I and Infinium II technologies.** Infinium I and Infinium II technologies use two different primer extension assays: allele-specific primer extension (ASPE, upper panel) and single-base extension (SBE, lower panel).

- **Scanning:** In order to scan the processed bead chips, an Illumina Stand-Alone BeadArray Reader was used. This reader is a two-color (543nm/643nm) confocal fluorescent scanner with 0.84  $\mu\text{m}$  pixel resolution. The scanner excites the fluorescent group generated during signal amplification of the allele-specific (one color) or single-base (two colors) extension products on the beads of the beadchip. Light emissions from these fluorescent groups are then recorded in high-resolution images of each beadchip section using BeadScan software.

. **Data analysis:** In order to extract genotypes from the raw image data, a normalization procedure must be undertaken. The BeadStudio software package, normalizes raw data from each BeadChip, generates genotype calls, performs clustering and analyzes data intensity using data generated from Infinium assays. Any normalization procedure applied to Illumina's genotyping data must be applied on the sub-bead pool level. A sub-bead pool is a group of beads that were manufactured together and are located in roughly the same beadchip analytical position (a stripe). Thus, for example Human-1 beadchips are divided into 12 stripes each of them containing two different pools. Therefore, Human-I beadchips contain a total of 24 different sub-bead pools. The sub-bead pool information is provided along with each beadchip as a bead pool manifest (\*.bpm) file.

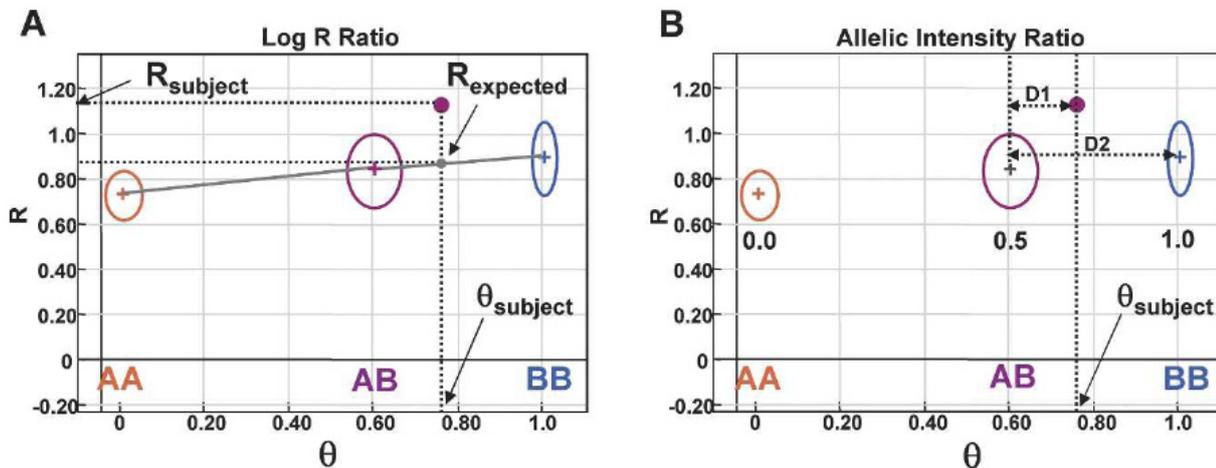
Because the performance of external controls can vary from sample to sample, Illumina's standard normalization is performed by means of a self-normalization algorithm which draws on information contained in the array itself. The normalization algorithm is designed to remove outliers, adjusts for channel-dependent background and global intensity differences, and also scales the data. During this process, the X and Y color channels (alleles A and B) undergo an affine coordinate transformation to make the data appear as accurate as possible with the homozygotes lying along the transformed X or Y axes. Five steps are performed: 1) Outlier removal; 2) translation correction in which the asymptotes are fitted to candidate AA or BB homozygotes. Two straight lines are fit into candidate homozygote A and B alleles. The intersection of these two lines is computed and defines the translated origin; 3) rotational correction: the angle between the AA homozygote asymptote and the translated X axis is used to define the rotational

correction. This angle corresponds to the theta ( $\theta$ ) parameter; 4) shear correction: the angle of the BB homozygote asymptote with respect to the translated and rotated Y axis is used to define the shear correction. This angle corresponds to the shear parameter; 5) scaling correction: statistical centroids are computed for the candidate AA homozygotes to define an X-axis scaling parameter, and for candidate BB homozygotes to define a Y-axis scaling parameter. The translated, rotated, shear-corrected data are normalized to a scale of  $\sim 1$  using the scaling parameters.

After normalization, the genotyping data are transformed to a polar coordinate plot of normalized intensity  $R = X_{\text{norm}} + Y_{\text{norm}}$  and allelic intensity ratio  $\theta_{\text{subject}} = (2/\pi) * \arctan(Y_{\text{norm}}/X_{\text{norm}})$ , where  $X_{\text{norm}}$  and  $Y_{\text{norm}}$  represent transformed normalized signals from alleles A and B for a particular *locus*. The log R ratio of signal intensities is shown as  $\log_2(R_{\text{subject}}/R_{\text{expected}})$  and is the base 2 logarithm of the ratio of observed intensity versus expected intensity.  $R_{\text{expected}}$  is computed from a linear interpolation of the observed allelic ratio ( $\theta_{\text{subject}}$ ) with respect to the canonical genotype clusters. The three canonical genotype clusters were generated at one point in time by training on 120 normal samples, and serve as standards for all future experiments. In addition to computing  $R_{\text{expected}}$ , the observed allelic intensity ratio ( $\theta_{\text{subject}}$ ) is used to estimate a quantitative B allele frequency for the particular SNP in the given sample by using interpolation of known B allele frequencies of the three canonical clusters (0, 0.5 and 1.0).

This two transformed parameters (log R ratio and B allele frequency) are then plotted along the entire genome for all SNPs in the array using the Illumina Genome viewer tool (IGV) within BeadStudio software (Illumina, <http://www.illumina.com>). This allows

identification of chromosomal aberrations and tracks of contiguous homozygous calls (figure 11).



**Figure 11. Normalization and transformation processes applied to the Infinium raw genotype data. A:** Normalization and The  $\log_2 R$  ratio compares the observed normalized intensity ( $R_{\text{subject}}$ ) of the subject sample to the expected intensity ( $R_{\text{expected}}$ ; gray dot) based on the observed allelic ratio,  $\theta_{\text{subject}}$ , through a linear interpolation (gray lines) of the canonical clusters AA, AB, and BB (shown as circles). The normalized intensity value obtained from a single SNP is represented as a purple dot. The  $R$  and  $\theta$  values for the subject are shown with thick black dotted lines. **B:** The canonical clusters (shown as circles) are also used to convert  $\theta$  values, that is,  $\theta_{\text{subject}}$ , to B allele frequency (allelic copy ratio). This is accomplished by a linear interpolation of the known allele frequencies assigned to each cluster (0.0, 0.5, and 1.0). The allele frequency for an observed  $\theta$  value falling between two clusters is also calculated by linear interpolation with lines D1 and D2.

### 3.5.3. Quality-Control procedures.

**3.5.3.1. Low quality genotyping:** Any sample with a call rate below 95% was repeated on a fresh DNA aliquot and if the call rate persisted below this level the sample was excluded from the analysis. Low-quality genotyping led us to repeat 11 individual samples, of which seven were ultimately excluded from the analysis.

**3.5.3.2. Population substructure:** In an attempt to detect the presence of significant population substructure or ethnically mismatched individuals we selected 267 random, unlinked SNPs from throughout the genome and ran STRUCTURE v2.2 (<http://pritch.bsd.uchicago.edu/structure.html>) on these data from all genotyped individuals. This software provides allows inferring the presence of distinct populations, assigning individuals to populations, studying hybrid zones, identifying migrants and admixed individuals, and estimating population allele frequencies in situations where many individuals are migrants or admixed. It can be applied to most of the commonly used genetic markers including SNPs.

#### **3.5.4. Statistical analysis.**

Power calculations were performed using the program PS v2.1.30 (Dupont and Plummer, 1990). Based on the average observed minor allele frequency (26%) and assuming that either the causal variant is typed or that there is complete and efficient tagging of common variation by the genotyped tSNPs, our data provides 80% power to detect an allelic association with an odds ratio of more than 2.09 and less than 0.40 at an uncorrected significance level of  $p = 0.000001$ .

For each SNP we computed a series of estimates and tests using a program developed at Wake Forest University called SNPgwa ([http://www.phs.wfubmc.edu/public\\_bios/home.cfm](http://www.phs.wfubmc.edu/public_bios/home.cfm)). Each SNP was tested for departures from Hardy-Weinberg equilibrium. Five tests of genotypic association were computed: two degrees of freedom overall test for 2×3 tables, dominant model, additive model (Cochran-Armitage trend test), recessive model, and lack of fit to an additive

model. We calculated odds ratios (ORs), 95% CIs, and p values for each of the association models. We used the program Dandelion, which ran within SNPgwa, to do two-marker and three marker moving-window haplotype-association analyses for those SNPs that were consistent with Hardy-Weinberg equilibrium in controls. For all p values with an uncorrected significance of less than 0.05 we did permutation tests within SNPgwa using a variable number of permutations based on the p value of the test. For each permutation, SNPgwa permutes the affection status (case or control) of the entire sample represented in the input file while preserving the total number of cases and total number of controls in each permutation. The permutation is done using a Wichman-Hill random number generator (Wichmann, 1982).

### 3.5.5. Beyond association

The genotypic data generated the illumine Infinium assays can be used to look at copy number variation and to score homozygosity. This can be done by direct visualization with the IGV tool of log R ratio and B allele frequency (see *data analysis* in section 3.5.2.2).

Because there is considerable redundancy at each *locus* interrogated (each SNP is assayed several times within a beadtype), both alleles of each SNP are assayed multiple times; accordingly, B allele value for an individual SNP in a single sample gives an estimate of the proportion of times an individual allele at each polymorphism was called A or B. This metric is simply viewed as B allele frequency, thus an individual homozygous for A allele would have a score close to 0, an individual homozygous for B allele a score close to 1, and a score approximately 0.5 would indicate a heterozygous A/B genotype.

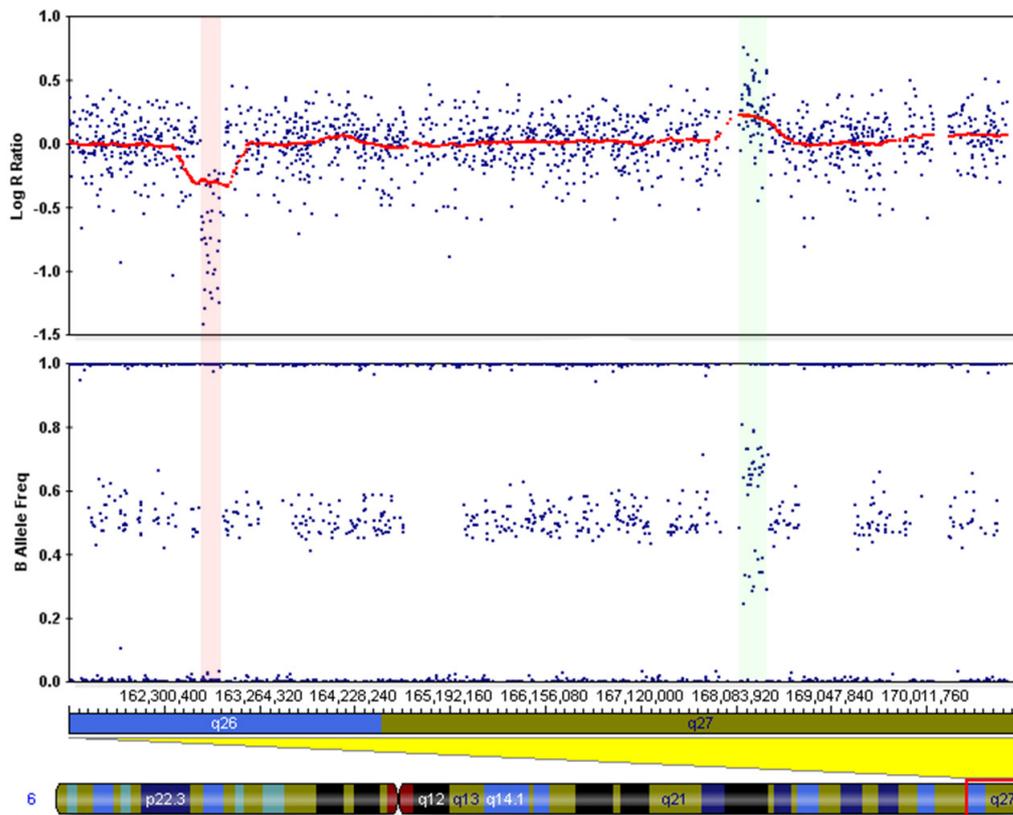
Significant deviation of these values in contiguous SNPs is indicative of an alteration in copy number.

$R_{\text{observed}}$  is a measure of the signal intensity for a *locus* and thus, when compared with an average expected value for that *locus* ( $R_{\text{expected}}$ ), the resulting log R ratio provides an indirect measure of copy number. A signal in a sample which is stronger than the expected signal is indicated by a log R ratio above 1.00 and indicates a copy number increase. On the other hand, a weaker signal than that expected for a certain SNP results in a decreased log R ratio and is suggestive of a deletion.

Considering these two metrics, a profile indicative of genomic duplication is an increase in log R ratio in the presence of four allele clusters outside the expected three. These clusters correspond to A/A/A, A/A/B, A/B/B and B/B/B genotypes (figure 12).

A profile indicative of heterozygous duplications would be a decrease in log R ratio in the presence of a lack of heterozygous calls (figure 12).

Resolution of this technique is sensitive to SNP density of the assay at any particular genomic *locus*. For all platforms this varies throughout the genome. However, typically we observe copy number variations as small as 200 kb using the Infinium I assays (109,365 SNPs) and as small as 50 kb using the Infinium II assays (317,511 SNPs).

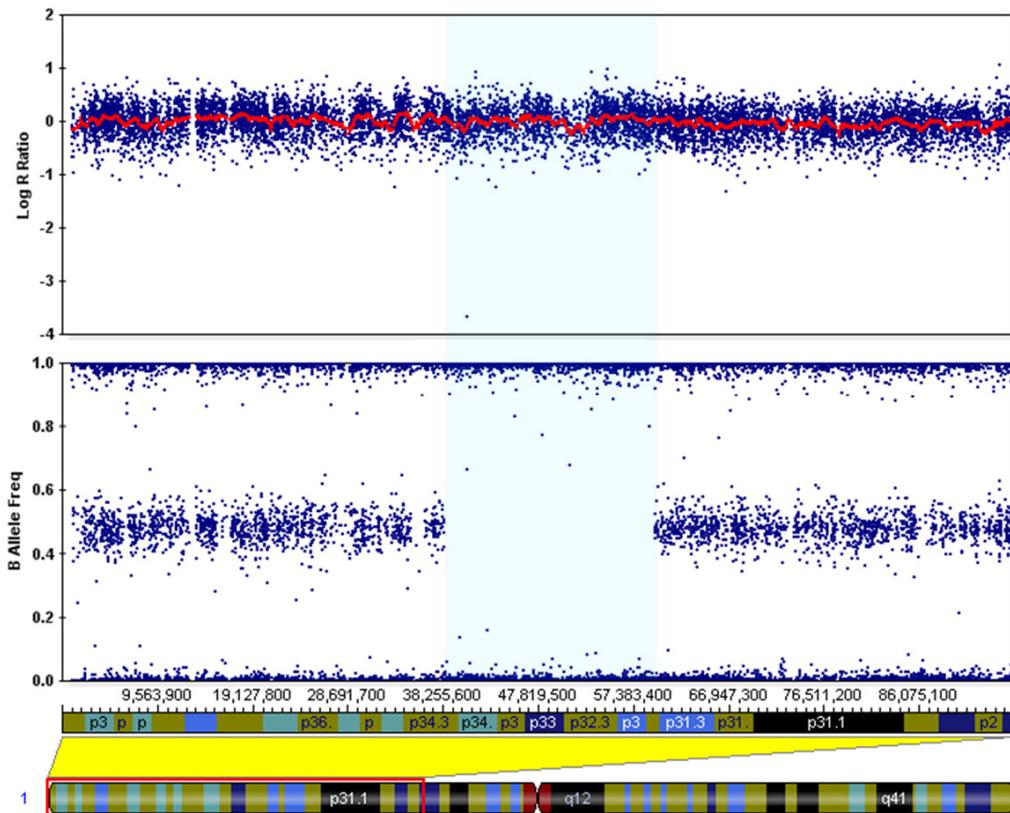


**Figure 12: Illumina Genome Viewer visualization of Infinium genotypic data.** Log R ratio (upper panel) and B allele frequency (lower panel) visualization with IGV for sample *ND04177*. A drop in log R ratio along with lack of heterozygous calls in B allele frequency plot shows a heterozygous 0.17 Mb heterozygous deletion in chromosome 6q26 (shaded in red). An increase of log R ratio coupled with the appearance of four clusters in the B allele frequency plot (corresponding with A/A/A, A/A/B, A/B/B and B/B/B genotypes) indicates a 0.24 Mb heterozygous duplication in chromosome 6q27 of this same sample (shaded in green).

Visualization of B allele frequency and log R ratio was performed by two individuals blinded to each others results. All apparent discrepancies were assessed by a third person. For both Human-1 and HumanHap300 BeadChips individually, homozygosity was scored based on a simple presence/absence basis; the areas of extended homozygosity were first highlighted by a simple script that searched for contiguous tracks of homozygous genotypes greater than 5 Mb, this was then followed by visual examination to rule out a drop in the log R ratio value across the same region, which would indicate

genomic deletion. Once the Human-1 and HumanHap300 data was merged, we also defined a homozygous track as a continuous track of homozygous genotypes where the track had to be at least 1 Mb in length and contain at least 10 SNPs (figure 13).

In order to test whether the extended homozygosity and structural genomic alterations were a result of, or amplified by, the LCL creation and passage process we repeated the genome-wide SNP genotyping in DNA extracted directly from the blood samples initially used for LCL creation. We assessed six out of the 26 control samples identified with one or more regions of contiguous homozygosity greater than 5 Mb in size, and all 24 control samples that harbor putative genomic deletions or multiplications.



**Figure 13: Illumina Genome Viewer visualization of Infinium genotypic data.** Log R ratio (upper panel) and B allele frequency (lower panel) visualization with IGV for sample 2035-44. Lack of heterozygous calls in the B allele frequency plot, along with no copy variation changes, indicates an extended homozygous track (shaded in blue) spanning more than 19 Mb in chromosome 1.

### 3.5.6. Confirmation of structural alterations by Real-time PCR

In order to confirm that the structural alterations found in these samples were not artifacts of the methodological or analytical process, quantitative PCR was performed in all 272 control samples using 14 TaqMan MGB probes targeting a 0.5 Mb duplication at chromosome 7q11 a 0.019 Mb deletion in chromosome 11q21, a 33 Mb heterosomic deletion in chromosome 13p12-q21 and a 0.11 Mb duplication in chromosome 16q23.2. In addition 10 individual probes across the *PARK2 locus* were used to verify the presence and size of copy number variation observed in this region in both control and PD individuals.

Primer and probes (sequence in table 20) were designed with Primer Express v2.0.0 software (Applied Biosystems, <http://www.appliedbiosystems.com>). TaqMan MGB probe for reference gene encoding  $\beta$ -globin was labeled with 6-FAM whereas probes against each target region in chromosomes 7, 11, 13 and 16, as well as those 10 targeting the *PARK2 locus*, on chromosome 6 were labeled with VIC. Real-time PCR was performed on a ABI7900Ht Sequence Detector system (Applied Biosystems).

PCR was carried out with TaqMan Universal PCR Master Mix (Applied Biosystems) using 25 ng of genomic DNA, 900 nmol/L primers and 250 nmol/L probes on a total volume of 20  $\mu$ l. PCR cycling conditions were 95°C for 10 min, 95°C for 15 s and 60°C for 1 min (40 cycles). The plates contained four replicates of each genomic DNA sample, control DNA and a no-template water control. The cycle in the log phase of PCR amplification at which a significant fluorescent threshold was reached (Ct) was used to quantify each amplicon. The dosage of each amplicon relative to the reference gene ( $\beta$ -globin) and normalized control DNA was determined using the  $2^{-\Delta\Delta C_t}$  method

(Corporation, 1997). To be considered valid, the requirements were a standard deviation (SD) of 0.16 and threshold values reached before 28 cycles. A value was considered a heterozygous deletion between 0.3 and 0.6, normal between 0.8 and 1.2 and heterozygous duplication between 1.3 and 1.6. Interestingly, individual *ND04946*, carrying a heterosomic deletion, showed a  $2^{-\Delta\Delta Ct}$  value between heterozygous deletion and normal, 0.62 (+0.08 SD).

**Table 20:** Primers and probes sequences utilized to perform Real-time to verify the presence and size of copy number variation observed in these regions in both control and PD individuals

Oligo name	Target gene-exon	Chromosome	Sequence 5'-3'	5' label
<b>Gene: <math>\beta</math>-Globin</b>				
Beta globin-F	$\beta$ -Globin-Exon 2	11p15.4	TGGGCAACCCTAAGGTGAAG	
Beta globin-R	$\beta$ -Globin-Exon 2	11p15.4	GTGAGCCAGGCCATCACTAAA	
Beta globine Probe	$\beta$ -Globin-Exon 2	11p15.4	CTCATGGCAAGAAAGTGCTCGGTGC	FAM
<b>Gene: <i>PAPK2</i></b>				
Parkin Exon 1-F	PARK2-Exon 1	6q25.2-q27	CCACCTACCCAGTGACCATGA	
Parkin Exon 1-R	PARK2-Exon 1	6q25.2-q27	CGGCGCAGAGAGGCTGTA	
Parkin Exon 1 Probe	PARK2-Exon 1	6q25.2-q27	TACGTGGGTACCTGCC	VIC
Parkin Exon 2-F	PARK2-Exon 2	6q25.2-q27	CCCAGTGGAGGTTCGATTCTG	
Parkin Exon 2-R	PARK2-Exon 2	6q25.2-q27	CCCCTGTCGCTTAGCAA	
Parkin Exon 2 Probe	PARK2-Exon 2	6q25.2-q27	CCAGCATCTTCCAGCTCAAGGAGGTG	VIC
Parkin Exon 4-F	PARK2-Exon 4	6q25.2-q27	TTCTTCTCCAGCAGGTAGATCAATC	
Parkin Exon 4-R	PARK2-Exon 4	6q25.2-q27	TTTTCCCGGCTGCACTCTT	
Parkin Exon 4 Probe	PARK2-Exon 4	6q25.2-q27	TTTTATGTGTATTGCAAAGGCCCTGTCA	VIC
Parkin Exon 5-F	PARK2-Exon 5	6q25.2-q27	CCGGATGAGTGGTG	
Parkin Exon 5-R	PARK2-Exon 5	6q25.2-q27	AGAGGAATGAATGTGACCAGGTACT	
Parkin Exon 5 Probe	PARK2-Exon 5	6q25.2-q27	CCACACTGCCCTGGGACTAGTGCA	VIC
Parkin Exon 6-F	PARK2-Exon 6	6q25.2-q27	GCACACCCACCTCTGACA	
Parkin Exon 6-R	PARK2-Exon 6	6q25.2-q27	TGCAAGTGATGTCCGACTATTTG	
Parkin Exon 6 Probe	PARK2-Exon 6	6q25.2-q27	AAACATCAGTAGCTTTGCACCTGATCGCA	VIC

**Table 20** (continued)

Oligo name	Target gene-exon	Chromosome	Sequence 5'-3'	5' label
<b>Gene: <i>PARK2</i></b>				
Parkin Exon 7-F	PARK2-Exon 7	6q25.2-q27	CCGCCACGTGATTTGCTTA	
Parkin Exon 7-R	PARK2-Exon 7	6q25.2-q27	CTGCCGATCATTGAGTCTTGTC	
Parkin Exon 7 Probe	PARK2-Exon 7	6q25.2-q27	CTGTTTCCACTTATACTGTG	VIC
Parkin Exon 8-F	PARK2-Exon 8	6q25.2-q27	GCAGCCTTTGAGATGCTCACT	
Parkin Exon 8-R	PARK2-Exon 8	6q25.2-q27	AGAGCTCCATCACTTCAGGATTCT	
Parkin Exon 8 Probe	PARK2-Exon 8	6q25.2-q27	ACCTGCTCTTCTCC	VIC
Parkin Exon 9-F	PARK2-Exon 9	6q25.2-q27	GGACACACTCCTCTGCACCAT	
Parkin Exon 9-R	PARK2-Exon 9	6q25.2-q27	CAATCTGCTTTTTGGGTTTTGC	
Parkin Exon 9 Probe	PARK2-Exon 9	6q25.2-q27	CTGCTGGTACCGGTTG	VIC
Parkin Exon 11-F	PARK2-Exon 11	6q25.2-q27	GCTCGGCGGCTCTTCA	
Parkin Exon 11-R	PARK2-Exon 11	6q25.2-q27	ACGCCTTTCCTCTTTGTTTCC	
Parkin Exon 11 Probe	PARK2-Exon 11	6q25.2-q27	CGACTCTGTAGGCCTG	VIC
Parkin Exon 12-F	PARK2-Exon 12	6q25.2-q27	CGAACCCACCACACCTTTGT	
Parkin Exon 12-R	PARK2-Exon 12	6q25.2-q27	TGCGGACACTTCATGTGCAT	
Parkin Exon 12 Probe	PARK2-Exon 12	6q25.2-q27	TTCTGCCCCCAACAGGAGGCTG	VIC
<b>Gene: <i>MAML2</i></b>				
ND01496-F	MAML2- Intron 1-2	11q21	CCTTTCCTACCCAGCAAGTT	
ND01496-R	MAML2- Intron 1-2	11q21	GGGCACATGCTTTCAGATCTG	
ND01496-Probe	MAML2- Intron 1-2	11q21	TCTCTGCTTAGCTATCATG	VIC
<b>Gene: <i>ZNF273</i></b>				
ND02214-F	ZNF273-Intron 3-4	7q11.21	GCACCCGACCTGCAATAGAC	
ND02214-R	ZNF273-Intron 3-4	7q11.21	GTGCCCCAGGTAAGCACAAAT	
ND02214-Probe	ZNF273-Intron 3-4	7q11.21	TTTCCTTCTATATGGCTCTTT	VIC
<b>Gene: <i>BCMO1</i></b>				
ND03628-F	BCMO1-Intron 2-3	16q21-23	GTTGGGATCATTGGAAGTATTGG	
ND03628-R	BCMO1-Intron 2-3	16q21-23	TCTCTTAAGCATCTACCAGGTGTCA	
ND03628-Probe	BCMO1-Intron 2-3	16q21-23	CGTTCTAGCAAGGGC	VIC
<b>Gene: <i>ZC3H13</i></b>				
ND04946-F	ZC3H13-Intron 1-2	13q14.12	TGTCCTCTGTTGCCTGATTCG	
ND04946-R	ZC3H13-Intron 1-2	13q14.12	AGCTCTCAGACCAGCAAAGCA	
ND04946-Probe	ZC3H13-Intron 1-2	13q14.12	ACTACTCTCCTCCCTTGAC	VIC

### **3.6. WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 2.**

#### **3.6.1. Subject collection.**

The sample size used for *Stage 1* of this project was increased to a total of 4134 (3581 more than in stage 1, see section 3.5.1), by adding 604 late-onset PD, 108 YOPD and 552 neurologically normal control samples from the NINDS-funded Neurogenetics repository hosted by the Coriell Institute for Medical research (<http://ccr.coriell.org>); 2,243 control samples from the Cancer Genetic Markers of Susceptibility initiative (CGEMS); and 75 YOPD samples collected from different parts of North America by one of our specialists.

**3.6.1.1. Late-onset PD samples:** Besides *NDPT001*, *NDPT005* and *NDPT007*, panels *NDPT017*, *NDPT018* and 250 non-paneled samples from the Coriell institute were included in these experiments, making a total of 880.

All patients were Caucasian individuals with idiopathic Parkinson's disease from the United States. The mean age at onset of parkinsonian syndrome was 62 years, ranging from 41 to 98 years. Age at onset was defined as when symptom(s) of PD were first noted (including at least one: resting tremor, rigidity, bradykinesia, gait disorder, postural instability).

In order for subject inclusion, complete NINDS Repository Clinical Data Elements were required. Furthermore, only subjects who met the UK Brain Bank Criteria idiopathic Parkinson's disease were included (Hughes, Daniel *et al.*, 1992). By those criteria, all

subjects had bradykinesia, and at least one of the following: muscular rigidity, 4-6 Hz resting tremor, postural instability (not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction). None had exclusionary features. All had documentation of sustained, excellent response to anti-parkinsonian therapy. All samples and data were collected with informed consent under local approved protocols.

All subjects were queried regarding family history of parkinsonism, dementia, tremor, gait disorders, and other neurological dysfunction. Subjects both with and without a reported family history of Parkinson's disease were included on this panel. None were included who had three or more relatives with parkinsonism, nor with apparent Mendelian inheritance of PD.

**3.6.1.2. YOPD samples:** Here we define YOPD as when parkinsonian symptom(s) appear before or at age of 40. Following this criterion, only samples included in Coriell panel *NDPT014*, fourteen from panel *NDPT015*, those two erroneously included in *NDPT007* (see *structural variation in PD-related loci* in section 4.5.2.3) and those 75 collected from North America by one of our specialists, were considered as YOPD. The mean age at onset for these samples was 32.7 years, ranging from 7 to 40 years. For those samples from the Coriell institute. In order for subject inclusion, complete NINDS Repository Clinical Data Elements (CDEs) were required. Furthermore, only subjects who met the UK Brain Bank Criteria idiopathic Parkinson's disease were included (Hughes, Daniel *et al.*, 1992). By those criteria, all subjects had bradykinesia, and at least one of the following: muscular rigidity, 4-6 Hz rest tremor, or postural instability (not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction). None had

exclusionary features. All had documentation of sustained, excellent response to dopamine replacement therapy.

**3.6.1.3. Coriell Institute neurologically normal control samples:** Besides those samples included in *NDPT002*, *NDPT006* and *NDPT009*, those samples from Coriell the Institute panels *NDPT019*, *NDPT020*, *NDPT021*, *NDPT022*, *NDPT023* and *NDPT024*, were used for these set of experiments.

Each of these panels contains DNA of 92 unique and unrelated Caucasian individuals free from neurological disorders. All were asked specifically regarding the following disorders, which were not present: Alzheimer's, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, cerebrovascular disease, dementia, dystonia, Parkinson's, and schizophrenia. None had any first-degree relative with a known primary neurological disorder and the mean age of participants was 58 years, ranging from 15 to 98 years.

**3.6.1.4. CGEMS initiative control samples:** The National Cancer Institute-funded Cancer Genetic Markers of Susceptibility (CGEMS, <http://cgems.cancer.gov>) is a three-year, \$14 million initiative aiming to identify genetic alterations that make people susceptible to prostate and breast cancer. For this purpose they have collected a total of 1,177 males with prostate cancer and 1,200 females with breast cancer, as well as 1,101 male and 1,142 female controls. Genotyping data from all these 2,243 control samples was generously shared by the National Cancer Institute and included in our study.

### 3.6.2. Genotyping

Samples within DNA panels *NDPT014*, *NDPT015*, *NDPT016*, *NDPT017*, *NDPT018*, *NDPT019*, *NDPT020*, *NDPT021*, *NDPT022*, *NDPT023*, *NDPT024* and those 252 not-paneled samples; as well as those 75 collected from one the specialists at our laboratory, were assayed with Infinium II HumanHap550 genotyping beadchips version 3 and version 1 respectively. These two different versions of this product enable whole-genome genotyping of 555,363 and 561,467 SNPs respectively derived from the recently completed Phase I and Phase II International HapMap Project (<http://www.hapmap.org>). The Phase I + II HapMap data set contains over 2 million common SNPs (those with a minor allele frequency (MAF)  $\geq 0.05$  in each population studied (Caucasian [CEU], Han Chinese/Japanese [CHB+JPT], and Yoruba [YRI])). To capture this variation, Illumina scientists used an algorithm for the linkage disequilibrium (LD) statistic, “ $r^2$ ”, to select tagSNPs in all HapMap populations (Carlson, Eberle *et al.*, 2004). This was supplemented with additional SNPs to achieve even spacing across genome. On average there is one common SNP (minor allele frequency  $> 0.05$ ) every 5.5, 6.5 and 6.3 kb on the autosomes of the CEU, CHB + JPT and YRI respectively. In addition there are 4,300 SNPs from recently reports on copy number polymorphism regions of the genome, 7,800 non-synonymous SNPs, 1,800 tagSNPs of the Major Histocompatibility Complex, 177 mitochondrial SNPs and 11 Y-chromosome SNPs. Coverage using HumanHap550 is equivalent to that of over one million randomly selected SNP markers.

Those samples previously assayed with Infinium Human-1 and Infinium HumanHap300 SNP chips for *Stage I* of this project (samples within DNA panels *NDPT001*, *NDPT002*, assayed with *NDPT005*, *NDPT006*, *NDPT007* and *NDPT009*) were assayed with

Infinium II HumanHap 240S genotyping beadchips. This array utilizes the Infinium II assay to interrogate an additional 241,846 tagSNPs *loci* derived from the Phase I and Phase II of the International HapMap project and provides tag SNP coverage in regions of lower linkage disequilibrium (LD) in the genome. Combination of this array with the HumanHap300 genotyping bead chip provides the same genotyping information as Infinium II HumanHap550 v.1 beadchips.

Using these genotyping platforms, more than 545,066 unique SNPs were genotyped for each sample of our cohort. Given that most of these were tagSNPs selected from all HapMap populations, it is equivalent to over one million randomly-selected SNPs.

For more details about these genotyping platforms and infinium workflow, see section 3.5.2.

### **3.6.3. Quality-control procedures.**

Besides providing the opportunity to scan the whole genome in a relative short period of time, this approach also has a major problem: the high rate of false positive results. Thus, eliminating any systematic bias like population stratification (existing when the case and control groups are not well-matched genetically or where there are several distinct, but unrecognized, sub-populations in a cohort) or not documented relatedness is required to minimize the rate of false positives.

**3.6.3.1. Low quality genotyping:** Any sample with a call rate below 95% was repeated on a fresh DNA aliquot and if the call rate persisted below this level, the sample was excluded from the analysis. Low-quality genotyping led us to repeat 57 individual

samples, of which 40 were ultimately excluded from the analysis, including 11 late-onset PD, 5 YOPD, 13 neurologically normal control samples from the Coriell Institute and 12 controls from CGEMS initiative.

**3.6.3.2. Gender problems:** PLINK v1.01 (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (Purcell, Neale *et al.*, 2007) was utilized to flag individuals whose reported sex did not match with the estimated sex based on heterozygosity on chromosome X genotypes (inbreeding coefficient [F] in this chromosome). A male call is made if F is more than 0.8 and a female call if F is less than 0.2. A problem arose if two sexes did not match or if the SNP data and pedigree data are ambiguous with regard to the sex. All problematic samples were analyzed by visual examination of log R ratio and B allele frequency metrics with the IGV tool within BeadStudio v3.1.12 (Illumina, <http://www.illumina.com>) to rule out whether this discrepancy was caused because of copy number variation or extended homozygosity in chromosome X. These analyses led to the exclusion of 15 samples, including 9 late-onset PD, 2 YOPD and 4 neurologically normal control samples from the Coriell Institute.

**3.6.3.3. Population substructure:** In an attempt to detect the presence of significant population substructure or ethnically mismatched individuals, 10,000 random SNPs in linkage equilibrium with each other were selected. These SNPs were selected from a list of ~50,000 obtained after LD-pruning with PLINK v1.01 (<http://pngu.mgh.harvard.edu/~purcell/plink/>).

Subsequent analysis of these 10,000 SNPs in all genotyped individuals along with 30 trios from Yoruba (Nigeria), 45 unrelated individual from the Tokyo area in Japan, 45 unrelated individuals from Beijing (China) and US-resident trios with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH, Paris, France); all downloaded from the HapMap website (<http://www.hapmap.org>), allowed inference of the presence of individuals belonging to a distinct population. This approach was carried out using STRUCTURE v2.2 software (<http://pritch.bsd.uchicago.edu/structure.html>).

These results were confirmed by plotting a matrix of pairwise Identity By State (IBS) distances created with PLINK v1.01 (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (Purcell, Neale *et al.*, 2007). Thus, IBS distance to its “nearest neighbor” was calculated for each individual. This distribution was standardized (by the sample mean and variance of nearest neighbor) and inspected for outliers. For this last purpose Multidimensional scaling (MDS) was performed. This scaling is a set of techniques often used for data visualization for exploring similarities or dissimilarities in data.

**3.6.3.4. Non-reported relatedness:** The pairwise clustering based on IBS distances (see previous section) is useful for making estimations of pairwise Identity by Descent (IBD) to find pairs of individuals who look more similar than expected by chance in a random sample. By estimating the probability of sharing 0, 1, or 2 alleles IBD for any two individuals, a proportion of IBD can be calculated ( $PI-HAT = P [IBD = 2] + 0.5 \times P [IBD = 1]$ ).

Using 0.2 as a threshold for PI-HAT, 17 sample pairs were considered too similar to each other. Thus, one member of each pair was removed from further SNP association tests. These were 6 late-onset PD, 5 YOPD and 6 neurologically normal control samples from the Coriell Institute.

Besides, PI-HAT data revealed 50 replicates within our dataset including 9 late-onset PD, 40 YOPD and 1 neurologically normal control sample from the Coriell Institute. These were removed from further analyses.

#### **3.6.4. Statistical analysis.**

Power calculations were performed using the program PS v2.1.30 (Dupont and Plummer, 1990). Based on the average observed minor allele frequency (26%) and assuming that either the causal variant is typed or that there is complete and efficient tagging of common variation by the genotypes tagSNPs, our data provides 80% power to detect an allelic association with an odds ratio of more than 1.38 and less than 0.69 at an uncorrected significance level of  $p = 0.000001$ .

All estimates and tests were performed with the PLINK v1.01 toolset, developed by Shaun Purcell and co-workers at the Center for Human Genetic Research, Massachusetts General Hospital, and the Broad Institute of Harvard and Massachusetts Institute of Technology (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (Purcell, Neale *et al.*, 2007).

Only those SNPs genotyped in at least 95% of our control cohort as well as those with a minimum allele frequency (MAF) above 5% and with no extreme departure from Hardy-Weinberg equilibrium (HWE) in controls ( $p > 0.01$ ) were included in further statistical analysis.

For each SNPs that successfully passed these three filters (474,995 SNPs), five tests were applied including the basic allele test, Cochran-Armitage trend test, genotypic test (two degrees of freedom), dominant gene action test (one degree of freedom) and recessive gene action test (one degree of freedom). We calculated odds ratios (ORs), 95% CIs, and p values for each of the association models. All calculation were performed with plink v1.01 software.

For multiple test correction, 1000 max T permutations were performed. For each permutation, PLINK v.1.01 permutes the affection status (case or control) of the entire sample represented in the input file while preserving the total number of cases and total number of controls in each permutation. The difference between an adaptive permutation analyses, is that max T doesn't drop early clearly nonsignificant SNPs. Thus, if 1000 permutations are specified, then all 1000 will be performed for all SNPs. The benefit of doing this is that two sets of empirical significance values can then be calculated and also a value that controls for that fact that thousands of other SNPs were tested. This is achieved by comparing each observed test statistic against the maximum of all permuted statistics (i.e. over all SNPs) for each single replicate.

Multiple test correction based on Bonferroni and FDR-BH methods (based on Bonferroni) were also performed.

### **3.6.5. Combined analysis.**

In order to add power to our genome-wide association study and further confirm the results we obtained, we performed a collaborative analysis with Drs. Thomas Gasser (Hartie-Institute for Clinical Brain Research, Tuebingen, Germany) and Tatsushi Toda

(Osaka University Graduate School of Medicine, Suita, Japan) who have performed similar approaches in a German and a Japanese cohort respectively.

**3.6.5.1. German cohort:** This cohort consisted of a total of 1,686 samples including 742 PD cases and 944 controls from Germany. All samples with a genotyping success below 95%, gender mismatch and with a standard deviation of  $\pm 4$  from the mean heterozygosity value were excluded. After IBS distance clustering, samples with a non-European ancestry were also removed.

Since only genotypic counts for each SNP were shared with us, population stratification analysis could not be assessed between our and the German cohort. For further analysis, genotypic counts for all the SNPs that successfully passed the HWE and MAF filters in the German cohort (498,560) were merged with that of our successful SNPs 474,995 giving a total of 465,684 successful SNPs common to both cohorts. For this purpose it was assumed that the minor allele in both populations are the same. Besides, since gender can not be accounted if using genotype counts only, results for non-autosomal SNPs were removed.

All tests and corrections used for our cohort (see section 3.6.4) were applied for each of these SNPs.

**3.6.5.2. Japanese cohort:** This cohort consisted on 1,000 PD cases and 2,500 neurologically normal controls from Japan. All samples with a genotyping success below 95%, gender mismatch and with a standard deviation of  $\pm 4$  from the mean heterozygosity value were excluded. After IBS distance clustering, samples with a non-

Japanese ancestry were also removed. After HWE and MAF filtering, 398,416 successful SNPs were left for analysis. Given the existing differences in the genetic background of Asian and Caucasian populations genetic counts were not merged for analysis. Thus, this population was used for side by side comparison of results.

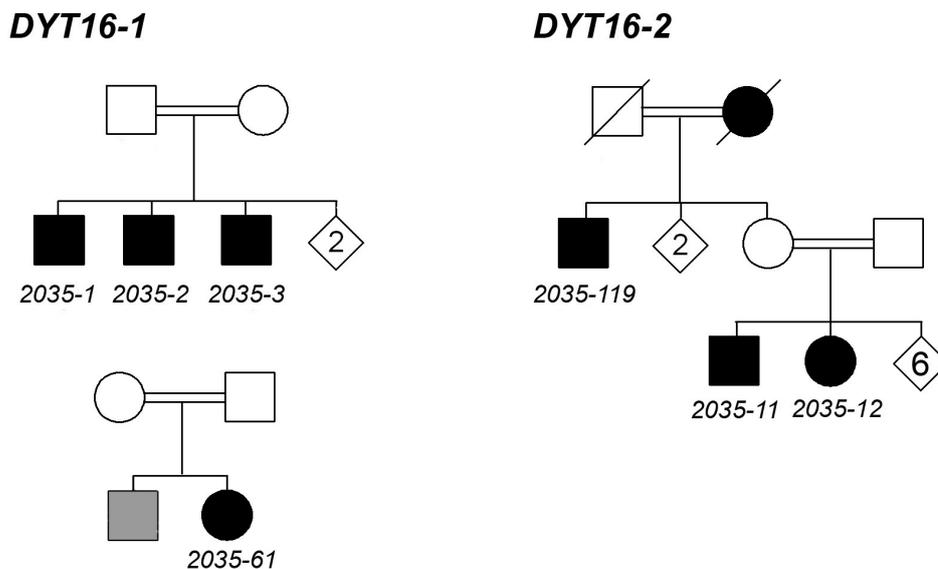
### **3.7. AUTOZYGOSITY MAPPING IN BRAZILIAN FAMILIES**

#### **3.7.1. Subject collection.**

All subjects gave informed consent and the study was approved by a local ethics board. Families *DYT16-1* and *DYT16-2* (figure 14) were identified by Drs. Sarah Camargos and Francisco Cardoso at the Movement Disorder Clinic at the Federal University of Minas Gerais, Brazil. These families shared singular dystonic features (table 21) and both stated parental consanguinity. The family members were questioned regarding the presence of additional affected members outside of the nuclear family; however no additional affected family members were reported and building of a family tree failed to link these two families genealogically. The inheritance pattern observed in these families was consistent with an autosomal recessive mode of inheritance. Thus, there is a high possibility that affected individuals had inherited both copies (paternal and maternal) of the mutated gene from a common ancestor. As a consequence of this principle (Lander and Botstein, 1987) the chromosomal region surrounding the mutated gene may be expected to be homozygous and identical by descent (IBD) among affected individuals (autozygous). In order to identify the disease-causing gene, these individuals were

selected for autozygosity mapping, in which identical by state (IBS) homozygous segments between affected individuals are searched.

In addition, Sarah Camargos and Francisco Cardoso, collected DNA samples from 45 young-onset PD patients (age at onset ranging from 18 to 40 years), 12 apparently unrelated young-onset dystonia patients, including individual 2035-61 (age at onset ranging from 8 to 42 years), and 83 neurologically normal controls (age at sampling ranging from 25 to 50 years). All these additional samples belonged to Brazilian individuals collected at the Movement Disorder Clinic at the Federal University of Minas Gerais, Brazil.



**Figure 14: Schematic representation of *DYT16-1* and *DYT16-2* pedigrees.** A black symbol indicates an affected family member, an open symbol indicates an unaffected family member, and a grey symbol indicates a family member reported as affected by history but unexamined. Squares are males, circles are females. Multiple siblings are denoted by a diamond; the number of individuals is indicated in the diamond. Individual 2035-61 does not belong to either of these families (see results section 4.7.3).

**Table 21:** Summary of clinical features of affected members of families *DYT16-1* and *DYT16-2* as well as individual 2035-61.

Patient	Age at onset	First symptom	Generalized dystonia	Fahn Marsden Scale	Parkinsonism	UPDRS motor score	Pyramidal signs
2035-1	11	Lower limbs dystonia	Severe	96	Absent	0	No
2035-2	12	Lower limbs dystonia	moderate	58	Present	21	Yes
2035-3	2	Lower limbs dystonia	moderate	36	Absent	0	Yes
2035-11	11	Upper limbs dystonia	Severe	64	Present	21	No
2035-12	2	Spasmodic dystonia	Slight	14	Present	30	No
2035-119	18	Lower limbs dystonia	Moderate	44	Present	18	No
2035-61	7	Upper limbs dystonia	Severe	103	Absent	19	Yes

UPDRS: Unified Parkinson's disease rating scale.

A detailed description of patients in which mutations in *PRKRA* were described (see results section 4.7) can be found below. Brain CT scans of all affected family members of *DYT16-1* and *DYT16-2* and brain MRI scans on family members 2035-1 and 2035-11 of *DYT16-1* and *DYT16-2*, respectively, showed no specific abnormalities are shown in figure 15. No abnormalities were seen.

### 3.7.1.1. Family DYT16-1:

. **Individual 2035-1:** This patient is currently 32 years-old. Until 11 years of age he was described as neurologically normal, when he developed a gait disturbance and pain in the lower limbs. After two years, the symptoms spread causing a swallowing disturbance. He was placed on medication (biperiden up to 24 mg a day and levodopa/carbidopa 125/12.5 mg three times a day) without significant improvement. Physical exam showed mildly high blood pressure. Neurological exam showed a marked generalized dystonia with severe retrocollis and opisthotonic posture. Currently the patient scores 96 on the Burk-Fahn-Marsden Scale with a disability score of 26 (Burke, Fahn *et al.*, 1985). Botulinum toxin injection of the muscles of the neck (*splenii capitis*)

was attempted with no response. There were no cortico-spinal tracts signs or parkinsonism.

. **Individual 2035-2:** Currently aged 35 years, this patient had an uneventful early childhood although he did not start to speak until the age of 2 years. A movement disorder began at the age of 12 with pain in the lower limbs and involuntary leg adduction when walking. Shortly thereafter the voice became spasmodic and he developed oromandibular dystonia, laterocollis and upper limb dystonia. Reflexes were brisk with ankle clonus. Bradykinesia was marked with a UPDRS part III score of 21 (Fahn, Elton *et al.*, 1987). The patient scored 58 on the Burke-Fahn-Marsden Scale with a disability score of 23. His generalized dystonia was mainly characterized by involvement of legs, hands and voice. The outcome with anticholinergic drugs (biperiden, 24mg per day) and levodopa/carbidopa (375/37.5 mg per day) treatment was poor although the bradykinesia had mild improvement with levodopa.

. **Individual 2035-3:** This patient is currently 19 years old. Despite normal pregnancy and normal delivery, this patient exhibited delay of developmental motor and cognitive milestones. He started to walk at age 2 years when relatives noted feet inversion and knee flexion, suggestive of spasticity. He has never had intelligible language and speech is apparently affected by a spasmodic dysphonia. Lower limb dystonia spread to superior limbs and orofacial muscles and trunk. At age 14 he had an appropriate behavior when seen at the UFMG Movement Disorders Clinic despite the mention of occasional outbursts of aggressiveness by his relatives. The neurological

examination showed generalized dystonia characterized by involvement of legs, hands, trunk and voice with a Burk-Fahn-Marsden scale score of 36 on the objective section and 16 on the disability part. There was severe bradykinesia but no tremor. His deep reflexes were brisk. Use of levodopa/carbidopa (375/37.5 mg per day, six months) and biperiden (24 mg per day) did not benefit the dystonia but there was moderate improvement of the bradykinesia without significant UPDRS change.

### **3.7.1.2. Family DYT16-2:**

. **Individual 2035-119:** This patient is 64 years old and has only been followed up at the UFMG Movement Disorders Clinic for the past year. He has no recollection of any neurological symptom until the age of 18, when he noticed an involuntary contraction of the left leg during walking. The symptoms progressed slowly, involving the right leg and the left arm. Ten years later, the patient developed a right laterocollis and a dysarthric speech. The trunk at this time also developed contractions in the same direction of the neck. Neurological examination showed right laterocollis, right laterotrunk, spasmodic dysphonia, extension of the left leg when walking, slight bradykinesia but normal tendon reflexes. He has been treated with anticholinergic drugs (biperiden 12mg per day) and baclofen (30 mg per day) with modest, if any, improvement.

. **Individual 2035-11:** This patient is 34 years old and has been followed for the past 14 years. The mother reported no problems during pregnancy or delivery. There was, however, a slight delay in the developmental milestones (he started to walk at

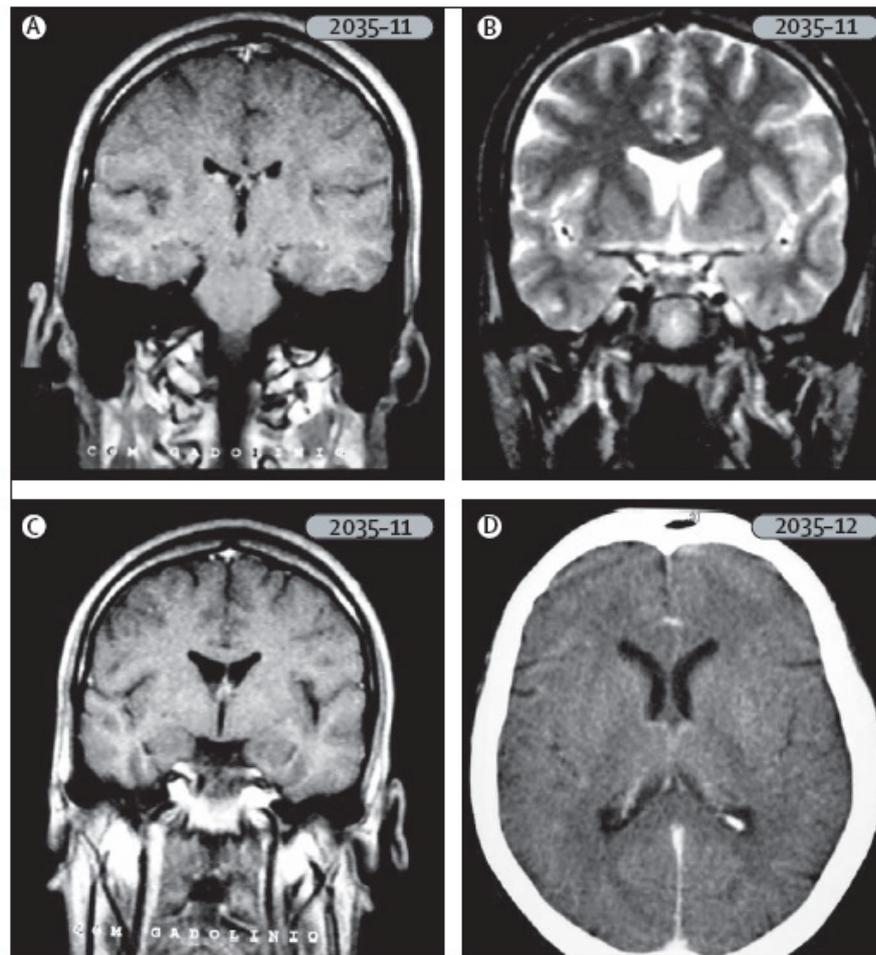
age 18 months and did not speak before 5 years of age). At the age of 11 he noticed an involuntarily closing of the left hand and he needed to use his right hand to open the left. In the same year the movement disorder deteriorated, with involuntary pushing of the left shoulder backwards and downwards. At age of 13 years, the right shoulder was also involved and there was hoarseness of the voice. Three years later, the patient reported that his neck moved backwards involuntarily. The symptoms ameliorated when he touched his head. At this time, aged 20 years, he began to display impaired walking with right foot inversion and academic problems; at this point he sought medical assistance. Neurological exam at that time showed no Kayser-Fleischer ring, a brisk mentonian reflex and facial grimacing. The voice was dysarthric, with tongue dystonia. There was retrocollis and a laterocollis with contraction of both splenius capitis and scapular elevator. There was protrusion of left shoulder, hyperextension of left elbow, hyperflexion of the left wrist and fingers and hyperextension of the right wrist. The gait was dystonic with inversion of the right foot. Deep tendon reflexes, cerebellar tests and sensitivity were normal. There was no response to levodopa/carbidopa (375/37.5 mg per day, 6 months). Baclofen (60mg per day, 4 months) and anticholinergic drugs (biperiden, 24 mg per day until the date) also failed to provide meaningful benefit. In 1993 the patient underwent botulinum toxin injection with excellent improvement of laterocollis and retrocollis. Despite this, the dystonia of inferior limbs and voice worsened. One year later the patient developed opisthotonic posturing and writer's cramp. At age 29 there was deterioration of the movement disorders rendering the patient unable to walk, dress and feed without assistance. At this time, the patient lost weight and there was worsening of the arm and trunk dystonia. When the patient was 31-years-old speech was

unintelligible most of the time and there was moderate neck pain caused by the dystonia. He continued to receive periodic injections of botulinum toxin but increasing dosages (maximum of 600U units of Botox per session) were necessary to alleviate the retrocollis. Even during the peak of the effect, although the neck contractions improved at rest, there was intense overflow with return of the dystonic posturing of the neck during the gait. The only side effect was worsening of the dysphagia. The Burke-Fahn-Marsden scale rating during botulinum toxin effect score was 64 on the objective part and 21 on the disability section.

. **Individual 2035-12:** This patient is currently 48 years old and has been followed for 3 years. She was born by natural delivery, following an uneventful pregnancy. Motor development was normal but her mother mentioned that she started to speak just at age 2 years and that her speech has always been difficult to understand. At the age 10 years, her mother noticed slowness of the hands although the patient became aware of this symptom as well as of tremor in the hands only when she turned 20. One year later she noticed involvement of the right leg. Since then there has been a slow worsening of the movement disorder. Currently the patient mentions freezing of her gait and writing as well as dysphagia. Neurological examination at age 45 years showed facial grimacing, severe hypomimia, and a slight oromandibular dystonia. The voice was characterized by a combination of lower volume and strangled character suggestive of a mixed form of spasmodic dysphonia. The patient also had dystonic posturing in the hands at rest (flexion of the fingers) and when writing (flexion of the wrist) as well as left foot eversion when walking. There was bradykinesia, slight postural tremor of the arms and

rigidity in the upper and lower limbs. There was no abnormality of deep tendon reflexes, cerebellar or sensory functions. 5 mg per day, six months) and biperiden (24 mg per day) did not benefit the dystonia but there was moderate improvement of the bradykinesia without significant UPDRS change.

**3.7.1.3. Individual 2035-61:** This 42 year-old woman has been followed up for 12 years at the UFMG Movement Disorders Clinic. The symptoms began at the age of 7 years with difficulty in writing. Five years later she developed gait disturbance (extension of the left leg) followed by rapid spread to upper limbs, trunk, face and voice. There is a report of one of her brothers having had similar symptoms with onset at age 8. Unfortunately he died at the age of 34 years from pneumonia before we could examine him. At the time of his death he was severely disabled and wheelchair-bound. The parents were cousins and died from cardiac problems. Neurological examination of our patient at age 30 years showed generalized dystonia, with facial grimacing, protrusion of tongue and lips, anarthria, left scoliosis, extension of both wrists with flexion of fingers, flexion of knees with feet inversion. The deep tendon reflexes were brisk but there were no parkinsonian features. The Burke-Fahn-Marsden scale score was 103 with disability of 19. Levodopa (375/37.5mg per day), baclofen (30mg per day) and trihexyphenidyl (20mg daily) resulted in no improvement. Botulinum toxin injection in the submental area has consistently improved the dysphagia (brief summary in table 21).



**Figure 15: Brain MRI and CT scans for affected individuals of families DYT16-1 and DYT16-2.** (A, B) Coronal T1-weighted MRI after gadolinium contrast injection in patient 2035-11. (C) Coronal T2-weighted MRI in patient 2035-11. (D) CT scan of patient 2035-12. No specific abnormalities can be appreciated.

### 3.7.2. Exclusion of known genes

Affected family members from both families (*DYT16-1* and *DYT16-2*) were screened for missense mutations in *PRKN* (linked to young-onset Parkinson's disease; *PARK2*), *ATPIA3* (linked to rapid-onset dystonia parkinsonism; *DYT12*) and *GCHI* (linked to dopa responsive dystonia; *DYT5*) by direct sequencing of exons and flanking intronic sequence (Ichinose, Ohye *et al.*, 1994; Kitada, Asakawa *et al.*, 1998; De Carvalho

Aguiar, Sweadner *et al.*, 2004). The *TOR1A*  $\Delta$ GAG mutation (linked to primary torsion dystonia; *DYT1*) was screened in both families by sequencing of the affected exon (Ozelius, Hewett *et al.*, 1997).

In each PCR reaction, 20 ng of genomic DNA was amplified with 5 pmol of each primer (primers sequence in table 22) and 2U of FastStart TaqDNA polimerase (Roche Diagnostics Corporation, <http://www.roche.com>). The final volume was of 15  $\mu$ l. The thermal cycling program used was a 60td50 program (table 4).

The PCR products were verified in a 2% ethidium-bromide-containing agarose-TBE gel and visualized via ultraviolet transillumination. Prior to loading each PCR product on a gel, it was mixed with 6X Orange G Loading Dye Solution.

After verification, PCR products were purified with MultiScreen 96-Well Filter Plates (Millipore, <http://www.millipore.com>) as explained in section 3.1.2.

Each purified product was sequenced using forward or reverse primers with Applied Biosystems BigDye terminator v3.1 sequencing chemistry (for more details see section 3.1.2). Sequencing products were then purified with MultiScreen 96-Well Filter Plates (section 3.1.2) and electrophoresed in an ABI3730 XL genetic analyzer (Applied Biosystems, <http://www.appliedbiosystems.com>). Sequencher v4.1.4 software (Gene Codes, <http://www.genecodes.com>) was used to analyze results. All changes that derived from the wild type sequence were verified by PCR amplification of a fresh DNA aliquot and sequencing in both forward and reverse directions.

Copy number mutations in *PRKN* were screened by means of real-time PCR as described in section 3.5.5.

An affected family member of DYT16-2 had previously been screened for missense mutations in *LRRK2*, *PINK1* and *SNCA* in addition to whole gene multiplication mutation of *SNCA* as previously described (Johnson, Hague *et al.*, 2004).

**Table 22:** Forward and reverse primers sequences utilized for exclusion of *ATPIA3*, *TORIA*, *GCHI* and *PRKN* as candidates to cause disease in families *DYT16-1* and *DYT16-2*.

Exon	Forward primer (5'-3')	Reverse primer (5'-3')
<b>Gene: <i>ATPIA3</i></b>		
Exon 1	TATTGCGGAGGTCTCTGGG	GACCACATGGATTGGCTGG
Exon 2	GCATGGAGGGCCGGTGTC	GACATCTTGTGCTCTGTCTG
Exon 3	GAGGTGGCTATGGTAAGCC	CCTTCACCAGACCCCCAG
Exon 4	TTCTGGGGGTCTGGTGAAG	CTCTTGCTGGGTGATTGTGA
Exon 5	TCAGACACACCAACCCTCAT	CGTTCACCTGCATCTTCTCA
Exon 6	AGGAGGCCAAGAGCTCCAAG	GGGGGCCTGGACTCCTG
Exon 7	TTCTAGCTGTGATCTCCAGG	GATGCCAGGGTGGCGATAC
Exon 8	CCACCAACTGTGTGGAAGG	TCCAGGCCTCTAGCCCCT
Exon 9	CTCATTCTTTCCAGGCCCT	GCCACCCAGGTGTGCGAA
Exon 10	AGAACCCGCATGACAGTCGC	GGTTGTGAGAACAGGGAC
Exon 11	CGCTGGGCTCTGCAATCG	ATTATCATTCCCATTTTACAGAC
Exon 12	AGGTCTCTGTGAGAGAGTGT	AAGACGGCCAGTCAGCATT
Exon 13	GAATGCTGACTGGCCGTCTT	CGTTGCCCTCAGAGATGATG
Exon 14	ACTTCACCACGGACAACCTC	TCCCAGAAAGAATGGGACAG
Exon 15	GTCTCTGGCTCTCTCGG	TCAGACCCAGGGGTCCAG
Exon 16	GTCTGAGGGAGGAAGGTCT	CAGGGGAGGAAGACAGAAAA
Exon 17	AAAAAAAAAAGGACGTTGGATGA	CTGTCTTTCATGATGTGCG
Exon 18	CTCTGCATCGATCTGGGC	GTGCAGACCCCCCCCAC
Exon 19	AGGGCCAGGAGGGCATA	GGGCCATCGTAGGAAGTGG
Exon 20	AGGCCCTGAGGACCCCAT	GAGCTATGATTACACCCTGA
Exon 21	GGGTGTCTGCACTGTGCC	ATGGGGACTGCAGTGGGG
Exon 22	CACCTGTGAGCACGAAGGA	ACCCGTGAGAAGACAGAGTG
<b>Gene: <i>TORIA</i></b>		
Exon5	AATGTGTATCCGAGTGAAATG	TGCCAATCATGACTGTCAATC
<b>Gene: <i>GCHI</i></b>		
Exon1	GTTTAGCCGCAGACCTCGAAGCG	GAGGCAACTCCGAAACTTCTG
Exon 2	GTAACGCTCGTTATGTTGACTGTC	ACCTGAGATATCAGCAATTGGCAGC
Exon 3	AGATGTTTTCAAGGTAATACATTGTCG	CTCGCAAGTTTGCTGAGGC
Exon 4	GTCCTTTTGTGTTTATGAGGAAGGC	GGTGTGCACTTATAATCTCAGC
Exon 5	GTGTCAGACTCTCAAAGTACGCTC	TCACTTCTAGTGCACCATTATGACG
Exon 6	ACCAAACCAGCAGCTGTACTCC	AATGCTACTGGCAGTACGATCGG
<b>Gene: <i>PRKN</i></b>		
Exon 1	GCGCGGCTGGCGCCGCTGCGCGCA	GCGGCGCAGAGAGGCTGTAC
Exon 2	ATGTTGCTATCACCATTTAAGGG	AGATTGGCAGCGCAGGCGGCATG
Exon 3	ACATGTCACTTTTGCTTCCCT	AGGCCATGCTCCATGCAGACTGC
Exon 4	ACAAGCTTTTAAAGAGTTTCTTGT	AGGCAATGTGTTAGTACACA
Exon 5	ACATGTCTTAAGGAGTACATTT	TCTCTAATTTCTGGCAAACAGTG
Exon 6	AGAGATTGTTTACTGTGAAACA	GAGTGATGCTATTTTTAGATCCT
Exon 7	TGCCTTTCCACACTGACAGGFACT	TCTGTTCTTATTAGCATTAGAGA
Exon 8	TGATAGTCATAACTGTGTGTAAG	ACTGTCTCATTAGCGTCTATCTT
Exon 9	GGGTGAAATTTGCAGTCAGT	AATATAATCCCAGCCCATGTGCA
Exon 10	ATTGCCAAATGCAACCTMTGTC	TTGGAGGAATGAGTAGGGCATT
Exon 11	ACAGGGAACATAAACTCTGATCC	CAACACACCAGGCACCTTCAGA
Exon 12	GTTTGGGAATGCGTGTITT	AGAATTAGAAAATGAAGGTAGACA

### 3.7.3. Genome-wide SNP genotyping.

Based on number of affected family members and DNA availability, families *DYT16-1* and *DYT16-2* were selected from a collection of 58 Brazilian young-onset PD and generalized dystonia cases for autozygosity mapping. Thus, DNA from all affected available members of these families was genotyped with Illumina's HumanHap 550 genotyping chips (version 1) as previously described in this work (sections 3.5.2 and 3.6.2). Data was analyzed using BeadStudio v3 (Illumina, <http://www.illumina.com>).

### 3.7.4. Autozygosity mapping.

As previously explained (section 3.5.5), genotypic data generated by Infinium technology has numerous applications. The profile indicative of the existence of a homozygous track is lack of heterozygosity in the B allele frequency plot with no change at that same region of the log R ratio plot. In kindreds with an apparently recessive disorder, particularly in those where consanguinity is suspected, this approach can be used to score and compare extended regions of homozygosity across family members. Whether this approach will lead to the identification of single or multiple regions of interest and the size of these, relies on several factors: the degree of parental consanguinity, the number of informative family members and the relative stochastic nature of recombination.

Homozygous tracks for each sample were scored using an in-house tool developed to help visualization of contiguous tracks of homozygosity (tracker v0.99, available at <http://www.neurogenetics.org/XXXX>).

Following the identification of all homozygous tracks within affected individuals, genotypes were exported from BeadStudio v3.1 and compared for identity by state across

all six affected family members (i.e. stretches of identical genotype calls across all affected persons from both apparently unrelated families).

All regions in the genome that were homozygous at greater than 50 contiguous SNPs and identical by state between family members both within and between families were considered as candidate disease gene regions.

### 3.7.5. Sequencing of genes within critical intervals.

All genes and transcripts included in candidate regions flanked by markers rs1434087 to rs10497541, rs13405069 to rs7581560 and rs1518709 to rs10930936, were identified using the Map Viewer tool (Build 36.2) within the National Center for Biotechnology Information (NCBI) website (<http://ncbi/mapview/>) and the ENSEMBL (based on the NCBI 36 assembly) data set ([http://www.ensembl.org/Homo\\_sapiens/index.html](http://www.ensembl.org/Homo_sapiens/index.html)).

Primers were designed to allow amplification and sequencing of all coding exons and at least 30bp of flanking intronic sequence with the online tool ExonPrimer (<http://ihg.gsf.de/ihg/ExonPrimer.html>). Thus, the coding exons of *OSBPL6*, *PRKRA*, *DFNB59*, *FKBP7*, *PLEKHA3*, *TTN*, *FLJ39502*, *SESTD1*, *LOC728984*, *LOC644776*, *ZNF533* and *LOC72900* were sequenced (Supplementary material table S1).

### 3.7.6. Assay for the c.665C>T (P222L) mutation.

We analyzed this mutation by direct sequencing of exon 7 of *PRKRA*. This exon was amplified and sequenced with primer pair x7F-*PRKRA* (5'-AATGTTGTCTTGTTTAAATTG-3') and x7R-*PRKRA* (5'-TACTATCCACAAGAATGGG-3'). The PCR amplification was carried out in a final

volume of 15µl containing 10ng of genomic DNA, 10pmol of forward and reverse primers and 10µl of FastStart PCR Master mix (Roche Diagnostics Corporation, <http://www.roche.com>). The thermal cycling program used was a 57td50 program (table 5).

The PCR products were verified in a 2% ethidium-bromide-containing agarose-TBE gel and visualized via ultraviolet transillumination. Prior to loading each PCR product on a gel, it was mixed with 6X Orange G Loading Dye Solution.

After verification, PCR products were purified with MultiScreen 96-Well Filter Plates (Millipore, <http://www.millipore.com>) following the protocol displayed in section 3.1.2.

Each purified product was sequenced using forward or reverse primers with BigDye terminator v3.1 chemistry and purified with MultiScreen 96-Well Filter Plates (see section 3.1.2 for details).

Purified sequencing reactions were then electrophoresed on an ABI3730 XL genetic analyzer (Applied Biosystems, <http://www.appliedbiosystems.com>) and analyzed with Sequencher software v4.1.4 (Gene Codes, <http://www.genecodes.com>). All changes that deviated from the wild type sequence were verified by PCR amplification of a fresh DNA aliquot and sequencing in both forward and reverse directions.

This exon was thus sequenced in all samples collected by Drs. Sarah Camargos and Francisco Cardoso (all available *DYT16-1* and *DYT16-2* family members as well as those 45 young-onset PD patients, 12 apparently unrelated young-onset dystonia patients, and 83 Brazilian neurologically normal controls), 439 North American Caucasian neurologically normal controls from panels *NDPT002*, *NDPT006*, *NDPT009*, *NDPT022*, *NDPT023* and *NDPT024* as well as 249 North American Caucasian patients with young-

onset PD from panels *NDPT014*, *NDPT015* and *NDPT016* from the Coriell Institute repository (see sections 3.5.1 and 3.6.1 for details), 426 neurologically normal controls from Portugal collected by Jose Braas and Rita Guerreiro, and 738 samples from the Human Genome Diversity Project DNA panels that includes samples representing diverse populations worldwide, including 44 Brazilian subjects (<http://www.cephb.fr/HGDP-CEPH-Panel>). In addition all coding exons of *PRKRA* were sequenced in all young-onset dystonia patients as described above (section 3.7.5).

## 4. RESULTS

### 4.1. GENETIC ANALYSIS OF PARK8-LINKED PARKINSON'S DISEASE PATIENTS.

#### 4.1.1. Minimal haplotype shared by UGM03, UGM04, UGM05 and UGM06.

A rare variant segregating with the disease in UGM03, UGM04 and UGM06 (Basque allele frequency of 0.01) indicated that these four kindreds might be ancestrally related (Paisan-Ruiz, Jain *et al.*, 2004). In an effort to examine the existence of a minimal interkindred disease haplotype, 23 SNPs contained in the region with strongest linkage to disease (D12S331 to D12S1668) were sequenced in all affected members from the four families. This analysis showed a shared haplotype and that rs10876410 delimited the critical interval on the p arm in families UGM03, UGM05 and UGM06. Interestingly, UGM04 also shared a haplotype limiting the minimal interkindred disease haplotype at rs4548690 in the p arm, narrowing down the critical region to a 2.6 Mb region flanked by this SNP and D12S1653 (figure 16). This region contained only 11 genes and predicted transcripts. Systematic sequence analysis of these 11 genes (experiments performed by Dr. Coro Paisan-Ruiz) showed segregating mutations within a putative kinase domain containing transcript, DKFZp434H2111, which contains a 7,449 open reading frame encoding a 2,482 amino acid protein that includes a leucine-rich repeat, a kinase domain, a RAS domain and a WD40 domain. The change identified was a p.R1441G, not identified in 1300 chromosomes from North American controls and 160 chromosome from the Basque country.

This gene was later given the name of *LRRK2* and the encoding protein dardarin.

Dardarin comes from the Basque word *dardara*, meaning tremor.

IDENTIFIER	bp	CONSENSUS	BASQUE FAMILY ID			
			UGM3	UGM4	UGM5	UGM6
D12S1698	30855986	-	122	126	118	124
D12S1621	31754700	-	191	191	191	191
rs1523118	37515966	-	T	T	C/T <sup>a</sup>	T
D12S331	37547321	-	177	177	177	177
rs11169992	37603474	-	C		C	C
rs10876410	37708557	-	T	A/T <sup>a</sup>	T	A
rs10876646	37887093	-	T	T	T	T
rs10747736	37912177	-	T	C	T	T
rs10747736	37912177	-	T		T	T
rs10876876 <sup>b</sup>	38011263	-	A	-	A	A
rs11171789 <sup>b</sup>	38024258	-	T	C	T	T
K543R <sup>b</sup>		-	G	A	G	G
rs10876886 <sup>b</sup>	38035530	-	C	A/C <sup>a</sup>	C	C
rs11172282	38161804	-	G	C/G <sup>a</sup>	G	G
rs11172541	38229025	-	A	-	A	A
rs10877201	38298564	-	C	T/C <sup>a</sup>	C	C
rs4548690	38475137	-	T	C	T	T
rs7294916	38494630	T	T	T	T	T
rs4423249	38554997	T	T	T	T	T
rs515205	38689229	A	A	A/G <sup>a</sup>	A	A
rs937110	38815159	C	C	C/G <sup>a</sup>	C	C
SNP1	38815163	T	T	T	T	T
rs4768224	38947670	T	T	T	T	T
IVS13-54 <sup>c</sup>	38943803	C	C/G <sup>a</sup>	C	C	C
IVS13+68 <sup>c</sup>	38944038	G	G/A <sup>a</sup>	G	G	G
<b>R1396G<sup>c</sup></b>	<b>38990503</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>
M1601T <sup>c</sup>	39011294	G	G	G	G	G
rs12423567	39063583	G	G	-	-	G
rs12423567	39063583	G	G	G	G	G
rs10784616	39117987	C	C	C	C	C
rs11612876	39256712	T	T	-	T	T/C <sup>a</sup>
rs10784800	39386364	C	C	C	C	C
rs10879192	39471471	C	C	C/T <sup>a</sup>	C	C
D12S1668	39489795	235	235	235	235	235
D12S1653	41093561	-	215	215	203	215

Figure 16: Markers used to fine-map the candidate interval and determine interfamily shared haplotype and the boundaries of this haplotype. The black outline indicates the extent of the haplotype common between each Basque family. <sup>a</sup> Phase not determined. <sup>b</sup> Within *KIF21A*. <sup>c</sup> Within *LRRK2*.

#### 4.1.2. A founder effect of p.R1441G in Basque population.

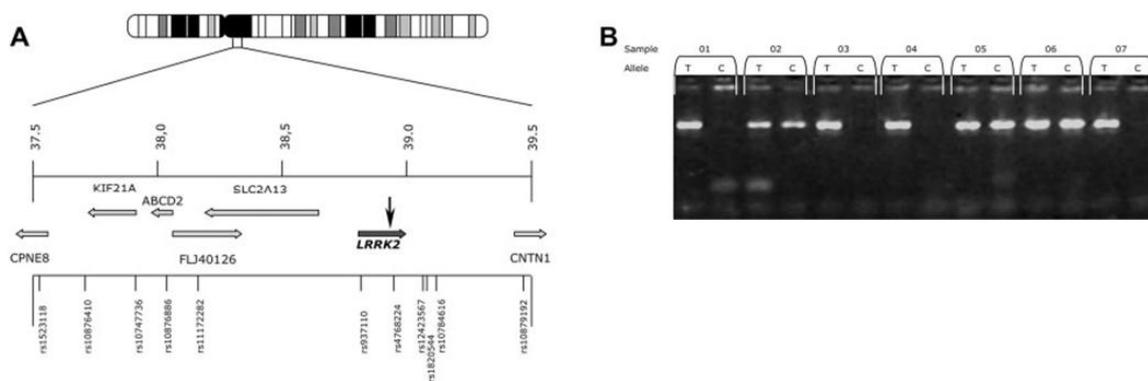
With our allele-specific PCR assay (figure 17), we were able to identify 17 patients with this mutation, 10 of which were considered as sporadic (~7% of the 238 analyzed). The

relatively high frequency of this mutation in the Basque population, can be interpreted in two different ways, either this position is susceptible to several mutational events, as supported by the fact that p.R1441C and p.R1441H have been found in families with different ethnicities (Zimprich, Biskup *et al.*, 2004; Mata, Kachergus *et al.*, 2005), or all Basque carriers of the p.R1441G mutation could have inherited this alteration from a single founder. To differentiate between these two possibilities, we analyzed the chromosomal region surrounding *LRRK2* in carriers of the mutation using the same 11 SNPs that defined the critical haplotype in the four original Basque families (figure 17). We found that all carriers showed genotypes consistent with the haplotypes found in those families (table 23), thus supporting the idea that p.R1441G is a founder mutation.

The minimal shared haplotype (656 Kb) is derived from a single individual, 01-407, who appeared to have a high degree of homozygosity for the rarer alleles at several of the SNPs assayed. Without this individual, the minimal shared haplotype would span 1.08 Mb from rs10876886 to rs10784616. Thus, to rule out the presence of a microdeletion in this patient leading to this high degree of homozygosity, we performed a semi quantitative analysis of SNPs rs1047736 and rs937110, using a non-related genomic area on 11p11.2 as an endogenous control. With this approach, we determined that there was no deletion in this area.

One alternative explanation to the fact that all carriers of the p.R1441G mutation shared the same haplotype, was that the haplotype where the mutation lies may be extremely frequent in the general Basque population. In order to rule out this possibility, we determined the frequency of this haplotype (that composed of 11 SNPs) in the general population, 80 control individuals from the Basque Country and 82 non-Basque Spanish

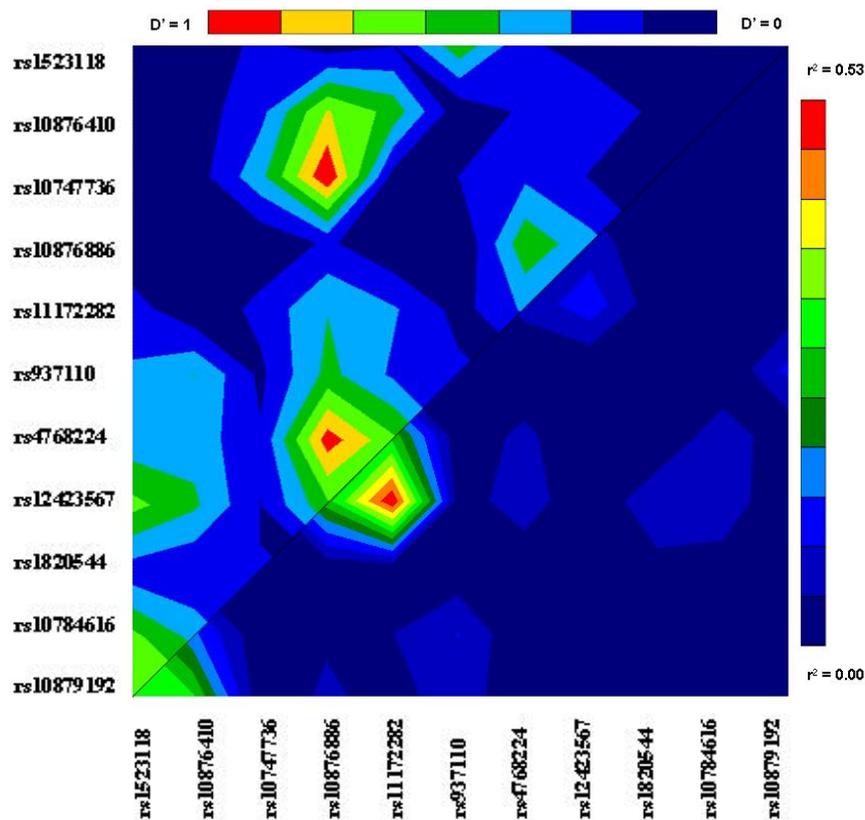
controls were genotyped. We failed to find this haplotype in the Basque population, indicating that it is less frequent than the haplotype with the lowest frequency found (0.625%). Curiously enough, we found more than 70 different haplotypes in the general Basque population, indicating that there is weak linkage disequilibrium in this region. Proof of this can be seen in figure 18, showing  $r^2$  and  $D'$  values (two linkage disequilibrium measures) between those 11 SNPs that defined the critical disease interval. In the non-Basque Spanish individuals analyzed, the frequency of this same haplotype was up to 3% (frequency of the less represented haplotype in this population).



**Figure 17: Chromosomal region surrounding *LRRK2*.** A: Schematic representation of the area covered with the single nucleotide polymorphisms (SNPs) assayed in this work. Below the chromosome 12 ideogram, the different genes are represented by gray arrows with the names of the genes shown. Distances are in Mb. The position of each SNP is indicated at the bottom. *LRRK2* is indicated by a darker arrow. The vertical arrow points to the position of the R1441G mutation within *LRRK2*. B: Sample of the genotyping of one of the SNPs in panel B using ARMS and showing several homozygous for the T allele (samples 01, 03, 04, and 07) and several heterozygous genotypes. No homozygous for the C allele were observed in the population.

Interestingly, typing rs4768224, rs12423567, rs1820544 and rs10784616 in five p.R1441G carriers from Asturias (PE008, PE155, PE040, PE066 and PE139) derived from a study performed by Mata and collaborators (Mata, Ross *et al.*, 2006), showed that they share the same ancestral haplotype as the Basque carriers (table 23).

Finally, table 24 shows a summary of clinical features characterizing the disease in the mutation carriers. As shown, there are no differences in any of the clinical parameters analyzed between mutation carriers and PD patients without this mutation. Thus, it can be considered that PD due to the p.R1441G mutation in Dardarin is clinically indistinguishable from idiopathic PD. When considering the age at onset of those carriers among the Basque mutation in *LRRK2*, all but one of the carriers had a late-onset disease; the only case with an young-onset disease showed initiation of the disease at 44 years of age.



**Figure 18: Linkage disequilibrium across the chromosomal region surrounding LRRK2.** Graphical Overview of Linkage Disequilibrium (GOLD) plots showing linkage disequilibrium of those 11 SNPs that defined the critical disease interval in 4 Basque families, in Basque population. Pairwise LD analysis was performed using EMLD;  $D'$  is shown in the upper left,  $r^2$  in the lower right.

Table 23: Genotypes of SNPs surrounding the *LRRK2* locus in carriers of the p.R1441G mutation

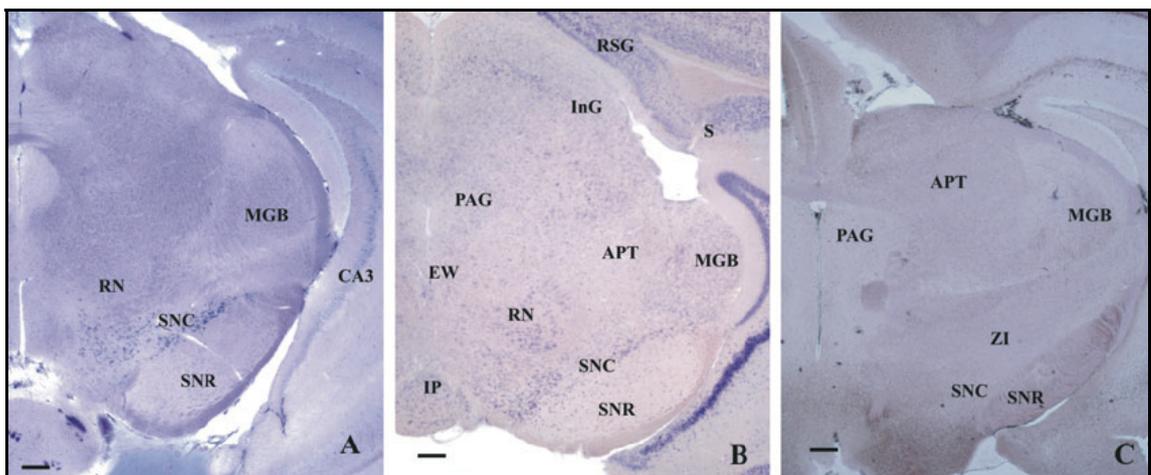
	rs1523118	rs10876410	rs10747736	rs10876886	rs11172282	rs937110	rs4768224	p.R1441G	rs12423567	rs1820544	rs10784616	rs10879192
01-102	T/T	A/A	T/T	C/C	G/G	C/G	A/T	C/G	G/T	C/C	C/C	C/C
01-113	C/T	T/T	C/T	C/C	G/G	C/C	T/T	C/G	G/T	C/T	C/G	C/T
01-142	T/T	T/T	C/T	C/C	G/G	C/G	T/T	C/G	G/T	C/C	C/C	C/C
01-407	T/T	A/A	C/C	C/C	G/G	G/G	T/T	C/G	G/T	C/T	C/G	T/T
01-492	T/T	A/T	T/T	A/C	C/G	C/G	T/T	C/G	G/G	C/T	G/C	C/C
01-517	T/T	A/A	T/T	C/C	G/G	C/C	T/T	C/G	G/G	C/T	C/G	C/C
01-371	C/T	A/A	C/T	A/C	C/G	C/G	T/T	C/G	G/G	C/C	C/C	C/T
01-226	C/T	A/T	T/T	C/C	G/G	C/C	T/T	C/G	G/G	C/C	C/C	C/C
02-143	T/T	A/T	T/T	C/C	G/G	C/G	T/T	C/G	G/G	C/C	G/C	C/C
02-209	C/T	A/A	C/T	A/C	C/G	C/G	T/T	C/G	G/T	C/T	C/C	C/C
02-254	T/T	A/T	T/T	A/C	C/G	C/C	A/T	C/G	G/T	C/C	C/G	C/T
04-259	C/T	T/T	T/T	C/C	G/G	C/C	T/T	C/G	G/T	C/C	C/C	C/T
04-400	C/T	T/T	T/T	C/C	G/G	C/C	T/T	C/G	G/T	C/T	C/G	C/T
04-409	C/T	A/A	C/C	A/C	C/G	C/C	T/T	C/G	G/T	C/T	C/G	C/T
04-410	C/T	A/A	C/T	A/C	C/G	C/C	T/T	C/G	G/T	C/T	C/G	C/C
04-419	C/T	A/A		A/C	C/G	C/G	A/T	C/G	G/T	C/C	C/C	C/C
04-466	C/T	T/T	C/T	C/C	G/G	C/C	T/T	C/G	G/G	C/T	C/G	C/T
PE008							T/A	C/G	G/T	C/C	C/G	
PE155							T/A	C/G	G/G	C/C	C/G	
PE040							T/T	C/G	G/T	C/T	C/C	
PE066							T/A	C/G	G/T	C/C	C/G	
PE139							T/T	C/G	G/T	C/C	C/G	
Families	T	T	T	C	G	C	T	G	G	C	C	C

Table 24: Summary of clinical features of Basque p.R1441G mutation carriers.

	Whole sample	Basque origin	Not Basque	Familial		Sporadic	
				Basque	Not Basque	Basque	Not Basque
N of R1441G carriers	17	16	1	6	1	10	-
AAO*-carriers	63.5±8.5	63.8±8.7	59	65.0±11.0	59	65.7±10.2	-
AAO*-non carriers	64.7±10.4	65.6±10.4	63.1±10.3	65.8±12.2	62.4±9.6	62.6±6.3	-
Sex ratio (M:F)	0.643	0.583	1	0.5	1	0.75	-
Years of evolution	10.2±6.7	9.6±6.3	21	7.3±6.4	21	10.9±6.1	-
Response to L-Dopa	Good	Good	Good	Good	Good	Good	-
Presenting symptom in carriers	UT	UT	UT	UT	UT	CI	-
	0.538	0.5		0.625		0.5	
Presenting symptom in non-carriers	UT	UT	UT	UT	UT	UT	-
(% of cases)	0.602	0.559	0.667	0.556	0.723	0.563	

#### 4.2. *IN SITU* HYBRIDIZATION OF *LRRK2* mRNA IN ADULT MOUSE BRAIN

We studied *LRRK2* mRNA expression in adult mouse brains through *in situ* hybridization analysis using digoxigenin-labelled RNA probes. We first confirmed the specificity of the probe by comparing the staining pattern of the antisense probe with that of the sense probe. The antisense probe showed specific signal patterns, whereas no signal was observed with the sense probe. This difference between both probes indicates that the antisense probe can specifically hybridize with its corresponding transcript (figure 19). As a result we found that *LRRK2* mRNA was expressed throughout the mouse brain. As summarized in table 25, signal intensities varied from undetected (–), faint (+), moderate (++) and strong (+++). A detailed description of the most relevant findings organized into brain areas is as follows.

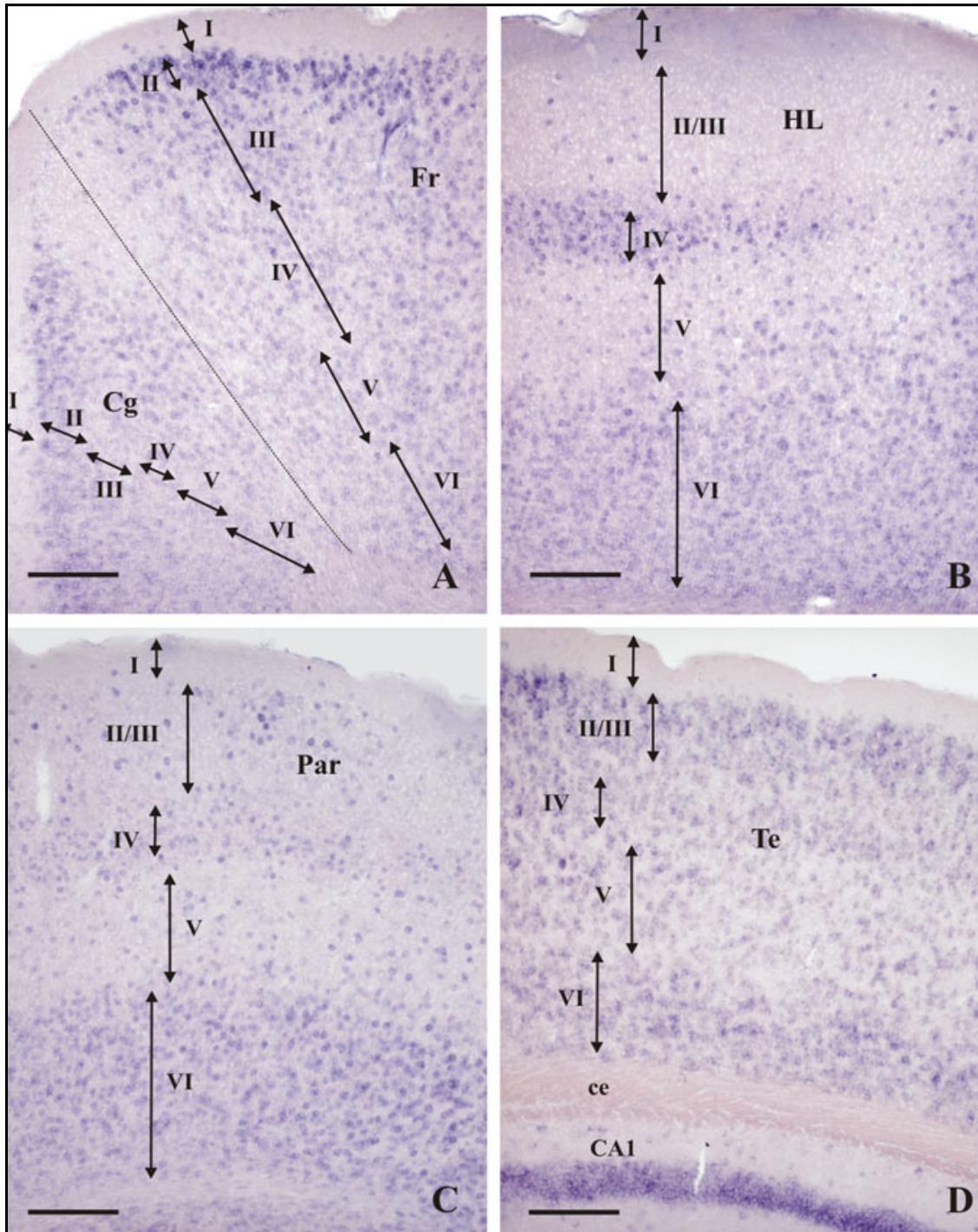


**Figure 19:** *LRRK2* mRNA expression in the mouse brain. Note that the antisense probe (A) shows a specific hybridization pattern and that the sense probe (C) show no hybridization signal. The hybridization pattern observed after 2 h of developing time is shown in A. In (A) and (B) note the strong reaction in the substantia nigra pars compacta and in some midbrain nuclei, including the interpeduncular and red nuclei as well as the periaqueductal gray and superior colliculus. Calibration bars, 200  $\mu$ m.

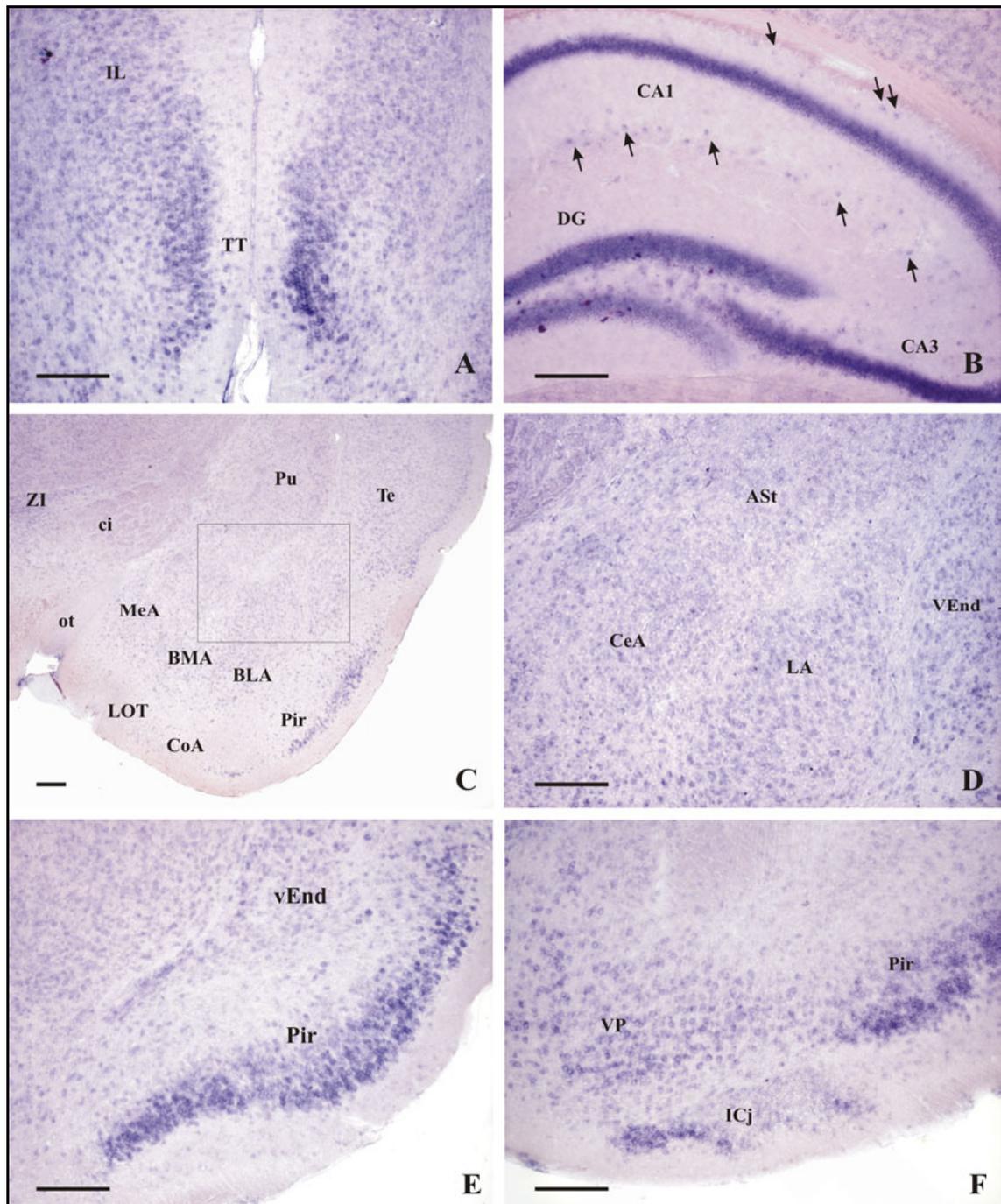
#### 4.2.1. Telencephalon.

Varying *LRRK2* mRNA expression patterns were observed in different cortical areas. In some areas, *LRRK2* was expressed throughout the six cortical areas and the signal intensity depended on the cell density in each layer. This was the case of cingular-frontal (figure 20A), temporal (figure 20D) or occipital and the medial prefrontal cortex. In other areas, *LRRK2* was absent or in disperse neurons in specific cortical layers. In the hindlimb area of the cortex (figure 20B) and in the parietal cortex (figure 20C), labelling was specifically scattered into the layers II / III and in layer V. Within the septal area, *LRRK2* mRNA was not expressed in the nucleus triangularis septalis and was weakly expressed in the dorsal, ventral and intermediate parts of the lateral septal nucleus. However, a strong signal was observed in the horizontal limb of the medial septal nucleus, and a less but consistent labelling appeared in its vertical division dorsal tenia tecta. Furthermore, this gene was also highly expressed in the adjacent septohippocampal nuclei (figure 21A). Within the hippocampus, *LRRK2* mRNA was abundantly expressed in the pyramidal cell layer of Ammon's horn (CA1, CA2 and CA3) as well as in the granular cell layer of the dentate gyrus (figure 21B). A strong expression was also observed in disperse neurons either located in the molecular or in the radiatum-oriens layers. Strong labelling was also observed in the hilar region of the dentate gyrus. *LRRK2* mRNA was strongly expressed in almost every part of the amygdala except for the cortical-superficial area (figure 21C). Thus, it was strongly expressed in the amygdaloid intercalary grey area, the amygdalostriatal transition area, and it was moderately expressed in the amygdalopiriform transition area. With regard to the basal amygdaloid nucleus, very strong *LRRK2* expression was observed in the lateral, basomedial,

basolateral and posterior basal divisions (figure 21D). Strong *LRRK2* expression was noted in the medial amygdaloid nucleus in both the posterodorsal and posteroventral parts. *LRRK2* expression was also detected in the intra-amygdaloid division of the bed nucleus of the stria terminalis close to the posterodorsal area of the medial amygdaloid nucleus and also in the central amygdala. The cortical nucleus, located proximally to the medial amygdaloid nucleus, also expresses *LRRK2* mRNA in both the posterolateral and posteromedial parts. No signal was seen in the nucleus of the lateral olfactory tract or in the cortical amygdala (figure 21C). *LRRK2* mRNA expression was high in the most superficial layers of the piriform cortex (figure 21E). The olfactory tubercle and the islands of Calleja are located close to the piriform cortex, which also showed strong *LRRK2* expression (figure 21F). Within the basal ganglia, *LRRK2* mRNA was detected throughout the striatum, where no patchy differences were observed that can indicate a possible differential reaction between striosome and matrix. Interestingly, the *globus pallidus* was nearly devoid of any *LRRK2* expression (figure 22A).



**Figure 20: *LRRK2* mRNA expression in different cortical areas.** (A) cingular / frontal cortical areas (B) hindlimb cortical area (C) parietal cortex, and (D) temporal cortex. Note that in both the hindlimb and parietal cortices, layers II/III and V showed scattered labeling. A long (12 h) chromogenic reaction was performed in all these cases. Calibration bars, 200  $\mu$ m.



**Figure 21: *LRRK2* mRNA distribution in telencephalic areas.** (A) Rostral telencephalic areas tenia tecta and infralimbic medial prefrontal cortex. (B) Hippocampus, note that the nonpyramidal neurons of the Ammons horn (arrows) as well as the hilar neurons also express *LRRK2*. (C and D) Amygdala, note the scattered labeling in cortical amygdala and the nucleus of the lateral olfactory tract the square in C as magnified in D. (E) Piriform cortex, note the strong reaction of the superficial layer. (F) Rostral ventral telencephalon showing the islands of Calleja and the ventral striatum. A long (12 h) chromogenic reaction was performed in all these cases. Calibration bars, 200  $\mu$ m.

#### 4.2.2. Diencephalon.

With the exception of the paraventricular thalamic nucleus (figure 22B) and the reticular nucleus (figure 22C), which strongly expresses *LRRK2* mRNA, all the nuclei within the thalamus moderately expressed this gene (figure 22C). Some labelled cells were observed in both the medial and lateral geniculate nuclei. A strong *LRRK2* mRNA expression was also observed in the medial habenular nucleus. However, only a very faint expression could be seen in the lateral nucleus (figure 22B). The *zona incerta* also contained strongly labeled neurons (figure 22C). On regard to the hypothalamus, a *LRRK2* mRNA expression was observed in the posterior and dorsal areas, although its expression was very weak in the ventrolateral nucleus. The ventromedial hypothalamic nucleus showed a moderate-to-high expression in the dorsomedial and central parts but not in the ventrolateral one (figure 22D). The dorsomedial hypothalamic nucleus showed a high *LRRK2* mRNA expression in compact, ventral and dorsal parts (figure 22D). A scattered reaction was observed in the lateral hypothalamic area.

In addition to these above-mentioned diencephalic areas, the geniculate, zona incerta and subthalamic nucleus also expressed *LRRK2* moderately.

#### 4.2.3. Brainstem.

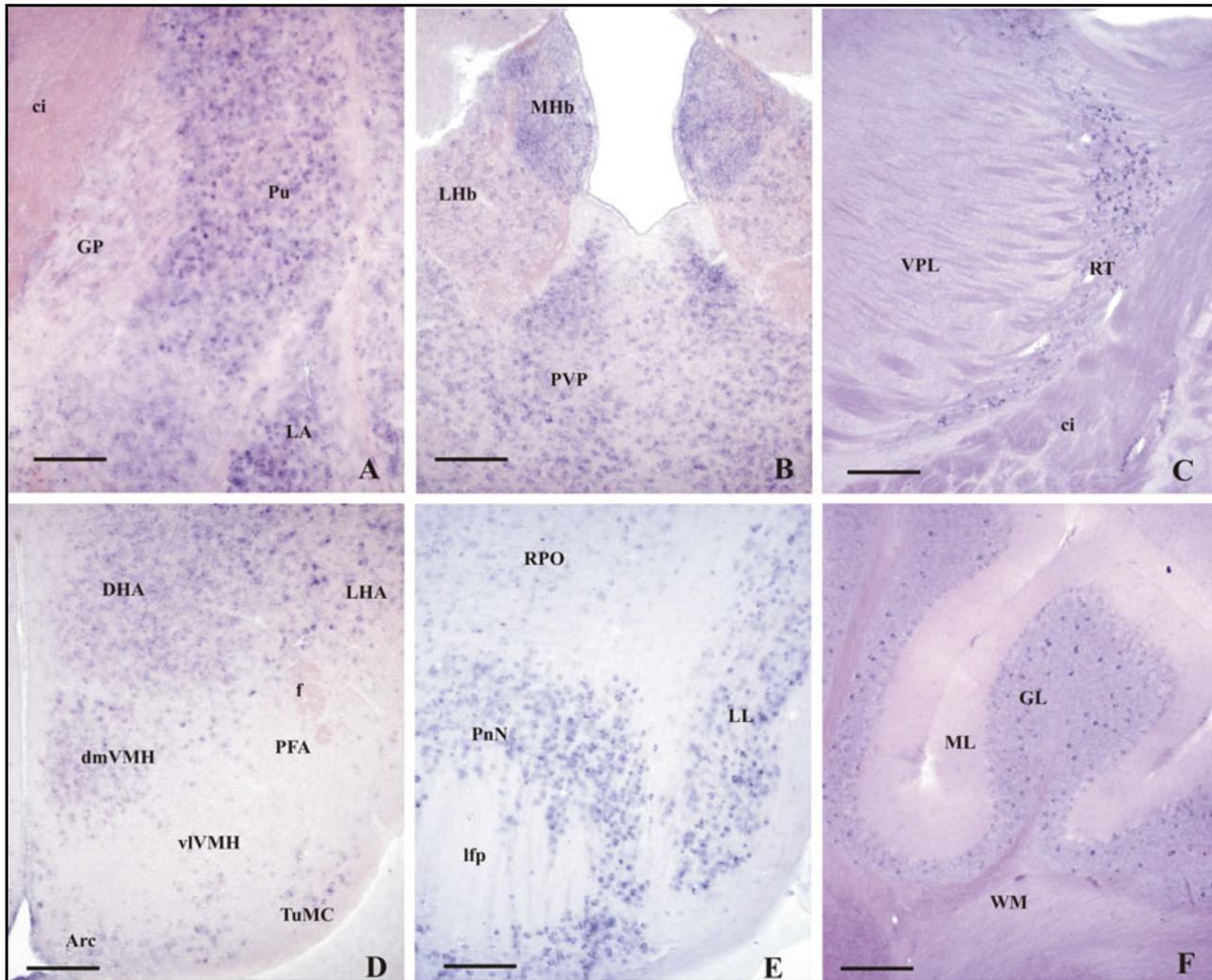
Strong *LRRK2* mRNA expression was observed in both the dorsal and ventral tiers of the *substantia nigra pars compacta*. In contrast, weak *LRRK2* mRNA expression was detected throughout the *substantia nigra pars reticulata* (figure 19B). The ventral tegmental area and the nucleus of Darkschewitsch moderately expressed *LRRK2* mRNA. A strong reaction was observed in the intermediate grey layer of the superior colliculus

(figure 19B) and the red nucleus. Moreover, dispersely labeled neurons also occurred in the inferior colliculus and the nuclei of the lateral lemniscus (figure 22E). The periaqueductal grey also showed disperse staining (figure 19B). Moderate labeling was detected in the dorsal and in the median raphe nucleus. In contrast, the dorsal tegmental nuclei were almost devoid of expression.

Faint *LRRK2* mRNA expression was seen in the anterior, posterior and medial pretectal nuclei. Disperse labeling was observed along the reticular formation, whereas some cells of the pontine nuclei also expressed *LRRK2* (figure 22E).

#### **4.2.4. Cerebellum.**

As seen in figure 22F the granular cell layer of the cerebellum exhibited a moderate *LRRK2* mRNA expression. On the other hand, neither the Purkinje cell layer nor the basal nuclei showed *LRRK2* mRNA expression.



**Figure 22: *LRRK2* mRNA distribution in subcortical areas.** (A) Striatum, note the strong labelling in the putamen and contrastingly, the specific scattered reaction in the globus pallidus. (B) Habenula and paraventricular thalamic nucleus, contrasting the high signal in the medial habenula, the lateral habenula only showed disperse cells. (C) Reticular thalamic nucleus in the case of a short (2 h) chromogenic reaction. (D) Ventral tuberal region of the hypothalamus, note the scattered labelling in the ventrolateral division of the ventromedial hypothalamic nucleus contrasting with the rest of the hypothalamic region and areas. (E) Ventral pons, note the strong reaction in the pontine nuclei and the ventral division of the lateral lemniscus. (F) Cerebellum, note the strong reaction in the granular layer. A long (12 h) chromogenic reaction was performed in all the cases except C. Calibration bars, 200  $\mu$ m.

**Table 25:** Relative levels of expression of *LRRK2* in adult mouse brain.

<b>Brain area</b>	<b><i>LRRK2</i> mRNA</b>	<b>Brain area</b>	<b><i>LRRK2</i> mRNA</b>
<b>Telencephalon</b>		<b>Diencephalon</b>	
Cerebral cortex	++	Centrolateral nucleus	+
Olfactory tubercle	++	Paraventricular nucleus	+++
Piriform cortex	+++	Geniculate	++
Clastrum	+	Habenular complex	
Hippocampal formation		Medial nucleus	+++
Ammon's horn	+++	Lateral nucleus	-
Dentate gyrus	+++	Subthalamus	++
Septal area		Zona incerta	++
Tenia tecta	++	Hypothalamus	
Vertical medial septal nucleus	+	Preoptic nuclei	-
Horizontal medial septal nucleus	++	Posterior area	+
Lateral septal nucleus	+	Dorsal area	+
Septohippocampal nucleus	++	Ventrolateral nucleus	+
Nucleus Triangularis septalis	-	Ventromedial nucleus	++
Amygdala and extended amygdala		Dorsomedial nucleus	++
Amygdaloid intramedullary grey	+++	<b>Brainstem</b>	
Amygdalopiriform transition area	++	Pretectum	+
Amygdalostriatal transition area	+++	Ventral tegmental area	+
Basolateral amygdaloid nucleus	+++	Red nucleus	++
Basomedial amygdaloid nucleus	+++	Nucleus of Darkschewitsch	++
Central amygdaloid nucleus	+	Ventral tegmental area	+
Medial amygdaloid nucleus	+++	Periaqueductal gray	++
Lateral amygdaloid nucleus	++	Median raphe	++
Nucleus of the lateral olfactory tract	-	Dorsal raphe	++
Cortical amygdala	-	Dorsal tegmental nuclei	-
Bed nucleus of the stria terminalis	++	Pontine nuclei	++
Basal ganglia		Superior colliculus	++
Globus pallidus	-	Inferior colliculus	+
Caudate putamen	+++	Substantia nigra	
Nucleus accumbens	-	Pars compacta	+++
<b>Diencephalon</b>		Pars reticulata	-
Thalamus		<b>Cerebellum</b>	
Anterodorsal nucleus	++	Cerebellar nuclei	-
Reticular nucleus	++	Cortex	
Parafascicular nucleus	+	Purkinje cell layer	-
Mediodorsal nucleus	++	Granule cell layer	++

### 4.3. SEQUENCING ANALYSIS OF *OMI/HTRA*

After analyzing the entire protein coding region of *OMI/HTRA2* (RefSeq NM\_013247) in a total of 644 PD patients and 828 neurologically normal controls we identified the c. 1195G>A substitution in exon 7 (leading to p.G399S), previously identified as a mutation associated with PD by Strauss and collaborators (Strauss, Martins *et al.*, 2005), in a total of 5 PD patients (0.77%) and 6 neurologically normal control individuals (0.72%) (figure 23). The PD patients carrying this substitution were *ND00428*, *ND00154*, *ND01247*, *ND09816* and *ND00148*.

*ND00428* and *ND00154* were both diagnosed with YOPD with ages at onset of 38 and 30 years respectively. *ND01247*, *ND09816* and *ND00148* were diagnosed with late-onset PD with ages at onset of 56, 65 and 77 years respectively.

The control samples carrying p.G399S were *ND01699*, *ND05065*, *ND10345*, *ND03376*, *ND10860* and *ND08025*. These samples were derived from panels *NDPT002*, *NDPT006*, *NDPT019*, *NDPT020*, *NDPT0 21* and *NDPT023* with ages at sampling of 85, 70, 25, 42, 46 and 62 years respectively. Fisher's exact association test showed no association between p.G399S and PD (table 26).

Fisher's exact association tests considering either YOPD or late-onset PD samples along with their corresponding age-matched controls were performed. No association was seen after these analyses (table 26).

**Table 26:** Minor allele frequencies (MAF), Fisher's exact test p values, odds ratios (OR), and 95% confidence intervals (C.I 95%), for variants p.G399S and p.A141S.

Variant	Samples included in analysis	Minor Allele (Cases-Controls)	MAF Cases - Controls	Fisher's p value	OR	C.I 95%
p.G399S	PD: all, n = 644 Controls: all, n = 828	A (5-6)	0.388%-0.363%	1	1.069	0.3256-3.511
p.G399S	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	A (2-1)	0.944% - 0.354%	0.5794	2.676	0.2411 - 29.71
p.G399S	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	A (3-5)	0.278% - 0.365%	1	0.7633	0.182 - 3.201
p.A141S	PD: all, n = 644 Controls: all, n = 828	T (28-46)	2.215% - 2.991%	0.2365	0.7348	0.4566 - 1.182
p.A141S	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	T (7-9)	3.398% - 3.214%	1	1.059	0.3879 - 2.892
p.A141S	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	T (21-37)	1.985% - 2.941%	0.1815	0.6683	0.3887 - 1.149

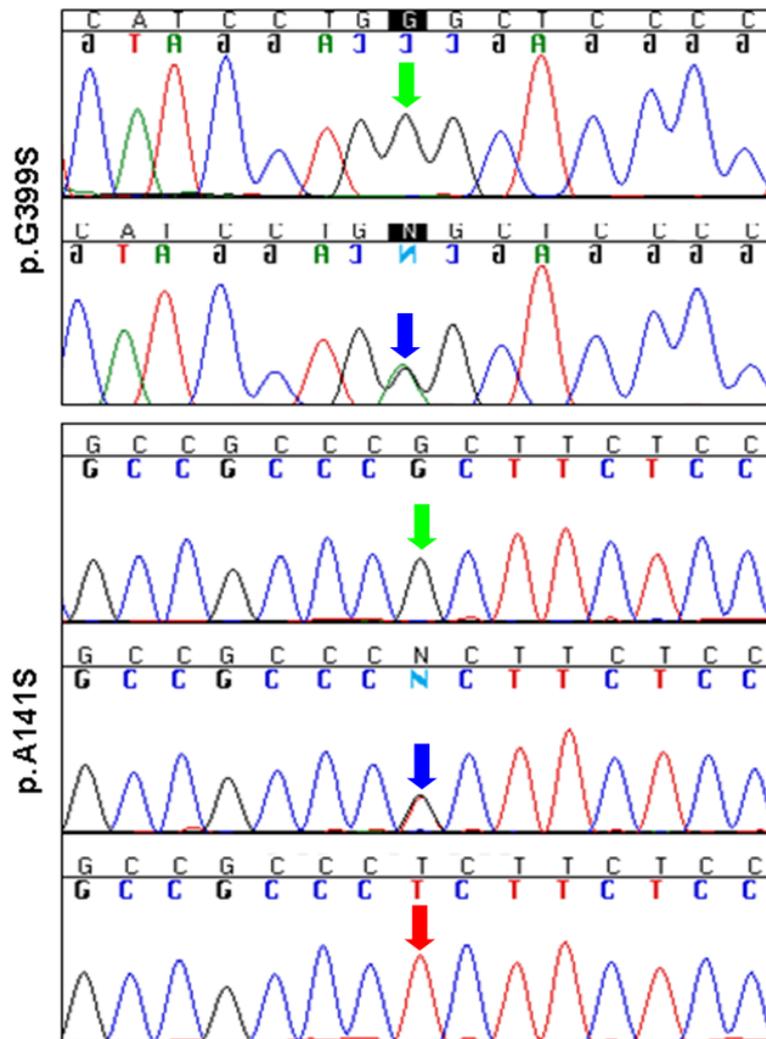
Table 26 also shows the frequency of genotypes of the variant c.421G>T in exon 1 of *OMI/HTRA2* (leading to p.A141S). The T allele of this variant has previously been associated with PD (Strauss, Martins *et al.*, 2005); however, we failed to find such an association in our dataset performing Fisher's exact association tests considering all samples and only YOPD and late-onset PD cases independently, along with their corresponding age-matched neurologically normal controls (table 26). Only *ND4314* and *ND04275* (two neurologically normal control individuals of 49 and 55 years respectively) were homozygous for this change (figure 22).

We failed to find the synonymous p.F149F variant, previously identified by Strauss and colleagues, in our population.

In addition to these variants, we identified 8 novel heterozygous coding variants, both synonymous (p.V109V, p.L118L, p.R209R and p.L367L) and non-synonymous

(p.W12C, p.P128L, p.F172V and p.A227S), in both cases and control samples. None of these seem to be causative of disease in our population (table 27).

In addition to these variants, 8 SNPs have been described within the *OMI/HTRA2* coding region, two being exonic and six residing within introns. Of these only rs2231249, rs2231248, rs11538692, rs2241027 and rs2241028 were within our sequencing boundaries and thus, were genotyped. Variants rs2231248 (exon 1), rs2241027 (intron 4) and rs2241028 (intron 5) had at least one sample with a variant allele; however, none of these SNPs showed association with PD after Fisher's exact association tests (table 27).



**Figure 23: Chromatogram showing p.G399S and p.A141S variants in *OMI/HTRA2*. Green arrows indicate wild-type sequence, blue arrows heterozygous changes, and the red arrow in the bottom panel shows a homozygous change for p.A141S.**

**Table 27:** Minor allele frequencies (MAF), Fisher’s exact test p values, odd ratios (OR) and 95% confidence intervals (C.I 95%) for all new variants found in our data set as well as 3 SNPs within *OMI/HTRA2*.

Variant	Samples included in analysis	Minor Allele (Cases-Controls)	MAF Cases-Controls	Fisher's p value	OR	C.I 95%
p.V109V	PD: all, n = 644 Controls: all, n = 828	A (2-0)	0158% - 0%	-	-	-
p.V109V	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	A (1-0)	0.4854% - 0%	-	-	-
p.V109V	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	A (1-0)	0.0945% - 0%	-	-	-
p.L118L	PD: all, n = 644 Controls: all, n = 828	G (0-2)	0% - 0.130%	-	-	-
p.L118L	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (0-2)	0 - 0.714%	-	-	-
p.L118L	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G (0-0)	0% - 0%	-	-	-
p.R209R	PD: all, n = 644 Controls: all, n = 828	G(2-0)	0.156% - 0%	-	-	-
p.R209R	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (0-0)	0% - 0%	-	-	-
p.R209R	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G (2-0)	0.187% - 0%	-	-	-
p.L367L	PD: all, n = 644 Controls: all, n = 828	C (0-1)	0% - 0.063%	-	-	-
p.L367L	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	C (0-0)	0% - 0%	-	-	-
p.L367L	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	C (0-1)	0% - 0.077%	-	-	-
p.W12C	PD: all, n = 644 Controls: all, n = 828	G (0-1)	0% - 0.064%	-	-	-
p.W12C	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (0-0)	0% - 0%	-	-	-
p.W12C	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G (0-1)	0% - 0.078%	-	-	-
p.P128L	PD: all, n = 644 Controls: all, n = 828	C (0-1)	0% - 0.065%	-	-	-
p.P128L	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	C (0-1)	0% - 0.357%	-	-	-
p.P128L	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	C (0-0)	0% - 0%	-	-	-

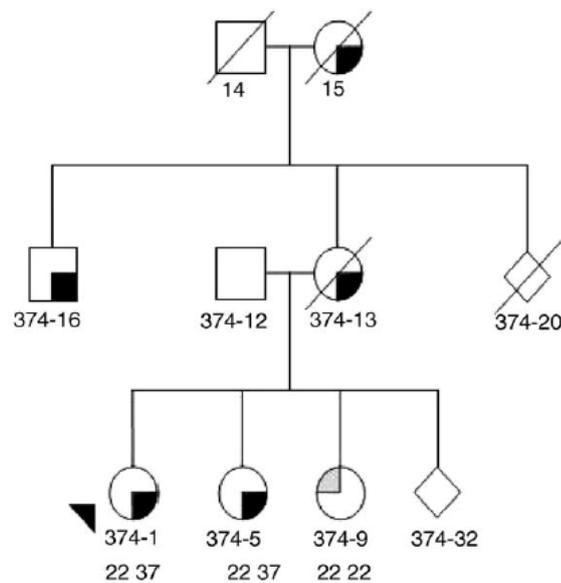
MAFs, p values, ORs and C.I 95% were calculated in both the entire group (644 PD cases and 828 neurologically normal controls) and only testing YOPD and Late-onset PD cases independently with corresponding age-matched control groups.

Table 27 (continued)

<b>Variant</b>	<b>Samples included in analysis</b>	<b>Minor Allele (Cases-Controls)</b>	<b>MAF Cases-Controls</b>	<b>Fisher's p value</b>	<b>OR</b>	<b>C.I 95%</b>
p.F172V	PD: all, n = 644 Controls: all, n = 828	G (1-0)	0.0782% - 0%	-	-	-
p.F172V	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (1-0)	0.4762% - 0%	-	-	-
p.F172V	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G (0-0)	0 - 0	-	-	-
p.A227S	PD: all, n = 644 Controls: all, n = 828	T (1-0)	0.078% - 0%	-	-	-
p.A227S	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	T (0-0)	0% - 0%	-	-	-
p.A227S	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	T (1-0)	0.093% - 0%	-	-	-
rs2231248	PD: all, n = 644 Controls: all, n = 828	G (1-0)	0.079% - 0%	-	-	-
rs2231248	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (0-0)	0% - 0%	-	-	-
rs2231248	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G (1-0)	0.094% - 0%	-	-	-
rs2241027	PD: all, n = 644 Controls: all, n = 828	G (6-8)	0.465% - 0.493%	1	0.9442	0.3268-2.728
rs2241027	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (3-1)	1.415% - 0.354%	0.3187	4.033	0.4166-39.05
rs2241027	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G(3-7)	0.278% - 0.522%	0.5269	0.5324	0.1374-20.64
rs2241028	PD: all, n = 644 Controls: all, n = 828	G (60-85)	4.73% - 5.400%	0.4412	0.8701	0.6199-1.221
rs2241028	PD: age at onset ≤ 40, n = 106 Controls: age at sampling ≤ 40, n = 141	G (13-20)	6.311% - 7.140%	0.8556	0.8756	0.4251-1.804
rs2241028	PD: age at onset > 40, n = 538 Controls: age at sampling > 40, n = 687	G (47-65)	4.420% - 5.020%	0.5596	08755	0.5961-1.286

#### 4.4. ANALYSIS OF SCA-2 AND SCA-3 REPEATS IN PARKINSONISM

Analysis of *ATXN-2* and *ATXN-3* expansions in 280 probands with PD (both with and without positive family history) with different ethnic backgrounds (see methods section 3.4), showed no evidence of *ATXN-3* expansion in any of the patients analyzed (repeat sizes ranged from 11 to 44). However, one Caucasian proband (patient 374-1) had an expanded *ATXN-2* allele with 37 repeats. The remaining individuals had *ATXN-2* trinucleotide repeat lengths within the normal range (17–29). Assessing the *ATXN-2* locus in available family members of patient 374-1 demonstrated that the expanded allele was also found in the other family member with parkinsonism (individual 374-5, with repeat sizes of 22 and 37) but not in subject 374-9, who had essential tremor (figure 24). These results show that this expansion mutation within *ATXN-2* is the cause of disease in this family (figure 24).



**Figure 24: Pedigree of family 374.** Number pairs underneath each sampled individual represent *SCA-2* repeat sizes. A black section represents that parkinsonism is present. Gray section on individual 374-9 indicates essential tremor.

#### 4.5. WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 1

##### 4.5.1. Whole-genome association

We genotyped 276 samples from patients with idiopathic PD and 276 from unrelated neurologically normal controls with Illumina's Infinium I and Infinium II HumanHap 300 genotyping beadchips.

Genotyping of the control and the three PD replicate samples with the Infinium I assay gave genotype concordance rates greater than 99.99%. Besides, analysis of the 18,073

SNPs that overlap between Infinium I and Infinium II platforms revealed genotype concordance rates of 99.94% between the assays across samples.

All samples with a call rate below 95% were re-assayed using a new aliquot of DNA extracted directly from blood. Those samples that did not reach the 95% threshold after repeated assay were dropped from further analysis. Thus, four control samples and three PD samples were dropped from our study. Since two YOPD samples had erroneously been included in panel *NDPT007* (see *structural variation in PD-related loci* in section 4.5.2.3) the final number of fully genotyped samples was 271 controls and 267 PD patients. For those 408,803 SNPs assayed, the genotype call rate was greater than 99% for 395,275 SNPs (96.6%) and greater than 95% for 406,312 (99.4%). The Hardy-Weinberg equilibrium p value was higher than 0.01 for 395,493 SNPs and higher than 0.05 for 375,527 SNPs (figure 25). The average minor allele frequency for autosomes was 26.47%. A total of 219,577,497 unique genotype calls were made and the average call rate across all samples was 99.6%.

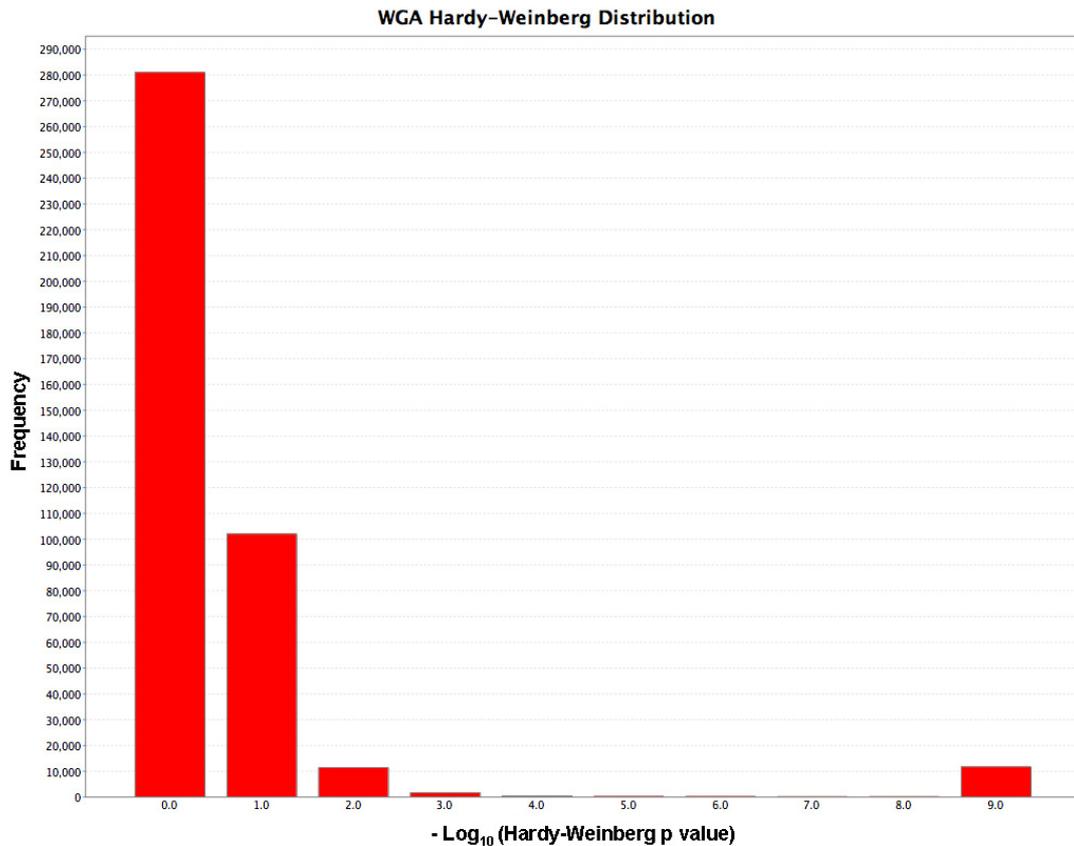
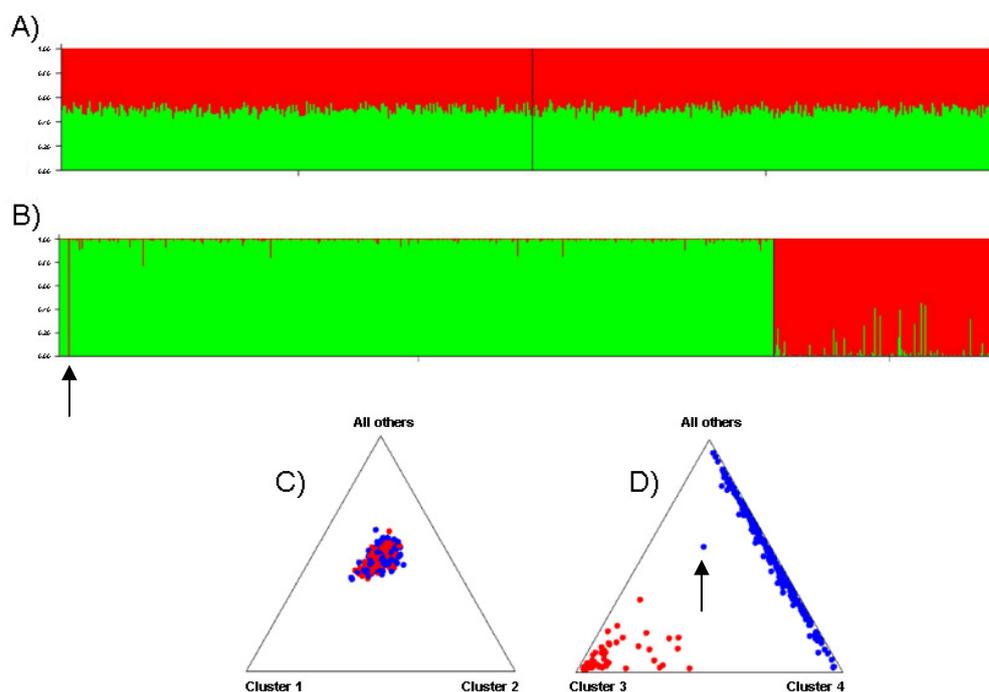


Figure 25: Hardy-Weinberg p value distribution in our final set of samples.

Analysis with STRUCTURE v2.2 (<http://pritch.bsd.uchicago.edu/structure.html>) showed that there is no discernible difference in the population substructure between cases and controls (figure 26). Furthermore, comparison of the cases and controls pooled together versus genotypes from a cohort of 173 non-white participants showed clear separation of the Parkinson's disease and control group from the non-white group, with the exception of a single patient from the former cohort, who, based on these analyses, had significant non-white genetic background. This individual was removed from the association analysis.

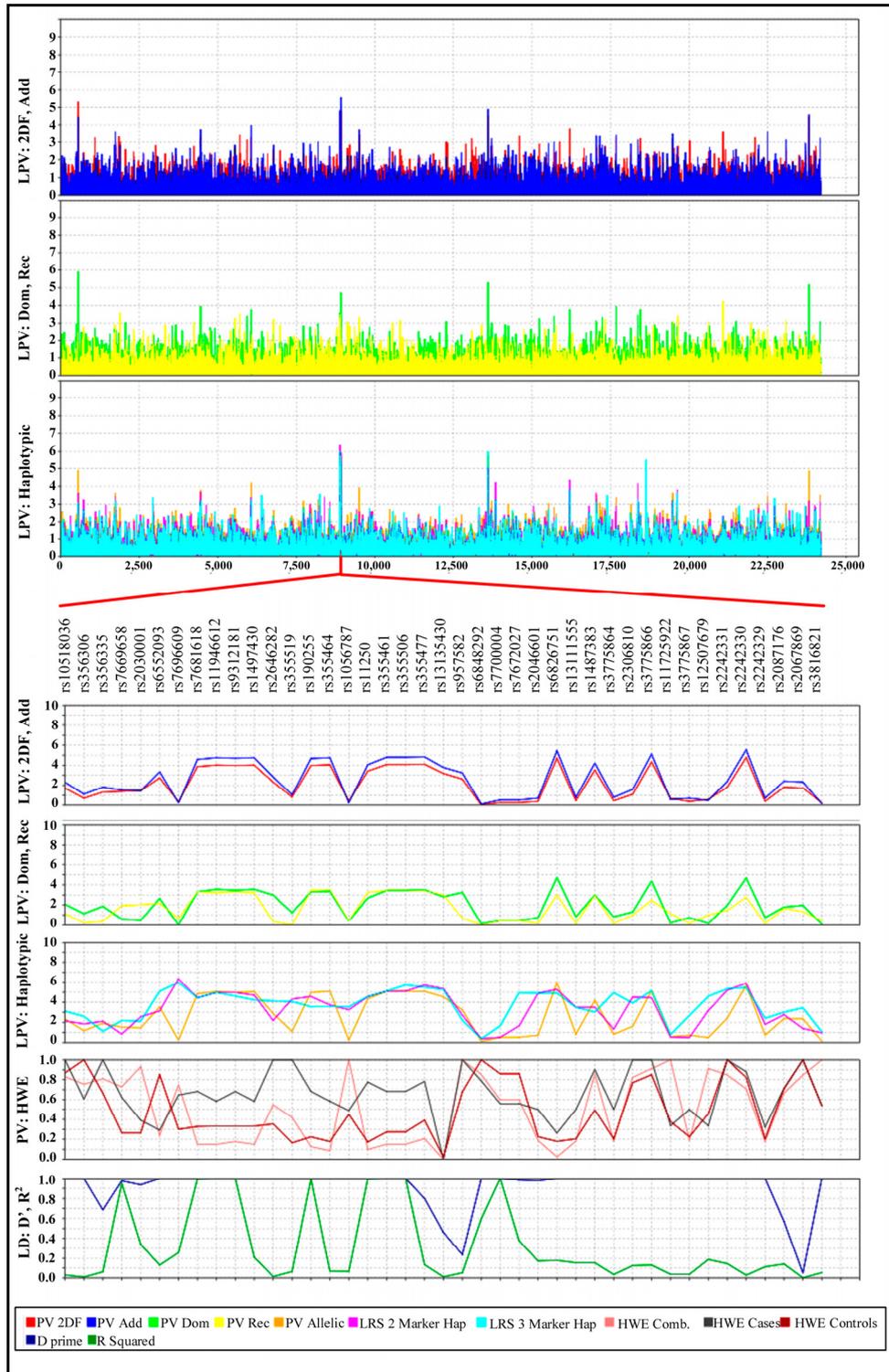


**Figure 26: Bar and Triangle Plots from Structure using 267 random Autosomal SNPs. A)** Bar plot for  $K = 2$  sorted by putative population where Population 1 consists of 271 Caucasian controls and population 2 consists of 267 sporadic PD patients. **B)** Bar plot for  $K = 2$ , sorted by putative population using the same set of 267 SNPs where population 1 consists of 538 Caucasians (sporadic PD case/control series) and Population 2 consists of 173 non-Caucasian subjects. **C)** Triangle plot with same putative populations as bar plot **A** but with  $K = 4$  where blue dots are population 1 (Controls) and red dots are population 2 (PD cases). **D)** Triangle plot with same putative populations as bar plot **B** but with  $K = 4$  where blue dots are Population 1 (Caucasian sporadic PD cases and controls) and red dots are population 2 (non-Caucasian subjects). Arrows indicate non-white individual dropped from further analyses.

Statistical analysis of association was done for all genotypes regardless of Hardy-Weinberg disequilibrium or minor allele frequency (figure 27). Analysis of our data showed 26 SNPs with a two-degree of freedom p value below 0.0001 (table 28), with ORs ranging from 0.2 (95% CI 0.04-0.5) to 0.6 (0.5-0.8) and from 1.7 (1.3-2.2) to 2.2 (1.6-3.2). However, a stringent Bonferroni correction based on 408,803 independent tests means that a precorrection p value of less than  $1.2 \times 10^{-7}$  would be needed to provide a

corrected significant p value of less than 0.05. Thus, none of the values listed were significant after correction.

A secondary objective of this project was the release of genome-wide data from a well-characterized publicly available cohort of both controls and PD patients. Thus, genotyping data obtained in our experiments can be downloaded from (<http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap>).



**Figure 27:** Example of the plots obtained after statistical analysis with SNPgwa software. Log p values are plotted across chromosome 4 (upper panel) and across a significant region (lower panel). A high degree of linkage disequilibrium can be seen across the significant region, which explains the significant association of many neighboring SNPs. LPV=log p value. PV=p value. 2DF=two degree of freedom genotype test. Add=additive model. Rec=recessive model. Dom=dominant model. Allelic=allelic association. LRS 2 marker=two marker haplotype association (two contiguous markers). LRS 3 marker=three marker haplotype association (three contiguous markers). HWE=p value indicating deviation from Hardy Weinberg equilibrium. HW Comb=for cases and controls. HW cases=for cases. HW controls=controls. D' and R<sup>2</sup> squared are both measures of linkage disequilibrium.

**Table 28:** *p* values with uncorrected significance greater than  $1 \times 10^{-5}$ .

Chr. Location	SNP ID	Location bp	No. Geno	Gene	HWE <i>p</i> -value	2DF <i>p</i> -value	2DF Empirical <i>p</i> -value	Dom/Add/Rec <i>p</i> -value	OR (95% CI)	Dom/Add/Rec Empirical <i>p</i> -value
11q14	rs10501570	84095494	536	<i>DLG2</i>	0.396	$7.3 \times 10^{-06}$	$2.0 \times 10^{-06}$	$5.3 \times 10^{-04}$ R	0.2 (0.0-0.5)	$4.9 \times 10^{-04}$ R
17p11.2	rs281357	19683106	537	<i>ULK2</i>	0.852	$9.8 \times 10^{-06}$	$4.0 \times 10^{-06}$	0.0002R	0.4 (0.2-0.6)	$1.5 \times 10^{-05}$ R
4q13.2	rs2242330 <sup>+</sup>	68129844	537	<i>BRDG1</i>	0.708	$1.7 \times 10^{-05}$	$1.2 \times 10^{-05}$	$2.9 \times 10^{-06}$ A	0.5 (0.4-0.7)	$<1 \times 10^{-06}$ A
10q11.21	rs1480597*	44481115	525	Intergenic	1.000	$1.9 \times 10^{-05}$	$7.0 \times 10^{-06}$	$3.2 \times 10^{-06}$ D	0.4 (0.3-0.6)	$2.0 \times 10^{-06}$ D
4q13.2	rs6826751 <sup>+</sup>	68116450	536	<i>BRDG1</i>	0.024	$2.0 \times 10^{-05}$	$1.8 \times 10^{-05}$	$3.5 \times 10^{-06}$ A	0.6 (0.4-0.7)	$5.0 \times 10^{-06}$ A
16q23.1	rs4888984	78066835	537	Intergenic	1.000	$2.7 \times 10^{-05}$	$1.1 \times 10^{-05}$	$4.6 \times 10^{-06}$ A	0.5 (0.3-0.7)	$3.0 \times 10^{-06}$ A
4q35.2	rs4862792	188438344	511	Intergenic	0.358	$3.5 \times 10^{-05}$	$8.0 \times 10^{-06}$	$6.8 \times 10^{-06}$ D	2.9 (1.8-4.6)	$7.0 \times 10^{-06}$ D
4q13.2	rs3775866 <sup>+</sup>	68126775	537	<i>BRDG1</i>	0.911	$4.6 \times 10^{-05}$	$3.3 \times 10^{-05}$	$7.8 \times 10^{-06}$ A	0.5 (0.4-0.7)	$8.0 \times 10^{-06}$ A
20q13.13	rs2235617 <sup>‡</sup>	47988384	530	<i>ZNF313</i>	0.034	$4.7 \times 10^{-05}$	$4.7 \times 10^{-05}$	$8.8 \times 10^{-06}$ D	0.4 (0.3-0.6)	$1.2 \times 10^{-05}$ D
1p31	rs988421	72322424	536	<i>NEGR1</i>	0.667	$4.9 \times 10^{-05}$	$4.3 \times 10^{-05}$	$7.0 \times 10^{-04}$ R	2.0 (1.3-3.0)	$8.2 \times 10^{-04}$ R
10q11.21	rs7097094*	44530696	537	Intergenic	0.294	$5.0 \times 10^{-05}$	$2.7 \times 10^{-05}$	$8.9 \times 10^{-06}$ D	0.5 (0.3-0.7)	$8.0 \times 10^{-06}$ D
10q11.21	rs999473*	44502322	537	Intergenic	0.294	$5.0 \times 10^{-05}$	$3.8 \times 10^{-05}$	$8.9 \times 10^{-06}$ D	2.2 (1.5-3.1)	$8.0 \times 10^{-06}$ D
11q11	rs1912373	56240441	537	Intergenic	0.375	$5.6 \times 10^{-05}$	$6.1 \times 10^{-05}$	$9.7 \times 10^{-06}$ D	2.2 (1.6-3.2)	$1.2 \times 10^{-05}$ D
1q25	rs1887279 <sup>#</sup>	182176783	537	<i>GLT25D2</i>	0.424	$5.7 \times 10^{-05}$	$3.5 \times 10^{-05}$	$1.2 \times 10^{-05}$ A	0.5 (0.4-0.7)	$6.0 \times 10^{-06}$ A
1q25	rs2986574 <sup>#</sup>	182173237	536	<i>GLT25D2</i>	0.350	$6.3 \times 10^{-05}$	$2.4 \times 10^{-05}$	$1.3 \times 10^{-05}$ A	2.0 (1.4-2.7)	$6.0 \times 10^{-06}$ A
22q13	rs11090762	46133989	536	Intergenic	0.730	$6.3 \times 10^{-05}$	$4.2 \times 10^{-05}$	$1.2 \times 10^{-05}$ D	0.4 (0.3-0.6)	$8.0 \times 10^{-06}$ D
20q13.13	rs6125829 <sup>‡</sup>	48002336	509	<i>ZNF313</i>	0.004	$6.6 \times 10^{-05}$	$7.2 \times 10^{-05}$	$1.4 \times 10^{-05}$ D	2.2 (1.6-3.2)	$1.8 \times 10^{-05}$ D
7p12	rs7796855	49627992	537	Intergenic	0.931	$6.6 \times 10^{-05}$	$7.2 \times 10^{-05}$	$1.3 \times 10^{-05}$ D	0.4 (0.3-0.6)	$1.2 \times 10^{-05}$ D
4q13.2	rs355477 <sup>+</sup>	68079120	533	<i>BRDG1</i>	0.207	$7.9 \times 10^{-05}$	$7.4 \times 10^{-05}$	$1.5 \times 10^{-05}$ A	0.6 (0.5-0.8)	$1.7 \times 10^{-05}$ A
1q25	rs3010040 <sup>#</sup>	182174845	537	<i>GLT25D2</i>	0.421	$8.0 \times 10^{-05}$	$6.2 \times 10^{-05}$	$1.6 \times 10^{-05}$ A	0.5 (0.4-0.7)	$1.2 \times 10^{-05}$ A
1q25	rs2296713 <sup>#</sup>	182176340	537	<i>GLT25D2</i>	0.421	$8.0 \times 10^{-05}$	$6.2 \times 10^{-05}$	$1.6 \times 10^{-05}$ A	2.0 (1.4-2.7)	$1.2 \times 10^{-05}$ A
4q13.2	rs355461 <sup>+</sup>	68063319	537	<i>BRDG1</i>	0.150	$8.3 \times 10^{-05}$	$6.0 \times 10^{-05}$	$1.6 \times 10^{-05}$ A	1.7 (1.3-2.2)	$1.9 \times 10^{-05}$ A
4q13.2	rs355506 <sup>+</sup>	68068677	537	<i>BRDG1</i>	0.150	$8.3 \times 10^{-05}$	$6.0 \times 10^{-05}$	$1.6 \times 10^{-05}$ A	1.7 (1.3-2.2)	$1.9 \times 10^{-05}$ A
4q13.2	rs355464 <sup>+</sup>	68061719	531	<i>BRDG1</i>	0.086	$8.9 \times 10^{-05}$	$9.3 \times 10^{-05}$	$1.7 \times 10^{-05}$ A	1.7 (1.3-2.2)	$2.1 \times 10^{-05}$ A
4q13.2	rs1497430 <sup>+</sup>	68040409	535	<i>BRDG1</i>	0.150	$9.7 \times 10^{-05}$	$8.0 \times 10^{-05}$	$1.8 \times 10^{-05}$ A	1.7 (1.3-2.2)	$1.9 \times 10^{-05}$ A
4q13.2	rs11946612 <sup>+</sup>	68018566	535	<i>BRDG1</i>	0.150	$9.7 \times 10^{-05}$	$8.5 \times 10^{-05}$	$1.8 \times 10^{-05}$ A	0.6 (0.5-0.8)	$2.1 \times 10^{-05}$ A

Only data from SNPs that gave successful genotypes in >95% of samples (406,312 of 408,803 SNPs) is included in this table. While the SNPs outlined here remain candidates, an appropriate replication or joint-analysis follow up is required and would include genotyping of *loci* that are significant down to a less stringent *p* value. \*, #, ‡, + each of these indicate closely associated SNPs

#### 4.5.2. Beyond association

We embarked on a whole genome SNP genotyping project, primarily to generate publicly available genotyping data for PD patients and neurologically normal controls derived from the NINDS-funded open-access Neurogenetics repository at Coriell Cell Repositories so that these data could be mined and augmented by other researchers, and also to undertake a preliminary analysis in an attempt to localize common genetic variation exerting a large effect on risk of PD in a cohort of white North Americans. However, analysis of B allele frequency and log R ratio metrics can be used to visualize copy number variation and tracks of homozygosity in the analyzed samples. As explained in section 3.6.5, the log R ratio gives an indirect measure of copy number of each SNP by plotting the ratio of observed to expected hybridization intensity. B allele frequency plots the proportion of times an allele is called A or B at each genotype: thus the expected ratios are 1.0 (B/B), 0.5 (A/B) and 0.0 (A/A). Using these two metrics, we scored copy number variation and tracks of contiguous homozygosity both in the PD cohort and in the neurologically normal controls.

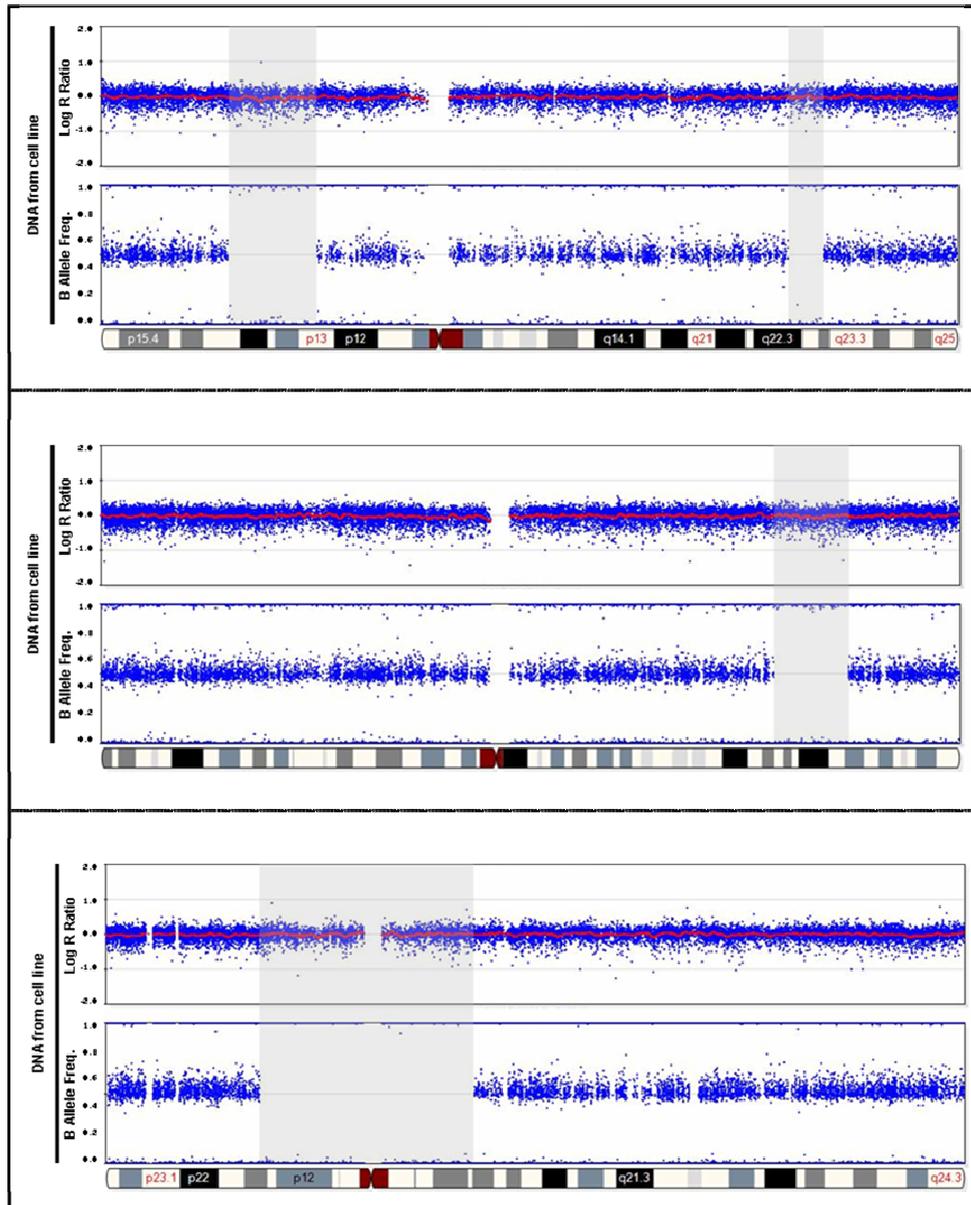
**4.5.2.1. Extended homozygosity in control population:** Using a cut-off of 5Mb, 26 neurologically normal control samples were found to harbor at least one track of consecutive homozygous SNPs (9.5% of the neurologically normal control cohort). Out of these 26 samples, 11 (42.3%) showed more than one region of homozygosity greater than 5 Mb in size. If this co-occurrence was due to chance alone, we would expect a total of  $\sim 2.44$  control samples ( $0.095 \times 0.095 \times 271$  control samples) with more than one homozygous track bigger than 5Mb. Thus, the co-occurrence detected here is 4.5-fold

greater than expected by chance alone. This supports the idea that parental consanguinity is the cause (table 29, figure 28).

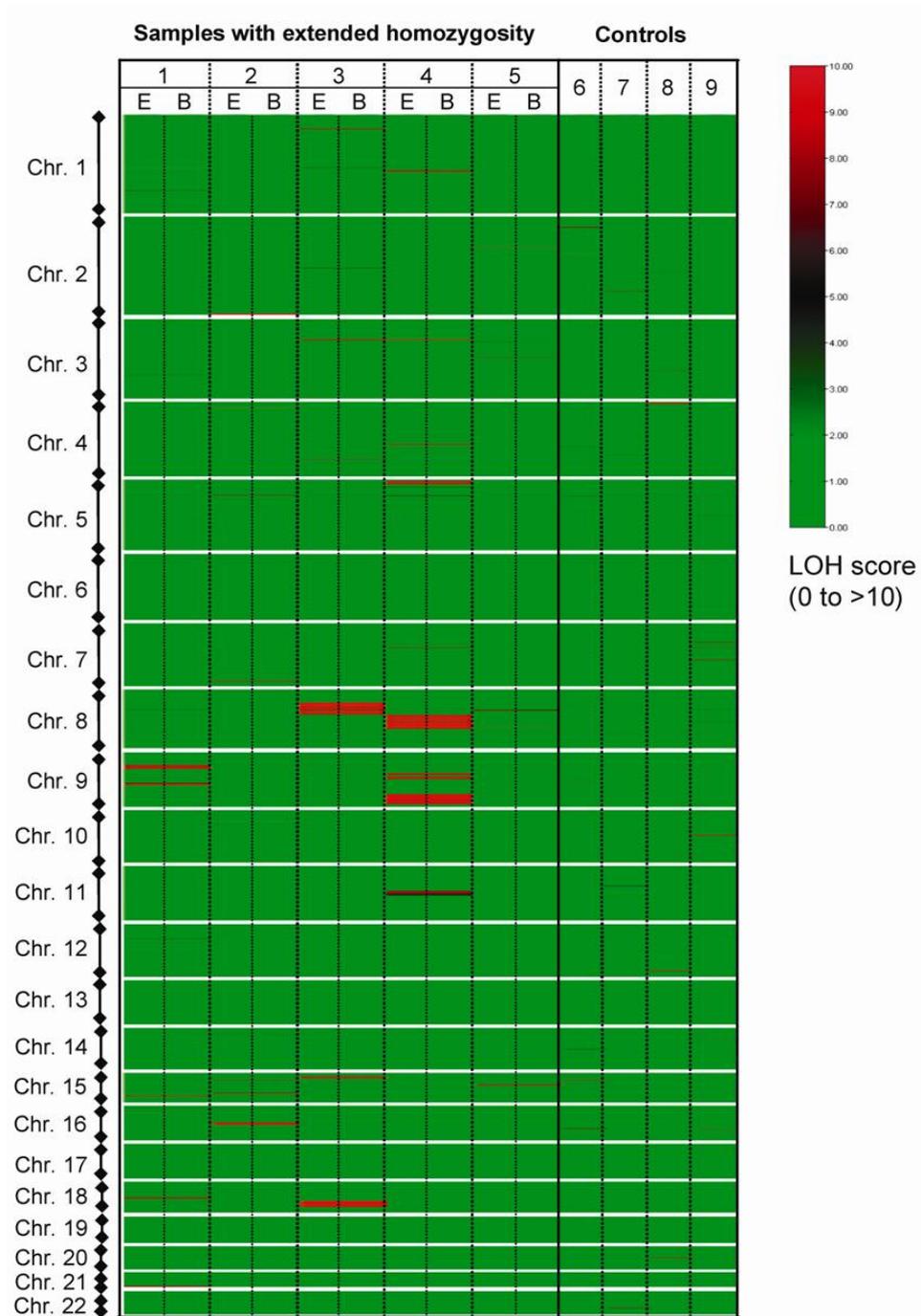
Genotyping six of these samples using DNA extracted directly from blood, revealed identical patterns of homozygosity in the replicate DNA samples, discarding the possibility that lymphoblast cell line (LCL) creation and cell culture is the cause (figure 29).

**Table 29:** Regions of contiguous extended homozygosity >5 Mb identified in 26 of 272 individuals analyzed (9.5%).

Member	Chromosome	Size (Mb)	Member	Chromosome	Size (Mb)
ND01217	3p12.2-12.1	5.45	ND04275* <sup>+</sup>	5p15.2-p15.3	9.66
ND01317	3p12.2-12.1	5.23		8q12-q22	39.68
ND01565	3q25.1-25.32	5.37		9q21	6.21
ND01567	3q25-q26.3	17.47		9q21-q22	8.38
	5q35.2-35.3	5.93		9q33-q34	15.85
	11p13-p15	14.10	11q11-q13	11.05	
	11q22-q23	5.58	7q31	5.29	
	12q23.2-23.3	6.05	ND04276	15q11.2-q14	14.32
ND02422	18q21-q22	23.66	20p11.2-q12	15.28	
	21q21-q22	5.99	ND04305* <sup>+</sup>	9p13-p21	9.24
	11p11.2-q11	5.83		9q22-q31	5.15
ND03549	1p35.1-p36.11	9.74	ND04312	6q22-q23	5.53
	3q26.31-26.33	6.83	11p11.1-p14	28.15	
ND03713	6p11.1-p12	21.72	ND04404	9q32-33.1	6.59
	4q12-q13	13.08	ND05283	3p21.31-21.2	5.24
	17p11.1-p11.2	7.64	ND05296	1p31	14.91
ND03715	22q12-q13	9.83	ND05369	2q36-q37	6.47
	3q25-q26.1	6.10	ND05372*	1p35	6.26
ND04016* <sup>+</sup>	15q24-q25	4.96	ND00674* <sup>+</sup>	8p21-q12	36.74
	16q12.1-q13	7.06		15q14-q15	5.83
ND04104* <sup>+</sup>	15q21	5.37	18q21-q22	14.58	
	1p22-p31	19.70	ND00677	11p11.1-p11.2	8.31
ND04178	8p11.1-p21	17.96	ND00689	3p21.31-21.2	5.11
	21q21	12.54	ND00707	10q26	8.57
ND04214	14q21	6.62		17q24-q25	5.82
ND04240	8q24.2-q24.3	5.24			



**Figure 28: Data from Human-1 BeadChips revealing extended homozygosity.** Extended homozygosity detected on chromosome 3 (top panel), chromosome 8 (middle panel) and chromosome 11 (bottom panel) in samples *ND00674* (chromosome 8) and *ND01567* (chromosomes 3 and 11). Each panel includes two plots. The upper plot is the log R ratio, which is a measure of copy number for each SNP. The mean of the log R ratio (over a 50 kb sliding window) is denoted by a red line. The lower plot shows B allele frequency, and shows genotypes for BB (B allele frequency = 1), AB (B allele frequency = 0.5) and AA (B allele frequency = 0). A lack of AB genotypes across a large contiguous area, in the absence of an alteration in copy number is indicative of a large area of homozygosity. The homozygous regions are indicated by gray shading



**Figure 29: Composite heat map of Loss Of Homozygosity (LOH) statistics calculated using BeadStudio.** LOH statistics were calculated for a moving window of >1Mb. Lanes 1-5 each contain duplicate samples designated with homozygous regions >5Mb, each lane shows results from a pair of samples, one from DNA extracted from EBV immortalized cell lines (E) and one from DNA extracted directly from blood (B). Lane 1 *ND04305*; lane 2 *ND04016*; lane 3 *ND00674*; lane 4 *ND04275*; lane 5 *ND04104*. Lanes 6-9 are 4 representative samples without extended homozygosity.

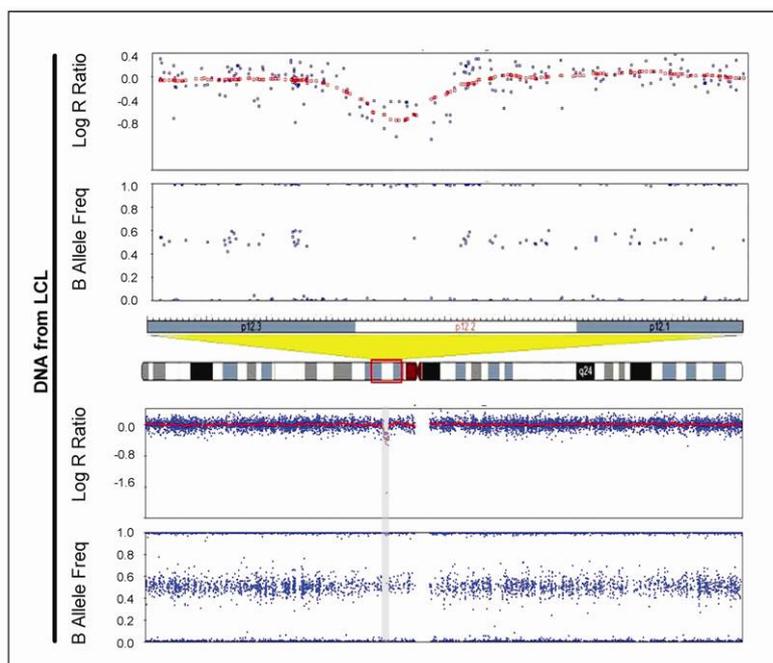
#### 4.5.2.2. Copy number variation in control population:

• **Genomic deletions:** Analysis of log R ratio and B allele frequency in each of the 271 samples assayed with the Human-1 BeadChips revealed data consistent with genomic deletion in nine samples, six sub-chromosomal (ranging in size from 0.6 to 32.9 Mb) and three affecting the whole chromosome X (table 30).

Of those sub-chromosomal deletions, only one (2.6M), on chromosome 3 of *ND01493* was consistent with simple heterozygous deletion. This was indicated by a decrease in the log R ratio and a lack of heterozygous SNP calls (figure 30). For the remaining five, we observed contiguous drops in log R ratio, consistent with a decrease in copy number. However, the observed drop in log R ratio was not coupled with an absence of heterozygous calls in these regions, thus excluding simple, heterozygous genomic deletion (table 30). Analysis at the sites of interests of B allele frequency revealed four clusters of contiguous genotypes, rather than the expected three. These were at 1 (100% B allele), 0 (100% A allele) and at two other locations, one situated between 0.5 and 1 (cluster c) and one between 0 and 0.5 (cluster d). Within each sample, the average position for cluster c was the same as 1 minus the average position for cluster d, and these cluster positions ranged approximately from B allele frequency 0.15 and 0.85 to 0.45 and 0.55. The observed alterations are consistent with heterosomic deletions, where a heterozygous genomic deletion is present in only a proportion of the cells contained within a sample (figure 31).

**Table 30:** Location, size and replication of detected structural alterations.

<b>Individual</b>	<b>Chr.</b>	<b>Approximate Size (Mb)</b>	<b>Alteration in EBV DNA</b>	<b>Alteration confirmed in DNA from blood</b>
<i>ND04019</i>	11p12	0.2	Duplication	Confirmed
<i>ND03628</i>	16q23	0.2	Duplication	Confirmed
<i>ND05067</i>	9q32	0.3	Duplication	Confirmed
<i>ND01525</i>	14q21	0.3	Duplication	Confirmed
<i>ND05016</i>	1q44	0.4	Duplication	Confirmed
<i>ND02214</i>	7q11	0.5	Duplication	Confirmed
<i>ND05052</i>	2p22	0.7	Duplication	Confirmed
<i>ND04274</i>	17q25	0.8	Duplication	Confirmed
<i>ND01678</i>	12p11-q12	0.9	Duplication	Confirmed
<i>ND01702</i>	15q13	1.0	Duplication	Confirmed
<i>ND01934</i>	13q12	1.3	Duplication	Confirmed
<i>ND01493</i>	3p12	2.6	Deletion	Confirmed
<i>ND04946</i>	13q12-q21	32.9	Partial Deletion	Confirmed
<i>ND01577</i>	22q11.2	0.5	Partial Deletion	Not confirmed
<i>ND03792</i>	22q11.2	0.5	Partial Deletion	Not confirmed
<i>ND01354</i>	8q21.3	1.1	Duplication	Not confirmed
<i>ND01684</i>	16p13.3	2.8	Duplication	Not confirmed
<i>ND02760</i>	5p14-p15.1	6.6	Partial Deletion	Not confirmed
<i>ND01569</i>	15q26	10.1	Partial Deletion	Not confirmed
<i>ND00412</i>	6p22-25	26.6	Duplication	Not confirmed
<i>ND00674</i>	12	Entire Chr.	Duplication	Not confirmed
<i>ND01552</i>	X	Entire Chr.	Partial Deletion	Not confirmed
<i>ND02212</i>	X	Entire Chr.	Partial Deletion	Not confirmed
<i>ND01424</i>	X	Entire Chr.	Partial Deletion	Not confirmed



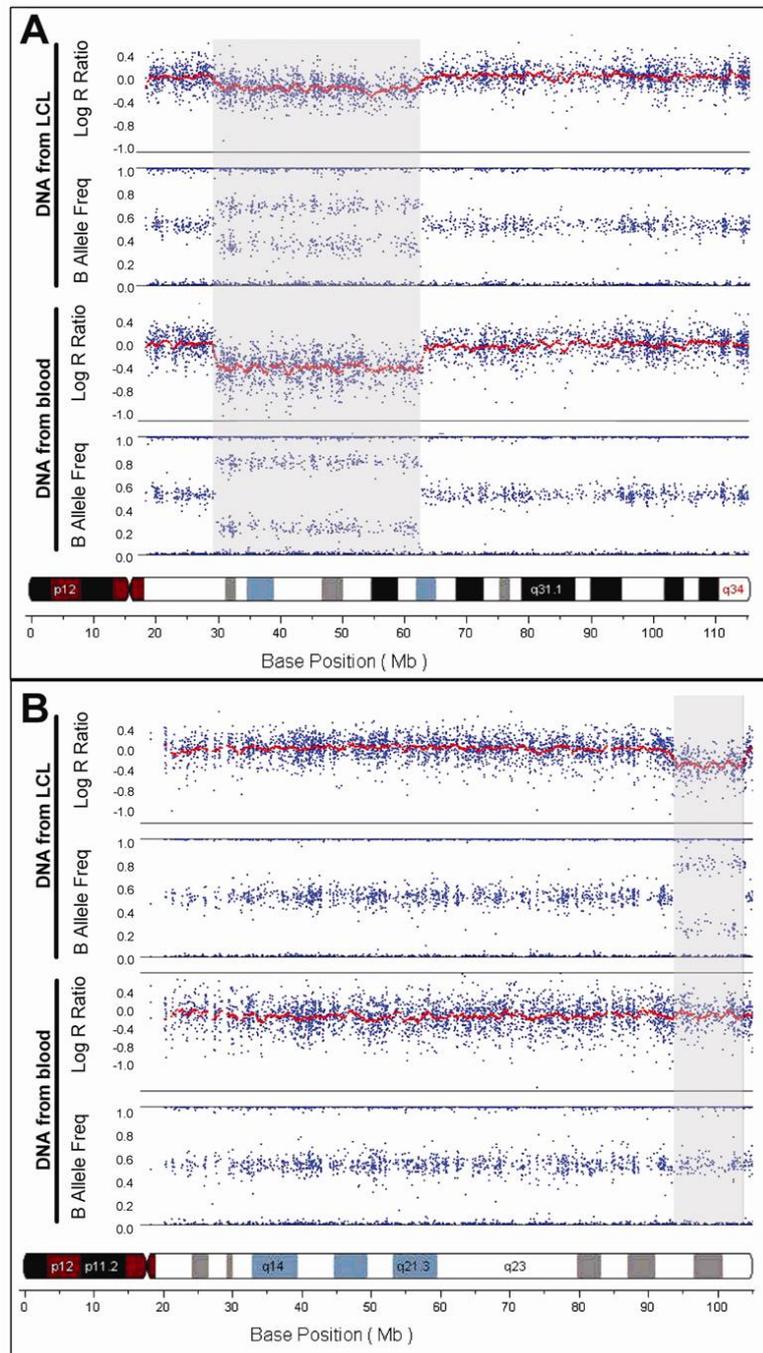
**Figure 30: Data from Human-1 BeadChips revealing genomic deletion in LCL *ND01493*.** This deletion consisted of a 2.6 Mb region at chromosome 3p12. A high resolution view of the deleted region is shown in the upper panel, a lower resolution whole chromosome view is shown in the low panel. The deletion is identified by a drop in the log R ratio (indicating a decrease in copy number) and a lack of heterozygous genotype calls in the area, as seen in the B allele frequency plot. This deletion was also identified in the source blood used for LCL creation.

Analysis of the scan data obtained from each of the 271 samples using the HumanHap300 beadchips confirmed all nine deletions and revealed the existence of 119 more, giving a total of 128, of which 28 were heterosomic (supplementary material, table S2). Comparison of observed deletion events with those published in the database for genomic variants (<http://projects.tcag.ca/variation/>) revealed that 34 of the 100 identified genomic deletions overlapped with previously reported structural variants.

In an attempt to examine whether the LCL creation and culture process lead to the appearance of the genomic deletions noted, we re-assayed DNA extracted directly from blood for all samples in which this type of mutation was noted in the Human-1 BeadChip

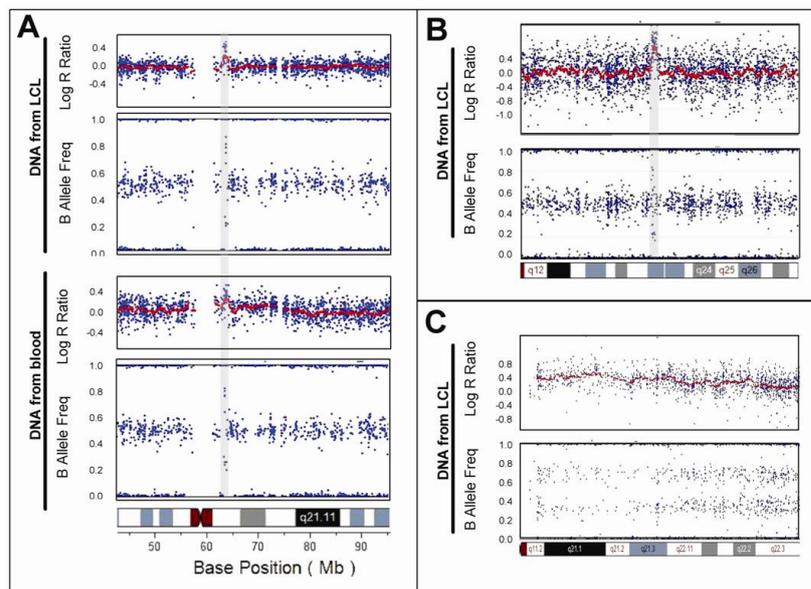
data (table 30). The blood sample corresponding to *ND01493* confirmed the presence of the simple heterozygous deletion in the source tissue.

Only one of the eight heterosomic deletions, detected in *ND04946*, was detected in these replicate experiments (figure 30). Notably the mean cluster positions for clusters c and d described above differed between the DNA extracted from blood and the DNA extracted from the cell line for this subject. These data indicate that this heterosomic deletion was present in the blood of this patient, and that the EBV immortalization process and subsequent culture have altered the ratio of cells with and without the deletion. In this specific instance there is decreased heterosomy in the DNA extracted from the cell line, indicating preferential immortalization and/or culture of cells with a full diploid genome. This region of chromosome 13 covers 32.9 Mb, contains ~189 genes, and encompasses the minimal common deleted region described for multiple myeloma and chronic lymphocytic leukemia (Migliazza, Bosch *et al.*, 2001; Elnenaei, Hamoudi *et al.*, 2003). Also of note, two of the heterosomic deletions identified in LCLs but not in blood, were apparently identical in size and position. These were identified in *ND01577* and *ND03792* at 22q11.2 and span a 0.5 Mb region containing the immunoglobulin lambda gene cluster, a region previously described to be duplicated (Sebat, Lakshmi *et al.*, 2004). These events probably represent V(D)J-type recombination either in the LCLs or as a rare event in blood, amplified by the process of LCLs moving toward a clonal state. The five remaining deletions, all apparently heterosomic, were not identified in the original blood samples, suggesting that they were either an artifact of the EBV immortalization process, or a biased representation of the donor tissues state at immortalization, caused by LCL creation and culture.



**Figure 31: Log R ratio and B allele frequency plots for observed heterosomic deletions using Human-1 BeadChips.** Each panel consists of a log R ratio plot (top) and a B allele frequency plot (bottom) across a single chromosome for DNA extracted from LCLs and directly from blood. The regions containing apparent deletions are shaded gray. **A:** An apparent heterosomic deletion of 32.9 Mb in chromosome 13 in *ND04946*, present both in DNA extracted from EBV immortalized lymphocytes and from blood. **B:** An apparent heterosomic deletion of 10.1 Mb in chromosome 15 in *ND01569*, present in DNA extracted from EBV immortalized lymphocytes but not visible in DNA extracted from the blood sample used for the creation of this LCL.

• **Genomic multiplications:** Results indicative of genomic multiplication were identified in 15 LCL-derived samples after analysis with the Human-1 BeadChip, ranging in size from 0.2 Mb to the entire chromosome. The profile indicative of a genomic multiplication is an increase in log R ratio in the presence of allele clusters outside the expected homozygous or heterozygous range for the value of B allele frequency (figure 32A). As positive controls for the utility of this approach to detect genomic copy number changes, we assayed DNA samples derived from a subject who carries the *SNCA locus* triplication, ~1.6 Mb in length (Singleton, Farrer *et al.*, 2003) and a subject with trisomy 21. Data from these samples clearly showed the triplication mutation and the presence of an additional copy of chromosome 21 (figures 32B and C).



**Figure 32: Analysis of subjects with genomic multiplication mutations using Human-1 BeadChips.** Each panel contains log R ratio and B allele frequency plots across a single chromosome. **A:** Increased log R ratio and divergence of B allele frequency away from normal clusters for A/A (0), A/B (0.5) and B/B (1) genotypes reveals a 0.5 Mb duplication in chromosome 7 in LCL *ND02214* (gray shading). The duplication is clearly present in DNA extracted from both the cell line and the blood sample used to create the line. **B:** Analysis of a positive control DNA known to harbor a 1.6 Mb triplication mutation across the *SNCA locus* on chromosome 4 (gray shading) indicated by an increase in log R ratio and divergence of B allele frequency away from normal clusters for A/A (0), A/B (0.5) and B/B (1). **C:** Analysis of a positive control DNA extracted from the blood of a person with trisomy 21. The chromosomal-wide increase in log R ratio and the presence of 4 genotype clusters (A/A/A, A/A/B, A/B/B and B/B/B) is consistent with trisomy.

Again, analysis of the scan data of the HumanHap300 BeadChip, confirmed all 15 alterations and revealed 197 more (supplementary material, table S3).

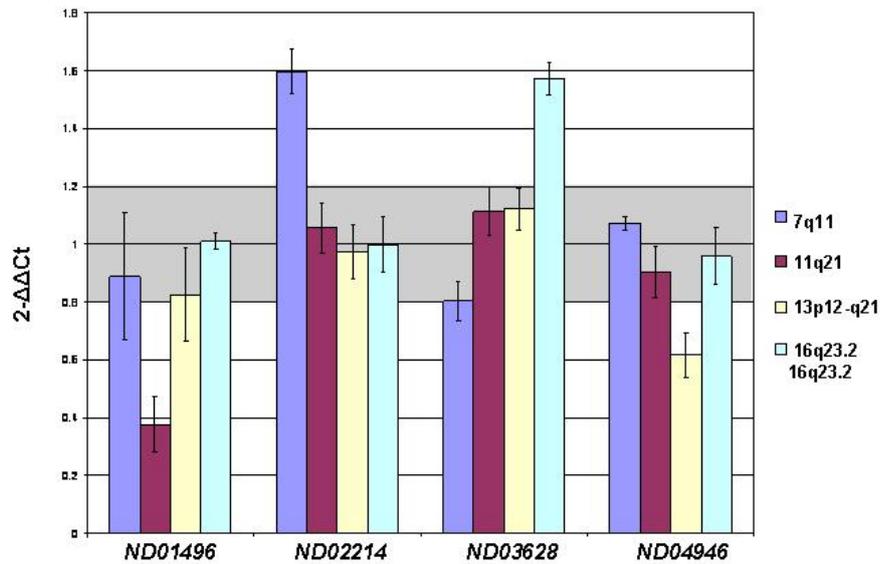
Comparison of observed deletion and duplication events with those published in the database for genomic variants (<http://projects.tcag.ca/variation/>) revealed 77 of 212 observed sub-chromosomal multiplications overlapped with previously reported structural mutations.

Analysis of DNA extracted directly from the blood used for the creation of the LCLs in which genomic duplication was noted revealed 11 of the 15 alterations were also present in the source tissue. The genomic multiplications with discordant results between LCLs and blood comprised four of the five largest multiplications identified (table 30).

Interestingly, four duplications affecting the *PARK2 locus* were found in four different control samples, affecting exon 2 in *ND04788* and *ND01583*, exon 7 of *ND05536*, and exons 2, 3 and 4 of individual *ND05093* (table 31).

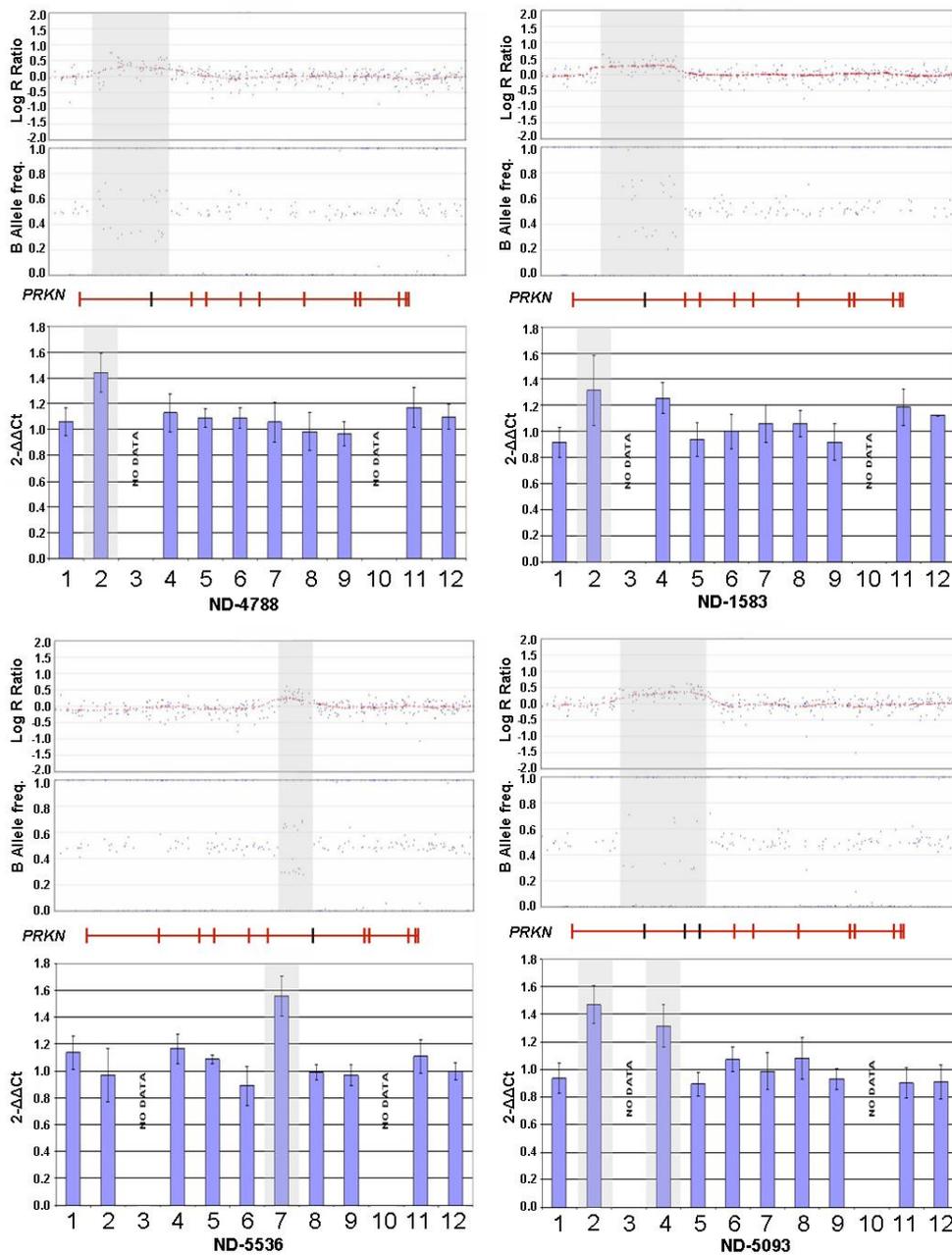
. **Confirmation of structural changes by real-time PCR:** In order to confirm that the structural alterations found in these samples were not artifacts of the methodological or analytical processes, quantitative PCR was performed in all 271 samples using 14 TaqMan probes. This was performed using probes for a 0.5 Mb duplication at chromosome 7q11 (not previously described), a 0.019 Mb deletion in chromosome 11q21 (not previously described), a 33 Mb heterosomic deletion in chromosome 13p12-q21 (overlapping with TCAG's variant number 0313) and a 0.11 Mb duplication in chromosome 16q23.2 (not previously described). In all instances, structural changes predicted by Infinium analyses were confirmed according to the  $2^{-\Delta\Delta C_t}$  value

obtained (Corporation, 1997) (figure 33). No additional carriers of any of these structural changes were found, supporting the sensitivity of this approach.



**Figure 33: Data from quantitative real-time PCR experiments demonstrating genomic structural changes in control subjects.**  $2^{-\Delta\Delta C_t}$  value obtained after performing real-time PCR for target regions in chromosomes 7q11, 11q21, 13p12-q21 and 16q23.2. Grey shading indicates the  $2^{-\Delta\Delta C_t}$  value expected from a normal diploid genome; a value below this range indicates loss of copy number and a value above this range an increase in copy number. Data shown is assay for these four loci in four samples, each of which contained a copy number change in one of these loci as indicated by initial GW-SNP assay. ND01496 shows deletion at 11q21; ND02214 shows duplication at 7q11; ND03628 shows duplication at 16q23.2; ND04946 shows a value indicative of heterosomic deletion at 13p12-q21. In all instances the duplication or deletion events predicted by Infinium data were confirmed by real-Time PCR.

In addition, 10 individual probes across *PARK2* were used to verify the presence and size of the four unique duplications observed in this region in individuals ND01583, ND05536, ND04788 and ND05093. Structural changes were confirmed in all instances. No additional carriers were found, supporting the sensitivity of this approach (figure 34).



**Figure 34: Data from Infinium II and quantitative real-time PCR experiments demonstrating genomic duplication events in control subjects.** An increase in log R ratio (upper panels) indicates an increase in the gene dosage of the highlighted area (grey shading). The B allele frequency plots show the appearance of unexpected allele clusters indicative of genotypes A/A/A, A/A/B, A/B/B and B/B/B, indicative of duplication. The lower panel plots the  $2^{-\Delta\Delta C_t}$  value obtained after performing real-time PCR for exons of *PARK2* in these samples. In all instances the duplication events predicted by the log R ratio and B allele frequency metrics produced from analysis of Infinium II data were confirmed by real time PCR (indicated by grey shading of  $2^{-\Delta\Delta C_t}$  values in each plot).

#### 4.5.2.3. Copy number variation in PD population:

. **Overall structural variation:** Analysis of the B allele frequency and log R ratio metrics from both arrays, revealed both subchromosomal deletion and multiplications, and whole chromosome duplications. Thus, sample *ND01532* showed heterosomic deletion of the entire chromosome 9, samples *ND00200* and *ND02978* heterosomic deletion of chromosome 12, and samples *ND00171*, *ND00740* and *ND00410* heterosomic deletion of chromosome X.

As for the subchromosomal changes, different number and size of CNVs was detected depending on the beadchip utilized. Thus, Infinium Human-I analysis revealed data consistent with subgenomic multiplications in 35 different PD samples, ranging in size from a 0.074 Mb duplication in chromosome 6q27 of individual *ND04177* to a 88.4 Mb one in chromosome 5q15-35.3 of individual *ND01216*. Results indicative of subgenomic deletion were identified in 14 North American PD samples, 10 heterosomic deletions and 4 simple heterozygous deletions ranging in size from 1.23 Mb on chromosome 18q22.1 of individual *ND00216* to 3.36 Mb on chromosome 18q22.2-q22.3 of individual *ND02854*.

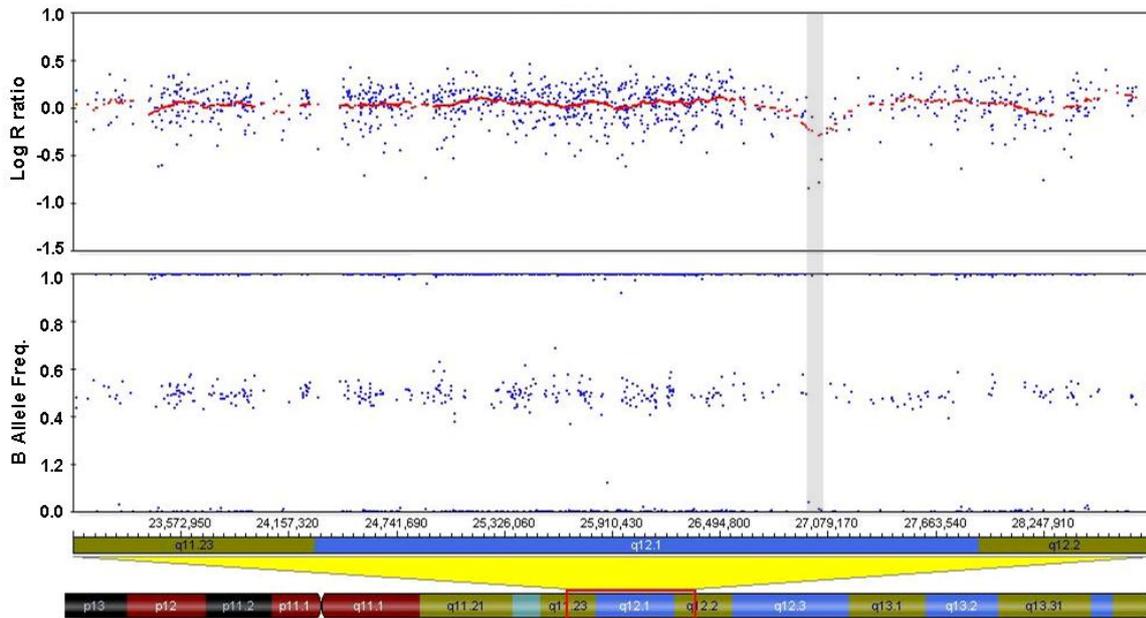
Analysis of the B allele frequency and log R ratio metrics corresponding to the Illumina Infinium HumanMap300 data, confirmed all the subgenomic structural changes found using the Illumina Infinium Human-I platform, in addition to 147 new multiplication events (182 in total in 130 different samples) and 157 new heterozygous deletions (161 in total in 114 different samples) (supplementary material, tables S4 and S5).

The smallest events found were a 0.03 Mb duplication on chromosome 22q13.33 of individual *ND02575* and a 0.016 Mb deletion on chromosome 20p11.21 of *ND02918*.

Thirty seven new heterosomic deletions were identified. As we demonstrated in the results we obtained in our control population (previous section) this is likely to result from the Epstein-Barr virus (EBV) transformation and cell culture process, being absent in the source DNA. Of these 47 heterosomic deletions, 23 were located on chromosome 22q11.22 and span a region containing the immunoglobulin lambda gene cluster, a region previously described to be duplicated (Sebat, Lakshmi *et al.*, 2004). These events probably represent V(D)J-type recombination either in the LCLs or as a rare event in blood, amplified by the process of LCLs moving toward a clonal state. Likewise the additional heterosomic deletions, as well as the largest structural variants, are likely to be artifacts of the LCL creation and culture process.

Comparison of observed heterozygous deletion and duplication events with the changes previously identified in our control population and those published in the Database for Genomic Variants (<http://projects.tcag.ca/variation>) revealed that 144 out of 182 (79.1%) identified duplications overlapped with previously reported structural variants and 117 out of 161 (72.6%) observed multiplications overlapped with previously reported structural mutations. Since the Database of Genomic Variants publishes only variation data identified in control populations, it is unlikely, although not impossible, that these changes are directly related to the pathogenesis of PD in this population. While we can not rule out that the remaining alterations (38 duplications and 44 deletions) harbor genes related to the etiology of PD in this population only one region in chromosome 22q12.1, had structural changes affecting more than one single sample and absent from controls. These changes were three different heterozygous deletions in chromosome 22q12.1 in individuals *ND01683*, *ND04397*, and *ND02631*. These deletions spanned 0.05, 0.06, and

0.26 Mb, respectively, and involved exons 2, 3, and 4 of *TCC28* (Tetratricopeptide RepeatDomain 28) predicted transcript (figure 35). To check whether these deletions were actually heterozygous deletions and not artifact, we performed real-time PCR at exons 2 and 4 of *TTC28* in new stock DNA from individuals *ND01683*, *ND04397*, and *ND02631*. The  $2^{-\Delta\Delta C_t}$  values obtained did not confirm the results expected, ranging in all three samples from 0.8 to 1.2 (Corporation, 1997). Analysis of new DNA stocks for samples *ND01683*, *ND04397*, and *ND02631* using Infinium HumanHap240S arrays, confirmed that no heterozygous deletion events were present at this *locus* in any of these three new sample aliquots; however, the genotype concordance between overlapping SNPs on the HumanHap240S and Human-I arrays reassured us that these samples were derived from the same individuals. Thus the observed *TTC28* deletions were likely an artifact resulting from the LCL creation and culture process (figure 35).



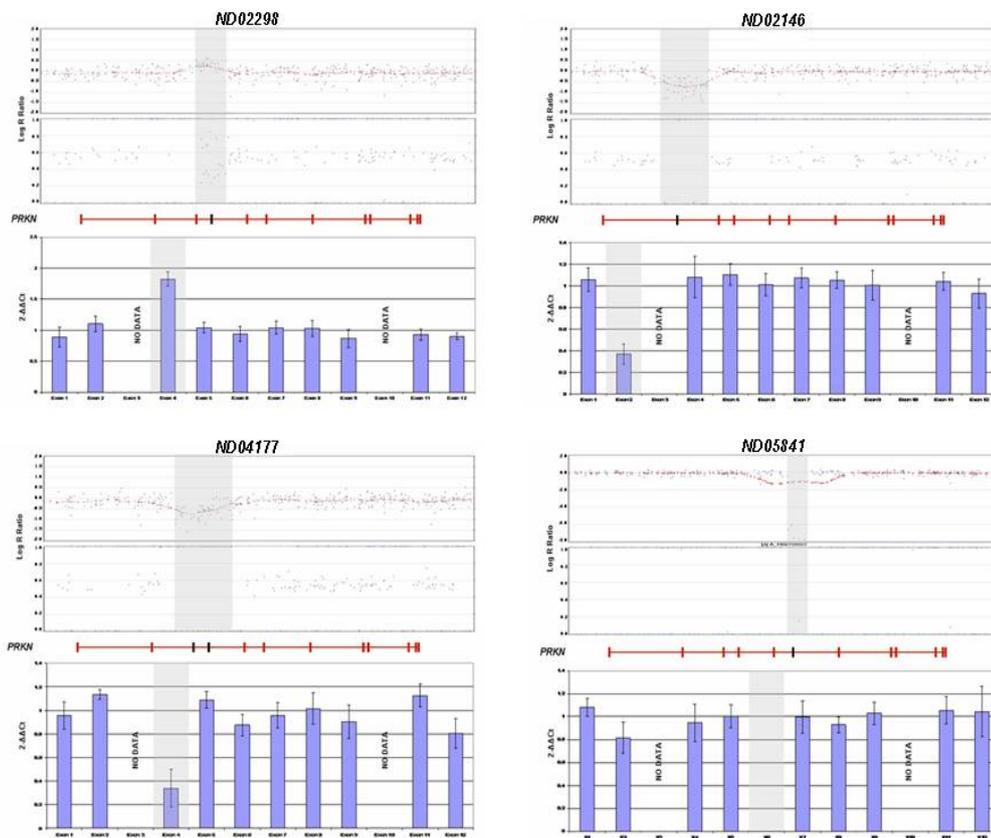
**Figure 35: Data from HumanHap300 beadchips revealing heterozygous deletions affecting *TCC28* in *ND01683*.** The upper plot show B allele frequency, and shows genotypes for BB (B allele frequency = 1), AB (B allele frequency = 0.5) and AA (B allele frequency = 0) in three different PD samples. The lower plot is the log R ratio, which is a measure of copy number for each SNP. The mean of the log R ratio is denoted by a grey line. The deletion is identified by a drop in the log R ratio and a lack of heterozygous genotype calls.

• **Structural variation in PD-related loci:** No structural changes were found in *SNCA*, *PARK7*, *PINK1*, or *LRRK2*. However, structural variation was found in *PARK2* (table 31). A single PD sample (*ND02298*) showed a duplication within *PARK2*, involving exon 4. Besides, data consistent with genomic heterozygous deletion was found in PD samples *ND02146* and *ND04177* (figure 36, table 31). The former indicated a 0.125 Mb deletion involving exons 3 and 4, the latter a 0.218 Mb deletion involving only exon 2. To note, no deletions were found in DNA samples from the 271 control samples analyzed (see previous section).

Interestingly, we also identified a homozygous deletion affecting *PARK2* in sample *ND05841* (figure 35, table 31). Because this mutation would typically cause YOPD, we

reexamined the subject details for the samples included within this cohort, this revealed that two young-onset patients, including *ND05841* (age at onset 13 years), had erroneously been included within *NDPT007*.

All copy number variation events affecting *PARK2* were confirmed by means of quantitative PCR assay according to the  $2^{-\Delta\Delta C_t}$  value obtained (Corporation, 1997).



**Figure 36: Data from HumanHap300 and quantitative real-time PCR experiments demonstrating genomic duplication and deletion events in PD patients. In all instances the duplication and deletion events predicted by the log R ratio and B allele frequency metrics produced from analysis of HumanHap300 data were confirmed by real time PCR (see figure 33).**

**Table 31:** Structural Genomic Copy Number Variants Identified in *PARK2* in both control and PD samples. Only deletion affecting ND02146 was in frame (F).

Sample ID	Status	Change	SNP 3'	Position 3'	SNP 5'	Position 5'	Size (Mb)	Exons involved
ND05841	Affected	Deletion	rs1790024	162283949	rs10945791	162336195	0.052	6
ND02146	Affected	Deletion	rs2023075	162534963	rs9356000	162659755	0.125	3 and 4 (F)
ND04177	Affected	Deletion	rs9364646	162629938	rs2023001	162847793	0.218	2
ND02298	Affected	Duplication	rs1954926	162436616	rs9458486	162547280	0.111	4
ND01583	Not affected	Duplication	rs7739802	162623409	rs9456785	162863945	0.241	2
ND05536	Not affected	Duplication	rs11968578	162122933	rs6455783	162243987	0.09	7
ND04788	Not affected	Duplication	rs7752854	162714014	rs2803108	162985009	0.271	2
ND05093	Not affected	Duplication	rs4605857	162485708	rs2023001	162847793	0.362	2, 3 and 4

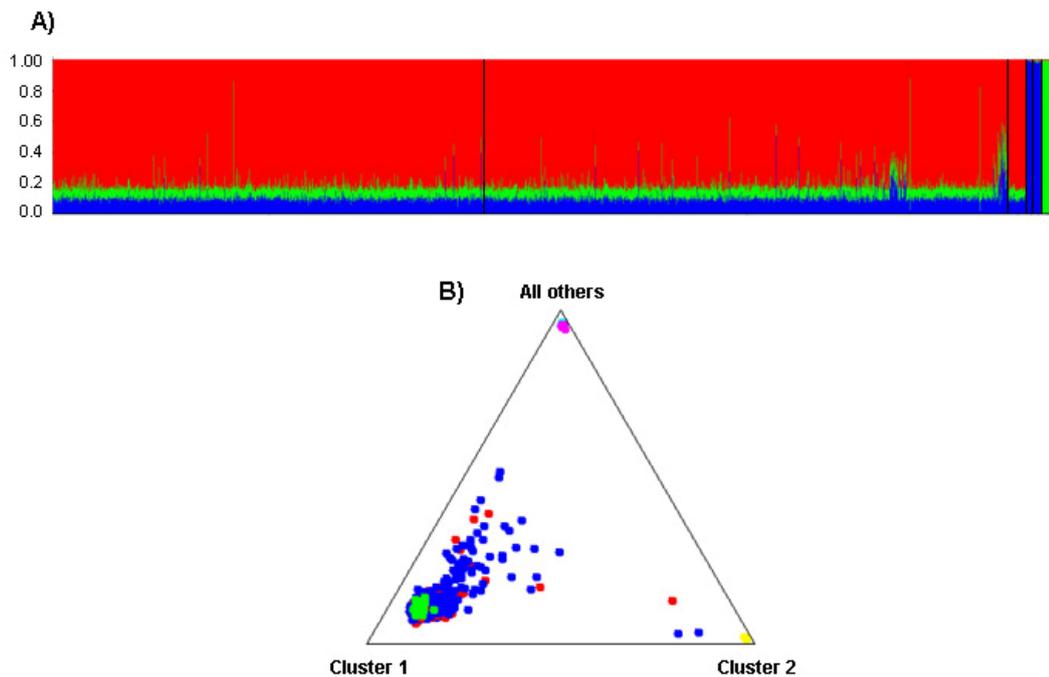
#### 4.6. WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 2

By means of combining genotyping results from different Infinium platforms (see section 3.7.2) we obtained a total of 545,066 unique SNPs genotyped in a total of 880 patients with late-onset PD, 183 with YOPD and 828 neurologically normal controls from the NINDS-funded Coriell institute for research. This dataset was increased with genotypes from 2,243 control individuals generously shared by the Cancer Genetic Markers of Susceptibility initiative, hosted by the National Cancer Institute (CGEMS, <http://cgems.cancer.gov/>), giving us a total of 3071 control samples.

##### 4.6.1. Quality-control approaches.

Low call rate genotyping led us to remove 41 individual samples, including 11 late-onset PD, 5 YOPD, 13 neurologically normal controls from the Coriell Institute and 12 controls from the CGEMS initiative.

Analysis of 10,000 SNPs in linkage equilibrium using STRUCTURE v2.2 (<http://pritch.bsd.uchicago.edu/structure.html>) showed that except for three individuals with genetic background indicative of African ancestry, the rest of our sample clearly shared Caucasian ancestry. Since population stratification and population admixture are one of the most common confounds in genome-wide studies, those three samples were dropped from further analysis (figure 37). These results were confirmed by plotting IBS distances of each sample to its “nearest neighbor” along with that same value for each of the samples included in the HapMap project. In this case, control samples from the CGEMS initiative were also included (figure 38).



**Figure 37: Bar and triangle plots from Structure v2.2 using 10,000 pruned autosomal SNPs.** A) Bar plot for K = 3 sorted by putative population where population 1 consists of all Coriell PD cases and controls plus those 30 Caucasian trios from the HapMap project; population 2 consists of 45 Chinese and 45 Japanese individuals from the HapMap project, and population 3 in 30 Yoruban trios also from the HapMap project. This plot shows how our cohort (except for the three outliers) genetically resembles those Caucasian individuals from HapMap and differs from both the Asian and the African samples. B) Triangle plot with same putative populations as bar plot A where red dots are controls, blue dots are PD cases and green, pink, light blue and yellow are Caucasian, Japanese, Chinese and Yoruban samples from the HapMap project. This plot clearly shows how two cases and one control sample from our cohort, is closer genetically to African than Caucasian population.



worldwide laboratories (including ours) by the Coriell Institute. All these samples were also dropped from further analysis.

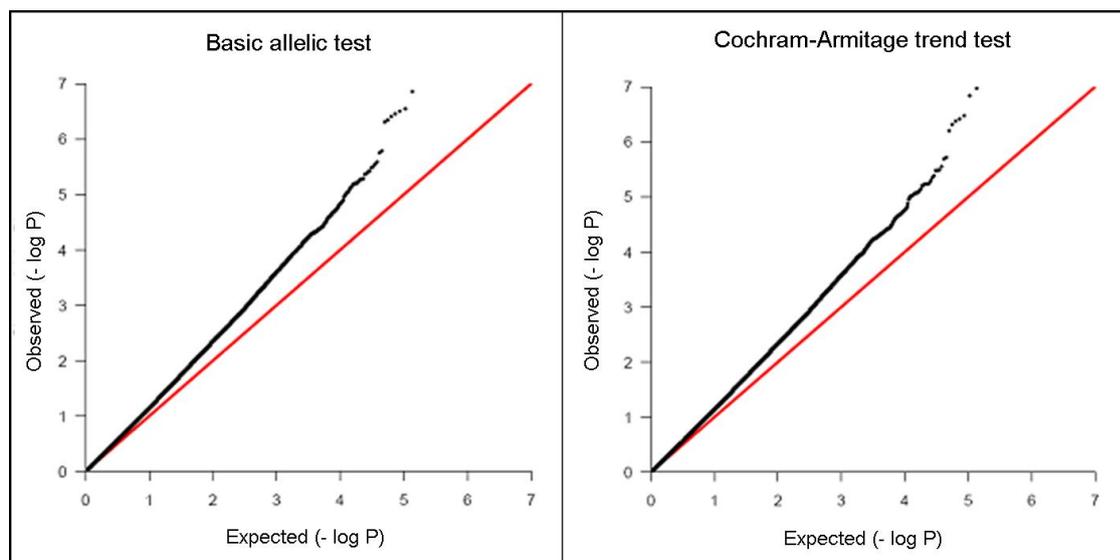
Last, 15 samples including 9 late-onset PD, 2 YOPD and 4 neurologically normal controls from the Coriell Institute, were removed because of a mismatch between reported and genotype gender.

After this extensive quality-control phase, the final number of fully genotyped samples was 4,005 including 840 late-onset PD, 131 YOPD, 803 neurologically normal controls from the Coriell institute and 2,231 CGEMS controls.

For those 545,066 SNPs assayed in these samples the genotyping call rate was 98.7%. Those SNPs with a Hardy-Weinberg equilibrium p value smaller than 0.01, MAF smaller than 5% or missing in more than 5% in our control cohort, were excluded from further analysis, giving a total of 474,995 SNPs with a total genotyping rate of 99.82 %.

#### **4.6.2. Statistical analysis**

We assessed evidence for association in several ways including 1 degree of freedom and 2 degrees of freedom test (for more details see section 3.7.4). Corrections, including 1000 max T permutation, Bonferroni and FDR-BH methods were applied for multiple test correction. Quantile-quantile plots for the allelic association test and Cochran-Armitage trend test show a minor deviation from the null distribution, except in the extreme tails which correspond to the association reported below (figure 39). These plots, along with association in *loci* previously linked with PD (reported below) give confidence in the quality of our data and the robustness of our analysis.



**Figure 39: Quantile-quantile plots for the allelic and the trend p values distribution obtained.** Results of these two statistical approaches are shown in black for all 474,995, showing a minor deviation from the null distribution (in red) except for those SNPs at the extreme tail.

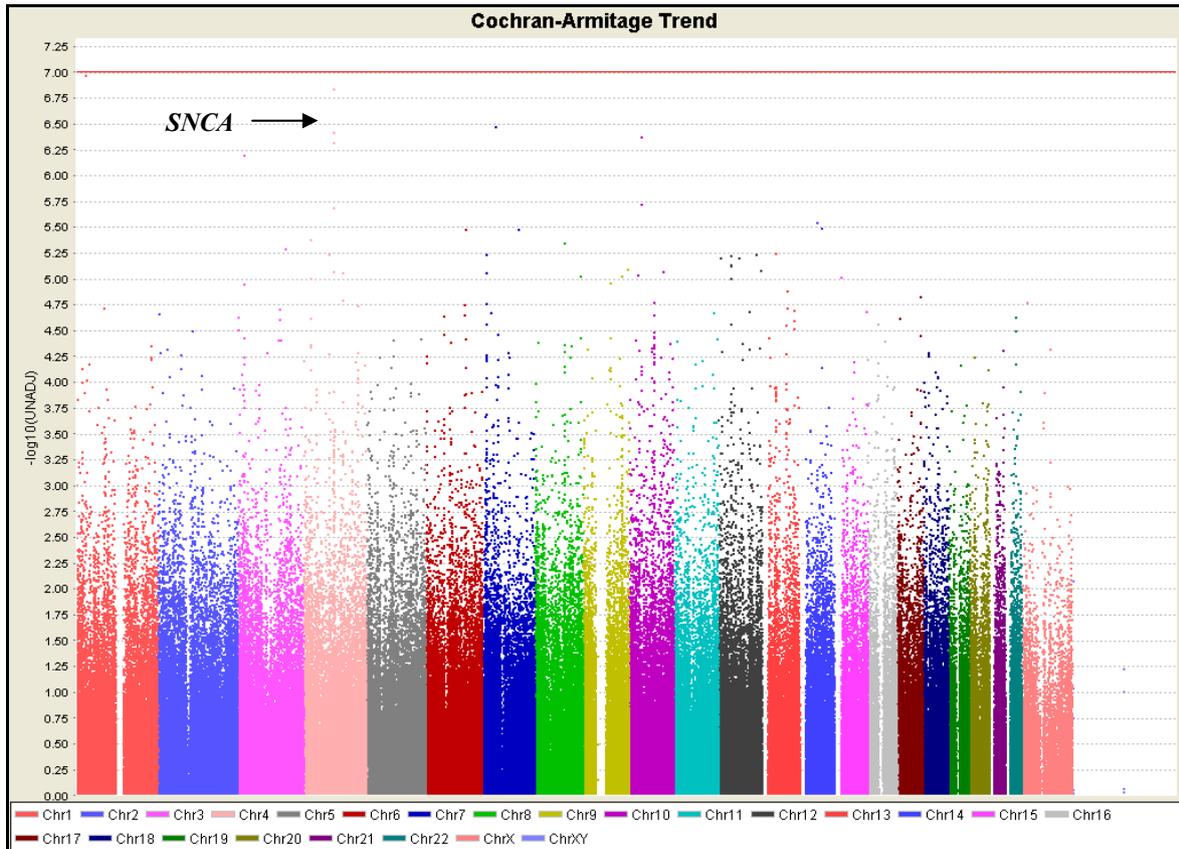
The highest association detected was for rs4889730, rs2627426 and rs3784847 with trend p values of  $1.6 \times 10^{-37}$ ,  $7.45 \times 10^{-15}$  and  $2.75 \times 10^{-12}$ . However, reexamination of B allele frequency and log R ratio clusters with Beadstudio v3.1 (see *data analysis* in section 3.6.2.2) showed that bad genotyping of these three SNPs had occurred. Thus, association results in these SNPs will be disregarded. The next 30 “hits” we obtained are displayed in table 32. This table shows p values obtained after basic allelic, Cochran-Armitage trend, genotypic and dominant/recessive tests. Interestingly 5 of these hits are in SNPs within *SNCA locus*, described as the gene most strongly associated with PD pathology.

The totality of Cochran-Armitage Trend p values for each of the 474,995 SNPs analyzed (excluding those with poor clustering after visual inspection), are plotted in figure 40.

**Table 32:** 30 lowest uncorrected p values obtained in our analysis.

Chr. Location	SNP ID	Location bp	Gene	HWE p value		Trend p value	Allelic p value	Genotypic p value	Dom/Rec p value	OR (95%CI)
				Controls	Cases					
1	rs4949254	30,567,547	<i>MATN1</i>	0.01116	3.13 x 10 <sup>-12</sup>	1.06x 10 <sup>-07</sup>	3.50x 10 <sup>-07</sup>	3.68 x 10 <sup>-11</sup>	5.86 x 10 <sup>-12</sup> R	0.740 (0.660 - 0.8317)
<b>4</b>	<b>rs2736990</b>	<b>90,897,564</b>	<b><i>SNCA</i></b>	<b>0.6878</b>	<b>0.4391</b>	<b>1.45x 10<sup>-07</sup></b>	<b>1.38x 10<sup>-07</sup></b>	<b>5.12x 10<sup>-07</sup></b>	<b>8.68 x 10<sup>-07</sup> R</b>	<b>1.317 (1.189 - 1.459)</b>
7	rs859522	38,936,691	<i>LOC400750</i>	1	0.3151	3.34x 10 <sup>-07</sup>	2.83x 10 <sup>-07</sup>	1.29x 10 <sup>-06</sup>	1.94 x 10 <sup>-06</sup> D	1.573 (1.322 - 1.872)
<b>4</b>	<b>rs11931074</b>	<b>90,858,538</b>	<b><i>BARHL2</i></b>	<b>0.5741</b>	<b>0.04077</b>	<b>3.83x 10<sup>-07</sup></b>	<b>3.13x 10<sup>-07</sup></b>	<b>1.21x 10<sup>-07</sup></b>	<b>9.77 x 10<sup>-06</sup> R</b>	<b>1.575 (1.322 - 1.877)</b>
10	rs11591754	35,247,159	<i>ZMYM6</i>	0.3753	0.8941	4.18x 10 <sup>-07</sup>	4.90x 10 <sup>-07</sup>	2.52x 10 <sup>-06</sup>	7.61 x 10 <sup>-07</sup> D	0.692 (0.600 - 0.799)
<b>4</b>	<b>rs3857059</b>	<b>90,894,261</b>	<b><i>SNCA</i></b>	<b>0.5736</b>	<b>0.03905</b>	<b>4.81x 10<sup>-07</sup></b>	<b>3.91x 10<sup>-07</sup></b>	<b>1.44x 10<sup>-07</sup></b>	<b>1.22 x 10<sup>-05</sup> D</b>	<b>1.57 (1.318 - 1.872)</b>
3	rs1605527	21,351,855	<i>ECE1</i>	0.1022	0.8599	6.26x 10 <sup>-07</sup>	4.53x 10 <sup>-07</sup>	2.73x 10 <sup>-06</sup>	4.27 x 10 <sup>-07</sup> D	1.582 (1.322 - 1.892)
10	rs2492448	35,235,412	<i>ZMYM6</i>	0.9306	0.3808	1.91 x 10 <sup>-06</sup>	1.76x 10 <sup>-06</sup>	7.22x 10 <sup>-06</sup>	1.97 x 10 <sup>-06</sup> D	0.750 (0.667 - 0.844)
<b>4</b>	<b>rs3775439</b>	<b>90,928,764</b>	<b><i>SNCA</i></b>	<b>1</b>	<b>0.09297</b>	<b>2.04 x 10<sup>-06</sup></b>	<b>1.63x 10<sup>-06</sup></b>	<b>3.01x 10<sup>-06</sup></b>	<b>4.05 x 10<sup>-05</sup> D</b>	<b>1.422 (1.23 - 1.642)</b>
14	rs12431733	53,360,580	<i>SLC1A7</i>	0.1843	0.07237	2.81 x 10 <sup>-06</sup>	2.86x 10 <sup>-06</sup>	8.60x 10 <sup>-07</sup>	1.98 x 10 <sup>-07</sup> R	1.277 (1.153 - 1.415)
14	rs1950712	68,550,317	<i>LOC149224</i>	0.9064	1	3.23x 10 <sup>-06</sup>	3.16x 10 <sup>-06</sup>	NaN	NaN	0.596 (0.478 - 0.742)
7	rs10487302	110,478,664	<i>IMMP2L</i>	1	0.1516	3.29x 10 <sup>-06</sup>	3.81x 10 <sup>-06</sup>	9.23x 10 <sup>-06</sup>	4.89 x 10 <sup>-07</sup> D	1.458 (1.242 - 1.713)
6	rs7756430	228,03,604	<i>NOTCH2</i>	0.6077	0.1044	3.30x 10 <sup>-06</sup>	2.58x 10 <sup>-06</sup>	7.58x 10 <sup>-06</sup>	1.52 x 10 <sup>-06</sup> D	0.756 (0.673 - 0.850)
4	rs6830403	22,803,604	<i>EPHA8</i>	0.5468	0.2279	4.13x 10 <sup>-06</sup>	3.30x 10 <sup>-06</sup>	1.53x 10 <sup>-05</sup>	7.01 x 10 <sup>-06</sup> D	0.776 (0.698 - 0.864)
8	rs2616510	89,087,241	<i>GTF2B</i>	0.09716	0.005525	4.47x 10 <sup>-06</sup>	4.23x 10 <sup>-06</sup>	3.13x 10 <sup>-08</sup>	1.35 x 10 <sup>-08</sup> R	1.329 (1.177 - 1.500)
3	rs7622285	146,367,022	<i>FLJ39739</i>	0.06545	0.8832	5.10x 10 <sup>-06</sup>	6.40x 10 <sup>-06</sup>	1.15x 10 <sup>-05</sup>	4.66 x 10 <sup>-05</sup> R	1.292 (1.156 - 1.444)
13	rs9525776	42,928,966	<i>LOC647049</i>	0.1421	0.791	5.62x 10 <sup>-06</sup>	4.31x 10 <sup>-06</sup>	3.17x 10 <sup>-05</sup>	3.33 x 10 <sup>-05</sup> D	1.278 (1.151 - 1.418)
12	rs12425761	114,119,258	<i>RSBNI</i>	0.755	0.7824	5.81x 10 <sup>-06</sup>	5.30x 10 <sup>-06</sup>	3.43x 10 <sup>-05</sup>	1.25 x 10 <sup>-05</sup> D	1.435 (1.227 - 1.677)
4	rs7655536	77,395,792	<i>PIGK</i>	0.8206	0.6678	5.81x 10 <sup>-06</sup>	5.75x 10 <sup>-06</sup>	3.21x 10 <sup>-05</sup>	1.15 x 10 <sup>-05</sup> R	1.321 (1.171 - 1.491)
7	rs1371608	12,373,351	<i>FLJI4712</i>	0.6851	0.1398	5.84x 10 <sup>-06</sup>	6.89x 10 <sup>-06</sup>	2.32x 10 <sup>-05</sup>	1.82 x 10 <sup>-05</sup> D	1.265 (1.142 - 1.401)
12	rs7485262	37,088,013	<i>GRIK3</i>	0.5481	0.7556	5.89x 10 <sup>-06</sup>	5.16x 10 <sup>-06</sup>	3.50x 10 <sup>-05</sup>	1.03 x 10 <sup>-05</sup> D	1.469 (1.244 - 1.735)
12	rs12366742	64,313,976	<i>RORI</i>	0.3037	0.7978	6.23x 10 <sup>-06</sup>	5.31x 10 <sup>-06</sup>	3.22x 10 <sup>-05</sup>	9.95 x 10 <sup>-05</sup> D	0.763 (0.679 - 0.857)
12	rs797770	6,173,418	<i>RPL22</i>	0.6059	0.678	6.23x 10 <sup>-06</sup>	5.44x 10 <sup>-06</sup>	3.65x 10 <sup>-05</sup>	1.07 x 10 <sup>-05</sup> D	1.561 (1.287 - 1.894)
12	rs6582668	37,052,871	<i>GRIK3</i>	0.675	0.4822	7.22x 10 <sup>-06</sup>	6.19x 10 <sup>-06</sup>	3.76x 10 <sup>-05</sup>	2.97 x 10 <sup>-05</sup> D	1.387 (1.203 - 1.6)
12	rs4508240	37,186,069	<i>GRIK3</i>	0.6143	0.4803	7.47x 10 <sup>-06</sup>	6.34x 10 <sup>-06</sup>	3.88x 10 <sup>-05</sup>	3.08 x 10 <sup>-05</sup> D	1.387 (1.203 - 1.6)
9	rs11244079	135174347	LOC653163	0.08245	0.1166	5.89 x 10 <sup>-06</sup>	5.16 x 10 <sup>-06</sup>	3.50 x 10 <sup>-05</sup>	NA	1.469(1.244-1.735)
12	rs2170504	128432246	TMEM132D	0.3047	0.2508	6.23 x 10 <sup>-06</sup>	5.31 x 10 <sup>-06</sup>	3.22 x 10 <sup>-05</sup>	6.14 x 10 <sup>-05</sup> R	0.763 (0.679-08579)
10	rs11191064	103357814	intergenic	0.0562	0.08505	6.23 x 10 <sup>-06</sup>	5.44 x 10 <sup>-06</sup>	3.65 x 10 <sup>-05</sup>	2.18 x 10 <sup>-05</sup> D	1.561 (1.287-1.894)
<b>4</b>	<b>rs356220</b>	<b>90860363</b>	<b>intergenic</b>	<b>0.1237</b>	<b>0.1637</b>	<b>7.22 x 10<sup>-06</sup></b>	<b>6.19 x 10<sup>-06</sup></b>	<b>3.76 x 10<sup>-05</sup></b>	<b>2.20 x 10<sup>-04</sup>D</b>	<b>1.387 (1.203-1.600)</b>
4	rs13130144	120092696	<i>SYNPO2</i>	0.1233	0.1632	7.47 x 10 <sup>-06</sup>	6.34 x 10 <sup>-06</sup>	3.88 x 10 <sup>-05</sup>	2.98 x 10 <sup>-05</sup> D	1.387 (1.203-1.600)

P values corresponding to basic allelic, Cochran-Armitage trend, genotypic and dominant/recessive tests are displayed. For the dominant/recessive tests, D or R indicates that the displayed p value was obtained after dominant or recessive test respectively. Besides, Hardy-Weinberg equilibrium p values obtained in both cases and controls, as well as odds ratios (OR) and CI 95% are shown. Bolded are those SNPs in *SNCA* locus.



**Figure 40:** - log p values of the Cochran-Armitage for all SNPs except for those excluded because of poor clustering after visual inspection. These values are plotted against position on each chromosome. Chromosomes are shown in alternating colours. Red line at  $-\log = 7$  indicates the significant p value threshold of  $10^{-7}$ . This plot was generated with Haploview 4.0RC2 software.

### 4.6.3. Combined association analysis.

In order to add power to our genome-wide association study and further confirm the results we obtained, we undertook a collaborative analysis with Drs. Thomas Gasser (Hartie-Institute for Clinical Brain Research, Tuebingen, Germany) and Tatsushi Toda (Osaka University Graduate School of Medicine, Suita, Japan).

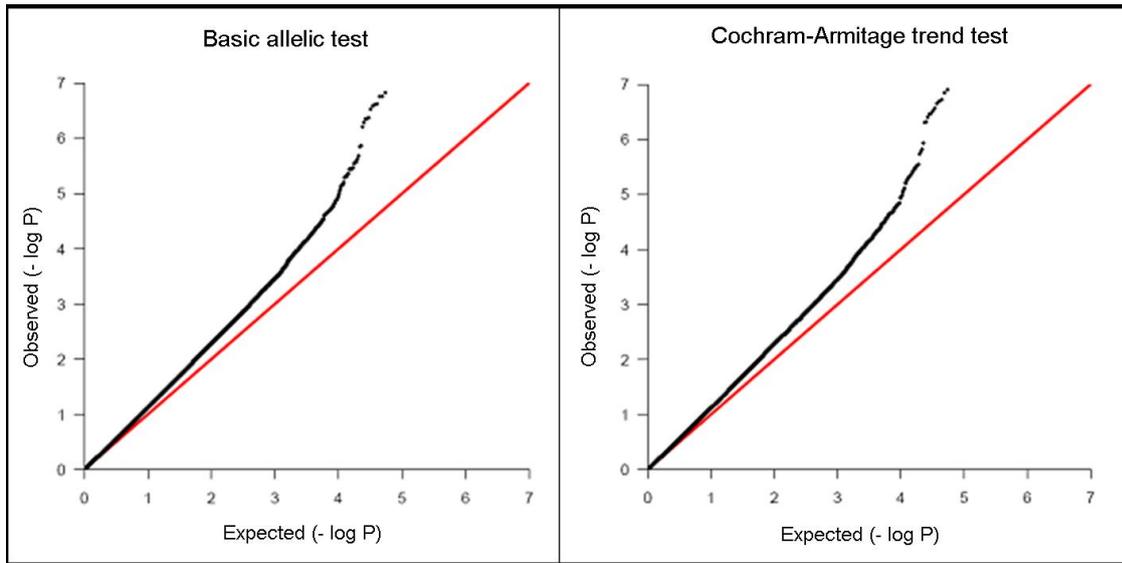
**4.6.3.1. German cohort:** After removing samples that did not reach a success rate of 95%, gender mismatches and population outliers, 1,686 samples remained (including 742

cases and 944 controls). Those SNPs with a Hardy-Weinberg equilibrium p value smaller than 0.01 in controls and a MAF smaller than 5% in controls were excluded from further analysis. Genotypic counts for all those SNPs that successfully passed these filters (498,560) were merged with those of our successful 474,995 SNPs, giving a total of 463,185 SNPs common to both cohorts (see section 3.7.5.1).

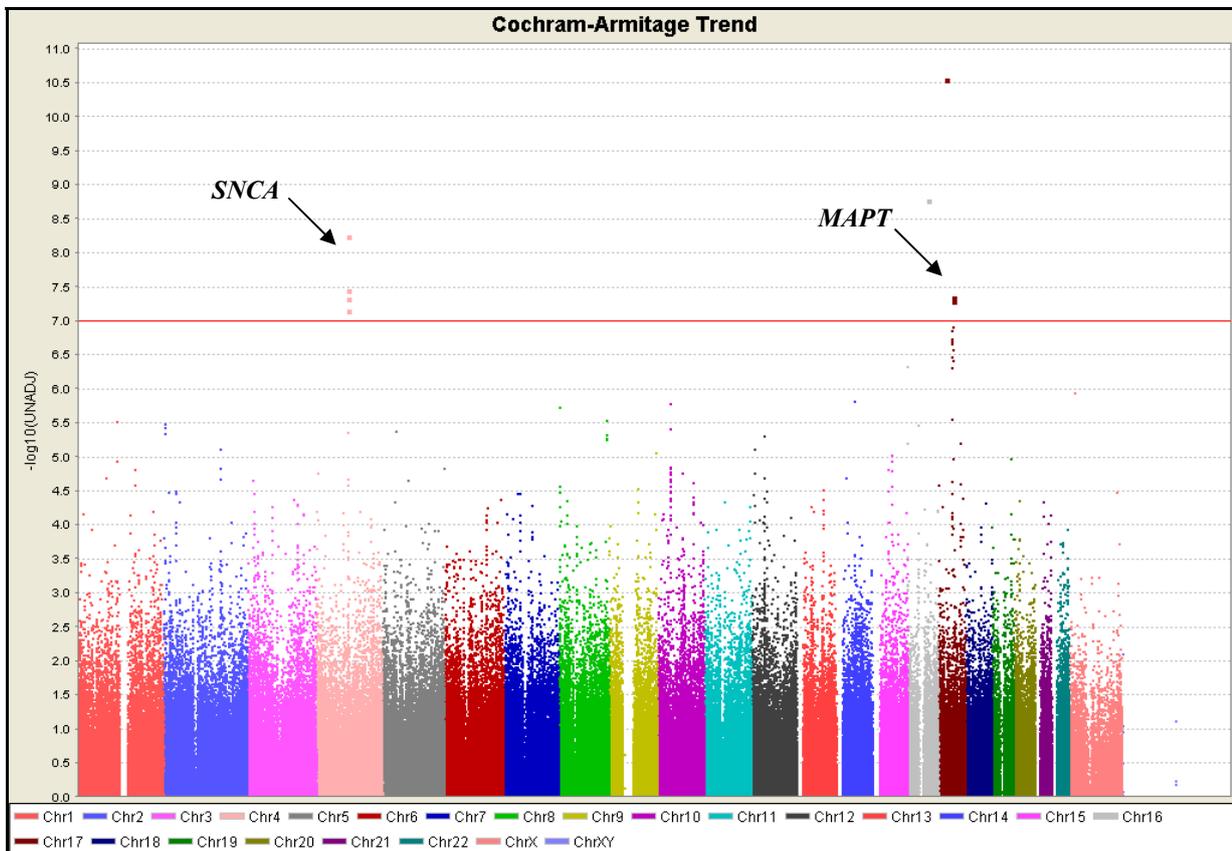
All tests and corrections used for our cohort (see section 3.6.4) were applied for each of these SNPs.

Association of *SNCA* and disease found in our previous analysis was confirmed after this joint analysis. Moreover, association of SNPs within *MAPT* (highly associated with PD, see section 1.6.3.9) was obtained. These results can be used as positive controls for our study and add strength to the association reported in other genes not previously related to PD. These, along with the comparison of a null distribution of p with that obtained in our experiments (figure 41) supports the robustness of our analysis.

Table 33 shows results obtained for the basic allelic and Cochran-Armitage trend, genotypic and dominant/recessive tests for those 30 SNPs with higher association in our analysis. The whole distribution of trend p values for this joint analysis is plotted in figure 42.



**Figure 41: Quantile-quantile plots for the allelic and the trend p values distribution in our combined analysis.** Results of these two statistical approaches are shown in black for all 465,684 SNPs common to both the U.S and the German population, showing a minor deviation from the null distribution (in red) except for those SNPs at the extreme tail.



**Figure 42: - log p values of the Cochran-Armitage trend test for the US-German those combined analysis.** Chromosomes are shown in alternating colors. Red line at  $-\log = 7$  indicates the significant p value threshold of  $10^{-7}$ . This plot was generated with Haploview 4.0RC2 software.

**Table 33:** 30 lowest uncorrected p values obtained in the USA-German combined analysis.

SNP	Chr. Location	Position bp	Gene	HWE p value (USA)		HWE p value (GER)		Trend p value	Allelic p value	Genotypic p value
				Controls	Cases	Controls	Cases			
rs2736990	4	90,897,564	<i>SNCA</i>	0.6878	0.4391	0.1474	0.05616	5.69 X 10 <sup>-09</sup>	3.99 X 10 <sup>-09</sup>	1.25 X 10 <sup>-08</sup>
rs3857059	4	90,894,261	<i>SNCA</i>	0.5736	0.03905	0.7916	0.04442	3.60 X 10 <sup>-08</sup>	2.61 X 10 <sup>-08</sup>	2.97 X 10 <sup>-09</sup>
rs415430	17	42,214,305	<i>WNT3</i>	0.2133	0.7377	0.4536	0.8841	4.50 X 10 <sup>-08</sup>	5.40 X 10 <sup>-08</sup>	1.50 X 10 <sup>-07</sup>
rs11931074	4	90,858,538	<i>SNCA</i> (3' UTR)	0.5741	0.04077	0.7922	0.04195	4.78 X 10 <sup>-08</sup>	3.50 X 10 <sup>-08</sup>	3.62 X 10 <sup>-09</sup>
rs199533	17	42,184,098	<i>NSF</i>	0.1861	0.8201	0.341	0.8841	5.05 X 10 <sup>-08</sup>	6.36 X 10 <sup>-08</sup>	1.67 X 10 <sup>-07</sup>
rs356220	4	90,860,363	<i>SNCA</i> (3' UTR)	0.8774	0.9478	0.5228	0.1994	6.99 X 10 <sup>-08</sup>	5.86 X 10 <sup>-08</sup>	3.45 X 10 <sup>-07</sup>
rs169201	17	42,145,386	<i>NSF</i>	0.2694	0.5747	0.3405	1	1.25 X 10 <sup>-07</sup>	1.49 X 10 <sup>-07</sup>	4.00 X 10 <sup>-07</sup>
rs393152	17	41,074,926	<i>CI7orf69</i>	0.3263	0.6815	0.3657	0.8906	1.42 X 10 <sup>-07</sup>	1.75 X 10 <sup>-07</sup>	8.12 X 10 <sup>-07</sup>
rs12185268	17	41,279,463	<i>IMP5</i>	0.1923	0.9171	0.3612	0.781	1.90 X 10 <sup>-07</sup>	2.38 X 10 <sup>-07</sup>	8.35 X 10 <sup>-07</sup>
rs1981997	17	41,412,603	<i>MAPT</i>	0.1581	0.917	0.3611	0.783	2.02 X 10 <sup>-07</sup>	2.57 X 10 <sup>-07</sup>	8.41 X 10 <sup>-07</sup>
rs2532274	17	41,602,941	<i>KIAA1267</i>	0.5999	1	0.3553	0.7854	2.21 X 10 <sup>-07</sup>	2.47 X 10 <sup>-07</sup>	1.17 X 10 <sup>-06</sup>
rs2532269	17	41,605,885	<i>KIAA1267</i>	0.5585	0.8304	0.3076	0.7799	2.70 X 10 <sup>-07</sup>	2.97 X 10 <sup>-07</sup>	1.13 X 10 <sup>-06</sup>
rs8070723	17	41,436,901	<i>MAPT</i>	0.1747	0.8357	0.3611	0.89	3.36 X 10 <sup>-07</sup>	4.23 X 10 <sup>-07</sup>	1.52 X 10 <sup>-06</sup>
rs17563986	17	41,347,100	<i>MAPT</i>	0.1925	0.7544	0.361	1	3.44 X 10 <sup>-07</sup>	4.38 X 10 <sup>-07</sup>	1.79 X 10 <sup>-06</sup>
rs2668692	17	41,648,797	<i>LOC644246</i>	0.3154	0.9157	0.5998	0.7788	3.94 X 10 <sup>-07</sup>	4.44 X 10 <sup>-07</sup>	1.64 X 10 <sup>-06</sup>
rs11648673	16	317,795	<i>AXIN1</i>	0.3955	1	1	0.9133	4.77 X 10 <sup>-07</sup>	5.16 X 10 <sup>-07</sup>	2.54 X 10 <sup>-06</sup>
rs12373139	17	41,279,910	<i>IMP5</i>	0.1912	0.917	0.3128	1	4.91 X 10 <sup>-07</sup>	6.19 X 10 <sup>-07</sup>	2.27 X 10 <sup>-06</sup>
rs239748	23	18,793,279	<i>LOC239748</i>	0.755	0.44	0.569	0.745	1.17 X 10 <sup>-06</sup>	1.35 X 10 <sup>-06</sup>	7.06 X 10 <sup>-06</sup>
rs12431733	14	53,360,580	<i>BMP4</i>	0.1843	0.07237	1	0.2112	1.52 X 10 <sup>-06</sup>	1.42 X 10 <sup>-06</sup>	2.98 X 10 <sup>-07</sup>
rs11591754	10	35,247,159	<i>LOC646213</i>	0.3753	0.8941	0.06475	0.9015	1.68 X 10 <sup>-06</sup>	2.08 X 10 <sup>-06</sup>	6.21 X 10 <sup>-06</sup>
rs7013027	8	2,911,376	<i>CSMD1</i>	0.2426	0.3683	0.5553	0.5569	1.85 X 10 <sup>-06</sup>	2.37 X 10 <sup>-06</sup>	1.15 X 10 <sup>-05</sup>
rs111012	17	40,869,224	<i>LOC201175</i>	0.06664	0.6256	0.7245	0.5354	2.85 X 10 <sup>-06</sup>	3.59 X 10 <sup>-06</sup>	1.16 X 10 <sup>-05</sup>
rs7004938	8	140,328,407	<i>COL22A1</i>	0.8821	0.3085	0.09952	0.3536	2.97 X 10 <sup>-06</sup>	2.69 X 10 <sup>-06</sup>	1.70 X 10 <sup>-05</sup>
rs10857899	1	111,929,241	<i>LOC643329</i>	0.1877	0.3208	0.514	0.3258	3.06 X 10 <sup>-06</sup>	2.71 X 10 <sup>-06</sup>	1.72 X 10 <sup>-05</sup>
rs6542651	2	3,737,705	<i>LOC728597</i>	0.2543	1	0.8108	0.3916	3.34 X 10 <sup>-06</sup>	2.85 X 10 <sup>-06</sup>	2.02 X 10 <sup>-05</sup>
rs2285459	16	30,402,913	<i>ITGAL</i>	0.186	0.8931	0.3903	0.8791	3.38 X 10 <sup>-06</sup>	3.64 X 10 <sup>-06</sup>	1.78 X 10 <sup>-05</sup>
rs13027881	2	6,473,384	<i>LOC391349</i>	0.9429	0.2869	0.3471	0.5692	3.76 X 10 <sup>-06</sup>	3.59 X 10 <sup>-06</sup>	1.99 X 10 <sup>-05</sup>
rs2492448	10	35,235,412	<i>LOC646213</i>	0.9306	0.3808	0.6477	0.5215	3.84 X 10 <sup>-06</sup>	3.39 X 10 <sup>-06</sup>	1.35 X 10 <sup>-05</sup>
rs4957473	5	39,378,920	<i>C9</i>	0.7893	0.8312	0.1133	0.683	4.23 X 10 <sup>-06</sup>	4.42 X 10 <sup>-06</sup>	2.30 X 10 <sup>-05</sup>
rs3775439	4	90,928,764	<i>SNCA</i>	1	0.09297	1	0.1991	4.42 X 10 <sup>-06</sup>	3.58 X 10 <sup>-06</sup>	3.94 X 10 <sup>-06</sup>

P values corresponding to basic allelic, Cochran-Armitage trend and genotypic tests are displayed. Hardy-Weinberg equilibrium p values obtained in both the US (USA) and the German (GER) cohort both cases and controls are shown

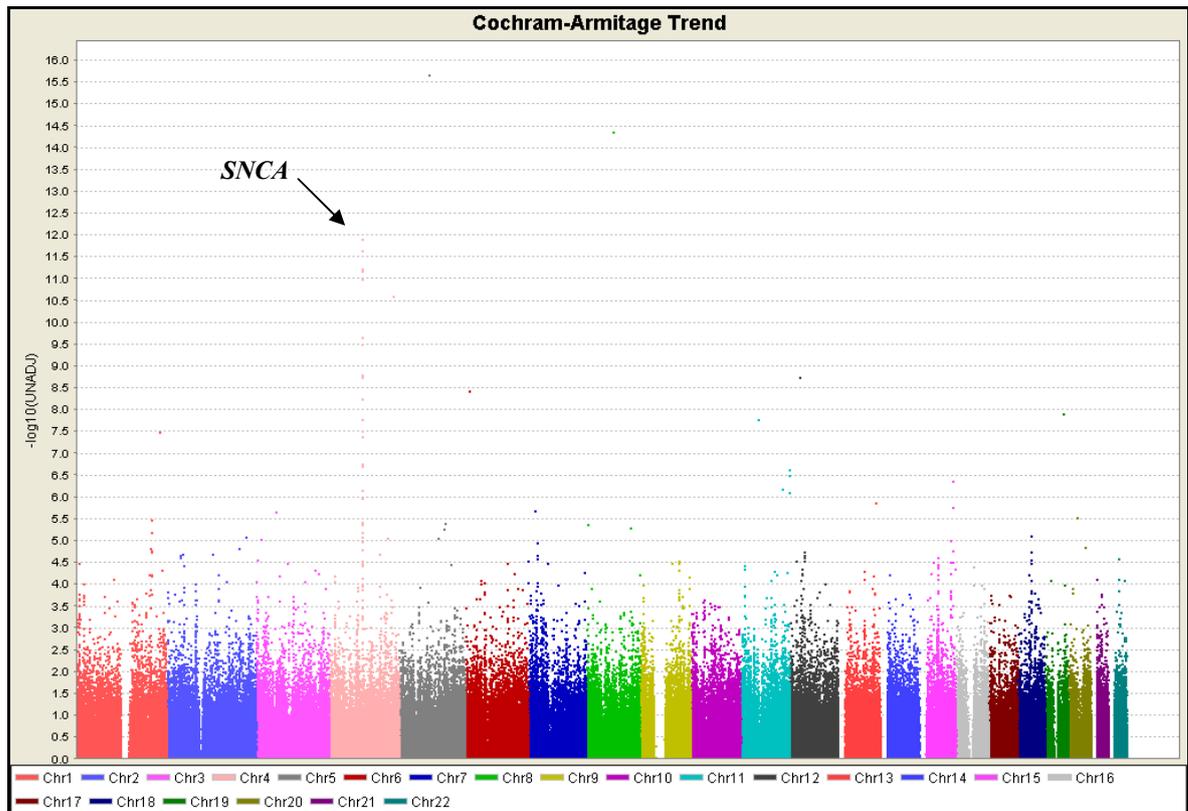
**4.6.3.2. Japanese cohort:** Because of population differences between the Japanese and the Caucasian cohorts (US population and German population), genotyping counts from Dr. Toda's experiments can not be merged with ours. Thus, side by side comparison of the results obtained was performed in this case. Table 34 shows the p values for the Cochran-Armitage test obtained in the Japanese cohort. The integrity of trend p values

obtained in the Japanese series is shown in figure 43. Note the strong association detected across *SNCA* locus.

**Table 34:** 30 lowest uncorrected p values in the Japanese population.

SNP ID	Chr. location	Position bp	Gene	Trend p value	USA min p value	GER min p value	Combined min p value
rs11931074	4	90,858,538	<i>SNCA</i>	1.25 X 10 <sup>-12</sup>	1.21 X 10 <sup>-07</sup>	8.95 X 10 <sup>-03</sup>	3.62 X 10 <sup>-09</sup>
rs3857059	4	90,894,261	<i>SNCA</i>	2.31 X 10 <sup>-12</sup>	1.44 X 10 <sup>-07</sup>	6.07 X 10 <sup>-03</sup>	2.97 X 10 <sup>-09</sup>
rs3775439	4	90,928,764	<i>SNCA</i>	6.20 X 10 <sup>-12</sup>	1.63 X 10 <sup>-06</sup>	-	3.58 X 10 <sup>-06</sup>
rs3796661	4	90,906,530	<i>SNCA</i>	6.74 X 10 <sup>-12</sup>	-	-	-
rs356220	4	90,860,363	<i>SNCA</i>	1.03 X 10 <sup>-11</sup>	8.40 X 10 <sup>-06</sup>	2.86 X 10 <sup>-04</sup>	5.86 X 10 <sup>-08</sup>
rs7655606	4	175,548,801	<i>KIAA1712</i>	2.54 X 10 <sup>-11</sup>	-	-	2.58 X 10 <sup>-02</sup>
rs6532194	4	90,999,925	<i>SNCA</i>	2.21 X 10 <sup>-10</sup>	4.69 X 10 <sup>-04</sup>	-	2.77 X 10 <sup>-03</sup>
rs2736990	4	90,897,564	<i>SNCA</i>	3.17 X 10 <sup>-10</sup>	1.38 X 10 <sup>-07</sup>	1.11 X 10 <sup>-03</sup>	3.99 X 10 <sup>-09</sup>
rs12644119	4	90,822,442	<i>SNCA</i>	3.23 X 10 <sup>-10</sup>	1.11 X 10 <sup>-04</sup>	1.99 X 10 <sup>-02</sup>	2.00 X 10 <sup>-05</sup>
rs3822095	4	90,955,540	<i>SNCA</i>	1.65 X 10 <sup>-09</sup>	2.66 X 10 <sup>-03</sup>	-	6.06 X 10 <sup>-03</sup>
rs12502363	4	90,932,087	<i>SNCA</i>	1.83 X 10 <sup>-09</sup>	5.03 X 10 <sup>-03</sup>	-	8.10 X 10 <sup>-03</sup>
rs9350203	6	10,335,181	<i>OFCCI</i>	3.77 X 10 <sup>-09</sup>	-	-	-
rs10516848	4	90,994,235	<i>SNCA</i>	5.84 X 10 <sup>-09</sup>	2.94 X 10 <sup>-02</sup>	1.83 X 10 <sup>-02</sup>	2.69 X 10 <sup>-02</sup>
rs13125108	4	90,764,236	<i>SNCA</i>	1.75 X 10 <sup>-08</sup>	1.35 X 10 <sup>-03</sup>	2.56 X 10 <sup>-02</sup>	5.90 X 10 <sup>-04</sup>
rs9985612	4	90,810,912	<i>SNCA</i>	3.13 X 10 <sup>-08</sup>	2.52 X 10 <sup>-04</sup>	-	2.66 X 10 <sup>-04</sup>
rs12129824	1	228,973,714	<i>CAPN9</i>	3.26 X 10 <sup>-08</sup>	-	-	-
rs3733449	4	90,791,345	<i>SNCA</i>	4.18 X 10 <sup>-08</sup>	4.51 X 10 <sup>-04</sup>	-	2.06 X 10 <sup>-03</sup>
rs1390280	4	90,738,475	<i>SNCA</i>	1.78 X 10 <sup>-07</sup>	6.85 X 10 <sup>-04</sup>	-	6.68 X 10 <sup>-03</sup>
rs6850389	4	90,746,367	<i>SNCA</i>	1.96 X 10 <sup>-07</sup>	4.11 X 10 <sup>-02</sup>	-	-
rs7930275	11	134,117,017	<i>B3GATI</i>	2.46 X 10 <sup>-07</sup>	-	-	-
rs7929317	11	134,113,081	<i>B3GATI</i>	3.24 X 10 <sup>-07</sup>	-	-	-
rs12591171	15	90,920,803	<i>LOC400451</i>	4.47 X 10 <sup>-07</sup>	-	-	-
rs2364857	11	113,382,767	<i>HTR3A</i>	6.86 X 10 <sup>-07</sup>	-	-	-
rs12233759	4	91,099,426	<i>MMRNI</i>	7.12 X 10 <sup>-07</sup>	-	-	-
rs1318557	4	91,098,636	<i>MMRNI</i>	7.26 X 10 <sup>-07</sup>	-	-	-
rs7943044	11	134,113,328	<i>B3GATI</i>	7.81 X 10 <sup>-07</sup>	-	-	-
rs6854896	4	91,090,584	<i>MMRNI</i>	1.11 X 10 <sup>-06</sup>	-	-	-
rs4468458	13	101,546,407	<i>FGF14</i>	1.42 X 10 <sup>-06</sup>	-	-	-
rs285705	15	90,919,508	<i>LOC400451</i>	1.74 X 10 <sup>-06</sup>	-	-	-
rs2158250	7	20391971	<i>ITGB8</i>	2.07 X 10 <sup>-06</sup>	-	-	-

P values corresponding Cochran-Armitage in the Japanese population are displayed. These are compared with the minimum p value at that same *locus* obtained in the North American (USA) and German (GER) population. P values for those SNPs genotyped in the Japanes but not in the North American or German population are represented as -.



**Figure 43:**  $-\log p$  values of the Cochran-Armitage trend test for those SNPs assayed by Toda and collaborators. Chromosomes are shown in alternating colors. The peak in chromosome 4 corresponds to association in *SNCA* locus.

## 4.7. AUTOZYGOSITY MAPPING IN BRAZILIAN FAMILIES

### 4.7.1. Exclusion of known genes.

Screening for mutations in *PRKN*, *LRRK2*, *SNCA*, *GCHI*, *TOR1A*, *ATPIA3* and *PINK1* were negative in all patients. Serum and urine copper and ceruloplasmin tests ruled out Wilson’s disease (OMIM #277900).

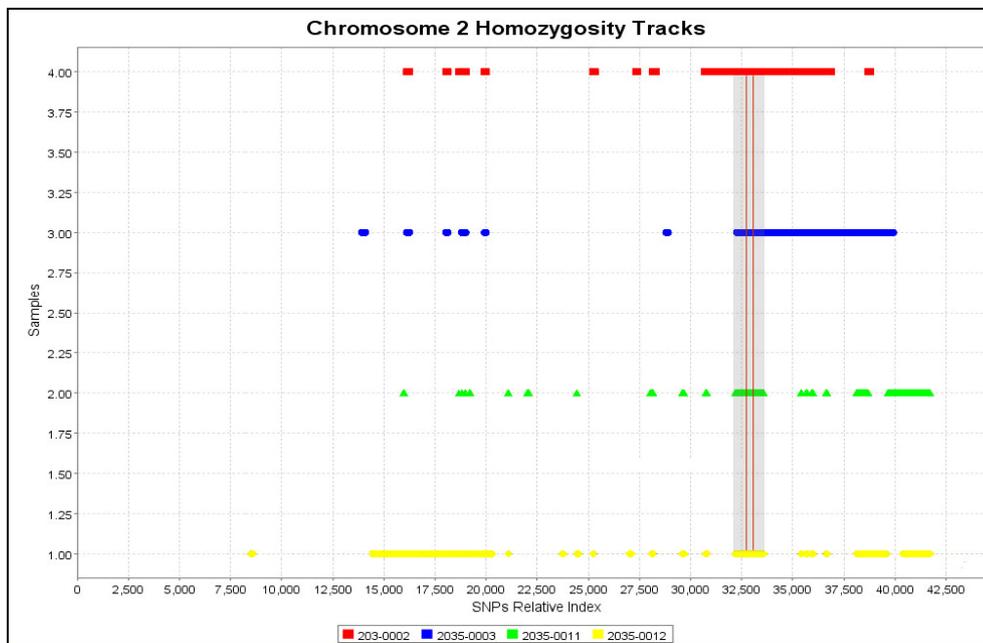
#### 4.7.2. Autozygosity mapping.

Analysis of Infinium genotyping data in samples 2035-2 and 2035-3 (*DYT16-1*), along with 2035-11 and 2035-12 (*DYT16-2*), revealed four regions in the genome that were homozygous at greater than 50 consecutive SNPs and identical by state between these four samples. Analysis of samples 2035-1 (affected brother of *DYT16-1*) and 2035-119 (affected uncle of *DYT16-2*) showed that one of these regions (containing 81 SNPs between markers rs2528907 and rs1695515 in chromosome 7) was not identical by state in these samples, therefore discarding it as a candidate *locus*.

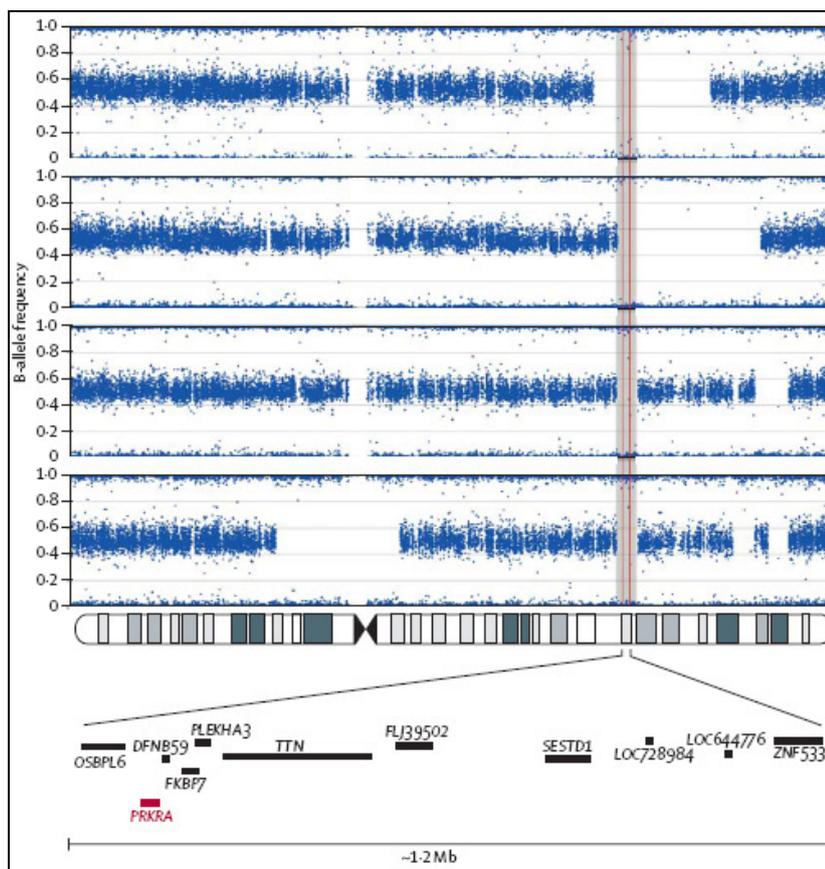
The remaining three regions were all located on chromosome 2q31.2 and encompassed 1.2 Mb from rs1434087 to rs10497541 (containing 278 SNPs), 0.13 Mb from rs13405069 to rs7581560 (containing 67 SNPs), and 0.4 Mb from rs1518709 to rs10930936 (containing 78 SNPs). Interestingly, 277 of those 278 SNPs in the first region were homozygous and identical by state within all affected family members. However, a single SNP (rs4897088), 0.23 Mb away from the telomeric edge of this critical region, although homozygous in all affected members had genotype AA in affected members from family *DYT16-1* and genotype GG in affected members of family *DYT16-2*. Resequencing of the region of this SNP from a PCR-amplified fragment using primers designed to amplify the region embedded between the two most closely flanking SNPs, showed that the genotypes were indeed correct and suggested that the localization of this SNP was correct (primer sequences in supplementary data table S1). Although we felt this result was perplexing, because it is not consistent with an ancestrally identical genomic segment, and perhaps only explained by a gene conversion event or complex structural rearrangement, we felt

that this large track of homozygous SNPs that are identical across families still implicated this region as a good candidate to contain the disease-causing mutation.

The primary candidate region (rs1434087- rs10497541) contains the following genes and predicted transcripts: *OSBPL6*, *PRKRA*, *DFNB59*, *FKBP7*, *PLEKHA3*, *TTN*, *FLJ39502*, *SESTD1*, *LOC728984*, *LOC644776* and the first four exons of fragment of *ZNF533*. The second candidate region contains the remaining portion of *ZNF533*, and the third region contains a single transcript, *LOC729001* (figures 44 and 45).



**Figure 44:** Tracks of contiguous homozygous SNPs were calculated with tracker v0.99. All homozygous regions in chromosome 2 are shown herein in red (2035-2), blue (2035-3), green (2035-11) and yellow (2035-12). All three candidate regions are located within the gray-shaded area. Bounded in red is the primary candidate interval identical by state between all members of families *DYT16-1* and *DYT16-2* (rs1434087 to rs10497541, 1.2 Mb in size).

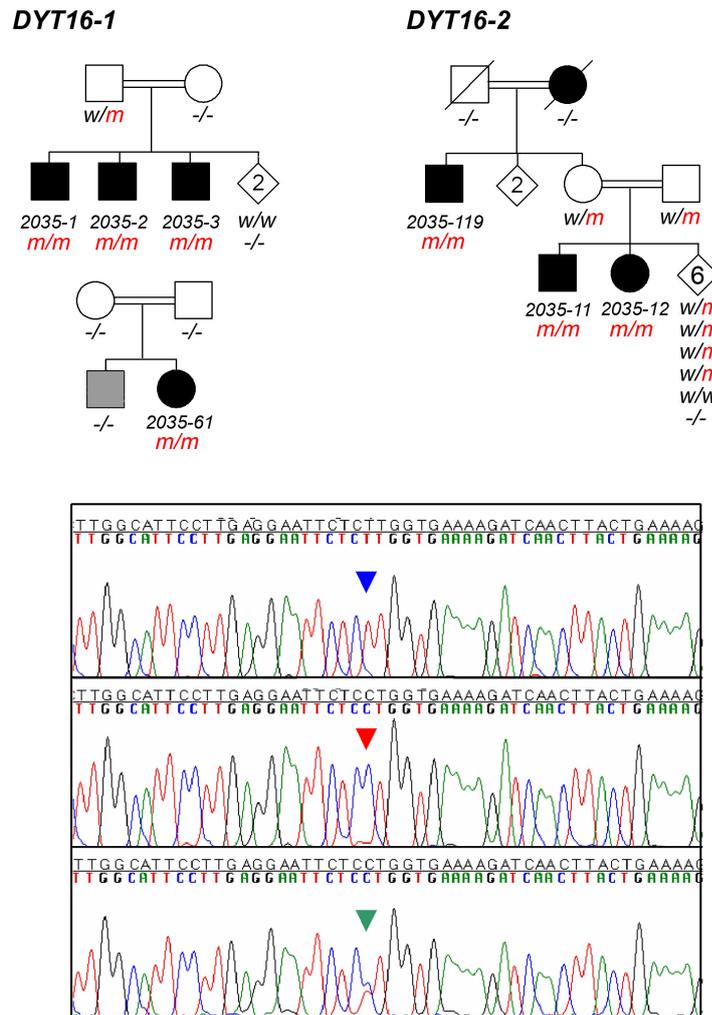


**Figure 45: Illumina Genome Viewer plot showing the disease-segregating homozygous region identified in the six affected members of families *DYT16-1* and *DYT16-2*.** Upper panels show the B allele frequency metrics across chromosome 2 for two affected individuals from *DYT16-1* (2035-1 and 2035-2) and *DYT16-2* (2035-11 and 2035-12). Stretches of homozygosity are denoted by a contiguous stretch of genotypes there is a lack of AB genotype calls (ie, B allele frequency of 0.5). The lower panel shows an ideogram of chromosome 2, the primary candidate interval that is identical by state between families *DYT16-1* and *DYT16-2* (rs1434087 to rs10497541; 1,2 Mb in size), and the genes in this primary critical interval.

Although we felt that the causal mutation would reside in the first candidate region, given that it was five times larger than the next largest region of contiguous homozygosity, we sequenced all 12 genes and transcript in these three largest candidate intervals. Sequence analysis of all coding exons and immediate flanking intronic sequence of all 12 genes (458 exons) revealed a single segregating homozygous variant (c.665C>T) in exon 7 of *PRKRA* (p.P222L) in all affected family members of *DYT16-1*

and *DYT16-2* (figure 46). Sequencing of this exon in all 1,992 samples described in the methods section 3.8.1 revealed a homozygous c.665C<T mutation in a single sample from a Brazilian Patient with generalized dystonia (patient 2035-61). Genotyping DNA from this patient with Illumina Infinium HumanHap550 revealed a region identical by state with all other affected members positive of c.665C>T between markers rs6738749 and rs10497541, limiting the disease segregating interval on the centromeric side by 12 SNPs (0.04 Mb).

The entire coding region of *PRKRA* was sequenced in 12 Brazilian patients with dystonia. No additional mutations were found after this analysis.



**Figure 46: Pedigrees of families *DYT16-1*, *DYT16-2*, and patient 2035-61.** A black symbol indicates an affected family member, an open symbol indicates an unaffected family member, and a grey symbol indicates a family member reported as affected by history but unexamined. Squares are males, circles are females. Multiple siblings are denoted by a diamond; the number of individuals is indicated in the diamond. Genotype at position c.665 (p.222) is shown below each family member being w = wild type (c.665C); m = mutant (c.665T); *-/-* = genotyping not possible, or attempted but failed. In the panel below, chromatograms show the mutation in *PRKRA*; the blue arrow in top panel indicates mutant homozygote, the red arrow in middle panel indicates wild type sequence, and the green arrow in bottom panel indicates heterozygous mutation.

## 5. DISCUSSION

### 5.1. GENETIC ANALYSIS OF PARK8-LINKED PARKINSON'S DISEASE PATIENTS.

In the course of sequencing *KIF21A* as candidate *PARK8* gene in a set of four families from the Basque region in Spain (UGM03, UGM04, UGM05 and UGM06), Paisan-Ruiz and collaborators (Paisan-Ruiz, Jain *et al.*, 2004) identified a rare variant segregating with the disease in three of the four families. Given the rarity of this variant in the general population from the Basque country, this suggested that these families may be ancestrally related and could share a minimal interkindred disease haplotype. In order to confirm the existence of this haplotype and confirm its limits, we chose to genotype informative SNPs in the region showing strongest linkage to disease in these families. This effort led us to reduce the critical disease interval to 2.6 Mb, a region containing 11 genes and predicted transcripts. Systematic sequence analysis developed by Dr. Paisan-Ruiz, showed segregating mutations within a putative kinase domain containing transcript, DKFZp434H2111. This gene was later given the name of *LRRK2* and the encoded protein was named dardarin, coming from the Basque word for tremor.

In order to assess the prevalence of this mutation (p.R1441G) in the Basque PD population, we genotyped a total of 238 PD patients presenting symptoms of typical late-onset PD. This analysis led us to find 17 patients (~ 7%), with and without positive family history of PD, carrying the p.R1441G mutation (clinical features on table 24).

Due to its orographical characteristics, the Basque Country has remained isolated from the different populations that invaded the Iberian Peninsula. Because of this, and because of an absence of immigration until the late seventies, the Basques are somewhat assimilated to a genetic isolate (Calafell and Bertranpetit, 1994; De Pancorbo, Lopez-Martinez *et al.*, 2001; Jakobsson, Scholz *et al.*, 2008). Thus, it is possible, as in the case for other diseases (Urtasun, Saenz *et al.*, 1998; Zivelin, Bauduer *et al.*, 2002; Cobo, Saenz *et al.*, 2004), that the most likely situation is that a founder effect in this population explains the high prevalence of the mutation. In order to investigate the possibility that all these patients have inherited p.R1441G from a common founder, we genotyped all SNPs used to narrow down the candidate region in the four Basque families, in these 17 carriers. This approach revealed genotypes consistent with the haplotypes found in the four families, supporting the idea that they all have inherited the mutation as part of an ancestral chromosome.

According to the data presented herein, the p.R1441G mutation in *LRRK2* accounts for 16.4% of familial and 4.0% of sporadic PD in the Basque population, excluding the members of the four original families reported by Paisan-Ruiz and collaborators (Paisan-Ruiz, Saenz *et al.*, 2005). Even though our sample was not collected as part of a population-based analysis, it could be considered as representative of the general population as it was collected through the combined effort of all Public Hospitals in the area of influence of San Sebastian. In this respect, the population served by those hospitals comprises 420,000 individuals. On the basis of epidemiological estimates, approximately 1,260 PD cases might exist (Bergareche, De La Puente *et al.*, 2004) of whom 238 (~ 18.8%) have been included in this study.

As for those p.R1441G non-Basques carriers found in our population, there is limited information about their ethnicity, as we were able to collect information only from the probands and their parents. Thus, the presence of the low-frequency of the disease haplotype seen in this population suggests that they have distant ancestors of Basque descent.

Different groups have looked for this mutation in populations across Spain. Thus, Mata and coworkers found this mutation in 2.7% of a PD population from Asturias (Mata, Taylor *et al.*, 2005), where 4 out of the 5 mutation carriers were considered as sporadic. Similar results were obtained by Gaig and collaborators in Catalonia, who found that 0.7% of the patients they analyzed carried the p.R1441G mutation (Gaig, Ezquerra *et al.*, 2006).

Gonzalez-Fernandez and collaborators did not find this change in any of the sporadic PD cases they analyzed. However, they found it in one family from Navarra and another from La Rioja (Gonzalez-Fernandez, Lezcano *et al.*, 2007). This mutation was not identified in 105 PD patients from Cantabria analyzed by Infante and colleagues (Infante, Rodriguez *et al.*, 2006).

These findings suggest the existence of a geographical gradient for this mutation, which goes from 4.0% of PD patients in the Basque Country to 2.7% in Asturias and 0.7% in Catalonia. If we collectively considered all this information, along with the fact that different aminoacid changes (p.R1441C and p.R1441H) have been described in this same codon elsewhere (Zimprich, Biskup *et al.*, 2004; Mata, Kachergus *et al.*, 2005) it would appear that a R>G change appeared in the hot-spot mutation site 1441 on a Basque chromosome, possibly in the valleys where the four original families were identified.

This mutation would later have spread through, at least, Northern Spain. In this respect, it is interesting to note that the group of p.R1441G carriers analyzed by Mata and coworkers share the same ancestral haplotype as the Basque patients we present in this study.

This geographical gradient of p.R1441G, is likely to be related to the global Basque *diaspora*. This diaspora refers to different migratory fluxes the Basque population has suffered along history for different reasons. This migration did not only proceed to other communities within Spain, but other countries of Europe, Central and South America (specially Argentina and Chile) and North America (Western United States and Canada) as well. Although several groups have tried to find p.R1441G in Portugal (Bras, Guerreiro *et al.*, 2005; Ferreira, Guedes *et al.*, 2007a), France (Funalot, Nichols *et al.*, 2006), Greece (Xiromerisiou, Hadjigeorgiou *et al.*, 2007), Italy (Goldwurm, Di Fonzo *et al.*, 2005), Chile (Perez-Pastene, Cobb *et al.*, 2007) and the United States (Paisan-Ruiz, Jain *et al.*, 2004; Mata, Kachergus *et al.*, 2005; Pankratz, Pauciulo *et al.*, 2006), only a single heterozygous carrier female was found by Deng and colleagues in North American population. Interestingly, although she apparently had no Spanish relatives she was Hispanic (Deng, Le *et al.*, 2006).

The fact that this mutation appears in sporadic patients in our population and elsewhere (Mata, Taylor *et al.*, 2005; Deng, Le *et al.*, 2006; Gaig, Ezquerra *et al.*, 2006) either suggests a lack of information on other family members of those sporadic patients who might be affected by a milder or subclinical form of PD, or that other environmental or genetic factors might greatly affect disease penetrance. Interestingly, Gonzalez-Fernandez and collaborators in a genetic study of *LRRK2*-associated parkinsonism, found

10 families from the Basque country with the p.R1441G substitution being present in 6 asymptomatic carriers with an age range of 32-75 years, indicating a low penetrance of this mutation (Gonzalez-Fernandez, Lezcano *et al.*, 2007). Moreover, for those 15 symptomatic carriers they found, clinical features differ from the first p.R1441G Basque families reported (Paisan-Ruiz, Jain *et al.*, 2004), where late-onset disease was common and tremor was the predominant initial symptom resulting in the term *dardarin* (Basque for tremor) to describe the protein encoded by *LRRK2*. This variation in the clinical phenotype of the p.R1441G pathogenic substitution must be influenced by other factors, genetic or environmental.

Collecting more family members and studying all available p.R1441G carriers is now crucial to better estimate the penetrance of this pathogenic substitution and provide appropriate genetic counseling to patients and their families.

## **5.2. *LRRK2* IS EXPRESSED IN BRAIN AREAS AFFECTED IN PD IN THE ADULT MOUSE BRAIN**

Previously, Northern blot analysis showed that *LRRK2* is expressed in the cerebellum, cerebral cortex, medulla, spinal cord, occipital pole, frontal lobe, temporal lobe and putamen (Paisan-Ruiz, Jain *et al.*, 2004). However, a complete and detailed map of *LRRK2* mRNA localization in the central nervous system was still lacking. Thus, with the objective of providing a complete atlas of the *LRRK2* mRNA distribution, we performed

an *in situ* hybridization analysis with digoxigenin-labeled RNA probes on mice brain coronal sections.

The results obtained, clearly show that *LRRK2* mRNA is distributed throughout the mouse brain. However, a stronger hybridization signal appears in specific brain regions such as the deep cerebral cortex layers, cingulate cortex, piriform cortex, Ammon's horn and dentate gyrus, caudate putamen, *substantia nigra*, some parts of the amygdala, reticular thalamic nucleus or cerebellar granular cell layer.

Among these high *LRRK2* expression areas, the caudate putamen, the *substantia nigra pars compacta* and the granular cell layer of the cerebellum, are involved in the regulation of motor activity. As most clinical PD symptoms are related to motor dysfunction, this expression data may indicate a role of *LRRK2* in motor activity regulation.

Regarding the appearance of *LRRK2* mRNA in the granular cell layer of the cerebellum, currently cerebellar symptoms have not been described in PD patients. However, several groups (including ourselves, see sections 1.6.3.8, 3.4, 4.4 and 5.4) have related the appearance of parkinsonian motor features (parkinsonism) to the expansion of CAG repeats encoding polyglutamine tracts within spinocerebellar ataxia-related genes (Gwinn-Hardy, Chen *et al.*, 2000; Wu, Lin *et al.*, 2004; Simon-Sanchez, Hanson *et al.*, 2005). The phenotype of these spinocerebellar ataxia patients is variable and may cause a disease that clinically overlaps with PD.

Other structures indirectly related to motor activity regulation also express *LRRK2* mRNA. We detected strong hybridization signals in the hippocampus, piriform cortex, cingulated cortex, olfactory tubercle, amygdala or reticular thalamic nucleus. Potential

relationships between these expression areas and PD pathology have been suggested by several groups studying PD and atrophy. Patients with PD and cognitive decline, show grey matter volume decreases in several brain regions such as the putamen, accumbens nuclei, some parts of the thalamus, hippocampus or cingulate cortex (Summerfield, Junque *et al.*, 2005). Moreover, Camicioli *et al.* found that a hippocampal atrophy was more often displayed in PD patients than in healthy subjects regardless of any existing cognitive dysfunction (Camicioli, Moore *et al.*, 2003). In relation to the amygdala, Junque and colleagues found that both demented and nondemented PD patients have clear amygdalar and hippocampal atrophies (Junque, Ramirez-Ruiz *et al.*, 2005). As the amygdala and the hippocampus are key structures of the limbic system, *LRRK2* may play a role in this function. In this respect, it is interesting to note the absence of cognitive dysfunction in most PD patients with dardarin mutations (see for example (Paisan-Ruiz, Saenz *et al.*, 2005).

As we have previously mentioned, we also found a strong *LRRK2* mRNA expression in the olfactory tubercle and some other olfactory related areas, such as the piriform cortex. It could be speculated from this finding that dardarin might play a role in the known loss of sense of smell in idiopathic PD (Berendse, Booij *et al.*, 2001).

One important finding from this work that may be relevant to the etiopathogenesis of PD is the strong *LRRK2* expression in the putamen which contrasts with a lower or non signal in the *globus pallidus*. We can speculate that compensatory mechanisms between both areas might be disrupted as a consequence of dardarin dysfunction.

Wszolek and colleagues reported the finding of different pathological signatures in brain tissue from four members of a PD family with the p.R1441C dardarin mutation (Wszolek,

Pfeiffer *et al.*, 2004). These authors found Lewy body pathology in two of the four individuals. One of the two Lewy body positive patients also had tau inclusions. A third patient of this family had solely tau inclusions whereas a fourth one only showed neuronal loss and gliosis in the *substantia nigra* as found in the other three cases. Our results are consistent with these observations in that *LRRK2* is expressed in all the areas where Wszolek and colleagues found evidence of pathological lesions.

Expression data from other studies published afterwards are generally consistent with ours, showing that *LRRK2* is expressed throughout the brain, especially in striatum and cortex (Galter, Westerlund *et al.*, 2006; Melrose, Lincoln *et al.*, 2006; Miklossy, Arai *et al.*, 2006; Taymans, Van Den Haute *et al.*, 2006; Higashi, Moore *et al.*, 2007; Melrose, Kent *et al.*, 2007). However, large controversy is seen on regard to *LRRK2* expression in dopaminergic neurons of the *substantia nigra*. On this regard, two reports using radioactive oligonucleotide probes concluded to the absence (Galter, Westerlund *et al.*, 2006; Melrose, Lincoln *et al.*, 2006) or extremely low presence (Higashi, Moore *et al.*, 2007) of *LRRK2* mRNA from the SNC. These reports concluded that dysfunction of *LRRK2* affects only dopamine-receptive brain regions such as caudate-putamen, leading to the degeneration of the nigrostriatal pathway.

On the other hand, our data are largely in agreement with the study of Taymans and colleagues (Taymans, Van Den Haute *et al.*, 2006) that also used non-radioactive probes for their *in situ* hybridization experiments. The reasons of this controversy may be driven by difference on the techniques used. On this regard, non-radioactive RNA probes create

more stable hybrids than radioactive oligonucleotide probes (shorter) making them more specific and sensitive (Tecott, Rubenstein *et al.*, 1989).

Expression of *LRRK2* in *substantia nigra* has been further supported by means of immunohistochemical analysis (Higashi, Biskup *et al.*, 2007; Melrose, Kent *et al.*, 2007).

In summary, we demonstrated the existence of *LRRK2* mRNA in many brain regions and nuclei. These results suggest that mutations in *LRRK2* could affect several motor structures that play an important role in PD development. In accord with this, most brain regions involved in the nigrostriatal pathway abundantly express *LRRK2* mRNA. The discovery of strong *LRRK2* expression in structures such as the hippocampus, amygdala or olfactory tubercle, may indicate that *LRRK2* dysfunction leads to the appearance of the non-motor symptoms that are characteristic of PD.

### **5.3. *OMI/HTRA2* IS NOT ASSOCIATED WITH PD IN A NORTH AMERICAN POPULATION.**

In molecular genetics, as important as the initial description of a genetic association between genetic variability and a trait is independent confirmation of such an association. In order to confirm the results published by Strauss and colleagues in which they described loss of function mutations in *OMI/HTRA2* in German PD patients (Strauss, Martins *et al.*, 2005), we sequenced the entire coding region of this gene in a large series of both young- and late-onset PD cases and age-matched neurologically normal controls

from the NINDS neurogenetics repository at Coriell (<http://ccr.coriell.org/Sections/Collections/NINDS/?SsId=10>). The data here shows that the p.G399S variant, previously described as a PD-causative mutation in the German study, is not associated with the disease in our population and was found at the same frequency in controls as cases (Fisher's p value = 1, OR=1.069, CI 95% = 0.3256-3.511). Moreover, a rare polymorphism previously described to be associated with PD (Strauss, Martins *et al.*, 2005) was not associated with disease in our cohort (Fisher's p value = 0.2365, OR = 0.7348, CI = 0.4566-1.182).

We also performed single marker association between 3 SNPs (exon 1 and introns 4 and 5) and disease. Results derived from this analysis are consistent with a lack of contribution of the *OMI/HTRA2 locus* to the risk of sporadic PD.

Interestingly, we have identified 8 novel variants in our population in both cases and controls, 4 of which are non-synonymous changes: p.W12C, p.P128L, p.F172V and p.A227S. p.W12C and p.P128L were detected in two different control samples meanwhile p.F172V and p.A227S in two different PD cases. Although it can not be ruled out, it seems unlikely that these variants are related to disease in these patients and are likely rare benign alterations.

It is worth noting that while the p.G399S change was identified in 6 controls, half of these were relatively young (ages 25, 42, 46, 62, 70 and 85 years); however, we do not feel that it is reasonable to suppose that these patients could be control patients who would convert to disease at a later date; first the average age that patients who carried this alteration in our case cohort was 53.2 years (range 30-77 years) and thus we would have expected some of these controls to convert to PD cases already; second, a pathogenic

change should be enriched in a case cohort even when compared to a random population cohort and this is particularly true in a disease which is relatively rare in occurrence.

While we can not rule out a small genetic risk at this *locus* particularly one present only in a subset of patients such as those with young-onset disease, the present study did have sufficient power to identify an effect of the magnitude and type originally described; thus the weight of evidence presented here suggests that variability at *OMI/HTRA2* does not contribute to risk for PD.

Another report examining the contribution of genetic variability in *OMI/HTRA2* to PD risk has recently been published (Bogaerts, Nuytemans *et al.*, 2008). In this report, the authors sequenced the entire coding region of *OMI/HTRA2* and its 5' and 3' regulatory regions in a series of 266 PD cases and 273 controls from Flanders, Belgium. Results derived from this analysis showed a PD-specific p.A404W variant and 6 PD-specific variants in the 5' and 3' regulatory regions of *OMI/HTRA2*. Although it can not be ruled out that p.A404W is related to disease in this patient, further mutation analysis in a larger series of both cases and controls should be performed before establishing an association between this variant and PD risk. *In silico* analysis of the 4 variants they found in the 5' regulatory region of *OMI/HTRA2*, predicted that they all change binding sites of transcription factors. In order to investigate this possibility *in vitro*, the authors performed luciferase reporter gene analysis for all four 5' regulatory region variants identified in their population. Results derived from this analysis showed that two of these variants (g.53572436C4G and c.-442C4T) statistically decreased transcription activity in dopamine cells. Although modest effects seen between mutant and wild type alleles can be important *in vivo* (Knight, 2004), reporter analysis results can largely be affected by the

experimental design, tissue and cell type, genetic background and environmental context (Cirulli and Goldstein, 2007). Thus further analysis of *OMI/HTRA2* 5' regulatory variants should be done before drawing any conclusion about its contribution to PD pathology.

*OMI/HTRA2* was included in the original *PARK3* locus, spanning 10.3cM from marker D2S134 to D2S286 (Gasser, Muller-Myhsok *et al.*, 1998). Refinement of this locus 3 years later to 2.5 Mb by West and colleagues (West, Zimprich *et al.*, 2001) excluded *OMI/HTRA2*. However, since several independent linkage and association studies suggest the existence of a disease-modulating gene outside this refined *PARK3* region (Pankratz, Uniacke *et al.*, 2004), Strauss and colleagues unsuccessfully looked for mutations in the coding region of this gene in families B, C, D and K that defined *PARK3* locus. Since no mutations were found in these *PARK3*-linked families, *OMI/HTRA2* was later designated as *PARK13* (OMIM #610297). Along with *PARK3* this would be the second locus identified in the German population (3 out of the 6 families used to define *PARK3* had a German background) and not replicated elsewhere, maybe because of the design of the original studies.

The data presented here suggest that mutations of *OMI/HTRA2* are not a significant risk factor or cause of PD and show the necessity independent replication after an original description of association.

#### 5.4. ANALYSIS OF SCA-2 AND SCA-3 GENES IN PARKINSONISM

Here, we present the identification of a single case (374-1) with expansion at *ATXN-2*. Assessing relatives of 374-1 demonstrated an affected sibling, 374-5, who also showed pathogenic expansion at *ATXN-2* and an unaffected sibling (374-9) who did not have *ATXN-2* expansion, showing that this mutation is the cause of disease in this family. The clinical signs and symptoms associated with disease in these two affected members of family 374 are consistent with mild levodopa-responsive PD and are consistent with our hypothesis that mutations at *ATXN-2 locus* may cause typical PD.

To date, several families worldwide have been described to harbor expansion mutations within *ATXN-2* that co-segregate with a range of phenotypes including parkinsonism and pure PD with no cerebellar affection and the absence of any obvious ataxia (Gwinn-Hardy, Chen *et al.*, 2000; Gwinn-Hardy, Singleton *et al.*, 2001; Shan, Soong *et al.*, 2001; Furtado, Farrer *et al.*, 2002; Ragothaman, Sarangmath *et al.*, 2004; Wilkins, Brown *et al.*, 2004; Modoni, Contarino *et al.*, 2007). Our data, along with that published so far, indicates that *ATXN-2* expansion mutations could be a cause of familial parkinsonism ranging from 1.5 to 8% of cases depending on ethnicity, being the highest among in Chinese population (Gwinn-Hardy, Chen *et al.*, 2000; Shan, Soong *et al.*, 2001). Thus, this gene should be tested when positive family history is reported.

Finally, comprehension of factors determining a predominantly parkinsonian phenotype in SCA2 may shed light on the pathogenesis of PD.

## 5.5. WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 1

### 5.5.1. Whole-genome association

We have undertaken a preliminary analysis in an attempt to localize common genetic variation exerting a large effect on risk for PD in a series of 552 cases and neurologically normal controls from North Americans. Although we are aware that our sample size is of limited power, it is worth indicating that the description of a GWAS as unpowered is largely a misused term. The ability of GWAS to succeed depends on many factors, including the sample and individual effect size, and risk allele frequency. The size of each individual genetic effect within the sample being analyzed and the number of samples analyzed are primary limitations on any study. Thus, the smaller the risk associated with a common disease, the more samples will be required to identify this variant. Another factor to take into account is the risk allele frequency, as the more common the variant related to disease the more power to detect association at that variant. These factors are of course interrelated and thus, a risk variant with a high odds ratio that is relatively rare in the population is much more difficult to identify using genome-wide approaches than a variant linked to moderate or mild risk that is common within the population. The classic example of such a study happened to be one of the first GWAS performed, where variability at *CFH* was identified as a risk factor for age-related macular degeneration using only 96 cases and 50 controls. In this case, the risk allele was both common and exerted a substantial effect, increasing risk ~ 4-7 fold (Klein, Zeiss *et al.*, 2005).

Our data suggest that there are no common genetic variants that exert an effect of greater than an OR of 4 in PD. From the standpoint of experimental design this information is very useful; however, there are important drawbacks to this interpretation. First, these results can strictly only be applied to the current population. Second, analysis of YOPD cases, where a genetic effect is thought to be stronger, could reveal genetic variants with an effect of this size (Tanner, Ottman *et al.*, 1999; Payami, Zarepari *et al.*, 2002). Third, this statement is reliant on either genotyping the causal variant or efficient and complete tagging of the causal variant.

To date, the *locus* most robustly associated with risk of PD is the *SNCA* gene. We did not identify a significant association at this *locus* (table 35); however, given that the OR associated with this *locus* is estimated at 1.4 it is not surprising that we were unable to identify this association. The OR associated for this *locus* comes from a global genetics consortium that conducted a collaborative analysis of data on Rep1 polymorphism in *SNCA* promoter (see section 1.6.2.1) and its flanking markers (Maraganore, De Andrade *et al.*, 2006),

**Table 35:** p values across *SNCA* locus, previously describe to be associated with risk for Parkinson’s disease.

SNP ID	Position Genome build 36.1	Successful genotypes	HWE P value	Genotype P value	Dominant P value	Additive P value	Recessive P value	2-marker haploype P value	3-marker haplotype P value
rs356174	90,849,924	537	0.852	0.919	0.711	0.808	0.960	0.050	0.289
rs11931074	90,858,538	537	0.411	0.199	0.123	0.092	0.243	0.157	0.179
rs6842093	90,866,888	537	1.000	0.335	0.154	0.199	0.996	0.176	0.391
rs3775423	90,876,514	537	0.411	0.230	0.152	0.114	0.243	0.242	0.242
rs356204	90,882,565	537	0.262	0.334	0.143	0.186	0.510	0.220	0.228
rs356168	90,893,454	537	0.262	0.334	0.143	0.186	0.510	0.228	0.502
rs2736990	90,897,564	534	0.298	0.297	0.123	0.166	0.496	0.563	0.282
rs356191	90,907,143	537	0.161	0.334	0.473	0.756	0.269	0.227	0.253
rs356188	90,910,560	529	0.202	0.270	0.404	0.693	0.245	0.277	0.253
rs3775439	90,928,764	537	0.124	0.116	0.057	0.041	0.248	0.274	0.271
rs2197120	90,948,625	537	0.161	0.334	0.473	0.756	0.269	0.618	0.565
rs3822095	90,955,540	536	0.508	0.655	0.672	0.452	0.364	0.372	0.274
rs1812923	90,958,562	535	0.252	0.804	0.738	0.993	0.678	0.380	0.544
rs10005233	90,962,354	537	0.491	0.646	0.784	0.457	0.350	0.800	0.909
rs2583978	90,96,9349	537	0.262	0.469	0.473	0.698	0.419	0.950	0.758
rs1471483	90,973,315	536	0.489	0.650	0.842	0.492	0.356	0.814	0.948
rs2583985	90,974,962	537	1.000	0.895	0.639	0.685	0.916	0.873	0.661
rs1372519	90,97,6332	536	0.126	0.693	0.617	0.781	0.569	0.885	0.970
rs2301134	90,977,968	537	0.343	0.694	0.406	0.576	0.943	0.889	0.971
rs7687945	90,983,722	537	0.343	0.766	0.467	0.576	0.862	0.889	0.955
rs1372522	90,984,303	532	0.435	0.702	0.426	0.621	0.993	0.834	-
rs2736988	90,99,368	537	1.000	0.861	0.638	0.586	0.683	-	-

Analysis of our data showed 26 loci with a two-degree of freedom p value less than 0.0001 (table 28), with odds ratios ranging from 0.2 (95% CI 0.04–0.5) to 0.6 (0.5–0.8) and from 1.7 (95% CI 1.3–2.2) to 2.2 (95% CI 1.6–3.2). However, an important caveat to the power of GWAS is that genetic coverage comes at a statistical price; since hundred of thousand of statistical tests are applied to look for association, low p values can be seen as a result of chance only rather than because of biological association. Thus, while p values of 0.0001 would normally be convincing, in this context a p value of that order of magnitude is most likely a false positive association. In order to correct this caveat, different methods are applied. One of the most widely used is the Bonferroni correction, that states that if a experimenter is testing  $n$  independent hypotheses, the statistical

significance level that should be used for each hypothesis separately is  $1/n$  times what it would be if only one hypothesis were tested. Thus, a stringent Bonferroni correction based on 408,803 independent tests means that a precorrection p value of less than  $1.2 \times 10^{-7}$  would be needed to provide a corrected significant p value of less than 0.05. Thus, none of the values listed were significant after correction. Although speculation on the plausibility and biological significance of these candidate *loci* is tempting, we regard these data as hypothesis generating. Furthermore, given the inevitably high false-positive rate of genome-wide association studies (see below), the next step in these analyses should involve genotyping in additional sample series. In the first instance, this work should be done in a cohort comprising patients and controls of similar demographic characteristics to reduce the confounding factors of allelic and genetic heterogeneity between ethnic groups (see sections 3.6, 4.6 and 5.6).

A secondary objective of this experiment was to generate and release genome-wide genotype data for publicly available PD patients and controls so that these data could be mined and augmented by other researchers;

The release of genotype data and not just allele frequency data means that genotype data from additional samples can be added easily to the current set allowing investigators to undertake joint analysis rather than replication-based analysis. The former approach is more powerful than the latter in identifying common genetic risk factors (Skol, Scott *et al.*, 2006). It is worth noticing that the control samples in the current study have been specifically obtained so that they can be used for other neurological disorders, including but not restricted to stroke and amyotrophic lateral sclerosis, so these data will also be of use to other researchers outside of the PD specialty.

A GWAS of PD had been previously reported. Although this experiment consisted of fewer than half of the SNPs we typed in our experiments, the multistage design used added substantial power and sensitivity to the results (Maraganore, De Andrade *et al.*, 2005). The authors of these experiments suggested that their data revealed 13 SNPs associated with risk for Parkinson's disease. However, several studies failed to replicate such an association (Clarimon, Scholz *et al.*, 2006; Elbaz, Nelson *et al.*, 2006; Farrer, Haugarvoll *et al.*, 2006; Goris, Williams-Gray *et al.*, 2006; Li, Rowland *et al.*, 2006).

Side by side comparison of our data and the most significant SNPs, published by Maraganore and colleagues (Maraganore, De Andrade *et al.*, 2005), did not show a replication of any of these published associations (table 36).

Our calculations show that at least 32,127 SNPs are shared between Maraganore's and our study. However, at writing, they have not released raw genotype data, thus, a pooled analysis of the two studies is not possible.

**Table 36:** Comparison of significant *loci* identified by Maraganore *et al.*, with the current data set.

Chr. Location	SNP ID	Location bp	HWE <i>p</i> -value	2DF <i>p</i> value	Dominant Model <i>p</i> value	Additive Model <i>p</i> value	Recessive Model <i>p</i> value	2 marker Haplo.	3 marker Haplo							
5p15.2	<i>rs7702187</i>	9385281	0.411	0.818	0.972	0.834	0.535	0.172	0.291							
	rs3797980	9381316								0.375	0.279	0.132	0.110	0.352	0.132	0.095
	rs1205731	9388344														
2q36	<b><i>rs10200894</i></b>	228525376	0.600	0.918	0.994	0.821	0.694	0.815	0.868							
	rs1524029	228521122	0.664	0.549	0.785	0.721	0.360	0.686	0.535							
	rs1715826	228527516														
4q31.1	<b><i>rs2313982</i></b>	139007510	0.786	0.587	0.315	0.355	0.993	0.432	0.055							
7p14	<b><i>rs17329669</i></b>	36818454	0.754	0.012	0.910	0.312	0.004	0.296	0.509							
	rs9986865	36807127	0.247	0.135	0.046	0.107	0.667	0.306	0.408							
	rs6969810	36834535														
5p15.3	<b><i>rs7723605</i></b>	5407615	0.146	0.300	0.566	0.994	0.237	0.091	0.219							
	rs2964145	5401640	1.000	0.431	0.352	0.249	0.268	0.480	0.611							
	rs10059715	5413890														
2q24	<b><i>rs16851009</i></b>	166338953	0.284	0.278	0.186	0.490	0.704	0.499	0.198							
	rs4667836	166334852	0.314	0.744	0.462	0.442	0.693	0.122	0.126							
	rs4621175	166357038														
1p36.2	<b><i>rs2245218</i></b>	14012413	0.178	0.295	0.191	0.374	0.698	0.284	0.633							
	rs2245197	14011783	0.727	0.303	0.960	0.763	0.135	0.983	0.883							
	rs6663564	14017281														
Xq28	<b><i>rs7878232</i></b>	150597031	0.000	0.020	0.034	0.205	0.922	0.260	0.380							
	rs4363323	150595500	0.000	0.014	0.041	0.327	0.780	0.266	0.237							
	rs7879285	150597950														
4q31.1	<b><i>rs1509269</i></b>	138973174	0.836	0.008	0.094	0.344	0.031	0.457	0.193							
	rs978890	138963394	0.770	0.281	0.326	0.595	0.346	0.026	0.060							
	rs6815259	138982861														
4q27	<b><i>rs11737074</i></b>	125300823	0.859	0.380	0.218	0.173	0.354	0.395	0.198							
	rs11736972	125300772	0.095	0.864	0.668	0.617	0.666	0.295	0.209							
	rs2162138	125317106														
1p32	<b><i>rs682705</i></b>	54410005	0.897	0.740	0.467	0.534	0.996	0.622	0.622							
	rs3795360	54409790	0.456	0.836	0.578	0.549	0.710	0.622	0.622							
	<b><i>Rs7520966</i></b>	54417850														

Only two SNPs overlapped between the sets (rs2313982 and rs7520966), so in every other instance the closest flanking SNPs are shown. SNPs described as significant by Maraganore and colleagues are shown in bold italics. 2 marker haplotype or 3 marker haplotype *p* value of disease association to a two marker or three marker haplotype where the interrogated SNP is the first of two or three contiguous SNPs placed into a haplotype. Chrom = chromosome. Haplo = haplotype.

In summary, we present here the generation and release of genotype data derived from publicly available PD cases and neurologically normal control samples. All DNA samples, raw genotype data, and significance test results are publicly available at <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap>.

These data suggest that there is no common genetic variant that exerts a large genetic risk for late-onset PD in white North Americans. These data are now available for future mining and augmentation to identify common genetic variability that results in minor and moderate risk for disease.

### 5.5.2. Extended homozygosity in control population.

Analysis of B allele frequency and log R ratio metrics (see *data analysis* in section 3.5.2.2) allowed us to score tracks of contiguous homozygous calls in any given sample. Here, we describe a striking level of homozygosity in a cohort of aged, out-bred North American Caucasians. These data extend upon previous observations of large tracts of homozygosity in subjects typed as part of the HapMap initiative (Gibson, Morton *et al.*, 2006) and in subjects from reference families from the Centre d'Etude du Polymorphisme Humain (Broman and Weber, 1999). In the analysis of individuals genotyped as a part of the HapMap initiative Gibson and colleagues identified contiguous tracts of homozygosity >5 Mb in length in 3.8% of the subjects assessed (8 of 209) and multiple tracts >5 Mb in half of these subjects (Gibson, Morton *et al.*, 2006). Our data derived from a single population of aged individuals, shows a comparable prevalence of homozygous tracts of 5 Mb and larger (9.5%), and multiple homozygous tracts in almost half of these individuals (42.3%). While events such as segmental uniparental disomy could explain our data, the observation that individuals with at least one large tract of homozygosity are likely to harbor other large regions of homozygosity supports the idea that parental consanguinity is the cause. One would presume, given the increased mobility of North Americans in the last 50 years, that analysis of younger subjects would reveal a lower incidence of individuals with extended tracts of homozygosity. These data suggest the utility of whole genome SNP analysis to perform homozygosity mapping of disease *loci* not only in families in which a recessive mode of inheritance is detected, but in cohorts of apparently unrelated young-onset cases. Any indication of parental consanguinity (isolated population, etc.) would prioritize a sample for analysis. This

approach would allow rapid fine mapping of disease intervals, with the advantage of simultaneously revealing copy number changes often associated with recessive loss of function alleles. In samples from individuals with disease, one would compare the regions of coincident extended homozygosity among cases with those in neurologically normal controls. Clearly, successful application of this approach will require large data sets, or enrichment of mutations, either by using a population in which one suspects a founder mutation or by using young-onset PD patients, in which recessive disease and structural genomic alteration is believed to play a larger role than in late-onset PD. Notwithstanding these limitations, the relative speed and ease of performing GWAS experiments suggests that the next several years will see leverage of this technique in the identification of recessive *loci* underlying disease.

### **5.5.3. Copy number variation in neurologically normal controls**

Until recently, the use of genome-wide SNP assays as a direct method of determining copy number variation (CNV) has not been widely appreciated (Peiffer, Le *et al.*, 2006). In our experiments, we identified genomic copy number changes in 182 control samples (340 changes in total). In an effort to demonstrate whether these changes were a result of the creation or passage of LCLs or not, we repeated the genotyping experiment in those 24 samples in which alterations were detected with the Human-1 array. Thirteen of the identified changes were readily detected in the source tissue used for immortalization, and concordance rates were higher for smaller structural alterations (75% for those changes <5 Mb; 13% for those changes >5 Mb). Notably, this method discriminates heterosomic and heterozygous genomic deletions, as each allele for each SNP is assayed

multiple times. This redundancy can be used to provide a proportion for allele calls at each SNP, thus significant deviation from homozygous (100% allele A or allele B) and heterozygous (50% allele A and 50% allele B) clusters, particularly when observed in multiple contiguous SNPs not only indicates structural alteration but gives a measure of the proportion of cells containing this genomic change. The single simple deletion that we observed with the Human-1 array, was present in both tissues examined.

In the case of the identified heterosomic deletions, we can establish these events as somatic, occurring in the source tissue in one instance, with the remaining events resulting from, or emphasized by, the process of LCL creation. An analogous process can be used to assess heterosomic duplications although in this case, the resolution is more difficult because one needs to interpret deviations from a 66% / 33% read which is more challenging. Real-time PCR analysis of sample *ND04946*, which whole-genome genotyping assay suggested a heterosomic deletion resulting in a ~30% drop in copy number, also revealed a drop in copy number of ~30%. Previous reports have identified structural alterations in the human genome as small as 1 kb; however, these approaches rely either on the availability of high density parent offspring genotype data (Rosenberg, Vaske *et al.*, 2000; Conrad, Andrews *et al.*, 2006; Mccarroll, Hadnott *et al.*, 2006), or the use of custom manufactured arrays (Iafrate, Feuk *et al.*, 2004; Sebat, Lakshmi *et al.*, 2004; Sharp, Locke *et al.*, 2005).

The lower limit in terms of size of structural genomic variability detected in the current study is ~ 0.02 Mb, a limit which is sensitive to the relative density and informativity of SNPs assessed in any one region; thus, the average limit of resolution is likely to be in the range of hundreds of kilobases as is most often seen in the data described here. Although

the inability of this method to detect balanced translocations would suggest that this will not yet replace current approaches aimed at defining chromosomal abnormalities, the simple and transferable nature of this technique means it may augment these methods.

In addition to the usefulness of this technique for quickly mapping structural variability with good resolution, this methodology is appealing because it is easily transferable between laboratories. This would greatly facilitate the rapid accumulation of equivalent data to produce an encyclopedia of normal and abnormal structural genomic variation, similar to previous efforts for gross chromosomal abnormalities (<http://jws-edck.wiley.com:8096/>) (Borgaonkar, Bolling *et al.*, 1975; Rosenberg, Vaske *et al.*, 2000; Iafrate, Feuk *et al.*, 2004; Sebat, Lakshmi *et al.*, 2004; Sharp, Locke *et al.*, 2005; Conrad, Andrews *et al.*, 2006; Mccarroll, Hadnott *et al.*, 2006; Peiffer, Le *et al.*, 2006) and enable standardized comparison between laboratories. For this purpose, The Database of Genetic Variants (<http://www.tcag.ca>) gathers detailed information about all the copy number variation identified so far (including ours) in control population.

The utility of this methodology to address other research questions is striking. One can envisage using these assays to catalog the occurrence of both germ-line and somatic structural mutation, both in disease and healthy states. This method also promises to enable genetic characterization of many of the key tools used within laboratories. For example, chromosomal rearrangements have been described in cell lines commonly used in molecular biology laboratories. This may be an evolving process, rather than a stable alteration, and if so, would presumably affect experimental outcome. Routine assay of standard cell lines would allow inter-laboratory standardization. The effect of creating stem cell lines on the genome has not been established; again this technique offers a

quick first-pass assay for analyzing the effects of clonally deriving stem-cell lines both for research and for therapies. In all instances the availability of a standardized product to perform these assessments is of great importance. Clearly, a weakness of the current approach is that scoring of structural variability was performed using a largely subjective measure; while we attempted to minimize the false positive and false negative rate by independent calling of structural variation by two experienced data analysts it is likely that some variability was missed. The development of a publicly available tool that uses a statistical basis for the detection of structural variation that can be applied to these data would greatly enhance the ease with which these analyses can be performed. Wang and colleagues have recently developed PennCNV, an algorithm for copy number variation detection demonstrated to be reliable for finemapping of copy number variation through high-density SNP genotyping. (Wang, Li *et al.*, 2007) (<http://www.neurogenome.org/cnv/penncnv>).

Our study shows that the relative effects of LCL creation and passage on the genomic architecture of both individual samples and the cohort as a whole are minimal and that the genetic content of LCLs is relatively faithful to that of the starting material used for LCL creation. In the context of GWAS the variation which appears to have resulted from, or have been amplified by, the process of creating LCLs is unlikely to significantly interfere with allelic or genotypic associations. This variability is limited to a relatively small number of samples and in these samples the genomic region affected represents a small percentage of the genome. Thus, the aggregate effect of this variability will be to alter <0.1% genotype calls (~25,000 individual genotypes potentially affected, from a total of 30 million surveyed). While this bodes well for the use of these data for genetic

association studies, further experiments defining the effect of extended culture and passage, if any, will be required to see if this increases structural alteration and potentially effects allele calling. In conclusion, this study provides public access to whole genome array data in a large series of Caucasians and illustrates that such data is useful, not only for GWAS for which it was planned, but also for determining the extent of inbreeding and structural mutations. Its simplicity suggests unexpected uses for this technology in the assessment of standard laboratory cell lines, stem cells and high resolution cytogenetic testing. Its speed and ease may lead it to replace traditional linkage analysis strategies, especially for recessive diseases but also because it will allow easy assessment of linkage disequilibrium between laboratories without the need for uncertain standardization against CEPH controls. The data presented here also show that the effects of LCL creation and passage on genotypes and genetic architecture are minimal and unlikely to confound the results of a genomewide case–control study.

#### 5.5.4. Copy number variation in Parkinson's disease patients

The aim of this analysis was to assess the role of structural genetic variation in risk for PD both genomewide, and at previously described PD *loci*. Most of the changes described herein (79.1% of the multiplications and 72.6% of the deletions found) had previously been described in control samples (<http://www.tcag.ca>). Although we can not rule out the remaining alterations (38 duplications and 44 deletions) as genetic *loci* related to PD etiology; it is likely, given the large number of unique CNVs seen in control samples, that the majority, or all, are benign variants; however, we have provided a

comprehensive list of these alterations (see supplementary material, table S4 and S5) to serve as the foundation for future work looking at structural genomic alterations in PD.

While this method was applied primarily to detect novel regions related to sporadic PD, we have identified some biologically interesting results in previously PD-linked *loci*. Not only did this technique reveal a homozygous deletion, which is likely pathogenic, but it also showed the presence of other structural mutations in *PARK2*, in both cases and controls. The pathological relevance of single heterozygous *PARK2* mutations, particularly to late-onset PD, has been an area of much discussion (Klein, Pramstaller *et al.*, 2000; Hoenicka, Vidal *et al.*, 2002; Oliveira, Scott *et al.*, 2003; Sun, Latourelle *et al.*, 2006; Kay, Moran *et al.*, 2007). In contrast to the data presented here there is considerable evidence, primarily from family based studies, that single *PARK2* mutations may be a risk factor for late-onset disease. However, as discordant penetrance rates in families versus populations attest, the ascertainment bias in family-based studies is considerable and these should not necessarily be relied upon to assess the relative contribution of genetic risk factors to disease. One point of interest, however, is the observation that in control samples, only *PARK2* duplications and not deletions have been identified. This is conceivably a result of chance but it could also indicate that these duplications do not necessarily interrupt the open reading frame of *PARK2*. Although heterozygous deletions at *PARK2* were only found in PD samples, lack of statistical evidence after Fisher's exact test prevents us on stating that these mutations are causal of disease in these samples.

## **5.6. WHOLE-GENOME SNP GENOTYPING IN PARKINSON'S DISEASE AND NEUROLOGICALLY NORMAL CONTROLS. STAGE 2.**

As discussed in section 5.5.1, in order to gain statistical power and be able to identify common genetic variation (SNP) related with a risk of PD, the GWAS should be complemented by genotyping a larger cohort comprising patients and controls of similar demographic characteristics to reduce the confounds of sample size, individual effect size and risk allele frequency. Besides, since in YOPD the genetic effect seems to be larger (Tanner, Ottman *et al.*, 1999; Payami, Zarepari *et al.*, 2002), individuals with this variant of the disease should be included as well. Taking into account these two points, we increased our sample size to 4,134 individuals, including 880 late-onset PD, 183 YOPD and 3,071 control samples (before quality control clean-up). Each of these samples was genotyped at greater as 545,066 unique SNPs genome-wide with Infinium technology. Genotyping data for all these samples will be released on the database of Genotype and Phenotype (dbGaP, <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap>) as part of a National Institutes of Health initiative to publish all genotypic data produced with intramural funding (data deposition in progress).

As explained in section 5.5, the data set utilized for stage 1 of this project (sections 3.5, 4.5 and 5.5) did not provide enough statistical power to detect association between *SNCA* and PD. However, data from this later stage of the project, allows the detection of variability exerting an OR as low as 1.3. Thus, it is not surprising that 3 of the 10 SNPs with the lowest p values ( $1.06 \times 10^{-7}$ ,  $4.8 \times 10^{-7}$  and  $2.04 \times 10^{-6}$ ) were obtained in this *locus*.

Although association of *SNCA* with PD is widely known, these results show the robustness of our dataset and can be used as a positive control for association seen in other *loci*. On this last regard, no evident association with any other *locus* was detected in our analysis. These data indicate that, at least in our population and considering that the causal variant is either typed or completely tagged with one of the SNPs assayed, there is not a genetic factor exerting a risk for PD as large as variability at *SNCA*.

In order to add power to our dataset, we undertook a close collaborative analysis with Drs. Gasser and Toda, who have performed a similar approach in German and Japanese populations respectively. Because of genetic background differences between these populations, different approaches were taken for the collaborative analysis. Thus, genotype counts from the results of genotyping the German series, were merged with ours. Although it could be argued that genetic heterogeneity (different risk genes in different populations) and allelic heterogeneity (different risk alleles in same gene in different populations) could mask association when pooling different populations for analysis, empirical evidence generated thus far supports the notion that this is not likely in relatively close populations such as North European and North American Caucasians (Saxena, Voight *et al.*, 2007; Scott, Mohlke *et al.*, 2007; Zeggini, Weedon *et al.*, 2007).

On the other hand, given the considerably genetic distance between Caucasian and Asian populations, this approach could not be performed for those results from Dr. Toda's lab. Thus, side by side comparison was performed in this case.

Combined results for the German and the US cohorts, confirmed the strong association with *SNCA*, with 4 out of the 6 most associated SNPs being located at this *locus* (p values of  $3.99 \times 10^{-9}$ ,  $2.61 \times 10^{-8}$ ,  $3.5 \times 10^{-8}$  and  $5.86 \times 10^{-8}$ ). Interestingly, another gene highly

related to the pathogenesis of PD (*MAPT*) showed a strong association to PD in this cohort (see section 1.6.3.9). Several studies suggest that protein tau can coaggregate with  $\alpha$ -synuclein in the Lewy bodies, and both may interact to promote the formation of synuclein fibrils (Ishizawa, Mattila *et al.*, 2003). Interestingly, this was genetically suggested by Goris and collaborators, who not only showed evidence for a contribution of common variation in *MAPT* and *SNCA* to PD, but that the combination of both risk genotypes doubled the risk for PD (Goris, Williams-Gray *et al.*, 2007).

Mutations in *LRRK2* are the most common cause of PD, accounting for about 1-2% (Gilks, Abou-Sleiman *et al.*, 2005) of sporadic and up to 13% of familial PD cases (Paisan-Ruiz, Jain *et al.*, 2004; Zimprich, Biskup *et al.*, 2004; Berg, Schweitzer *et al.*, 2005). However, whether *LRRK2* common variants (such as SNPs) are associated with common forms of PD is still controversial (Paisan-Ruiz, Lang *et al.*, 2005; Paisan-Ruiz, Evans *et al.*, 2006). Interestingly, although we did not find association in *LRRK2* locus, combined analysis of the German and the US samples, showed 23 SNPs with p values below 0.05 (ranging from in  $5.2 \times 10^{-6}$  to 0.05) in a gene  $\sim 0.5$  Mb apart from *LRRK2*. This gene's name is *SLC2A13* (solute carrier family 2, member 13) and is a facilitated glucose transporter. Interestingly, results obtained in the Japanese cohort showed that 33 SNPs within this locus had p values below 0.05 (ranging from  $1.82 \times 10^{-5}$  to 0.04). Whether this association is driven because of proximity to *LRRK2* or is a product of chance only, will be ruled out by follow-up studies in different populations.

Interestingly, 7 SNPs out of those 113 with trend p values below  $1 \times 10^{-5}$  in the combined Caucasian analysis were located in *CUL2*, a component of E3 ubiquitin ligase complexes, involved in degradation of multiple proteins. A neighboring gene (*CREM*, 0.1

Mb apart) also showed two hits below this threshold in the combined German and US analysis. Interestingly, although no strong association was seen in either of these genes in the Japanese cohort, a gene (*PAR3*) ~ 0.9 Mb had 11 SNPs with trend p values of the order  $1 \times 10^{-4}$ , showing different haplotype structures across populations exert different association to disease. Figure 47 shows association values for this region in chromosome 10 across all populations studied. Again, replication experiments in other populations are needed to establish any association.

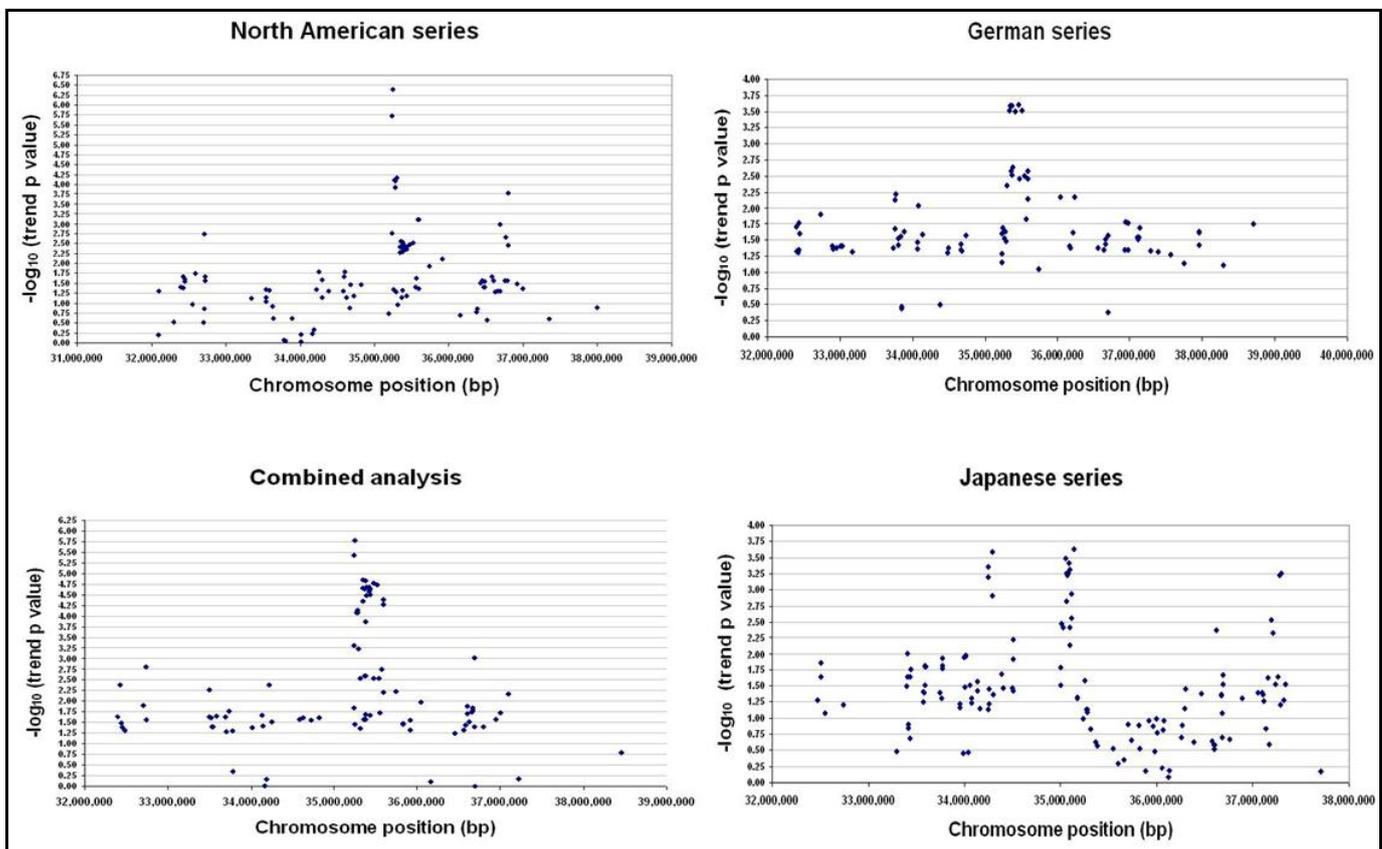


Figure 47:  $-\log$  trend p values of ~ 7 Mb region of interest in chromosome 10 in all populations assayed.

and control samples. Although association found should be confirmed by an appropriate follow-up study in different populations (ongoing project), several factors add robustness

to the results herein presented. First, a close collaboration with Dr. Gasser's group, allowed us to perform a combined analysis that substantially increased our dataset and thus, power to add association. Second, collaboration with Dr. Toda's group provided us the opportunity to side-by-side compare our results with a population with a notably different genetic background, adding considerable reliability to any association found across populations. Last, strong association signals found in genes previously linked to PD (*SNCA* and *TAU*) can be considered as positive controls, adding power to other results found genome-wide. These data suggest that there is no common genetic variant that exert a larger genetic risk to PD that *SNCA* or *TAU* do in the populations studied.

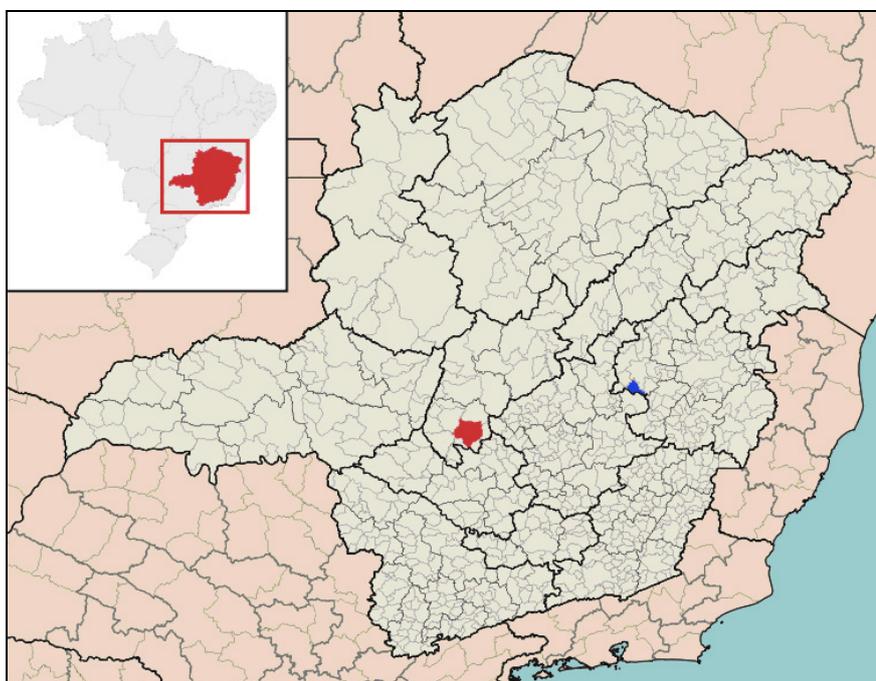
All DNA samples, raw genotype data, and significance test results are publicly available at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap> as part of a National Institutes of Health initiative.

On regard to the mentioned follow-up experiments, these will be done using Illumina's custom GoldenGate genotyping technology for VeraCode (Illumina, <http://www.illumina.com>). This technology is being used to genotype those 384 SNPs in which lowest p value has been detected in the US-German combined analysis. These genotyping experiments are being performed in ~ 3,000 PD cases and ~ 3,000 PD neurologically normal controls originally from the United Kingdom, Germany and North America.

### **5.7. AUTOZYGOSITY MAPPING IN BRAZILIAN FAMILIES**

We identified two consanguineous families from Southeast-central Brazil with a distinct young-onset movement disorder, most commonly presenting with dystonia, but also with

parkinsonism. The two families, *DYT16-1* and *DYT16-2*, were apparently unrelated and resided in two cities approximately 350 Km apart within Minas Gerais state, Brazil (figure 48). Given the resemblance of the clinical features between affected members across both families, we hypothesized that they were ancestrally related and shared a minimum interkindred haplotype leading to disease.



**Figure 48: Geographical location of cities where families *DYT16-1* (Dores de Guanhões, filled in blue) and *DYT16-2* (Bom Despacho, filled in red) are originally from**

Autozygosity mapping with a high-density whole-genome SNP analysis revealed a large disease-segregating region of homozygosity identical by state in all six affected family members. Sequence analysis of genes and predicted transcripts from this interval and two smaller disease-segregating intervals revealed a single disease-segregating mutation,

c.665C>T (p.P222L), in exon 7 of *PRKRA* (position based on NCBI accession numbers NM\_003690 and NP\_003681). This mutation was not identified in the NCBI SNP database or by sequence similarity search of the human nucleotide collection at NCBI (figure 49). Of note, in family *DYT16-2*, affected members have been confirmed in two generations (figure 46). In addition an affected family member (mother of 2035-119) was also reported to be affected, which may suggest additional parental consanguinity. However this member was not examined and we, therefore can not rule out the possibility of phenocopy or inaccurate reporting of the affected status. Interestingly, we found this same homozygous mutation in an additional patient with generalized dystonia from Parade Minas city, 276 and 80 km apart from Dores de Guanhães and Bom Despacho respectively within Minas Gerais state. This patient latter revealed to have a region spreading from rs6738749 to rs10497541 (~1.16 Mb), identical by state with all affected members of families *DYT16-1* and *DYT16-2*. Identification of such a segregating region, in a single rare disorder, strongly suggests that the genetic mutation causing this disease resides in this genomic region.

Screening for this mutation failed to reveal any carriers in 11 other dystonia cases from Brazil, 294 young-onset PD cases from the United States, 948 Neurologically normal controls from the United States, or 738 samples from diverse human populations including 44 Brazilian samples.

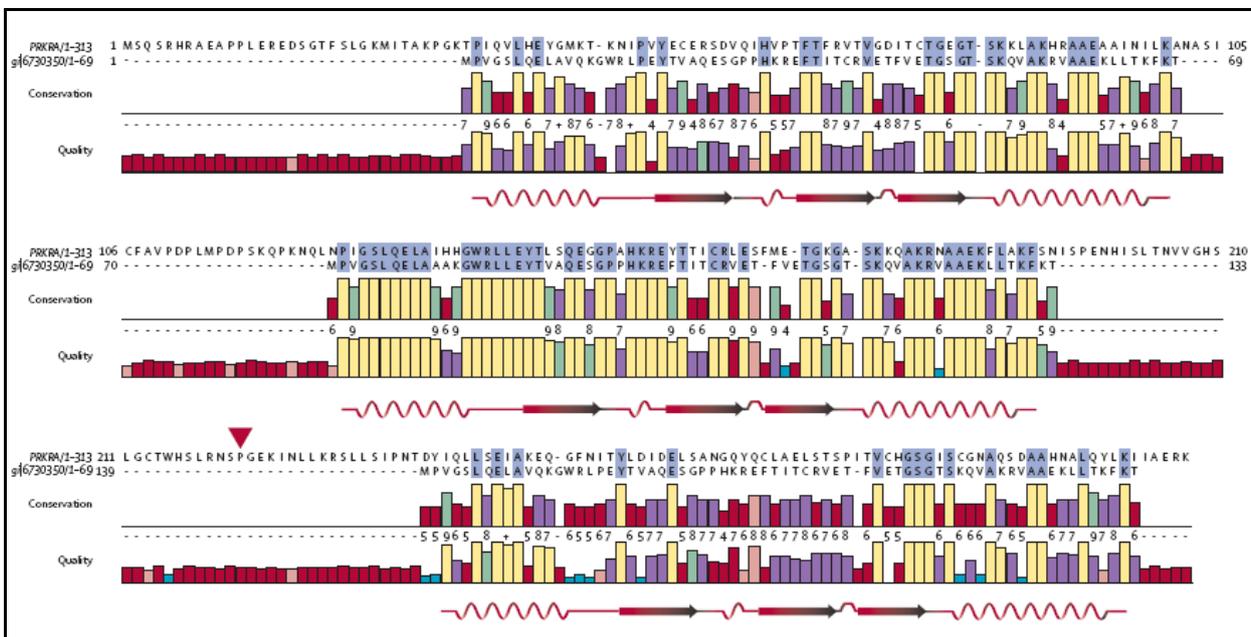
We have not examined 5' or 3' untranslated regions, or the predicted regulatory regions in these families, and we therefore can not rule out non-coding variability as a cause of the disease. However, absence of the specific p.P222L mutation from a large cohort of control samples indicates that this variant is the underlying causal mutation in these

families. This supposition is bolstered by the fact that no other segregating mutations were identified in the 457 other known and predicted exons of this region. However, to be absolutely confident that this mutation in *PRKRA* underlies this disease, independent identification of disease-segregating mutations is required. Interestingly, prompted by the fact that the mother of individual 2035-19 (a homozygous carrier of p.P222L mutation) was affected and could not be genotyped in our experiments, Seibler and coworkers hypothesized that heterozygous mutations in *PRKRA* may be related to dystonia (Seibler, Djarmati *et al.*, 2008). In order to investigate this possibility, they performed a mutation analysis of *PRKRA* in a series of cases and controls from Germany. Interestingly enough they found a heterozygous frameshift mutation in exon 3 of *PRKRA* (c.266\_267delAT; p.H89fsX20) in a patient with young-childhood-onset leg dystonia. His family history was unremarkable for any movement disorder, secondary causes of dystonia were excluded, and brain MRI was normal. At his last examination, the patient, who is now aged 9 years, had generalised dystonia that spared only the cranial and facial muscles, his walking was severely impaired, but his cognition was normal (Seibler, Djarmati *et al.*, 2008).

Although these results do not confirm that p.P222L mutation in *PRKRA* causes disease in our families, they support the idea of *PRKRA* as a risk factor for generalized dystonia.

The biological mechanism by which the p.P222L mutation in *PRKRA* may cause dystonia-parkinsonism is not clear. *PRKRA* encodes protein kinase, interferon-inducible double-stranded RNA-dependent activator (aliases PACT, RAX, HSD14). In response to extracellular stresses, *prkra* activates the latent protein kinase *pkp*, a protein involved in signal transduction, cell differentiation, cell proliferation, antiviral response, and

apoptosis (Patel, Handy *et al.*, 2000). More specifically, pkr is thought to inactivate the eukaryotic translation initiation factor 2 $\alpha$  (eif2 $\alpha$ ), which in turn inhibits protein synthesis (D'acquistio and Ghosh, 2001). Of note, wild-type prkra is a component of the human RNA-induced silencing complex, which also regulates protein synthesis via cleavage of mRNA (Lee, Hur *et al.*, 2006). The p.P222L mutation alters an aminoacid that is conserved across mammalian species at a residue between the second and third RNA-binding motifs (figure 49). This mutation may cause structural alteration of prkra and/or affect substrate affinity. Unfortunately, the crystal structure of this or related proteins has not been solved and thus we can not make any confident predictions about the likely effect of this variant on protein conformation or function.



**Figure 49: *In silico* sequence analysis of PRKRA.** Sequence analysis predicts that PRKRA encodes three copies of a double-stranded (ds) RNA-binding motif (DSRM). DSRM is a conserved domain in various proteins, including dsRNA-dependent protein kinase PKR, RNA helicases, *Drosophila melanogaster* staufen protein, *Escherichia coli* RNase III, RNases H1, and dsRNA-dependent adenosine deaminases. Upper panel shows the sequence-to-structure alignment between PRKRA sequence and the three-dimensional structure of a dsRNA-binding domain (gi:670350, NCBI Structure database). The yellow-brown bar plot shows conservation and alignment quality. The red swirl ribbons and arrowed ribbons show the  $\alpha$ -helix and  $\beta$ -strand, respectively, in the three dimensional structure of DSRM motif. The location of the p.P222L mutation is indicated with a red arrow.

On the basis of our data, *DYT16* disease should be considered when a patient presents with an apparent or suspected recessive pattern of inheritance of generalized dystonia, with involvement of the muscles of the neck and trunk to a greater extent than that of the limbs. Other typical features are oromandibular (sardonic smile), laryngeal dystonia, and parkinsonism, although the latter is less important than the dystonic features. All patients examined thus far have presented spasmodic dysphonia and sardonic smile. All patients failed to improve with any pharmacological treatment, including levodopa and high-dose anticholinergics. In contrast to *DYT16*, patients with *DYT1* first have limb involvement that rapidly spreads to other limbs and the trunk, ultimately becoming severe; furthermore, *DYT1* is a pure dystonia with an autosomal dominant mode of inheritance (although it is worth noting that the reduced penetrance could lead to suspicion of recessive inheritance) (De Carvalho Aguiar and Ozelius, 2002). The lack of response to levodopa in patients with *DYT16*, the absence of myoclonus, and sensitivity to alcohol differentiate this disease from *DYT5* and *DYT11*, respectively (De Carvalho Aguiar and Ozelius, 2002). There are many similarities between *DYT16* and *DYT12*, such as prominent bulbar signs and rostro-caudal gradient; however, *DYT12* has an abrupt onset and a clear autosomal dominant mode of inheritance (Brashear, Dobyns *et al.*, 2007). *DYT16* may also be easily differentiated from the *PARK* loci associated with parkinsonism and dystonia such as *PRKN* and *PARK9*-linked disease. The observed dystonia in *PRKN* disease is usually focal in nature and does not progress to a generalized dystonia, and patients with *PRKN* diseases are usually responsive to levodopa therapy. *PARK9* disease is associated with both dementia and a supranuclear gaze palsy, (Ramirez, Heimbach *et al.*, 2006) both of which are absent in *DYT16* disease.

In summary, we have described a novel autosomal recessive dystonia-parkinsonism syndrome in Brazilian patients that we have designated *DYT16*. Using high density SNP genotyping arrays, we have identified a genetic *locus* associated with disease. The homozygous track of SNPs shared between all mutation-positive samples clearly shows that this region is derived from a common ancestor. After complete sequence analysis of all coding regions within the *locus*, we show a single disease-segregating mutation within *PRKRA*. This mutation was present in three of 14 probands with generalized dystonia from the Brazilian population, representing 21.4% of independent cases. The P222L variant might be not pathogenic, and could be in linkage disequilibrium with the actual disease-causing variant. However, the absence of the mutation in such a large series of controls and our inability to identify other mutations, despite screening all other genes in the identified region, clearly supports our assertion that mutation in *PRKRA* is the causative genetic mutation in *DYT16*.

## 6. SUMMARY

Parkinson's disease (PD) is a heterogeneous disorder in which genetic factors play an important role. To date, six genes have unequivocally been linked to familial forms of the disease and a handful of genomic *loci* related to both the sporadic and the familial variants. This grouping into familial and sporadic variants does not imply that these two types of the disease should be considered as different entities. In fact, because of the clinical and pathological similarities existing between familial and non-monogenic PD, knowledge of the genes causing the familial disease could bring new ideas for the research of sporadic forms of the disease and vice-versa. Thus, in order to help understanding the etiology of the disease, we have analyzed different groups of patients in which different variants of PD had been diagnosed including autosomal dominant, autosomal recessive and sporadic.

The first approach we took to fulfill this objective was helping to determine the minimal common haplotype carrying the gene with mutations leading to autosomal dominant PD in four families from the Basque Country previously linked to *PARK8 locus*. This approach led to the discovery of *LRRK2* as the gene carrying the mutation that causes PD in these families. Interestingly, this mutation was also present in ~7% of PD patients from the Basque country including both familial and apparently sporadic cases. Haplotype analysis in all these patients led us to conclude that all had inherited the mutation as part of an ancestral chromosome. Further mutations have been identified in this gene being one of them (p.G2019S) the most common cause of both familial and apparently sporadic

PD. To further investigate the role of this gene in PD pathology, we thought that it would be interesting to provide a complete atlas of its mRNA distribution. With this objective we performed an *in situ* hybridization analysis with digoxigenin-labeled RNA probes on mice brain coronal sections. This approach showed that *LRRK2* was expressed in brain regions related with PD like the *substantia nigra pars compact* or the caudate putamen.

After helping to the identification of the gene carrying the mutation that cause the *PARK8* variant of autosomal PD, we thought it might be interesting to investigate if mutations in another autosomal dominant *locus* (*PARK13*) were actually involved in PD pathology. Mutation in *OMI/HTRA2* had recently identified to cause PD in German patients and this *locus* given the name of *PARK13*. Since in genetic research, as important as an original finding is replication of this finding in a different population, we thought it would be interesting to confirm *PARK13* results in a large series of Caucasian individuals from North America. Data obtained from these experiments showed that those variants previously described as pathogenic are present in neurologically normal control population and not associated with PD in our cohort. We feel this is a critical and important finding; this *locus* has already been given a *PARK* designation (<http://www.ncbi.nlm.nih.gov/entrez/dispmim.cgi?id=610297>) and we feel, given the evidence presented, that such a designation is probably misleading.

Another contribution to the understanding of autosomal PD was taken in a set of populations from different locations worldwide in which we analyzed expansion mutations in *ATXN2* and *ATXN3* genes. These genes are a known cause of

spinocerebellar ataxias type 2 and 3 respectively but had also been related to clinical features of PD. We failed on finding any mutations in *ATXN3*. However, we found segregating expansion mutations in *ATXN2* in a Caucasian family, showing the necessity of testing these expansion mutations when positive family history is reported.

On regard to the autosomal recessive variant of parkinsonism, we identified two consanguineous families from Southwest Brazil with a rare recessive dystonia-parkinsonism syndrome. Because of similarities in the phenotype they presented and because of the relatively proximity of the cities this families were from, there is a high possibility that affected individuals had inherited both copies (paternal and maternal) of the mutated gene from a common ancestor. Autozygosity mapping in these families led us to the identification of *PRKRA* as the gene carrying the mutations that cause disease in these two kindreds.

As for the non-Mendelian forms of PD, in order to unveil the role of common genetic variation in sporadic PD, we embarked a genome-wide association study in a population of North American Caucasians. In the first stage of this project, we genotyped more than 408,000 SNPs in a series of 552 idiopathic PD cases and neurologically normal controls from the National Institute of Neurological Disorders and Stroke (NINDS) funded Neurogenetics repository. Given the limited power of the cohort genotyped in this stage of the project, it is not surprising that we did not find any convincing disease-associated *loci*. Although it would be disingenuous to suggest that we were not aiming to identify genetic risk *loci*, an important aim of this work was also to provide a set of publicly

available genome wide SNP data for others to work with and augment. Interestingly, these data have been downloaded several hundred times and used in several publications originating from outside our group.

Besides SNPs, whole-genome genotyping technologies allow the identification of copy number variants (CNVs) as well as the detection of extended homozygosity. Extensive analysis of our data revealed 201 novel CNVs in control population that have been published at the database of genomic variants (<http://www.tcag.ca>) to help the construction of an encyclopedia of this type of variation. On regard to PD, we identified 82 novel CNVs not published in the database of genomic variants. Although we can not rule out that these are at genetic *loci* related with PD, it is likely, given the large number of unique CNVs seen in control population, that the majority, or all, are benign variants. However we provided a comprehensive list of these alterations to serve as a foundation of future works. While this method was applied primarily to detect novel regions related to sporadic PD, we have identified a homozygous deletion, which is likely pathogenic, and other structural mutations in *PARK2* in both cases and controls. Although heterozygous deletions at *PARK2* were only found in PD samples, lack of statistical evidence after Fisher's exact test prevents us on stating that these mutations are causal of disease in these samples.

Besides, we found striking levels of homozygosity in ostensibly control population. The observation that individuals with at least one large tract of homozygosity are likely to harbor other large regions of homozygosity supports the idea that parental consanguinity is the cause. These data suggest the utility of whole genome SNP analysis to perform

homozygosity mapping of disease *loci* not only in families in which a recessive mode of inheritance is detected, but in cohorts of apparently unrelated young-onset cases.

In the second stage of this project, our dataset was increased to ~545,000 unique SNPs typed in 880 PD patients and 3,071 controls. In order to add power to our genome-wide association study we undertook a collaborative analysis with a German (742 cases and 944 controls) and a Japanese group (1,000 cases and 2,500 controls). Results derived from these experiment showed strong association of two genes largely related to the etiology of PD: *SNCA* and *MAPT*. These results support the robustness of our analysis and serve as positive controls for other association detected in other *loci* not previously related to PD. On this regard, although should be replicated in other populations, interesting association was detected in several *loci* including *SLC2A13* (<0.5 Mb away from *LRRK2*) and a *locus* in chromosome 10 including *CUL2*, *CREM3* and *PARD3*. In order to verify the results obtained in this second stage of the genome-wide association project we are currently genotyping the top 384 “hits” obtained in the USA-German combined analysis in various populations worldwide.

## 7. CONCLUSIONS

1. Mutation p.R1441G in *LRRK2* is the cause of Parkinson's disease in ~7% of PD cases with and without positive family history from the Basque Country.
2. All p.R1441G carriers from the Basque Country share a common minimal disease haplotype suggesting that they have inherited the mutation from a common ancestral chromosome. This haplotype was shared by a group of p.R1441G carrier cases from Asturias, suggesting a gradient of this mutation in Northern Spain.
3. *LRRK2* is expressed throughout the mouse brain. However, a stronger expression is seen in specific brain areas such as the deep cerebral cortex, cingulate cortex, piriform cortex, Ammon's horn and dentate gyrus, caudate putame, *substantia nigra*, some parts of the amygdala, reticular thalamic nucleus or cerebellar granular cell layer.
4. Mutations in *OMI/HTRA2 (PARK13)* are not a significant risk factor or cause of Parkinson's disease a North American population.
5. Expansion mutations in *ATXN-2* are a cause of parkinsonism with clinical features overlapping those of Parkinson's disease in 1.5 to 8% of cases with positive family history. Thus, this gene should be tested when positive family history is suspected.

6. We have identified 201 novel copy number variants (CNVs) in control population that have been published at the database of genomic variants (<http://www.tcag.ca>) to help the construction of an encyclopedia of this type of variation.
7. We have identified 82 novel CNVs not published in the database of genomic variants. Although we can not rule out that these are at genetic *loci* related with Parkinson's disease, it is likely, given the large number of unique CNVs seen in control population, that the majority, or all, are benign variants. Type and location of these variants have been published so that they can serve as foundation of future works.
8. We have identified a homozygous *PARK2* deletion in a patient with young-onset Parkinson's disease. These finding suggests the utility of whole genome genotyping techniques to find structural changes causing disease. Although heterozygous deletions at *PARK2* were only found in Parkinson's disease samples, lack of statistical evidence after Fisher's exact test prevents us on stating that these mutations are causal of disease in these samples.
9. A genome-wide association study in a large series of Caucasian samples from North America suggests that, at least in this population and considering that the causal variant is either typed or completely tagged with one of the SNPs assayed, there is not a genetic factor exerting a risk for PD as large as variability at *SNCA*. This conclusion was expanded to *MAPT* in a combined analysis of US and German samples.

10. Although should be replicated in other populations, interesting association was detected in several *SLC2A13* (<0.5 Mb away from *LRRK2*) and a *locus* in chromosome 10 including *CUL2*, *CREM3* and *PARD3*.
  
11. p.P222L mutation in *PRKRA* is the cause of a weird syndrome presenting with juvenile dystonia and parkinsonism in two consanguineous families from Southeast Brazil.

## 8. BIBLIOGRAPHY

Abbas, N., C. B. Lucking, S. Ricard, A. Durr, V. Bonifati, G. De Michele, S. Bouley, J. R. Vaughan, T. Gasser, R. Marconi, E. Broussolle, C. Brefel-Courbon, B. S. Harhangi, B. A. Oostra, E. Fabrizio, G. A. Bohme, L. Pradier, N. W. Wood, A. Filla, G. Meo, P. Deneffe, Y. Agid and A. Brice. 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*, v.8, n.4, Apr, p.567-74. 1999.

Abou-Sleiman, P. M., D. G. Healy, N. Quinn, A. J. Lees and N. W. Wood. The role of pathogenic DJ-1 mutations in Parkinson's disease. *Ann Neurol*, v.54, n.3, Sep, p.283-6. 2003.

Agil, A., R. Duran, F. Barrero, B. Morales, M. Arauzo, F. Alba, M. T. Miranda, I. Prieto, M. Ramirez and F. Vives. Plasma lipid peroxidation in sporadic Parkinson's disease. Role of the L-dopa. *J Neurol Sci*, v.240, n.1-2, Jan 15, p.31-6. 2006.

Anderson, J. P., D. E. Walker, J. M. Goldstein, R. De Laat, K. Banducci, R. J. Caccavello, R. Barbour, J. Huang, K. Kling, M. Lee, L. Diep, P. S. Keim, X. Shen, T. Chataway, M. G. Schlossmacher, P. Seubert, D. Schenk, S. Sinha, W. P. Gai and T. J. Chilcote. Phosphorylation of Ser-129 is the dominant pathological modification of alpha-synuclein in familial and sporadic Lewy body disease. *J Biol Chem*, v.281, n.40, Oct 6, p.29739-52. 2006.

Arai, H., T. Muramatsu, S. Higuchi, H. Sasaki and J. Q. Trojanowski. Apolipoprotein E gene in Parkinson's disease with or without dementia. *Lancet*, v.344, n.8926, Sep 24, p.889. 1994.

Artazcoz Sanz, M. T. and J. J. Vines Rueda. [The estimation of prevalence of Parkinson disease in Navarra. An epidemiological study of the consumption of anti-parkinsonian drugs]. *Rev Esp Salud Publica*, v.69, n.6, Nov-Dec, p.479-85. 1995.

Auburger, G., G. O. Diaz, R. F. Capote, S. G. Sanchez, M. P. Perez, M. E. Del Cueto, M. G. Meneses, M. Farrall, R. Williamson, S. Chamberlain and Et Al. Autosomal dominant ataxia: genetic evidence for locus heterogeneity from a Cuban founder-effect population. *Am J Hum Genet*, v.46, n.6, Jun, p.1163-77. 1990.

Babinski, J. De l'asynergie cerebelleuse. *Rev neurol*, v.7, p.806-816. 1899.

Baker, M., I. Litvan, H. Houlden, J. Adamson, D. Dickson, J. Perez-Tur, J. Hardy, T. Lynch, E. Bigio and M. Hutton. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet*, v.8, n.4, Apr, p.711-5. 1999.

Baldereschi, M., A. Di Carlo, W. A. Rocca, P. Vanni, S. Maggi, E. Perissinotto, F. Grigoletto, L. Amaducci and D. Inzitari. Parkinson's disease and parkinsonism in a longitudinal study: two-fold higher incidence in men. ILSA Working Group. Italian Longitudinal Study on Aging. *Neurology*, v.55, n.9, Nov 14, p.1358-63. 2000.

Bandopadhyay, R., A. E. Kingsbury, M. R. Cookson, A. R. Reid, I. M. Evans, A. D. Hope, A. M. Pittman, T. Lashley, R. Canet-Aviles, D. W. Miller, C. Mclendon, C. Strand, A. J. Leonard, P. M. Abou-Sleiman, D. G. Healy, H. Ariga, N. W. Wood, R. De Silva, T. Revesz, J. A. Hardy and A. J. Lees. The expression of DJ-1 (PARK7) in normal human CNS and idiopathic Parkinson's disease. *Brain*, v.127, n.Pt 2, Feb, p.420-30. 2004.

Barrett, J. C., B. Fry, J. Maller and M. J. Daly. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, v.21, n.2, Jan 15, p.263-5. 2005.

Benito-Leon, J., F. Bermejo-Pareja, J. M. Morales-Gonzalez, J. Porta-Etessam, R. Trincado, S. Vega and E. D. Louis. Incidence of Parkinson disease and parkinsonism in three elderly populations of central Spain. *Neurology*, v.62, n.5, Mar 9, p.734-41. 2004.

Benito-Leon, J., F. Bermejo-Pareja, J. Rodriguez, J. A. Molina, R. Gabriel and J. M. Morales. Prevalence of PD and other types of parkinsonism in three elderly populations of central Spain. *Mov Disord*, v.18, n.3, Mar, p.267-74. 2003.

Bennett, P., V. Bonifati, U. Bonuccelli, C. Colosimo, M. De Mari, G. Fabbrini, R. Marconi, G. Meco, D. J. Nicholl, F. Stocchi, N. Vanacore, P. Vieregge and A. C. Williams. Direct genetic evidence for involvement of tau in progressive supranuclear palsy. European Study Group on Atypical Parkinsonism Consortium. *Neurology*, v.51, n.4, Oct, p.982-5. 1998.

Berendse, H. W., J. Booij, C. M. Francot, P. L. Bergmans, R. Hijman, J. C. Stoof and E. C. Wolters. Subclinical dopaminergic dysfunction in asymptomatic Parkinson's disease patients' relatives with a decreased sense of smell. *Ann Neurol*, v.50, n.1, Jul, p.34-41. 2001.

Berg, D., K. Schweitzer, P. Leitner, A. Zimprich, P. Lichtner, P. Belcredi, T. Brussel, C. Schulte, S. Maass and T. Nagele. Type and frequency of mutations in the LRRK2 gene in familial and sporadic Parkinson's disease\*. *Brain*, v.128, n.Pt 12, Dec, p.3000-11. 2005.

Bergareche, A., E. De La Puente, A. Lopez De Munain, C. Sarasqueta, A. De Arce, J. J. Poza and J. F. Marti-Masso. Prevalence of Parkinson's disease and other types of Parkinsonism. A door-to-door survey in Bidasoa, Spain. *J Neurol*, v.251, n.3, Mar, p.340-5. 2004.

Bogaerts, V., K. Nuytemans, J. Reumers, P. Pals, S. Engelborghs, B. Pickut, E. Corsmit, K. Peeters, J. Schymkowitz, P. P. De Deyn, P. Cras, F. Rousseau, J. Theuns and C. Van Broeckhoven. Genetic variability in the mitochondrial serine protease HTRA2 contributes to risk for Parkinson disease. *Hum Mutat*, Apr 9. 2008.

Bon, M. A., E. N. Jansen Steur, R. A. De Vos and I. Vermes. Neurogenetic correlates of Parkinson's disease: apolipoprotein-E and cytochrome P450 2D6 genetic polymorphism. *Neurosci Lett*, v.266, n.2, May 7, p.149-51. 1999.

Bonifati, V., G. J. Breedveld, F. Squitieri, N. Vanacore, P. Brustenghi, B. S. Harhangi, P. Montagna, M. Cannella, G. Fabbrini, P. Rizzu, C. M. Van Duijn, B. A. Oostra, G. Meco and P. Heutink. Localization of autosomal recessive early-onset parkinsonism to



Braak, H., E. Braak, D. Yilmazer, R. A. De Vos, E. N. Jansen, J. BohlandK. Jellinger. Amygdala pathology in Parkinson's disease. *Acta Neuropathol*, v.88, n.6, p.493-500. 1994a.

Braak, H., E. Braak, D. Yilmazer, R. A. De Vos, E. N. Jansen, J. BohlandK. Jellinger. Amygdala pathology in Parkinson's disease. *Acta Neuropathol (Berl)*, v.88, n.6, p.493-500. 1994b.

Braak, H., K. Del Tredici, U. Rub, R. A. De Vos, E. N. Jansen SteurandE. Braak. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging*, v.24, n.2, Mar-Apr, p.197-211. 2003.

Braak, H., E. Ghebremedhin, U. Rub, H. BratzkeandK. Del Tredici. Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res*, v.318, n.1, Oct, p.121-34. 2004.

Bras, J. M., R. J. Guerreiro, M. H. Ribeiro, C. Januario, A. Morgadinho, C. R. Oliveira, L. Cunha, J. HardyandA. Singleton. G2019S dardarin substitution is a common cause of Parkinson's disease in a Portuguese cohort. *Mov Disord*, v.20, n.12, Dec, p.1653-5. 2005.

Brashear, A., W. B. Dobyns, P. De Carvalho Aguiar, M. Borg, C. J. Frijns, S. Gollamudi, A. Green, J. Guimaraes, B. C. Haake, C. Klein, G. Linazasoro, A. Munchau, D. Raymond, D. Riley, R. Saunders-Pullman, M. A. Tijssen, D. Webb, J. Zaremba, S. B. BressmanandL. J. Ozelius. The phenotypic spectrum of rapid-onset dystonia-parkinsonism (RDP) and mutations in the ATP1A3 gene. *Brain*, v.130, n.Pt 3, Mar, p.828-35. 2007.

Broman, K. W.andJ. L. Weber. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet*, v.65, n.6, Dec, p.1493-500. 1999.

Brown, S. On hereditary ataxia, with a series of twenty-one cases. *Brain*, v.15, p.250-282. 1892.

Burke, R. E., S. Fahn, C. D. Marsden, S. B. Bressman, C. Moskowitz and J. Friedman. Validity and reliability of a rating scale for the primary torsion dystonias. *Neurology*, v.35, n.1, Jan, p.73-7. 1985.

Burn, D. J., M. H. Mark, E. D. Playford, D. M. Maraganore, T. R. Zimmerman, Jr., R. C. Duvoisin, A. E. Harding, C. D. Marsden and D. J. Brooks. Parkinson's disease in twins studied with 18F-dopa and positron emission tomography. *Neurology*, v.42, n.10, Oct, p.1894-900. 1992.

Burns, R. S., C. C. Chiueh, S. P. Markey, M. H. Ebert, D. M. Jacobowitz and I. J. Kopin. A primate model of parkinsonism: selective destruction of dopaminergic neurons in the pars compacta of the substantia nigra by N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. *Proc Natl Acad Sci U S A*, v.80, n.14, Jul, p.4546-50. 1983.

Calafell, F. and J. Bertranpetit. Principal component analysis of gene frequencies and the origin of Basques. *Am J Phys Anthropol*, v.93, n.2, Feb, p.201-15. 1994.

Camicioli, R., M. M. Moore, A. Kinney, E. Corbridge, K. Glassberg and J. A. Kaye. Parkinson's disease is associated with hippocampal atrophy. *Mov Disord*, v.18, n.7, Jul, p.784-90. 2003.

Carlson, C. S., M. A. Eberle, M. J. Rieder, Q. Yi, L. Kruglyak and D. A. Nickerson. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*, v.74, n.1, Jan, p.106-20. 2004.

Carmine, A., S. Buervenich, D. Galter, E. G. Jonsson, G. C. Sedvall, L. Farde, J. P. Gustavsson, H. Bergman, K. V. Chowdari, V. L. Nimgaonkar, M. Anvret, O. Sydow and L. Olson. NURR1 promoter polymorphisms: Parkinson's disease, schizophrenia, and personality traits. *Am J Med Genet B Neuropsychiatr Genet*, v.120, n.1, Jul 1, p.51-7. 2003.

Carmine Belin, A., M. Westerlund, O. Bergman, H. Nissbrandt, C. Lind, O. Sydow and D. Galter. S18Y in ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) associated with decreased risk of Parkinson's disease in Sweden. *Parkinsonism Relat Disord*, Feb 5. 2007.

Chartier-Harlin, M. C., J. Kachergus, C. Roumier, V. Mouroux, X. Douay, S. Lincoln, C. Levecque, L. Larvor, J. Andrieux, M. Hulihan, N. Waucquier, L. Defebvre, P. Amouyel, M. Farrer and A. Destee. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet*, v.364, n.9440, Sep 25-Oct 1, p.1167-9. 2004.

Chen, R. C., S. F. Chang, C. L. Su, T. H. Chen, M. F. Yen, H. M. Wu, Z. Y. Chen and H. H. Liou. Prevalence, incidence, and mortality of PD: a door-to-door survey in Ilan county, Taiwan. *Neurology*, v.57, n.9, Nov 13, p.1679-86. 2001.

Chiba-Falek, O., J. A. Kowalak, M. E. Smulson and R. L. Nussbaum. Regulation of alpha-synuclein expression by poly (ADP ribose) polymerase-1 (PARP-1) binding to the NACP-Rep1 polymorphic site upstream of the SNCA gene. *Am J Hum Genet*, v.76, n.3, Mar, p.478-92. 2005.

Chiba-Falek, O. and R. L. Nussbaum. Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. *Hum Mol Genet*, v.10, n.26, Dec 15, p.3101-9. 2001.

Chiba-Falek, O., J. W. Touchman and R. L. Nussbaum. Functional analysis of intra-allelic variation at NACP-Rep1 in the alpha-synuclein gene. *Hum Genet*, v.113, n.5, Oct, p.426-31. 2003.

Choi, P., H. Snyder, L. Petrucelli, C. Theisler, M. Chong, Y. Zhang, K. Lim, K. K. Chung, K. Kehoe, L. D'adamio, J. M. Lee, E. Cochran, R. Bowser, T. M. Dawson and B. Wolozin. SEPT5\_v2 is a parkin-binding protein. *Brain Res Mol Brain Res*, v.117, n.2, Oct 7, p.179-89. 2003.

Chung, K. K., Y. Zhang, K. L. Lim, Y. Tanaka, H. Huang, J. Gao, C. A. Ross, V. L. Dawson and T. M. Dawson. Parkin ubiquitinates the alpha-synuclein-interacting protein, synphilin-1: implications for Lewy-body formation in Parkinson disease. *Nat Med*, v.7, n.10, Oct, p.1144-50. 2001.

Ciechanover, A. and P. Brundin. The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron*, v.40, n.2, Oct 9, p.427-46. 2003.

Cirulli, E. T. and D. B. Goldstein. In vitro assays fail to predict in vivo effects of regulatory polymorphisms. *Hum Mol Genet*, v.16, n.16, Aug 15, p.1931-9. 2007.

Clarimon, J., S. Scholz, H. C. Fung, J. Hardy, J. Eerola, O. Hellstrom, C. M. Chen, Y. R. Wu, P. J. Tienari and A. Singleton. Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. *Am J Hum Genet*, v.78, n.6, Jun, p.1082-4; author reply 1092-4. 2006.

Clark, L. N., S. Afridi, H. Mejia-Santana, J. Harris, E. D. Louis, L. J. Cote, H. Andrews, A. Singleton, F. Wavrant De-Vrieze, J. Hardy, R. Mayeux, S. Fahn, C. Waters, B. Ford, S. Frucht, R. Ottman and K. Marder. Analysis of an early-onset Parkinson's disease cohort for DJ-1 mutations. *Mov Disord*, v.19, n.7, Jul, p.796-800. 2004.

Clark, L. N., P. Poorkaj, Z. Wszolek, D. H. Geschwind, Z. S. Nasreddine, B. Miller, D. Li, H. Payami, F. Awert, K. Markopoulou, A. Andreadis, I. D'souza, V. M. Lee, L. Reed, J. Q. Trojanowski, V. Zhukareva, T. Bird, G. Schellenberg and K. C. Wilhelmsen. Pathogenic implications of mutations in the tau gene in pallido-ponto-nigral degeneration and related neurodegenerative disorders linked to chromosome 17. *Proc Natl Acad Sci U S A*, v.95, n.22, Oct 27, p.13103-7. 1998.

Claveria, L. E., J. Duarte, M. D. Sevillano, A. Perez-Sempere, C. Cabezas, F. Rodriguez and J. De Pedro-Cuesta. Prevalence of Parkinson's disease in Cantalejo, Spain: a door-to-door survey. *Mov Disord*, v.17, n.2, Mar, p.242-9. 2002.

Cobo, A. M., A. Saenz, J. J. Poza, M. Urtasun, B. Indakoetxea, J. A. Urtizbera, A. Lopez De Munain and F. Calafell. A common haplotype associated with the Basque 2362AG --> TCATCT mutation in the muscular calpain-3 gene. *Hum Biol*, v.76, n.5, Oct, p.731-41. 2004.

Conrad, C., A. Andreadis, J. Q. Trojanowski, D. W. Dickson, D. Kang, X. Chen, W. Wiederholt, L. Hansen, E. Masliah, L. J. Thal, R. Katzman, Y. Xia and T. Saitoh. Genetic evidence for the involvement of tau in progressive supranuclear palsy. *Ann Neurol*, v.41, n.2, Feb, p.277-81. 1997.

Conrad, C., C. Vianna, M. Freeman and P. Davies. A polymorphic gene nested within an intron of the tau gene: implications for Alzheimer's disease. *Proc Natl Acad Sci U S A*, v.99, n.11, May 28, p.7751-6. 2002.

Conrad, D. F., T. D. Andrews, N. P. Carter, M. E. Hurles and J. K. Pritchard. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet*, v.38, n.1, Jan, p.75-81. 2006.

Conway, K. A., J. D. Harper and P. T. Lansbury. Accelerated in vitro fibril formation by a mutant alpha-synuclein linked to early-onset Parkinson disease. *Nat Med*, v.4, n.11, Nov, p.1318-20. 1998.

Conway, K. A., S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson and P. T. Lansbury, Jr. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc Natl Acad Sci U S A*, v.97, n.2, Jan 18, p.571-6. 2000.

Conway, K. A., J. C. Rochet, R. M. Bieganski and P. T. Lansbury, Jr. Kinetic stabilization of the alpha-synuclein protofibril by a dopamine-alpha-synuclein adduct. *Science*, v.294, n.5545, Nov 9, p.1346-9. 2001.

Corder, E. H., A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, G. W. Small, A. D. Roses, J. L. Haines and M. A. Pericak-Vance. Gene dose of apolipoprotein E

type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, v.261, n.5123, Aug 13, p.921-3. 1993.

Cormand, B., L. Vilageliu, J. M. Burguera, S. Balcells, R. Gonzalez-Duarte, D. Grinberg and A. Chabas. Gaucher disease in Spanish patients: analysis of eight mutations. *Hum Mutat*, v.5, n.4, p.303-9. 1995.

Corporation, A. B. Prism Sequence Detection System: relative quantitation of gene expression. User Bulletin 2. Foster City, CA. Applied Biosystems Corporation., p.11-16. 1997.

Corti, O., C. Hampe, H. Koutnikova, F. Darios, S. Jacquier, A. Prigent, J. C. Robinson, L. Pradier, M. Ruberg, M. Mirande, E. Hirsch, T. Rooney, A. Fournier and A. Brice. The p38 subunit of the aminoacyl-tRNA synthetase complex is a Parkin substrate: linking protein biosynthesis and neurodegeneration. *Hum Mol Genet*, v.12, n.12, Jun 15, p.1427-37. 2003.

Criado-Alvarez, J. J., C. Romo-Barrientos, J. Martinez-Hernandez and I. Gonzalez-Solana. [Use of antiparkinsonian agents in Castilla-La Mancha. Estimate of prevalence of Parkinson disease]. *Rev Neurol*, v.27, n.157, Sep, p.405-8. 1998.

Cuervo, A. M., L. Stefanis, R. Fredenburg, P. T. Lansbury and D. Sulzer. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. *Science*, v.305, n.5688, Aug 27, p.1292-5. 2004.

D'acquistio, F. and S. Ghosh. PACT and PKR: turning on NF-kappa B in the absence of virus. *Sci STKE*, v.2001, n.89, Jul 3, p.RE1. 2001.

Dale, G. E., A. Probst, P. Luthert, J. Martin, B. H. Anderton and P. N. Leigh. Relationships between Lewy bodies and pale bodies in Parkinson's disease. *Acta Neuropathol*, v.83, n.5, p.525-9. 1992.

Davis, G. C., A. C. Williams, S. P. Markey, M. H. Ebert, E. D. Caine, C. M. Reichert and I. J. Kopin. Chronic Parkinsonism secondary to intravenous injection of meperidine analogues. *Psychiatry Res*, v.1, n.3, Dec, p.249-54. 1979.

De Carvalho Aguiar, P., K. J. Sweadner, J. T. Penniston, J. Zaremba, L. Liu, M. Caton, G. Linazasoro, M. Borg, M. A. Tijssen, S. B. Bressman, W. B. Dobyns, A. Brashear and L. J. Ozelius. Mutations in the Na<sup>+</sup>/K<sup>+</sup> -ATPase alpha3 gene ATP1A3 are associated with rapid-onset dystonia parkinsonism. *Neuron*, v.43, n.2, Jul 22, p.169-75. 2004.

De Carvalho Aguiar, P. M. and L. J. Ozelius. Classification and genetics of dystonia. *Lancet Neurol*, v.1, n.5, Sep, p.316-25. 2002.

De La Fuente-Fernandez, R., M. A. Nunez and E. Lopez. The apolipoprotein E epsilon 4 allele increases the risk of drug-induced hallucinations in Parkinson's disease. *Clin Neuropharmacol*, v.22, n.4, Jul-Aug, p.226-30. 1999.

De La Fuente-Fernandez, R., A. Sellers, K. Beyer and J. I. Lao. Apolipoprotein E genotypes and age at onset of Parkinson's disease. *Ann Neurol*, v.44, n.2, Aug, p.294-5. 1998.

De Lau, L. M. and M. M. Breteler. Epidemiology of Parkinson's disease. *Lancet Neurol*, v.5, n.6, Jun, p.525-35. 2006.

De Lau, L. M., P. C. Giesbergen, M. C. De Rijk, A. Hofman, P. J. Koudstaal and M. M. Breteler. Incidence of parkinsonism and Parkinson disease in a general population: the Rotterdam Study. *Neurology*, v.63, n.7, Oct 12, p.1240-4. 2004.

De Michele, G., A. Filla, R. Marconi, G. Volpe, A. D'alessio, R. Scala, G. Ambrosio and G. Campanella. A genetic study of Parkinson's disease. *J Neural Transm Suppl*, v.45, p.21-5. 1995.

De Pancorbo, M. M., M. Lopez-Martinez, C. Martinez-Bouzas, A. Castro, I. Fernandez-Fernandez, G. A. De Mayolo, A. A. De Mayolo, P. A. De Mayolo, D. J. Rowold and R. J. Herrera. The Basques according to polymorphic Alu insertions. *Hum Genet*, v.109, n.2, Aug, p.224-33. 2001.

De Rijk, M. C., L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martinez-Lage, C. Trenkwalder and A. Hofman. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*, v.54, n.11 Suppl 5, p.S21-3. 2000.

Den, H. J. W. and J. Bethlem. The distribution of Lewy bodies in the central and autonomic nervous systems in idiopathic paralysis agitans. *J Neurol Neurosurg Psychiatry*, v.23, Nov, p.283-90. 1960.

Deng, H., W. Le, Y. Guo, C. B. Hunter, W. Xie, M. Huang and J. Jankovic. Genetic analysis of LRRK2 mutations in patients with Parkinson disease. *J Neurol Sci*, v.251, n.1-2, Dec 21, p.102-6. 2006.

Destefano, A. L., M. F. Lew, L. I. Golbe, M. H. Mark, A. M. Lazzarini, M. Guttman, E. Montgomery, C. H. Waters, C. Singer, R. L. Watts, L. J. Currie, G. F. Wooten, N. E. Maher, J. B. Wilk, K. M. Sullivan, K. M. Slater, M. H. Saint-Hilaire, R. G. Feldman, O. Suchowersky, A. L. Lafontaine, N. Labelle, J. H. Growdon, P. Vieregge, P. P. Pramstaller, C. Klein, J. P. Hubble, C. R. Reider, M. Stacy, M. E. Macdonald, J. F. Gusella and R. H. Myers. PARK3 influences age at onset in Parkinson disease: a genome scan in the GenePD study. *Am J Hum Genet*, v.70, n.5, May, p.1089-95. 2002.

Di Fonzo, A., C. F. Rohe, J. Ferreira, H. F. Chien, L. Vacca, F. Stocchi, L. Guedes, E. Fabrizio, M. Manfredi, N. Vanacore, S. Goldwurm, G. Breedveld, C. Sampaio, G. Meco, E. Barbosa, B. A. Oostra and V. Bonifati. A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. *Lancet*, v.365, n.9457, Jan 29-Feb 4, p.412-5. 2005.

Duffy, T. Phase and electron microscopic observations of Lewy bodies and melanin

granules in the substantia nigra and locus caeruleus in Parkinson's disease. *J Neuropathol Exp Neurol*, v.24, p.398-414. 1965.

Dupont, W. D. and W. D. Plummer, Jr. Power and sample size calculations. A review and computer program. *Control Clin Trials*, v.11, n.2, Apr, p.116-28. 1990.

Duvoisin, R. C., R. Eldridge, A. Williams, J. Nutt and D. Calne. Twin study of Parkinson disease. *Neurology*, v.31, n.1, Jan, p.77-80. 1981.

Duvoisin, R. C. and W. G. Johnson. Hereditary Lewy-body parkinsonism and evidence for a genetic etiology of Parkinson's disease. *Brain Pathol*, v.2, n.4, Oct, p.309-20. 1992.

Elbaz, A., C. Levecque, J. Clavel, J. S. Vidal, F. Richard, J. R. Correze, B. Delemotte, P. Amouyel, A. Alperovitch, M. C. Chartier-Harlin and C. Tzourio. S18Y polymorphism in the UCH-L1 gene and Parkinson's disease: evidence for an age-dependent relationship. *Mov Disord*, v.18, n.2, Feb, p.130-7. 2003.

Elbaz, A., L. M. Nelson, H. Payami, J. P. Ioannidis, B. K. Fiske, G. Annesi, A. Carmine Belin, S. A. Factor, C. Ferrarese, G. M. Hadjigeorgiou, D. S. Higgins, H. Kawakami, R. Kruger, K. S. Marder, R. P. Mayeux, G. D. Mellick, J. G. Nutt, B. Ritz, A. Samii, C. M. Tanner, C. Van Broeckhoven, S. K. Van Den Eeden, K. Wirdefeldt, C. P. Zabetian, M. Dehem, J. S. Montimurro, A. Southwick, R. M. Myers and T. A. Trikalinos. Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study. *Lancet Neurol*, v.5, n.11, Nov, p.917-23. 2006.

Eldridge, R. and S. E. Ince. The low concordance rate for Parkinson's disease in twins: a possible explanation. *Neurology*, v.34, n.10, Oct, p.1354-6. 1984.

Elnenaï, M. O., R. A. Hamoudi, J. Swansbury, A. M. Gruszka-Westwood, V. Brito-Babapulle, E. Matutes and D. Catovsky. Delineation of the minimal region of loss at 13q14 in multiple myeloma. *Genes Chromosomes Cancer*, v.36, n.1, Jan, p.99-106. 2003.

Errea, J. M., J. R. Ara, C. Aibar and J. De Pedro-Cuesta. Prevalence of Parkinson's disease in lower Aragon, Spain. *Mov Disord*, v.14, n.4, Jul, p.596-604. 1999.

Facheris, M., K. J. Strain, T. G. Lesnick, M. De Andrade, J. H. Bower, J. E. Ahlskog, J. M. Cunningham, S. Lincoln, M. J. Farrer, W. A. Rocca and D. M. Maraganore. UCHL1 is associated with Parkinson's disease: a case-unaffected sibling and case-unrelated control study. *Neurosci Lett*, v.381, n.1-2, Jun 10-17, p.131-4. 2005.

Fahn, S., R. L. Elton and The Updrs Development Committee. Unified Parkinson disease rating scale. *Recent developments in Parkinson's disease*, v.2, p.293-304. 1987.

Fahn, S. M., C. X-linked recessive dystonia and parkinsonism in Filipino males. *Ann. Neurol*, v.24, n.179. 1988.

Farrer, M., K. Gwinn-Hardy, M. Muentner, F. W. Devrieze, R. Crook, J. Perez-Tur, S. Lincoln, D. Maraganore, C. Adler, S. Newman, K. Macelwee, P. Mccarthy, C. Miller, C. Waters and J. Hardy. A chromosome 4p haplotype segregating with Parkinson's disease and postural tremor. *Hum Mol Genet*, v.8, n.1, Jan, p.81-5. 1999.

Farrer, M., J. Kachergus, L. Forno, S. Lincoln, D. S. Wang, M. Hulihan, D. Maraganore, K. Gwinn-Hardy, Z. Wszolek, D. Dickson and J. W. Langston. Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. *Ann Neurol*, v.55, n.2, Feb, p.174-9. 2004.

Farrer, M., D. M. Maraganore, P. Lockhart, A. Singleton, T. G. Lesnick, M. De Andrade, A. West, R. De Silva, J. Hardy and D. Hernandez. alpha-Synuclein gene haplotypes are associated with Parkinson's disease. *Hum Mol Genet*, v.10, n.17, Aug 15, p.1847-51. 2001.

Farrer, M. J., K. Haugarvoll, O. A. Ross, J. T. Stone, N. M. Milkovic, S. A. Cobb, A. J. Whittle, S. J. Lincoln, M. M. Hulihan, M. G. Heckman, L. R. White, J. O. Aasly, J. M. Gibson, D. Gosal, T. Lynch, Z. K. Wszolek, R. J. Uitti and M. Toft. Genomewide

association, Parkinson disease, and PARK10. *Am J Hum Genet*, v.78, n.6, Jun, p.1084-8; author reply 1092-4. 2006.

Feldman, B., J. Chapman and A. D. Korczyn. Apolipoprotein epsilon4 advances appearance of psychosis in patients with Parkinson's disease. *Acta Neurol Scand*, v.113, n.1, Jan, p.14-7. 2006.

Ferreira, J. J., L. C. Guedes, M. M. Rosa, M. Coelho, M. Van Doeselaar, D. Schweiger, A. Di Fonzo, B. A. Oostra, C. Sampaio and V. Bonifati. High prevalence of LRRK2 mutations in familial and sporadic Parkinson's disease in Portugal. *Mov Disord*, v.22, n.8, Jun 15, p.1194-201. 2007a.

Ferreira, J. J., L. C. Guedes, M. M. Rosa, M. Coelho, M. Van Doeselaar, D. Schweiger, A. Di Fonzo, B. A. Oostra, C. Sampaio and V. Bonifati. High prevalence of LRRK2 mutations in familial and sporadic Parkinson's disease in Portugal. *Mov Disord*, v.22, n.8, Apr 27, p.1194-1201. 2007b.

Foroud, T., S. K. Uniacke, L. Liu, N. Pankratz, A. Rudolph, C. Halter, C. Shults, K. Marder, P. M. Conneally and W. C. Nichols. Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. *Neurology*, v.60, n.5, Mar 11, p.796-801. 2003.

Frazer, K. A., D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, W. Sun, H. Wang, Y. Wang, X. Xiong, L. Xu, M. M. Waye, S. K. Tsui, H. Xue, J. T. Wong, L. M. Galver, J. B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J. F. Olivier, M. S. Phillips, S. Roumy, C. Sallee, A. Verner, T. J. Hudson, P. Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L. C. Tsui, W. Mak, Y. Q. Song, P. K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C. P. Bird, M. Delgado, E. T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B. E. Stranger, P. Whittaker, D. R. Bentley, M. J. Daly, P.

I. De Bakker, J. Barrett, Y. R. Chretien, J. Maller, S. Mccarroll, N. Patterson, I. Pe'er, A. Price, S. Purcell, D. J. Richter, P. Sabeti, R. Saxena, S. F. Schaffner, P. C. Sham, P. Varilly, L. D. Stein, L. Krishnan, A. V. Smith, M. K. Tello-Ruiz, G. A. Thorisson, A. Chakravarti, P. E. Chen, D. J. Cutler, C. S. Kashuk, S. Lin, G. R. Abecasis, W. Guan, Y. Li, H. M. Munro, Z. S. Qin, D. J. Thomas, G. Mcvean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L. R. Cardon, G. Clarke, D. M. Evans, A. P. Morris, B. S. Weir, J. C. Mullikin, S. T. Sherry, M. Feolo, A. Skol, H. Zhang, I. Matsuda, Y. Fukushima, D. R. Macer, E. Suda, C. N. Rotimi, C. A. Adebamowo, I. Ajayi, T. Aniagwu, P. A. Marshall, C. Nkwodimmah, C. D. Royal, M. F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I. F. Adewole, B. M. Knoppers, M. W. Foster, E. W. Clayton, J. Watkin, D. Muzny, L. Nazareth, E. Sodergren, G. M. Weinstock, I. Yakub, B. W. Birren, R. K. Wilson, L. L. Fulton, J. Rogers, J. Burton, N. P. Carter, C. M. Clee, M. Griffiths, M. C. Jones, K. Mclay, R. W. Plumb, M. T. Ross, S. K. Sims, D. L. Willey, Z. Chen, H. Han, L. Kang, M. Godbout, J. C. Wallenburg, P. L'archeveque, G. Bellemare, K. Saeki, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A. L. Holden, L. D. Brooks, J. E. Mcewen, M. S. Guyer, V. O. Wang, J. L. Peterson, M. Shi, J. Spiegel, L. M. Sung, L. F. Zacharia, F. S. Collins, K. Kennedy, R. Jamieson and J. Stewart. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, v.449, n.7164, Oct 18, p.851-61. 2007.

Fujiwara, H., M. Hasegawa, N. Dohmae, A. Kawashima, E. Masliah, M. S. Goldberg, J. Shen, K. Takio and T. Iwatsubo. alpha-Synuclein is phosphorylated in synucleinopathy lesions. *Nat Cell Biol*, v.4, n.2, Feb, p.160-4. 2002.

Funalot, B., W. C. Nichols, J. Perez-Tur, G. Mercier and G. Lucotte. Genetic screening for two LRRK2 mutations in French patients with idiopathic Parkinson's disease. *Genet Test*, v.10, n.4, Winter, p.290-3. 2006.

Funayama, M., K. Hasegawa, H. Kowa, M. Saito, S. Tsuji and F. Obata. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Ann Neurol*, v.51, n.3, Mar, p.296-301. 2002.

Furtado, S., M. Farrer, Y. Tsuboi, M. L. Klimek, R. De La Fuente-Fernandez, J. Hussey, P. Lockhart, D. B. Calne, O. Suchowersky, A. J. Stoessland and Z. K. Wszolek. SCA-2 presenting as parkinsonism in an Alberta family: clinical, genetic, and PET findings. *Neurology*, v.59, n.10, Nov 26, p.1625-7. 2002.

Gaig, C., M. Ezquerra, M. J. Martí, E. Muñoz, F. Valldeoriola and E. Tolosa. LRRK2 mutations in Spanish patients with Parkinson disease: frequency, clinical features, and incomplete penetrance. *Arch Neurol*, v.63, n.3, Mar, p.377-82. 2006.

Galter, D., M. Westerlund, A. Carmine, E. Lindqvist, O. Sydow and L. Olson. LRRK2 expression linked to dopamine-innervated areas. *Ann Neurol*, v.59, n.4, Apr, p.714-9. 2006.

Gandhi, S., M. M. Muqit, L. Stanyer, D. G. Healy, P. M. Abou-Sleiman, I. Hargreaves, S. Heales, M. Ganguly, L. Parsons, A. J. Lees, D. S. Latchman, J. L. Holton, N. W. Wood and T. Revesz. PINK1 protein in normal human brain and Parkinson's disease. *Brain*, v.129, n.Pt 7, Jul, p.1720-31. 2006.

Gasser, T., B. Müller-Muhsok, Z. K. Wszolek, R. Oehlmann, D. B. Calne, V. Bonifati, B. Bereznoi, E. Fabrizio, P. Vieregge and R. D. Horstmann. A susceptibility locus for Parkinson's disease maps to chromosome 2p13. *Nat Genet*, v.18, n.3, Mar, p.262-5. 1998.

Gelb, D. J., E. Oliver and S. Gilman. Diagnostic criteria for Parkinson disease. *Arch Neurol*, v.56, n.1, Jan, p.33-9. 1999.

Giasson, B. I., J. E. Duda, I. V. Murray, Q. Chen, J. M. Souza, H. I. Hurtig, H. Ischiropoulos, J. Q. Trojanowski and V. M. Lee. Oxidative damage linked to neurodegeneration by selective alpha-synuclein nitration in synucleinopathy lesions. *Science*, v.290, n.5493, Nov 3, p.985-9. 2000.

Giasson, B. I. and V. M. Lee. Parkin and the molecular pathways of Parkinson's disease. *Neuron*, v.31, n.6, Sep 27, p.885-8. 2001.

Gibb, W. R., T. Scott and A. J. Lees. Neuronal inclusions of Parkinson's disease. *Mov Disord*, v.6, n.1, p.2-11. 1991.

Gibson, J., N. E. Morton and A. Collins. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*, v.15, n.5, Mar 1, p.789-95. 2006.

Gilks, W. P., P. M. Abou-Sleiman, S. Gandhi, S. Jain, A. Singleton, A. J. Lees, K. Shaw, K. P. Bhatia, V. Bonifati, N. P. Quinn, J. Lynch, D. G. Healy, J. L. Holton, T. Revesz and N. W. Wood. A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet*, v.365, n.9457, Jan 29-Feb 4, p.415-6. 2005.

Gispert, S., R. Twells, G. Orozco, A. Brice, J. Weber, L. Heredero, K. Scheufler, B. Riley, R. Allotey, C. Nothens and Et Al. Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23-24.1. *Nat Genet*, v.4, n.3, Jul, p.295-9. 1993.

Goetz, C. G., P. F. Burke, S. Leurgans, E. Berry-Kravis, L. M. Blasucci, R. Raman and L. Zhou. Genetic variation analysis in parkinson disease patients with and without hallucinations: case-control study. *Arch Neurol*, v.58, n.2, Feb, p.209-13. 2001.

Golbe, L. I., G. Di Iorio, V. Bonavita, D. C. Miller and R. C. Duvoisin. A large kindred with autosomal dominant Parkinson's disease. *Ann Neurol*, v.27, n.3, Mar, p.276-82. 1990.

Golbe, L. I., G. Di Iorio, G. Sanges, A. M. Lazzarini, S. La Sala, V. Bonavita and R. C. Duvoisin. Clinical genetic analysis of Parkinson's disease in the Contursi kindred. *Ann Neurol*, v.40, n.5, Nov, p.767-75. 1996.

Golbe, L. I., A. M. Lazzarini, J. R. Szychala, W. G. Johnson, E. S. Stenroos, M. H. Mark and J. I. Sage. The tau A0 allele in Parkinson's disease. *Mov Disord*, v.16, n.3, May, p.442-7. 2001.

Goldwurm, S., A. Di Fonzo, E. J. Simons, C. F. Rohe, M. Zini, M. Canesi, S. Tesei, A. Zecchinelli, A. Antonini, C. Mariani, N. Meucci, G. Sacilotto, F. Sironi, G. Salani, J. Ferreira, H. F. Chien, E. Fabrizio, N. Vanacore, A. Dalla Libera, F. Stocchi, C. Diroma, P. Lamberti, C. Sampaio, G. Meco, E. Barbosa, A. M. Bertoli-Avella, G. J. Breedveld, B.

A. Oostra, G. Pezzoli and V. Bonifati. The G6055A (G2019S) mutation in LRRK2 is frequent in both early and late onset Parkinson's disease and originates from a common ancestor. *J Med Genet*, v.42, n.11, Nov, p.e65. 2005.

Gonzalez-Fernandez, M. C., E. Lezcano, O. A. Ross, J. C. Gomez-Esteban, F. Gomez-Busto, F. Velasco, M. Alvarez-Alvarez, M. B. Rodriguez-Martinez, R. Ciordia, J. J. Zarranz, M. J. Farrer, I. F. Mata and M. M. De Pancorbo. Lrrk2-associated parkinsonism is a major cause of disease in Northern Spain. *Parkinsonism Relat Disord*, v.13, n.8, Dec, p.509-15. 2007.

Gorell, J. M., C. C. Johnson, B. A. Rybicki, E. L. Peterson and R. J. Richardson. The risk of Parkinson's disease with exposure to pesticides, farming, well water, and rural living. *Neurology*, v.50, n.5, May, p.1346-50. 1998.

Goris, A., C. H. Williams-Gray, G. R. Clark, T. Foltynie, S. J. Lewis, J. Brown, M. Ban, M. G. Spillantini, A. Compston, D. J. Burn, P. F. Chinnery, R. A. Barker and S. J. Sawcer. Tau and alpha-synuclein in susceptibility to, and dementia in, Parkinson's disease. *Ann Neurol*, v.62, n.2, Aug, p.145-53. 2007.

Goris, A., C. H. Williams-Gray, T. Foltynie, D. A. Compston, R. A. Barker and S. J. Sawcer. No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. *Am J Hum Genet*, v.78, n.6, Jun, p.1088-90; author reply 1092-4. 2006.

Graeber, M. B., K. G. Kupke and U. Muller. Delineation of the dystonia-parkinsonism syndrome locus in Xq13. *Proc Natl Acad Sci U S A*, v.89, n.17, Sep 1, p.8245-8. 1992.

Gray, C. W., R. V. Ward, E. Karran, S. Turconi, A. Rowles, D. Viglienghi, C. Southan, A. Barton, K. G. Fantom, A. West, J. Savopoulos, N. J. Hassan, H. Clinkenbeard, C. Hanning, B. Amegadzie, J. B. Davis, C. Dingwall, G. P. Livi and C. L. Creasy. Characterization of human HtrA2, a novel serine protease involved in the mammalian cellular stress response. *Eur J Biochem*, v.267, n.18, Sep, p.5699-710. 2000.

Greenfield, J. The Spino-cerebellar Degenerations. CC Thomas, Springfield, IL, p.112. 1954.

Greggio, E., S. Jain, A. Kingsbury, R. Bandopadhyay, P. Lewis, A. Kaganovich, M. P. Van Der Brug, A. Beilina, J. Blackinton, K. J. Thomas, R. Ahmad, D. W. Miller, S. Kesavapany, A. Singleton, A. Lees, R. J. Harvey, K. Harvey and M. R. Cookson. Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiol Dis*, v.23, n.2, Aug, p.329-41. 2006.

Guidot, D. M., J. M. Mccord, R. M. Wright and J. E. Repine. Absence of electron transport (Rho 0 state) restores growth of a manganese-superoxide dismutase-deficient *Saccharomyces cerevisiae* in hyperoxia. Evidence for electron transport as a major source of superoxide generation in vivo. *J Biol Chem*, v.268, n.35, Dec 15, p.26699-703. 1993.

Gwinn-Hardy, K., J. Y. Chen, H. C. Liu, T. Y. Liu, M. Boss, W. Seltzer, A. Adam, A. Singleton, W. Koroshetz, C. Waters, J. Hardy and M. Farrer. Spinocerebellar ataxia type 2 with parkinsonism in ethnic Chinese. *Neurology*, v.55, n.6, Sep 26, p.800-5. 2000.

Gwinn-Hardy, K., A. Singleton, P. O'suilleabhain, M. Boss, D. Nicholl, A. Adam, J. Hussey, P. Critchley, J. Hardy and M. Farrer. Spinocerebellar ataxia type 3 phenotypically resembling parkinson disease in a black family. *Arch Neurol*, v.58, n.2, Feb, p.296-9. 2001.

Haberhausen, G., I. Schmitt, A. Kohler, U. Peters, S. Rider, J. Chelly, J. D. Terwilliger, A. P. Monaco and U. Muller. Assignment of the dystonia-parkinsonism syndrome locus, DYT3, to a small region within a 1.8-Mb YAC contig of Xq13.1. *Am J Hum Genet*, v.57, n.3, Sep, p.644-50. 1995.

Hague, S., E. Rogaeva, D. Hernandez, C. Gulick, A. Singleton, M. Hanson, J. Johnson, R. Weiser, M. Gallardo, B. Ravina, K. Gwinn-Hardy, A. Crawley, P. H. St George-Hyslop, A. E. Lang, P. Heutink, V. Bonifati and J. Hardy. Early-onset Parkinson's disease caused by a compound heterozygous DJ-1 mutation. *Ann Neurol*, v.54, n.2, Aug, p.271-4. 2003.

Hampshire, D. J., E. Roberts, Y. Crow, J. Bond, A. Mubaidin, A. L. Wriekat, A. Al-Din and C. G. Woods. Kufor-Rakeb syndrome, pallido-pyramidal degeneration with supranuclear upgaze paresis and dementia, maps to 1p36. *J Med Genet*, v.38, n.10, Oct, p.680-2. 2001.

Harding, A. E. Classification of the hereditary ataxias and paraplegias. *Lancet*, v.1, n.8334, May 21, p.1151-5. 1983.

Harhangi, B. S., M. C. De Rijk, C. M. Van Duijn, C. Van Broeckhoven, A. Hofman and M. M. Breteler. APOE and the risk of PD with or without dementia in a population-based study. *Neurology*, v.54, n.6, Mar 28, p.1272-6. 2000.

Harhangi, B. S., M. J. Farrer, S. Lincoln, V. Bonifati, G. Meo, G. De Michele, A. Brice, A. Durr, M. Martinez, T. Gasser, B. Bereznoi, J. R. Vaughan, N. W. Wood, J. Hardy, B. A. Oostra and M. M. Breteler. The Ile93Met mutation in the ubiquitin carboxy-terminal-hydrolase-L1 gene is not observed in European cases with familial Parkinson's disease. *Neurosci Lett*, v.270, n.1, Jul 23, p.1-4. 1999.

Hasegawa, E., K. Takeshige, T. Oishi, Y. Murai and S. Minakami. 1-Methyl-4-phenylpyridinium (MPP+) induces NADH-dependent superoxide formation and enhances NADH-dependent lipid peroxidation in bovine heart submitochondrial particles. *Biochem Biophys Res Commun*, v.170, n.3, Aug 16, p.1049-55. 1990.

Hasegawa, K. and H. Kowa. Autosomal dominant familial Parkinson disease: older onset of age, and good response to levodopa therapy. *Eur Neurol*, v.38 Suppl 1, p.39-43. 1997.

Hassan, H. M. Biosynthesis and regulation of superoxide dismutases. *Free Radic Biol Med*, v.5, n.5-6, p.377-85. 1988.

Hatano, Y., Y. Li, K. Sato, S. Asakawa, Y. Yamamura, H. Tomiyama, H. Yoshino, M. Asahina, S. Kobayashi, S. Hassin-Baer, C. S. Lu, A. R. Ng, R. L. Rosales, N. Shimizu, T. Toda, Y. Mizuno and N. Hattori. Novel PINK1 mutations in early-onset parkinsonism. *Ann Neurol*, v.56, n.3, Sep, p.424-7. 2004.

Hatano, Y., K. Sato, B. Elibol, H. Yoshino, Y. Yamamura, V. Bonifati, H. Shinotoh, M. Asahina, S. Kobayashi, A. R. Ng, R. L. Rosales, S. Hassin-Baer, Y. Shinar, C. S. Lu, H. C. Chang, Y. H. Wu-Chou, F. B. Atac, T. Kobayashi, T. Toda, Y. Mizuno and N. Hattori. PARK6-linked autosomal recessive early-onset parkinsonism in Asian populations. *Neurology*, v.63, n.8, Oct 26, p.1482-5. 2004.

Hattori, N., M. Tanaka, T. Ozawa and Y. Mizuno. Immunohistochemical studies on complexes I, II, III, and IV of mitochondria in Parkinson's disease. *Ann Neurol*, v.30, n.4, Oct, p.563-71. 1991.

Haugarvoll, K., M. Toft, O. A. Ross, J. T. Stone, M. G. Heckman, L. R. White, T. Lynch, J. M. Gibson, Z. K. Wszolek, R. J. Uitti, J. O. Aasly and M. J. Farrer. ELAVL4, PARK10, and the Celts. *Mov Disord*, v.22, n.4, Mar 15, p.585-7. 2007.

Healy, D. G., P. M. Abou-Sleiman, J. M. Gibson, O. A. Ross, S. Jain, S. Gandhi, D. Gosal, M. M. Muqit, N. W. Wood and T. Lynch. PINK1 (PARK6) associated Parkinson disease in Ireland. *Neurology*, v.63, n.8, Oct 26, p.1486-8. 2004a.

Healy, D. G., P. M. Abou-Sleiman, A. J. Lees, J. P. Casas, N. Quinn, K. Bhatia, A. D. Hingorani and N. W. Wood. Tau gene and Parkinson's disease: a case-control study and meta-analysis. *J Neurol Neurosurg Psychiatry*, v.75, n.7, Jul, p.962-5. 2004b.

Healy, D. G., P. M. Abou-Sleiman, E. M. Valente, W. P. Gilks, K. Bhatia, N. Quinn, A. J. Lees and N. W. Wood. DJ-1 mutations in Parkinson's disease. *J Neurol Neurosurg Psychiatry*, v.75, n.1, Jan, p.144-5. 2004c.

Hedrich, K., A. Djarmati, N. Schafer, R. Hering, C. Wellenbrock, P. H. Weiss, R. Hilker, P. Vieregge, L. J. Ozelius, P. Heutink, V. Bonifati, E. Schwinger, A. E. Lang, J. Noth, S. B. Bressman, P. P. Pramstaller, O. Riess and C. Klein. DJ-1 (PARK7) mutations are less frequent than Parkin (PARK2) mutations in early-onset Parkinson disease. *Neurology*, v.62, n.3, Feb 10, p.389-94. 2004.

Hedrich, K., C. Eskelson, B. Wilmot, K. Marder, J. Harris, J. Garrels, H. Meija-Santana, P. Vieregge, H. Jacobs, S. B. Bressman, A. E. Lang, M. Kann, G. Abbruzzese, P. Martinelli, E. Schwinger, L. J. Ozelius, P. P. Pramstaller, C. Klein and P. Kramer. Distribution, type, and origin of Parkin mutations: review and case studies. *Mov Disord*, v.19, n.10, Oct, p.1146-57. 2004.

Hedrich, K., J. Hagenah, A. Djarmati, A. Hiller, T. Lohnau, K. Lasek, A. Grunewald, R. Hilker, S. Steinlechner, H. Boston, N. Kock, C. Schneider-Gold, W. Kress, H. Siebner, F. Binkofski, R. Lencer, A. Munchau and C. Klein. Clinical spectrum of homozygous and heterozygous PINK1 mutations in a large German family with Parkinson disease: role of a single hit? *Arch Neurol*, v.63, n.6, Jun, p.833-8. 2006.

Helisalmi, S., K. Linnaranta, M. Lehtovirta, A. Mannermaa, O. Heinonen, M. Ryyanen, P. Riekkinen, Sr. and H. Soininen. Apolipoprotein E polymorphism in patients with different neurodegenerative disorders. *Neurosci Lett*, v.205, n.1, Feb 16, p.61-4. 1996.

Hering, R., S. Petrovic, E. M. Mietz, C. Holzmann, D. Berg, P. Bauer, D. Voitalla, T. Muller, K. Berger, R. Kruger and O. Riess. Extended mutation analysis and association studies of Nurr1 (NR4A2) in Parkinson disease. *Neurology*, v.62, n.7, Apr 13, p.1231-2. 2004.

Hernandez, A., C. Magarino, S. Gispert, N. Santos, A. Lunkes, G. Orozco, L. Heredero, J. Beckmann and G. Auburger. Genetic mapping of the spinocerebellar ataxia 2 (SCA2) locus on chromosome 12q23-q24.1. *Genomics*, v.25, n.2, Jan 20, p.433-5. 1995.

Hicks, A. A., H. Petursson, T. Jonsson, H. Stefansson, H. S. Johannsdottir, J. Sainz, M. L. Frigge, A. Kong, J. R. Gulcher, K. Stefansson and S. Sveinbjornsdottir. A susceptibility gene for late-onset idiopathic Parkinson's disease. *Ann Neurol*, v.52, n.5, Nov, p.549-55. 2002.

Higashi, S., S. Biskup, A. B. West, D. Trinkaus, V. L. Dawson, R. L. Faull, H. J. Waldvogel, H. Arai, T. M. Dawson, D. J. Moore and P. C. Emson. Localization of

Parkinson's disease-associated LRRK2 in normal and pathological human brain. *Brain Res*, v.1155, Jun 25, p.208-19. 2007.

Higashi, S., D. J. Moore, R. E. Colebrooke, S. Biskup, V. L. Dawson, H. Arai, T. M. Dawson and P. C. Emson. Expression and localization of Parkinson's disease-associated leucine-rich repeat kinase 2 in the mouse brain. *J Neurochem*, v.100, n.2, Jan, p.368-81. 2007.

Hoenicka, J., L. Vidal, B. Morales, I. Ampuero, F. J. Jimenez-Jimenez, J. Berciano, T. Del Ser, A. Jimenez, P. G. Ruiz and J. G. De Yébenes. Molecular findings in familial Parkinson disease in Spain. *Arch Neurol*, v.59, n.6, Jun, p.966-70. 2002.

Holzmann, C., R. Kruger, A. M. Saecker, I. Schmitt, L. Schols, K. Berger and O. Riess. Polymorphisms of the alpha-synuclein promoter: expression analyses and association studies in Parkinson's disease. *J Neural Transm*, v.110, n.1, Jan, p.67-76. 2003.

Huang, X., P. C. Chen and C. Poole. APOE-[epsilon]2 allele associated with higher prevalence of sporadic Parkinson disease. *Neurology*, v.62, n.12, Jun 22, p.2198-202. 2004.

Hughes, A. J., S. E. Daniel, L. Kilford and A. J. Lees. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*, v.55, n.3, Mar, p.181-4. 1992.

Hutton, M., C. L. Lendon, P. Rizzu, M. Baker, S. Froelich, H. Houlden, S. Pickering-Brown, S. Chakraverty, A. Isaacs, A. Grover, J. Hackett, J. Adamson, S. Lincoln, D. Dickson, P. Davies, R. C. Petersen, M. Stevens, E. De Graaff, E. Wauters, J. Van Baren, M. Hillebrand, M. Joosse, J. M. Kwon, P. Nowotny, L. K. Che, J. Norton, J. C. Morris, L. A. Reed, J. Trojanowski, H. Basun, L. Lannfelt, M. Neystat, S. Fahn, F. Dark, T. Tannenberg, P. R. Dodd, N. Hayward, J. B. Kwok, P. R. Schofield, A. Andreadis, J. Snowden, D. Craufurd, D. Neary, F. Owen, B. A. Oostra, J. Hardy, A. Goate, J. Van Swieten, D. Mann, T. Lynch and P. Heutink. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*, v.393, n.6686, Jun 18, p.702-5. 1998.

Huynh, D. P., D. R. Scoles, D. Nguyen and S. M. Pulst. The autosomal recessive juvenile Parkinson disease gene product, parkin, interacts with and ubiquitinates synaptotagmin XI. *Hum Mol Genet*, v.12, n.20, Oct 15, p.2587-97. 2003.

Iafraite, A. J., L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer and C. Lee. Detection of large-scale variation in the human genome. *Nat Genet*, v.36, n.9, Sep, p.949-51. 2004.

Ibanez, P., A. M. Bonnet, B. Debarges, E. Lohmann, F. Tison, P. Pollak, Y. Agid, A. Durrand and A. Brice. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet*, v.364, n.9440, Sep 25-Oct 1, p.1169-71. 2004.

Ibanez, P., G. De Michele, V. Bonifati, E. Lohmann, S. Thobois, P. Pollak, Y. Agid, P. Heutink, A. Durrand and A. Brice. Screening for DJ-1 mutations in early onset autosomal recessive parkinsonism. *Neurology*, v.61, n.10, Nov 25, p.1429-31. 2003.

Ibanez, P., S. Lesage, E. Lohmann, S. Thobois, G. De Michele, M. Borg, Y. Agid, A. Durrand and A. Brice. Mutational analysis of the PINK1 gene in early-onset parkinsonism in Europe and North Africa. *Brain*, v.129, n.Pt 3, Mar, p.686-94. 2006.

Ibanez, P., E. Lohmann, P. Pollak, F. Durif, C. Tranchant, Y. Agid, A. Durrand and A. Brice. Absence of NR4A2 exon 1 mutations in 108 families with autosomal dominant Parkinson disease. *Neurology*, v.62, n.11, Jun 8, p.2133-4. 2004.

Ibarreta, D., T. Gomez-Isla, A. Portera-Sanchez, R. Parrilla and M. S. Ayuso. Apolipoprotein E genotype in Spanish patients of Alzheimer's or Parkinson's disease. *J Neurol Sci*, v.134, n.1-2, Dec, p.146-9. 1995.

Ichinose, H., T. Ohye, E. Takahashi, N. Seki, T. Hori, M. Segawa, Y. Nomura, K. Endo, H. Tanaka, S. Tsuji and Et Al. Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase I gene. *Nat Genet*, v.8, n.3, Nov, p.236-42. 1994.

Ilic, T. V., M. Jovanovic, A. Jovicic and M. Tomovic. Oxidative stress indicators are elevated in de novo Parkinson's disease patients. *Funct Neurol*, v.14, n.3, Jul-Sep, p.141-7. 1999.

Imai, Y., M. Soda, H. Inoue, N. Hattori, Y. Mizuno and R. Takahashi. An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. *Cell*, v.105, n.7, Jun 29, p.891-902. 2001.

Infante, J., E. Rodriguez, O. Combarros, I. Mateo, A. Fontalba, J. Pascual, A. Oterino, J. M. Polo, C. Leno and J. Berciano. LRRK2 G2019S is a common mutation in Spanish patients with late-onset Parkinson's disease. *Neurosci Lett*, v.395, n.3, Mar 13, p.224-6. 2006.

Inzelberg, R., E. Schechtman and D. Paleacu. Onset age of Parkinson disease. *Am J Med Genet*, v.111, n.4, Sep 1, p.459-60; author reply 461. 2002.

Ishikawa, A. and S. Tsuji. Clinical analysis of 17 patients in 12 Japanese families with autosomal-recessive type juvenile parkinsonism. *Neurology*, v.47, n.1, Jul, p.160-6. 1996.

Ishizawa, T., P. Mattila, P. Davies, D. Wang and D. W. Dickson. Colocalization of tau and alpha-synuclein epitopes in Lewy bodies. *J Neuropathol Exp Neurol*, v.62, n.4, Apr, p.389-97. 2003.

Iwanaga, K., K. Wakabayashi, M. Yoshimoto, I. Tomita, H. Satoh, H. Takashima, A. Satoh, M. Seto, M. Tsujihata and H. Takahashi. Lewy body-type degeneration in cardiac plexus in Parkinson's and incidental Lewy body diseases. *Neurology*, v.52, n.6, Apr 12, p.1269-71. 1999.

Izumi, Y., H. Morino, M. Oda, H. Maruyama, F. Udaka, M. Kameyama, S. Nakamura and H. Kawakami. Genetic studies in Parkinson's disease with an alpha-

synuclein/NACP gene polymorphism in Japan. *Neurosci Lett*, v.300, n.2, Mar 9, p.125-7. 2001.

Jager, D. H. and J. Bethlem. The distribution of Lewy bodies in the central and autonomic nervous systems in idiopathic paralysis agitans. *J Neurol Neurosurg Psychiatry*, v.23, Nov, p.283-90. 1960.

Jakobsson, M., S. W. Scholz, P. Scheet, J. R. Gibbs, J. M. Vanliere, H. C. Fung, Z. A. Szpiech, J. H. Degnan, K. Wang, R. Guerreiro, J. M. Bras, J. C. Schymick, D. G. Hernandez, B. J. Traynor, J. Simon-Sanchez, M. Matarin, A. Britton, J. Van De Leemput, I. Rafferty, M. Bucan, H. M. Cann, J. A. Hardy, N. A. Rosenberg and A. B. Singleton. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, v.451, n.7181, Feb 21, p.998-1003. 2008.

Johnson, J., S. M. Hague, M. Hanson, A. Gibson, K. E. Wilson, E. W. Evans, A. A. Singleton, A. Mcinerney-Leo, R. L. Nussbaum, D. G. Hernandez, M. Gallardo, I. G. Mckeith, D. J. Burn, M. Ryu, O. Hellstrom, B. Ravina, J. Eerola, R. H. Perry, E. Jaros, P. Tienari, R. Weiser, K. Gwinn-Hardy, C. M. Morris, J. Hardy and A. B. Singleton. SNCA multiplication is not a common cause of Parkinson disease or dementia with Lewy bodies. *Neurology*, v.63, n.3, Aug 10, p.554-6. 2004.

Johnson, W. G., S. E. Hodge and R. Duvoisin. Twin studies and the genetics of Parkinson's disease--a reappraisal. *Mov Disord*, v.5, n.3, p.187-94. 1990.

Jones, A. C., Y. Yamamura, L. Almasy, S. Bohlega, B. Elibol, J. Hubble, S. Kuzuhara, M. Uchida, T. Yanagi, D. E. Weeks and T. G. Nygaard. Autosomal recessive juvenile parkinsonism maps to 6q25.2-q27 in four ethnic groups: detailed genetic mapping of the linked region. *Am J Hum Genet*, v.63, n.1, Jul, p.80-7. 1998.

Jones, J. M., P. Datta, S. M. Srinivasula, W. Ji, S. Gupta, Z. Zhang, E. Davies, G. Hajnoczky, T. L. Saunders, M. L. Van Keuren, T. Fernandes-Alnemri, M. H. Meisler and E. S. Alnemri. Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice. *Nature*, v.425, n.6959, Oct 16, p.721-7. 2003.

Junque, C., B. Ramirez-Ruiz, E. Tolosa, C. Summerfield, M. J. Martí, P. Pastor, B. Gomez-Anson and J. M. Mercader. Amygdalar and hippocampal MRI volumetric reductions in Parkinson's disease with dementia. *Mov Disord*, v.20, n.5, May, p.540-4. 2005.

Kachergus, J., I. F. Mata, M. Hulihan, J. P. Taylor, S. Lincoln, J. Aasly, J. M. Gibson, O. A. Ross, T. Lynch, J. Wiley, H. Payami, J. Nutt, D. M. Maraganore, K. Czyzewski, M. Styczynska, Z. K. Wszolek, M. J. Farrer and M. Toft. Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. *Am J Hum Genet*, v.76, n.4, Apr, p.672-80. 2005.

Kakita, A., H. Takahashi, Y. Homma and F. Ikuta. Lewy bodies in the cerebellar dentate nucleus of a patient with Parkinson's disease. *Pathol Int*, v.44, n.12, Dec, p.878-80. 1994.

Karamohamed, S., A. L. Destefano, J. B. Wilk, C. M. Shoemaker, L. I. Golbe, M. H. Mark, A. M. Lazzarini, O. Suchowersky, N. Labelle, M. Guttman, L. J. Currie, G. F. Wooten, M. Stacy, M. Saint-Hilaire, R. G. Feldman, K. M. Sullivan, G. Xu, R. Watts, J. Growdon, M. Lew, C. Waters, P. Viergge, P. P. Pramstaller, C. Klein, B. A. Racette, J. S. Perlmutter, A. Parsian, C. Singer, E. Montgomery, K. Baker, J. F. Gusella, S. J. Fink, R. H. Myers and A. Herbert. A haplotype at the PARK3 locus influences onset age for Parkinson's disease: the GenePD study. *Neurology*, v.61, n.11, Dec 9, p.1557-61. 2003.

Kawaguchi, Y., T. Okamoto, M. Taniwaki, M. Aizawa, M. Inoue, S. Katayama, H. Kawakami, S. Nakamura, M. Nishimura, I. Akiguchi and Et Al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet*, v.8, n.3, Nov, p.221-8. 1994.

Kay, D. M., P. Kramer, D. Higgins, C. P. Zabetian and H. Payami. Escaping Parkinson's disease: a neurologically healthy octogenarian with the LRRK2 G2019S mutation. *Mov Disord*, v.20, n.8, Aug, p.1077-8. 2005.

Kay, D. M., D. Moran, L. Moses, P. Poorkaj, C. P. Zabetian, J. Nutt, S. A. Factor, C. E. Yu, J. S. Montimurro, R. G. Keefe, G. D. Schellenberg and H. Payami. Heterozygous

parkin point mutations are as common in control subjects as in Parkinson's patients. *Ann Neurol*, v.61, n.1, Jan, p.47-54. 2007.

Khan, N., E. Graham, P. Dixon, C. Morris, A. Mander, D. Clayton, J. Vaughan, N. Quinn, A. Lees, S. Daniel, N. Wood and R. De Silva. Parkinson's disease is not associated with the combined alpha-synuclein/apolipoprotein E susceptibility genotype. *Ann Neurol*, v.49, n.5, May, p.665-8. 2001.

Khan, N. L., W. Horta, L. Eunson, E. Graham, J. O. Johnson, S. Chang, M. Davis, A. Singleton, N. W. Wood and A. J. Lees. Parkin disease in a Brazilian kindred: Manifesting heterozygotes and clinical follow-up over 10 years. *Mov Disord*, v.20, n.4, Apr, p.479-84. 2005.

Khan, N. L., S. Jain, J. M. Lynch, N. Pavese, P. Abou-Sleiman, J. L. Holton, D. G. Healy, W. P. Gilks, M. G. Sweeney, M. Ganguly, V. Gibbons, S. Gandhi, J. Vaughan, L. H. Eunson, R. Katzenschlager, J. Gayton, G. Lennox, T. Revesz, D. Nicholl, K. P. Bhatia, N. Quinn, D. Brooks, A. J. Lees, M. B. Davis, P. Piccini, A. B. Singleton and N. W. Wood. Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: clinical, pathological, olfactory and functional imaging and genetic data. *Brain*, v.128, n.Pt 12, Dec, p.2786-96. 2005.

Kitada, T., S. Asakawa, N. Hattori, H. Matsumine, Y. Yamamura, S. Minoshima, M. Yokochi, Y. Mizuno and N. Shimizu. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*, v.392, n.6676, Apr 9, p.605-8. 1998.

Klein, C., A. Djarmati, K. Hedrich, N. Schafer, C. Scaglione, R. Marchese, N. Kock, B. Schule, A. Hiller, T. Lohnau, S. Winkler, K. Wiegers, R. Hering, P. Bauer, O. Riess, G. Abbruzzese, P. Martinelli and P. P. Pramstaller. PINK1, Parkin, and DJ-1 mutations in Italian patients with early-onset parkinsonism. *Eur J Hum Genet*, v.13, n.9, Sep, p.1086-93. 2005.

Klein, C., P. P. Pramstaller, B. Kis, C. C. Page, M. Kann, J. Leung, H. Woodward, C. C. Castellan, M. Scherer, P. Vieregge, X. O. Breakefield, P. L. Kramer and L. J. Ozelius. Parkin deletions in a family with adult-onset, tremor-dominant parkinsonism: expanding the phenotype. *Ann Neurol*, v.48, n.1, Jul, p.65-71. 2000.

Klein, C., P. Vieregge, J. Hagenah, M. Sieberer, E. Doyle, H. Jacobs, T. Gasser, X. O. Breakefield, N. J. RischandL. J. Ozelius. Search for the PARK3 founder haplotype in a large cohort of patients with Parkinson's disease from northern Germany. *Ann Hum Genet*, v.63, n.Pt 4, Jul, p.285-91. 1999.

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. Sangiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. BarnstableandJ. Hoh. Complement factor H polymorphism in age-related macular degeneration. *Science*, v.308, n.5720, Apr 15, p.385-9. 2005.

Knight, J. C. Allele-specific gene expression uncovered. *Trends Genet*, v.20, n.3, Mar, p.113-6. 2004.

Kosaka, K. Lewy bodies in cerebral cortex, report of three cases. *Acta Neuropathol*, v.42, n.2, May 24, p.127-34. 1978.

Kruger, R., W. Kuhn, T. Muller, D. Voitalla, M. Graeber, S. Kosel, H. Przuntek, J. T. Epplen, L. ScholsandO. Riess. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat Genet*, v.18, n.2, Feb, p.106-8. 1998.

Kruger, R., A. M. Vieira-Saecker, W. Kuhn, D. Berg, T. Muller, N. Kuhn, G. A. Fuchs, A. Storch, M. Hungs, D. Voitalla, H. Przuntek, J. T. Epplen, L. ScholsandO. Riess. Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype. *Ann Neurol*, v.45, n.5, May, p.611-7. 1999.

Kupke, K. G., M. B. GraeberandU. Muller. Dystonia-parkinsonism syndrome (XDP) locus: flanking markers in Xq12-q21.1. *Am J Hum Genet*, v.50, n.4, Apr, p.808-15. 1992.

Kupke, K. G., L. V. LeeandU. Muller. Assignment of the X-linked torsion dystonia gene to Xq21 by linkage analysis. *Neurology*, v.40, n.9, Sep, p.1438-42. 1990.

Kuusisto, E., L. Parkkinen and I. Alafuzoff. Morphogenesis of Lewy bodies: dissimilar incorporation of alpha-synuclein, ubiquitin, and p62. *J Neuropathol Exp Neurol*, v.62, n.12, Dec, p.1241-53. 2003.

Kwok, J. B., E. T. Teber, C. Loy, M. Hallupp, G. Nicholson, G. D. Mellick, D. D. Buchanan, P. A. Silburn and P. R. Schofield. Tau haplotypes regulate transcription and are associated with Parkinson's disease. *Ann Neurol*, v.55, n.3, Mar, p.329-34. 2004.

Lander, E. S. and D. Botstein. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*, v.236, n.4808, Jun 19, p.1567-70. 1987.

Langston, J. W., P. Ballard, J. W. Tetrud and I. Irwin. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science*, v.219, n.4587, Feb 25, p.979-80. 1983.

Lazzarini, A. M., R. H. Myers, T. R. Zimmerman, Jr., M. H. Mark, L. I. Golbe, J. I. Sage, W. G. Johnson and R. C. Duvoisin. A clinical genetic study of Parkinson's disease: evidence for dominant transmission. *Neurology*, v.44, n.3 Pt 1, Mar, p.499-506. 1994.

Le, W. D., P. Xu, J. Jankovic, H. Jiang, S. H. Appel, R. G. Smith and D. K. Vassilatis. Mutations in NR4A2 associated with familial Parkinson disease. *Nat Genet*, v.33, n.1, Jan, p.85-9. 2003.

Lee, L. V., F. M. Pascasio, F. D. Fuentes and G. H. Viterbo. Torsion dystonia in Panay, Philippines. *Adv Neurol*, v.14, p.137-51. 1976.

Lee, Y., I. Hur, S. Y. Park, Y. K. Kim, M. R. Suh and V. N. Kim. The role of PACT in the RNA silencing pathway. *Embo J*, v.25, n.3, Feb 8, p.522-32. 2006.

Leroy, E., R. Boyer, G. Auburger, B. Leube, G. Ulm, E. Mezey, G. Harta, M. J. Brownstein, S. Jonnalagada, T. Chernova, A. Dehejia, C. Lavedan, T. Gasser, P. J. Steinbach, K. D. Wilkinson and M. H. Polymeropoulos. The ubiquitin pathway in Parkinson's disease. *Nature*, v.395, n.6701, Oct 1, p.451-2. 1998.

Lesage, S., A. Durr, M. Tazir, E. Lohmann, A. L. Leutenegger, S. Janin, P. Pollak and A. Brice. LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. *N Engl J Med*, v.354, n.4, Jan 26, p.422-3. 2006.

Leutenegger, A. L., M. A. Salih, P. Ibanez, M. M. Mukhtar, S. Lesage, A. Arabi, E. Lohmann, A. Durr, A. E. Ahmed and A. Brice. Juvenile-onset Parkinsonism as a result of the first mutation in the adenosine triphosphate orientation domain of PINK1. *Arch Neurol*, v.63, n.9, Sep, p.1257-61. 2006.

Levecque, C., A. Destee, V. Mouroux, P. Amouyel and M. C. Chartier-Harlin. Assessment of Nurr1 nucleotide variations in familial Parkinson's disease. *Neurosci Lett*, v.366, n.2, Aug 12, p.135-8. 2004.

Levecque, C., A. Destee, V. Mouroux, E. Becquet, L. Defebvre, P. Amouyel and M. C. Chartier-Harlin. No genetic association of the ubiquitin carboxy-terminal hydrolase-L1 gene S18Y polymorphism with familial Parkinson's disease. *J Neural Transm*, v.108, n.8-9, p.979-84. 2001.

Lewy, F. *Handbuch der Neurologie*. Berlin: Springer. 1912.

Li, Y., C. Rowland, S. Schrodi, W. Laird, K. Tacey, D. Ross, D. Leong, J. Catanese, J. Sninsky and A. Grupe. A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. *Am J Hum Genet*, v.78, n.6, Jun, p.1090-2; author reply 1092-4. 2006.

Li, Y., H. Tomiyama, K. Sato, Y. Hatano, H. Yoshino, M. Atsumi, M. Kitaguchi, S. Sasaki, S. Kawaguchi, H. Miyajima, T. Toda, Y. Mizuno and N. Hattori. *Clinicogenetic*

study of PINK1 mutations in autosomal recessive early-onset parkinsonism. *Neurology*, v.64, n.11, Jun 14, p.1955-7. 2005.

Li, Y. J., J. Deng, G. M. Mayhew, J. W. Grimsley, X. Huo and J. M. Vance. Investigation of the PARK10 Gene in Parkinson Disease. *Ann Hum Genet*, Mar 27. 2007.

Li, Y. J., M. A. Hauser, W. K. Scott, E. R. Martin, M. W. Booze, X. J. Qin, J. W. Walter, M. A. Nance, J. P. Hubble, W. C. Koller, R. Pahwa, M. B. Stern, B. C. Hiner, J. Jankovic, C. G. Goetz, G. W. Small, F. Mastaglia, J. L. Haines, M. A. Pericak-Vance and J. M. Vance. Apolipoprotein E controls the risk and age at onset of Parkinson disease. *Neurology*, v.62, n.11, Jun 8, p.2005-9. 2004.

Li, Y. J., W. K. Scott, D. J. Hedges, F. Zhang, P. C. Gaskell, M. A. Nance, R. L. Watts, J. P. Hubble, W. C. Koller, R. Pahwa, M. B. Stern, B. C. Hiner, J. Jankovic, F. A. Allen, Jr., C. G. Goetz, F. Mastaglia, J. M. Stajich, R. A. Gibson, L. T. Middleton, A. M. Saunders, B. L. Scott, G. W. Small, K. K. Nicodemus, A. D. Reed, D. E. Schmechel, K. A. Welsh-Bohmer, P. M. Conneally, A. D. Roses, J. R. Gilbert, J. M. Vance, J. L. Haines and M. A. Pericak-Vance. Age at onset in two common neurodegenerative diseases is genetically controlled. *Am J Hum Genet*, v.70, n.4, Apr, p.985-93. 2002.

Lim, S. W., Y. Zhao, E. Chua, H. Y. Law, Y. Yuen, R. Pavanni, M. C. Wong, I. S. Ng, C. S. Yoon, K. Y. Puong, S. H. Lim and E. K. Tan. Genetic analysis of SCA2, 3 and 17 in idiopathic Parkinson's disease. *Neurosci Lett*, v.403, n.1-2, Jul 31, p.11-4. 2006.

Lincoln, S., J. Vaughan, N. Wood, M. Baker, J. Adamson, K. Gwinn-Hardy, T. Lynch, J. Hardy and M. Farrer. Low frequency of pathogenic mutations in the ubiquitin carboxy-terminal hydrolase gene in familial Parkinson's disease. *Neuroreport*, v.10, n.2, Feb 5, p.427-9. 1999.

Lincoln, S. J., D. M. Maraganore, T. G. Lesnick, R. Bounds, M. De Andrade, J. H. Bower, J. A. Hardy and M. J. Farrer. Parkin variants in North American Parkinson's disease: cases and controls. *Mov Disord*, v.18, n.11, Nov, p.1306-11. 2003.

Liou, H. H., M. C. Tsai, C. J. Chen, J. S. Jeng, Y. C. Chang, S. Y. Chen and R. C. Chen. Environmental risk factors and Parkinson's disease: a case-control study in Taiwan. *Neurology*, v.48, n.6, Jun, p.1583-8. 1997.

Lipczynska-Lojkowska, W. [Etiology of parkinsonism. A verification of Poskanzer's and Schwab's hypothesis]. *Neurol Neurochir Pol*, v.13, n.6, Nov-Dec, p.593-9. 1979.

Lo Bianco, C., J. L. Ridet, B. L. Schneider, N. Deglon and P. Aebischer. alpha - Synucleinopathy and selective dopaminergic neuron loss in a rat lentiviral-based model of Parkinson's disease. *Proc Natl Acad Sci U S A*, v.99, n.16, Aug 6, p.10813-8. 2002.

Lockhart, P. J., R. Bounds, M. Hulihan, J. Kachergus, S. Lincoln, C. H. Lin, R. M. Wu and M. J. Farrer. Lack of mutations in DJ-1 in a cohort of Taiwanese ethnic Chinese with early-onset parkinsonism. *Mov Disord*, v.19, n.9, Sep, p.1065-9. 2004.

Lockhart, P. J., S. Lincoln, M. Hulihan, J. Kachergus, K. Wilkes, G. Bisceglia, D. C. Mash and M. J. Farrer. DJ-1 mutations are a rare cause of recessively inherited early onset parkinsonism mediated by loss of protein function. *J Med Genet*, v.41, n.3, Mar, p.e22. 2004.

Lowe, J., H. McDermott, M. Landon, R. J. Mayer and K. D. Wilkinson. Ubiquitin carboxyl-terminal hydrolase (PGP 9.5) is selectively present in ubiquitinated inclusion bodies characteristic of human neurodegenerative diseases. *J Pathol*, v.161, n.2, Jun, p.153-60. 1990.

Lucking, C. B., A. Durr, V. Bonifati, J. Vaughan, G. De Michele, T. Gasser, B. S. Harhangi, G. Meco, P. Deneffe, N. W. Wood, Y. Agid and A. Brice. Association between early-onset Parkinson's disease and mutations in the parkin gene. *N Engl J Med*, v.342, n.21, May 25, p.1560-7. 2000.

Lwin, A., E. Orvisky, O. Goker-Alpan, M. E. Lamarca and E. Sidransky. Glucocerebrosidase mutations in subjects with parkinsonism. *Mol Genet Metab*, v.81, n.1, Jan, p.70-3. 2004.

Macdonald, B. K., O. C. Cockerell, J. W. Sander and S. D. Shorvon. The incidence and lifetime prevalence of neurological disorders in a prospective community-based study in the UK. *Brain*, v.123 ( Pt 4), Apr, p.665-76. 2000.

Makino, S., R. Kaji, S. Ando, M. Tomizawa, K. Yasuno, S. Goto, S. Matsumoto, M. D. Tabuena, E. Maranon, M. Dantes, L. V. Lee, K. Ogasawara, I. Tooyama, H. Akatsu, M. Nishimura and G. Tamiya. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet*, v.80, n.3, Mar, p.393-406. 2007.

Maraganore, D. M., M. De Andrade, A. Elbaz, M. J. Farrer, J. P. Ioannidis, R. Kruger, W. A. Rocca, N. K. Schneider, T. G. Lesnick, S. J. Lincoln, M. M. Hulihan, J. O. Aasly, T. Ashizawa, M. C. Chartier-Harlin, H. Checkoway, C. Ferrarese, G. Hadjigeorgiou, N. Hattori, H. Kawakami, J. C. Lambert, T. Lynch, G. D. Mellick, S. Papapetropoulos, A. Parsian, A. Quattrone, O. Riess, E. K. Tan and C. Van Broeckhoven. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *Jama*, v.296, n.6, Aug 9, p.661-70. 2006.

Maraganore, D. M., M. De Andrade, T. G. Lesnick, K. J. Strain, M. J. Farrer, W. A. Rocca, P. V. Pant, K. A. Frazer, D. R. Cox and D. G. Ballinger. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet*, v.77, n.5, Nov, p.685-93. 2005.

Maraganore, D. M., M. J. Farrer, J. A. Hardy, S. J. Lincoln, S. K. McDonnell and W. A. Rocca. Case-control study of the ubiquitin carboxy-terminal hydrolase L1 gene in Parkinson's disease. *Neurology*, v.53, n.8, Nov 10, p.1858-60. 1999.

Maraganore, D. M., M. J. Farrer, J. A. Hardy, S. K. McDonnell, D. J. Schaid and W. A. Rocca. Case-control study of debrisoquine 4-hydroxylase, N-acetyltransferase 2, and apolipoprotein E gene polymorphisms in Parkinson's disease. *Mov Disord*, v.15, n.4, Jul, p.714-9. 2000.

Maraganore, D. M., T. G. Lesnick, A. Elbaz, M. C. Chartier-Harlin, T. Gasser, R. Kruger, N. Hattori, G. D. Mellick, A. Quattrone, J. Satoh, T. Toda, J. Wang, J. P. Ioannidis, M. De Andrade and W. A. Rocca. UCHL1 is a Parkinson's disease susceptibility gene. *Ann Neurol*, v.55, n.4, Apr, p.512-21. 2004.

Maraganore, D. M., D. J. Schaid, W. A. Rocca and A. E. Harding. Anticipation in familial Parkinson's disease: a reanalysis of 13 United Kingdom kindreds. *Neurology*, v.47, n.6, Dec, p.1512-7. 1996.

Marder, K., G. Maestre, L. Cote, H. Mejia, B. Alfaro, A. Halim, M. Tang, B. Tycko and R. Mayeux. The apolipoprotein epsilon 4 allele in Parkinson's disease with and without dementia. *Neurology*, v.44, n.7, Jul, p.1330-1. 1994.

Marder, K., M. X. Tang, H. Mejia, B. Alfaro, L. Cote, E. Louis, J. Groves and R. Mayeux. Risk of Parkinson's disease among first-degree relatives: A community-based study. *Neurology*, v.47, n.1, Jul, p.155-60. 1996.

Marie, P. Sur l'he´re´do-ataxie ce´re´belleuse. *Semaine me´d*, v.13, p.444-447. 1893.

Martin, E. R., W. K. Scott, M. A. Nance, R. L. Watts, J. P. Hubble, W. C. Koller, K. Lyons, R. Pahwa, M. B. Stern, A. Colcher, B. C. Hiner, J. Jankovic, W. G. Ondo, F. H. Allen, Jr., C. G. Goetz, G. W. Small, D. Masterman, F. Mastaglia, N. G. Laing, J. M. Stajich, R. C. Ribble, M. W. Booze, A. Rogala, M. A. Hauser, F. Zhang, R. A. Gibson, L. T. Middleton, A. D. Roses, J. L. Haines, B. L. Scott, M. A. Pericak-Vance and J. M. Vance. Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. *Jama*, v.286, n.18, Nov 14, p.2245-50. 2001.

Martinez-Suarez, M. M. and B. Blazquez-Menes. [Estimation of the prevalence of Parkinson's disease in Asturias, Spain. A pharmacoepidemiological study of the consumption of antiparkinson drugs]. *Rev Neurol*, v.31, n.11, Dec 1-15, p.1001-6. 2000.

Martins, L. M., A. Morrison, K. Klupsch, V. Fedele, N. Moiso, P. Teismann, A. Abuin, E. Grau, M. Geppert, G. P. Livi, C. L. Creasy, A. Martin, I. Hargreaves, S. J. Heales, H.

Okada, S. Brandner, J. B. Schulz, T. MakandJ. Downward. Neuroprotective role of the Reaper-related serine protease HtrA2/Omi revealed by targeted deletion in mice. *Mol Cell Biol*, v.24, n.22, Nov, p.9848-62. 2004.

Maslah, E., E. Rockenstein, I. Veinbergs, M. Mallory, M. Hashimoto, A. Takeda, Y. Sagara, A. SiskandL. Mucke. Dopaminergic loss and inclusion body formation in alpha-synuclein mice: implications for neurodegenerative disorders. *Science*, v.287, n.5456, Feb 18, p.1265-9. 2000.

Mata, I. F., V. Alvarez, E. Coto, M. Blazquez, L. M. Guisasola, C. Salvador, J. M. Kachergus, S. J. LincolnandM. Farrer. Homozygous partial genomic triplication of the parkin gene in early-onset parkinsonism. *Neurosci Lett*, v.380, n.3, Jun 3, p.257-9. 2005.

Mata, I. F., J. M. Kachergus, J. P. Taylor, S. Lincoln, J. Aasly, T. Lynch, M. M. Hulihan, S. A. Cobb, R. M. Wu, C. S. Lu, C. Lahoz, Z. K. WszolekandM. J. Farrer. Lrrk2 pathogenic substitutions in Parkinson's disease. *Neurogenetics*, v.6, n.4, Dec, p.171-7. 2005.

Mata, I. F., O. A. Ross, J. Kachergus, C. Huerta, R. Ribacoba, G. Moris, M. Blazquez, L. M. Guisasola, C. Salvador, C. Martinez, M. FarrerandV. Alvarez. LRRK2 mutations are a common cause of Parkinson's disease in Spain. *Eur J Neurol*, v.13, n.4, Apr, p.391-4. 2006.

Mata, I. F., J. P. Taylor, J. Kachergus, M. Hulihan, C. Huerta, C. Lahoz, M. Blazquez, L. M. Guisasola, C. Salvador, R. Ribacoba, C. Martinez, M. FarrerandV. Alvarez. LRRK2 R1441G in Spanish patients with Parkinson's disease. *Neurosci Lett*, v.382, n.3, Jul 15, p.309-11. 2005.

Mata, I. F., W. J. Wedemeyer, M. J. Farrer, J. P. TaylorandK. A. Gallo. LRRK2 in Parkinson's disease: protein domains and functional insights. *Trends Neurosci*, v.29, n.5, May, p.286-93. 2006.

Matsumine, H., M. Saito, S. Shimoda-Matsubayashi, H. Tanaka, A. Ishikawa, Y. Nakagawa-Hattori, M. Yokochi, T. Kobayashi, S. Igarashi, H. Takano, K. Sanpei, R. Koike, H. Mori, T. Kondo, Y. Mizutani, A. A. Schaffer, Y. Yamamura, S. Nakamura, S. Kuzuhara, S. Tsuji and Y. Mizuno. Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. *Am J Hum Genet*, v.60, n.3, Mar, p.588-96. 1997.

Matsumine, H., Y. Yamamura, N. Hattori, T. Kobayashi, T. Kitada, A. Yoritaka and Y. Mizuno. A microdeletion of D6S305 in a family of autosomal recessive juvenile parkinsonism (PARK2). *Genomics*, v.49, n.1, Apr 1, p.143-6. 1998.

Mayeux, R., K. Marder, L. J. Cote, J. Denaro, N. Hemenegildo, H. Mejia, M. X. Tang, R. Lantigua, D. Wilder, B. Gurland and Et Al. The frequency of idiopathic Parkinson's disease by age, ethnic group, and sex in northern Manhattan, 1988-1993. *Am J Epidemiol*, v.142, n.8, Oct 15, p.820-7. 1995.

Mccarroll, S. A., T. N. Hadnott, G. H. Perry, P. C. Sabeti, M. C. Zody, J. C. Barrett, S. Dallaire, S. B. Gabriel, C. Lee, M. J. Daly and D. M. Altshuler. Common deletion polymorphisms in the human genome. *Nat Genet*, v.38, n.1, Jan, p.86-92. 2006.

Mccormack, A. L., M. Thiruchelvam, A. B. Manning-Bog, C. Thiffault, J. W. Langston, D. A. Cory-Slechta and D. A. Di Monte. Environmental risk factors and Parkinson's disease: selective degeneration of nigral dopaminergic neurons caused by the herbicide paraquat. *Neurobiol Dis*, v.10, n.2, Jul, p.119-27. 2002.

Mckeran, R. O., P. Bradbury, D. Taylor and G. Stern. Neurological involvement in type 1 (adult) Gaucher's disease. *J Neurol Neurosurg Psychiatry*, v.48, n.2, Feb, p.172-5. 1985.

Mellick, G. D., D. M. Maraganore and P. A. Silburn. Australian data and meta-analysis lend support for alpha-synuclein (NACP-Rep1) as a risk factor for Parkinson's disease. *Neurosci Lett*, v.375, n.2, Feb 28, p.112-6. 2005.

Mellick, G. D. and P. A. Silburn. The ubiquitin carboxy-terminal hydrolase-L1 gene S18Y polymorphism does not confer protection against idiopathic Parkinson's disease. *Neurosci Lett*, v.293, n.2, Oct 27, p.127-30. 2000.

Melrose, H., S. Lincoln, G. Tyndall, D. Dickson and M. Farrer. Anatomical localization of leucine-rich repeat kinase 2 in mouse brain. *Neuroscience*, v.139, n.3, p.791-4. 2006.

Melrose, H. L., C. B. Kent, J. P. Taylor, J. C. Dachsel, K. M. Hinkle, S. J. Lincoln, S. S. Mok, J. G. Culvenor, C. L. Masters, G. M. Tyndall, D. I. Bass, Z. Ahmed, C. A. Andorfer, O. A. Ross, Z. K. Wszolek, A. Delldonne, D. W. Dickson and M. J. Farrer. A comparative analysis of leucine-rich repeat kinase 2 (Lrrk2) expression in mouse brain and Lewy body disease. *Neuroscience*, v.147, n.4, Jul 29, p.1047-58. 2007.

Migliazza, A., F. Bosch, H. Komatsu, E. Cayanis, S. Martinotti, E. Toniato, E. Guccione, X. Qu, M. Chien, V. V. Murty, G. Gaidano, G. Inghirami, P. Zhang, S. Fischer, S. M. Kalachikov, J. Russo, I. Edelman, A. Efstratiadis and R. Dalla-Favera. Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood*, v.97, n.7, Apr 1, p.2098-104. 2001.

Miklossy, J., T. Arai, J. P. Guo, A. Klegeris, S. Yu, E. G. McGeer and P. L. McGeer. LRRK2 expression in normal and pathologic human brain and in human cell lines. *J Neuropathol Exp Neurol*, v.65, n.10, Oct, p.953-63. 2006.

Miller, D. W., S. M. Hague, J. Clarimon, M. Baptista, K. Gwinn-Hardy, M. R. Cookson and A. B. Singleton. Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. *Neurology*, v.62, n.10, May 25, p.1835-8. 2004.

Miller, G., T. Shope, H. Lisco, D. Stitt and M. Lipman. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. *Proc Natl Acad Sci U S A*, v.69, n.2, Feb, p.383-7. 1972.

Miller, S. A., D. D. Dykes and H. F. Polesky. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, v.16, n.3, Feb 11, p.1215. 1988.

Mizuno, Y., S. Ohta, M. Tanaka, S. Takamiya, K. Suzuki, T. Sato, H. Oya, T. Ozawa and Y. Kagawa. Deficiencies in complex I subunits of the respiratory chain in Parkinson's disease. *Biochem Biophys Res Commun*, v.163, n.3, Sep 29, p.1450-5. 1989.

Mizuta, I., M. Nishimura, E. Mizuta, S. Yamasaki, M. Ohta and S. Kuno. Meta-analysis of alpha synuclein/ NACP polymorphism in Parkinson's disease in Japan. *J Neurol Neurosurg Psychiatry*, v.73, n.3, Sep, p.350. 2002.

Modoni, A., M. F. Contarino, A. R. Bentivoglio, E. Tabolacci, M. Santoro, M. L. Calcagni, P. A. Tonali, G. Neri and G. Silvestri. Prevalence of spinocerebellar ataxia type 2 mutation among Italian Parkinsonian patients. *Mov Disord*, v.22, n.3, Feb 15, p.324-7. 2007.

Momose, Y., M. Murata, K. Kobayashi, M. Tachikawa, Y. Nakabayashi, I. Kanazawa and T. Toda. Association studies of multiple candidate genes for Parkinson's disease using single nucleotide polymorphisms. *Ann Neurol*, v.51, n.1, Jan, p.133-6. 2002.

Moore, D. J., L. Zhang, T. M. Dawson and V. L. Dawson. A missense mutation (L166P) in DJ-1, linked to familial Parkinson's disease, confers reduced protein stability and impairs homo-oligomerization. *J Neurochem*, v.87, n.6, Dec, p.1558-67. 2003.

Morens, D. M., J. W. Davis, A. Grandinetti, G. W. Ross, J. S. Popper and L. R. White. Epidemiologic observations on Parkinson's disease: incidence and mortality in a prospective study of middle-aged men. *Neurology*, v.46, n.4, Apr, p.1044-50. 1996.

Mori, H., T. Kondo, M. Yokochi, H. Matsumine, Y. Nakagawa-Hattori, T. Miyake, K. Suda and Y. Mizuno. Pathologic and biochemical studies of juvenile parkinsonism linked to chromosome 6q. *Neurology*, v.51, n.3, Sep, p.890-2. 1998.

Morrison, P. J., R. B. Godwin-Austen and J. A. Raeburn. Familial autosomal dominant dopa responsive Parkinson's disease in three living generations showing extreme anticipation and childhood onset. *J Med Genet*, v.33, n.6, Jun, p.504-6. 1996.

Muller, U., G. Haberhausen, T. Wagner, N. D. Fairweather, J. Chelly and A. P. Monaco. DXS106 and DXS559 flank the X-linked dystonia-parkinsonism syndrome locus (DYT3). *Genomics*, v.23, n.1, Sep 1, p.114-7. 1994.

Munhoz, R. P., D. S. Sa, E. Rogaeva, S. Salehi-Rad, C. Sato, H. Medeiros, M. Farrer and A. E. Lang. Clinical findings in a large family with a parkin ex3delta40 mutation. *Arch Neurol*, v.61, n.5, May, p.701-4. 2004.

Najim Al-Din, A. S., A. Wriekat, A. Mubaidin, M. Dasouki and M. Hiari. Pallido-pyramidal degeneration, supranuclear upgaze paresis and dementia: Kufor-Rakeb syndrome. *Acta Neurol Scand*, v.89, n.5, May, p.347-52. 1994.

Narhi, L., S. J. Wood, S. Steavenson, Y. Jiang, G. M. Wu, D. Anafi, S. A. Kaufman, F. Martin, K. Sitney, P. Denis, J. C. Louis, J. Wypych, A. L. Biere and M. Citron. Both familial Parkinson's disease mutations accelerate alpha-synuclein aggregation. *J Biol Chem*, v.274, n.14, Apr 2, p.9843-6. 1999.

Nemeth, A. H., D. Nolte, E. Dunne, S. Niemann, M. Kostrzewa, U. Peters, E. Fraser, E. Bochukova, R. Butler, J. Brown, R. D. Cox, E. R. Levy, H. H. Ropers, A. P. Monaco and U. Muller. Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). *Genomics*, v.60, n.3, Sep 15, p.320-9. 1999.

Neudorfer, O., N. Giladi, D. Elstein, A. Abrahamov, T. Turezkite, E. Aghai, A. Reches, B. Bembi and A. Zimran. Occurrence of Parkinson's syndrome in type I Gaucher disease. *Qjm*, v.89, n.9, Sep, p.691-4. 1996.

Newton, C. R., A. Graham, L. E. Heptinstall, S. J. Powell, C. Summers, N. Kalsheker, J. C. Smith and A. F. Markham. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, v.17, n.7, Apr 11, p.2503-16. 1989.

Nichols, W. C., S. K. Uniacke, N. Pankratz, T. Reed, D. K. Simon, C. Halter, A. Rudolph, C. W. Shults, P. M. Conneally and T. Foroud. Evaluation of the role of Nurr1 in a large sample of familial Parkinson's disease. *Mov Disord*, v.19, n.6, Jun, p.649-55. 2004.

Nicklas, W. J., I. Vyas and R. E. Heikkila. Inhibition of NADH-linked oxidation in brain mitochondria by 1-methyl-4-phenyl-pyridine, a metabolite of the neurotoxin, 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine. *Life Sci*, v.36, n.26, Jul 1, p.2503-8. 1985.

Nolte, D., S. Niemann and U. Müller. Specific sequence changes in multiple transcript system DYT3 are associated with X-linked dystonia parkinsonism. *Proc Natl Acad Sci U S A*, v.100, n.18, Sep 2, p.10347-52. 2003.

Noureddine, M. A., X. J. Qin, S. A. Oliveira, T. J. Skelly, J. Van Der Walt, M. A. Hauser, M. A. Pericak-Vance, J. M. Vance and Y. J. Li. Association between the neuron-specific RNA-binding protein ELAVL4 and Parkinson disease. *Hum Genet*, v.117, n.1, Jun, p.27-33. 2005.

Ohama, E. and F. Ikuta. Parkinson's disease: distribution of Lewy bodies and monoamine neuron system. *Acta Neuropathol*, v.34, n.4, Apr 26, p.311-9. 1976.

Okazaki, H., L. E. Lipkin and S. M. Aronson. Diffuse intracytoplasmic ganglionic inclusions (Lewy type) associated with progressive dementia and quadriplegia in flexion. *J Neuropathol Exp Neurol*, v.20, Apr, p.237-44. 1961.

Oliveira, S. A., Y. J. Li, M. A. Noureddine, S. Zuchner, X. Qin, M. A. Pericak-Vance and J. M. Vance. Identification of risk and age-at-onset genes on chromosome 1p in Parkinson disease. *Am J Hum Genet*, v.77, n.2, Aug, p.252-64. 2005.

Oliveira, S. A., W. K. Scott, E. R. Martin, M. A. Nance, R. L. Watts, J. P. Hubble, W. C. Koller, R. Pahwa, M. B. Stern, B. C. Hiner, W. G. Ondo, F. H. Allen, Jr., B. L. Scott, C. G. Goetz, G. W. Small, F. Mastaglia, J. M. Stajich, F. Zhang, M. W. Booze, M. P. Winn, L. T. Middleton, J. L. Haines, M. A. Pericak-Vance and J. M. Vance. Parkin mutations and susceptibility alleles in late-onset Parkinson's disease. *Ann Neurol*, v.53, n.5, May, p.624-9. 2003.

Oliveri, R. L., M. Zappia, G. Annesi, D. Bosco, F. Annesi, P. Spadafora, A. A. Pasqua, C. Tomaino, G. Nicoletti, D. Pirritano, A. Labate, A. Gambardella, G. Logroscino, G. Manobianca, A. Epifanio, L. Morgante, G. Savettieri and A. Quattrone. The parkin gene is not involved in late-onset Parkinson's disease. *Neurology*, v.57, n.2, Jul 24, p.359-62. 2001.

Orozco, G., R. Estrada, T. L. Perry, J. Arana, R. Fernandez, A. Gonzalez-Quevedo, J. Galarraga and S. Hansen. Dominantly inherited olivopontocerebellar atrophy from eastern Cuba. Clinical, neuropathological, and biochemical findings. *J Neurol Sci*, v.93, n.1, Oct, p.37-50. 1989.

Orr, H. T., M. Y. Chung, S. Banfi, T. J. Kwiatkowski, Jr., A. Servadio, A. L. Beaudet, A. E. Mccall, L. A. Duvick, L. P. Ranum and H. Y. Zoghbi. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet*, v.4, n.3, Jul, p.221-6. 1993.

Oyanagi, K., K. Wakabayashi, E. Ohama, S. Takeda, Y. Horikawa, T. Morita and F. Ikuta. Lewy bodies in the lower sacral parasympathetic neurons of a patient with Parkinson's disease. *Acta Neuropathol*, v.80, n.5, p.558-9. 1990.

Ozaki, K., Y. Ohnishi, A. Iida, A. Sekine, R. Yamada, T. Tsunoda, H. Sato, M. Hori, Y. Nakamura and T. Tanaka. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet*, v.32, n.4, Dec, p.650-4. 2002.

Ozelius, L. J., J. W. Hewett, C. E. Page, S. B. Bressman, P. L. Kramer, C. Shalish, D. De Leon, M. F. Brin, D. Raymond, D. P. Corey, S. Fahn, N. J. Risch, A. J. Buckler, J. F. Gusella and X. O. Breakefield. The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. *Nat Genet*, v.17, n.1, Sep, p.40-8. 1997.

Ozelius, L. J., G. Senthil, R. Saunders-Pullman, E. Ohmann, A. Deligtisch, M. Tagliati, A. L. Hunt, C. Klein, B. Henick, S. M. Hailpern, R. B. Lipton, J. Soto-Valencia, N. Risch and S. B. Bressman. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *N Engl J Med*, v.354, n.4, Jan 26, p.424-5. 2006.

Paisan-Ruiz, C., E. W. Evans, S. Jain, G. Xiromerisiou, J. R. Gibbs, J. Eerola, V. Gourbali, O. Hellstrom, J. Duckworth, A. Papadimitriou, P. J. Tienari, G. M. Hadjigeorgiou and A. B. Singleton. Testing association between LRRK2 and Parkinson's disease and investigating linkage disequilibrium. *J Med Genet*, v.43, n.2, Feb, p.e9. 2006.

Paisan-Ruiz, C., S. Jain, E. W. Evans, W. P. Gilks, J. Simon, M. Van Der Brug, A. Lopez De Munain, S. Aparicio, A. M. Gil, N. Khan, J. Johnson, J. R. Martinez, D. Nicholl, I. M. Carrera, A. S. Pena, R. De Silva, A. Lees, J. F. Marti-Masso, J. Perez-Tur, N. W. Wood and A. B. Singleton. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron*, v.44, n.4, Nov 18, p.595-600. 2004.

Paisan-Ruiz, C., A. E. Lang, T. Kawarai, C. Sato, S. Salehi-Rad, G. K. Fisman, T. Al-Khairallah, P. St George-Hyslop, A. Singleton and E. Rogaeva. LRRK2 gene in Parkinson disease: mutation analysis and case control association study. *Neurology*, v.65, n.5, Sep 13, p.696-700. 2005.

Paisan-Ruiz, C., A. Saenz, A. Lopez De Munain, I. Marti, A. Martinez Gil, J. F. Marti-Masso and J. Perez-Tur. Familial Parkinson's disease: clinical and genetic analysis of four Basque families. *Ann Neurol*, v.57, n.3, Mar, p.365-72. 2005.

Pankratz, N., L. Byder, C. Halter, A. Rudolph, C. W. Shults, P. M. Conneally, T. Foroud and W. C. Nichols. Presence of an APOE4 allele results in significantly earlier onset of Parkinson's disease and a higher risk with dementia. *Mov Disord*, v.21, n.1, Jan, p.45-9. 2006.

Pankratz, N., W. C. Nichols, S. K. Uniacke, C. Halter, A. Rudolph, C. Shults, P. M. Conneally and T. Foroud. Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations. *Am J Hum Genet*, v.71, n.1, Jul, p.124-35. 2002.

Pankratz, N., W. C. Nichols, S. K. Uniacke, C. Halter, A. Rudolph, C. Shults, P. M. Conneally and T. Foroud. Significant linkage of Parkinson disease to chromosome 2q36-37. *Am J Hum Genet*, v.72, n.4, Apr, p.1053-7. 2003.

Pankratz, N., M. W. Pauciulo, V. E. Elsaesser, D. K. Marek, C. A. Halter, A. Rudolph, C. W. Shults, T. Foroud and W. C. Nichols. Mutations in LRRK2 other than G2019S are rare in a north American-based sample of familial Parkinson's disease. *Mov Disord*, v.21, n.12, Dec, p.2257-60. 2006.

Pankratz, N., S. K. Uniacke, C. A. Halter, A. Rudolph, C. W. Shults, P. M. Conneally, T. Foroud and W. C. Nichols. Genes influencing Parkinson disease onset: replication of PARK3 and identification of novel loci. *Neurology*, v.62, n.9, May 11, p.1616-8. 2004.

Pappolla, M. A., D. L. Shank, J. Alzofon and A. W. Dudley. Colloid (hyaline) inclusion bodies in the central nervous system: their presence in the substantia nigra is diagnostic of Parkinson's disease. *Hum Pathol*, v.19, n.1, Jan, p.27-31. 1988.

Park, H. J., Y. M. Seong, J. Y. Choi, S. Kang and H. Rhim. Alzheimer's disease-associated amyloid beta interacts with the human serine protease HtrA2/Omi. *Neurosci Lett*, v.357, n.1, Feb 26, p.63-7. 2004.

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

Parsian, A., B. Racette, L. J. Goldsmith and J. S. Perlmutter. Parkinson's disease and apolipoprotein E: possible association with dementia but not age at onset. *Genomics*, v.79, n.3, Mar, p.458-61. 2002.

Parsian, A., B. Racette, Z. H. Zhang, S. Chakraverty, M. Rundle, A. Goate and J. S. Perlmutter. Mutation, sequence analysis, and association studies of alpha-synuclein in Parkinson's disease. *Neurology*, v.51, n.6, Dec, p.1757-9. 1998.

Pastor, P., M. Ezquerro, E. Muñoz, M. J. Martí, R. Blesa, E. Tolosa and R. Oliva. Significant association between the tau gene A0/A0 genotype and Parkinson's disease. *Ann Neurol*, v.47, n.2, Feb, p.242-5. 2000.

Patel, C. V., I. Handy, T. Goldsmith and R. C. Patel. PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR. *J Biol Chem*, v.275, n.48, Dec 1, p.37993-8. 2000.

Paxinos G, F. K. The mouse brain in stereotaxic co-ordinates. Academic Press, San Diego, CA. 2001.

Payami, H., S. Bernard, K. Larsen, J. Kaye and J. Nutt. Genetic anticipation in Parkinson's disease. *Neurology*, v.45, n.1, Jan, p.135-8. 1995.

Payami, H., J. Nutt, S. Ganther, T. Bird, M. G. Mcneal, W. K. Seltzer, J. Hussey, P. Lockhart, K. Gwinn-Hardy, A. A. Singleton, A. B. Singleton, J. Hardy and M. Farrer. SCA2 may present as levodopa-responsive parkinsonism. *Mov Disord*, v.18, n.4, Apr, p.425-9. 2003.

Payami, H., S. Zarepari, D. James and J. Nutt. Familial aggregation of Parkinson disease: a comparative study of early-onset and late-onset disease. *Arch Neurol*, v.59, n.5, May, p.848-50. 2002.

Peiffer, D. A., J. M. Le, F. J. Steemers, W. Chang, T. Jenniges, F. Garcia, K. Haden, J. Li, C. A. Shaw, J. Belmont, S. W. Cheung, R. M. Shen, D. L. Barker and K. L. Gunderson. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res*, v.16, n.9, Sep, p.1136-48. 2006.

Perez-Pastene, C., S. A. Cobb, F. Diaz-Grez, M. M. Hulihan, M. Miranda, P. Venegas, O. T. Godoy, J. M. Kachergus, O. A. Ross, L. Layson, M. J. Farrer and J. Segura-Aguilar. Lrrk2 mutations in South America: A study of Chilean Parkinson's disease. *Neurosci Lett*, v.422, n.3, Jul 18, p.193-7. 2007.

Perez-Tur, J. Parkinson's disease genetics: a complex disease comes to the clinic. *Lancet Neurol*, v.5, n.11, Nov, p.896-7. 2006.

Petrovitch, H., G. W. Ross, R. D. Abbott, W. T. Sanderson, D. S. Sharp, C. M. Tanner, K. H. Masaki, P. L. Blanchette, J. S. Popper, D. Foley, L. Launer and L. R. White. Plantation work and risk of Parkinson disease in a population-based longitudinal study. *Arch Neurol*, v.59, n.11, Nov, p.1787-92. 2002.

Piccini, P., D. J. Burn, R. Ceravolo, D. Maraganore and D. J. Brooks. The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. *Ann Neurol*, v.45, n.5, May, p.577-82. 1999.

Piccini, P., P. K. Morrish, N. Turjanski, G. V. Sawle, D. J. Burn, R. A. Weeks, M. H. Mark, D. M. Maraganore, A. J. Lees and D. J. Brooks. Dopaminergic function in familial Parkinson's disease: a clinical and 18F-dopa positron emission tomography study. *Ann Neurol*, v.41, n.2, Feb, p.222-9. 1997.

Plante-Bordeneuve, V., D. Taussig, F. Thomas, M. Ziegler and G. Said. A clinical and genetic study of familial cases of Parkinson's disease. *J Neurol Sci*, v.133, n.1-2, Nov, p.164-72. 1995.

Poduslo, S. E., D. Riggs, T. Rolan and J. Schwankhaus. Apolipoprotein E and B alleles in Parkinson's patients. *Neurosci Lett*, v.194, n.1-2, Jul 14, p.145-7. 1995.

Polymeropoulos, M. H., J. J. Higgins, L. I. Golbe, W. G. Johnson, S. E. Ide, G. Di Iorio, G. Sanges, E. S. Stenroos, L. T. Pho, A. A. Schaffer, A. M. Lazzarini, R. L.

Nussbaum and R. C. Duvoisin. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science*, v.274, n.5290, Nov 15, p.1197-9. 1996.

Polymeropoulos, M. H., C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe and R. L. Nussbaum. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, v.276, n.5321, Jun 27, p.2045-7. 1997.

Poorkaj, P., T. D. Bird, E. Wijsman, E. Nemens, R. M. Garruto, L. Anderson, A. Andreadis, W. C. Wiederholt, M. Raskind and G. D. Schellenberg. Tau is a candidate gene for chromosome 17 frontotemporal dementia. *Ann Neurol*, v.43, n.6, Jun, p.815-25. 1998.

Poorkaj, P., L. Moses, J. S. Montimurro, J. G. Nutt, G. D. Schellenberg and H. Payami. Parkin mutation dosage and the phenomenon of anticipation: a molecular genetic study of familial parkinsonism. *BMC Neurol*, v.5, n.1, Feb 22, p.4. 2005.

Poskanzer, D. C. and R. S. Schwab. Studies in the epidemiology of Parkinson's disease predicting its disappearance as a major clinical entity by 1980. *Trans Am Neurol Assoc*, v.86, p.234-5. 1961.

Prestel, J., M. Sharma, P. Leitner, A. Zimprich, J. R. Vaughan, A. Durr, V. Bonifati, G. De Michele, H. A. Hanagasi, M. Farrer, A. Hofer, F. Asmus, G. Volpe, G. Meco, A. Brice, N. W. Wood, B. Muller-Myhsok and T. Gasser. PARK11 is not linked with Parkinson's disease in European families. *Eur J Hum Genet*, v.13, n.2, Feb, p.193-7. 2005.

Pulst, S. M., A. Nechiporuk, T. Nechiporuk, S. Gispert, X. N. Chen, I. Lopes-Cendes, S. Pearlman, S. Starkman, G. Orozco-Diaz, A. Lunke, P. Dejong, G. A. Rouleau, G. Auburger, J. R. Korenberg, C. Figueroa and S. Sahba. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet*, v.14, n.3, Nov, p.269-76. 1996.

Pulst, S. M., A. Nechiporuk and S. Starkman. Anticipation in spinocerebellar ataxia type 2. *Nat Genet*, v.5, n.1, Sep, p.8-10. 1993.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly and P. C. Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, v.81, n.3, Sep, p.559-75. 2007.

Ragothaman, M., N. Sarangmath, S. Chaudhary, V. Khare, U. Mittal, S. Sharma, S. Komatireddy, S. Chakrabarti, M. Mukerji, R. C. Juyal, B. K. Thelma and U. B. Muthane. Complex phenotypes in an Indian family with homozygous SCA2 mutations. *Ann Neurol*, v.55, n.1, Jan, p.130-3. 2004.

Rajput, A. H., K. P. Offord, C. M. Beard and L. T. Kurland. Epidemiology of parkinsonism: incidence, classification, and mortality. *Ann Neurol*, v.16, n.3, Sep, p.278-82. 1984.

Ramirez, A., A. Heimbach, J. Grundemann, B. Stiller, D. Hampshire, L. P. Cid, I. Goebel, A. F. Mubaidin, A. L. Wriekat, J. Roeper, A. Al-Din, A. M. Hillmer, M. Karsak, B. Liss, C. G. Woods, M. I. Behrens and C. Kubisch. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nat Genet*, v.38, n.10, Oct, p.1184-91. 2006.

Ramsay, R. R. and T. P. Singer. Energy-dependent uptake of N-methyl-4-phenylpyridinium, the neurotoxic metabolite of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine, by mitochondria. *J Biol Chem*, v.261, n.17, Jun 15, p.7585-7. 1986.

Ren, Y., J. Zhao and J. Feng. Parkin binds to alpha/beta tubulin and increases their ubiquitination and degradation. *J Neurosci*, v.23, n.8, Apr 15, p.3316-24. 2003.

Rohe, C. F., P. Montagna, G. Breedveld, P. Cortelli, B. A. Oostra and V. Bonifati. Homozygous PINK1 C-terminus mutation causing early-onset parkinsonism. *Ann Neurol*, v.56, n.3, Sep, p.427-31. 2004.

Rosenberg, M. J., D. Vaske, C. E. Killoran, Y. Ning, D. Wargowski, L. Hudgins, C. J. Tifft, J. Meck, J. K. Blancato, K. Rosenbaum, R. M. Pauli, J. Weber and L. G. Biesecker. Detection of chromosomal aberrations by a whole-genome microsatellite screen. *Am J Hum Genet*, v.66, n.2, Feb, p.419-27. 2000.

Rubinsztein, D. C., C. S. Hanlon, R. M. Irving, S. Goodburn, D. G. Evans, H. Kellar-Wood, J. H. Xuereb, O. Bandmann and A. E. Harding. Apo E genotypes in multiple sclerosis, Parkinson's disease, schwannomas and late-onset Alzheimer's disease. *Mol Cell Probes*, v.8, n.6, Dec, p.519-25. 1994.

Sack, G. H., Jr. Clinical diversity in Gaucher's disease. *Johns Hopkins Med J*, v.146, n.4, Apr, p.166-70. 1980.

Saggu, H., J. Cooksey, D. Dexter, F. R. Wells, A. Lees, P. Jenner and C. D. Marsden. A selective increase in particulate superoxide dismutase activity in parkinsonian substantia nigra. *J Neurochem*, v.53, n.3, Sep, p.692-7. 1989.

Sato, C., A. Morgan, A. E. Lang, S. Salehi-Rad, T. Kawarai, Y. Meng, P. N. Ray, L. A. Farrer, P. St George-Hyslop and E. Rogaeva. Analysis of the glucocerebrosidase gene in Parkinson's disease. *Mov Disord*, v.20, n.3, Mar, p.367-70. 2005.

Satoh, J. and Y. Kuroda. A polymorphic variation of serine to tyrosine at codon 18 in the ubiquitin C-terminal hydrolase-L1 gene is associated with a reduced risk of sporadic Parkinson's disease in a Japanese population. *J Neurol Sci*, v.189, n.1-2, Aug 15, p.113-7. 2001.

Saunders-Pullman, R. Estrogens and Parkinson disease: neuroprotective, symptomatic, neither, or both? *Endocrine*, v.21, n.1, Jun, p.81-7. 2003.

Saunders, A. M., K. Schmader, J. C. Breitner, M. D. Benson, W. T. Brown, L. Goldfarb, D. Goldgaber, M. G. Manwaring, M. H. Szymanski, N. McCown and Et Al.

Apolipoprotein E epsilon 4 allele distributions in late-onset Alzheimer's disease and in other amyloid-forming diseases. *Lancet*, v.342, n.8873, Sep 18, p.710-1. 1993.

Savettieri, G., E. V. De Marco, D. Civitelli, G. Salemi, G. Nicoletti, G. Annesi, I. C. Ciro Candiano and A. Quattrone. Lack of association between ubiquitin carboxy-terminal hydrolase L1 gene polymorphism and PD. *Neurology*, v.57, n.3, Aug 14, p.560-1. 2001.

Saxena, R., B. F. Voight, V. Lyssenko, N. P. Burtt, P. I. De Bakker, H. Chen, J. J. Roix, S. Kathiresan, J. N. Hirschhorn, M. J. Daly, T. E. Hughes, L. Groop, D. Altshuler, P. Almgren, J. C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H. N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E. K. Speliotes, M. R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S. B. Gabriel, G. W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke and S. Purcell. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, v.316, n.5829, Jun 1, p.1331-6. 2007.

Schapira, A. H., J. M. Cooper, D. Dexter, P. Jenner, J. B. Clark and C. D. Marsden. Mitochondrial complex I deficiency in Parkinson's disease. *Lancet*, v.1, n.8649, Jun 3, p.1269. 1989.

Scheet, P. and M. Stephens. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*, v.78, n.4, Apr, p.629-44. 2006.

Schlitter, A. M., M. Kurz, J. P. Larsen, D. Voitalla, T. Muller, J. T. Epplen and G. Dekomien. Parkin gene variations in late-onset Parkinson's disease: comparison between Norwegian and German cohorts. *Acta Neurol Scand*, v.113, n.1, Jan, p.9-13. 2006.

Schlossmacher, M. G., M. P. Frosch, W. P. Gai, M. Medina, N. Sharma, L. Forno, T. Ochiishi, H. Shimura, R. Sharon, N. Hattori, J. W. Langston, Y. Mizuno, B. T. Hyman,

D. J. Selkoe and K. S. Kosik. Parkin localizes to the Lewy bodies of Parkinson disease and dementia with Lewy bodies. *Am J Pathol*, v.160, n.5, May, p.1655-67. 2002.

Schut, J. Hereditary ataxia: clinical study through six generations. *Arch Neurol Psychiatr*, v.63, p.535-568. 1950.

Scott, L. J., K. L. Mohlke, L. L. Bonnycastle, C. J. Willer, Y. Li, W. L. Duren, M. R. Erdos, H. M. Stringham, P. S. Chines, A. U. Jackson, L. Prokunina-Olsson, C. J. Ding, A. J. Swift, N. Narisu, T. Hu, R. Pruim, R. Xiao, X. Y. Li, K. N. Conneely, N. L. Riebow, A. G. Sprau, M. Tong, P. P. White, K. N. Hetrick, M. W. Barnhart, C. W. Bark, J. L. Goldstein, L. Watkins, F. Xiang, J. Saramies, T. A. Buchanan, R. M. Watanabe, T. T. Valle, L. Kinnunen, G. R. Abecasis, E. W. Pugh, K. F. Doheny, R. N. Bergman, J. Tuomilehto, F. S. Collins and M. Boehnke. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, v.316, n.5829, Jun 1, p.1341-5. 2007.

Scott, W. K., M. A. Nance, R. L. Watts, J. P. Hubble, W. C. Koller, K. Lyons, R. Pahwa, M. B. Stern, A. Colcher, B. C. Hiner, J. Jankovic, W. G. Ondo, F. H. Allen, Jr., C. G. Goetz, G. W. Small, D. Masterman, F. Mastaglia, N. G. Laing, J. M. Stajich, B. Slotterbeck, M. W. Booze, R. C. Ribble, E. Rampersaud, S. G. West, R. A. Gibson, L. T. Middleton, A. D. Roses, J. L. Haines, B. L. Scott, J. M. Vance and M. A. Pericak-Vance. Complete genomic screen in Parkinson disease: evidence for multiple genes. *Jama*, v.286, n.18, Nov 14, p.2239-44. 2001.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg and M. Wigler. Large-scale copy number polymorphism in the human genome. *Science*, v.305, n.5683, Jul 23, p.525-8. 2004.

Seibler, P., A. Djarmati, B. Langpap, J. Hagenah, A. Schmidt, N. Bruggemann, H. Siebner, H. C. Jabusch, E. Altenmuller, A. Munchau, K. Lohmann and C. Klein. A heterozygous frameshift mutation in PRKRA (DYT16) associated with generalised dystonia in a German patient. *Lancet Neurol*, v.7, n.5, May, p.380-1. 2008.

Seidler, A., W. Hellenbrand, B. P. Robra, P. Vieregge, P. Nischan, J. Joerg, W. H. Oertel, G. UlmandE. Schneider. Possible environmental, occupational, and other etiologic factors for Parkinson's disease: a case-control study in Germany. *Neurology*, v.46, n.5, May, p.1275-84. 1996.

Semchuk, K. M., E. J. LoveandR. G. Lee. Parkinson's disease and exposure to agricultural work and pesticide chemicals. *Neurology*, v.42, n.7, Jul, p.1328-35. 1992.

Shan, D. E., B. W. Soong, C. M. Sun, S. J. Lee, K. K. LiaoandR. S. Liu. Spinocerebellar ataxia type 2 presenting as familial levodopa-responsive parkinsonism. *Ann Neurol*, v.50, n.6, Dec, p.812-5. 2001.

Sharp, A. J., D. P. Locke, S. D. Mcgrath, Z. Cheng, J. A. Bailey, R. U. Vallente, L. M. Pertz, R. A. Clark, S. Schwartz, R. Segraves, V. V. Oseroff, D. G. Albertson, D. PinkelandE. E. Eichler. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*, v.77, n.1, Jul, p.78-88. 2005.

Sherrington, R., E. I. Rogaev, Y. Liang, E. A. Rogaeva, G. Levesque, M. Ikeda, H. Chi, C. Lin, G. Li, K. HolmanandEt Al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*, v.375, n.6534, Jun 29, p.754-60. 1995.

Shi, Q.andE. Tao. An Ile93Met substitution in the UCH-L1 gene is not a disease-causing mutation for idiopathic Parkinson's disease. *Chin Med J (Engl)*, v.116, n.2, Feb, p.312-3. 2003.

Shimura, H., N. Hattori, S. Kubo, Y. Mizuno, S. Asakawa, S. Minoshima, N. Shimizu, K. Iwai, T. Chiba, K. TanakaandT. Suzuki. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat Genet*, v.25, n.3, Jul, p.302-5. 2000.

Shimura, H., M. G. Schlossmacher, N. Hattori, M. P. Frosch, A. Trockenbacher, R. Schneider, Y. Mizuno, K. S. KosikandD. J. Selkoe. Ubiquitination of a new form of alpha-synuclein by parkin from human brain: implications for Parkinson's disease. *Science*, v.293, n.5528, Jul 13, p.263-9. 2001.

Simon-Sanchez, J., M. Hanson, A. Singleton, D. Hernandez, A. Mcinerney, R. Nussbaum, J. Werner, M. Gallardo, R. Weiser, K. Gwinn-Hardy, A. B. Singleton and J. Clarimon. Analysis of SCA-2 and SCA-3 repeats in Parkinsonism: evidence of SCA-2 expansion in a family with autosomal dominant Parkinson's disease. *Neurosci Lett*, v.382, n.1-2, Jul 1-8, p.191-4. 2005.

Singleton, A. and K. Gwinn-Hardy. Parkinson's disease and dementia with Lewy bodies: a difference in dose? *Lancet*, v.364, n.9440, Sep 25-Oct 1, p.1105-7. 2004.

Singleton, A. B., M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum, S. Lincoln, A. Crawley, M. Hanson, D. Maraganore, C. Adler, M. R. Cookson, M. Muenter, M. Baptista, D. Miller, J. Blancato, J. Hardy and K. Gwinn-Hardy. alpha-Synuclein locus triplication causes Parkinson's disease. *Science*, v.302, n.5646, Oct 31, p.841. 2003.

Skipper, L., H. Shen, E. Chua, C. Bonnard, P. Kolatkar, L. C. Tan, R. D. Jamora, K. Puvan, K. Y. Puong, Y. Zhao, R. Pavanni, M. C. Wong, Y. Yuen, M. Farrer, J. J. Liu and E. K. Tan. Analysis of LRRK2 functional domains in nondominant Parkinson disease. *Neurology*, v.65, n.8, Oct 25, p.1319-21. 2005.

Skol, A. D., L. J. Scott, G. R. Abecasis and M. Boehnke. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*, v.38, n.2, Feb, p.209-13. 2006.

Spadafora, P., G. Annesi, A. A. Pasqua, P. Serra, I. C. Ciro Candiano, S. Carrideo, P. Tarantino, D. Civitelli, E. V. De Marco, G. Nicoletti, F. Annesi and A. Quattrone. NACP-REP1 polymorphism is not involved in Parkinson's disease: a case-control study in a population sample from southern Italy. *Neurosci Lett*, v.351, n.2, Nov 13, p.75-8. 2003.

Spillantini, M. G., J. R. Murrell, M. Goedert, M. R. Farlow, A. Klug and B. Ghetti. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc Natl Acad Sci U S A*, v.95, n.13, Jun 23, p.7737-41. 1998.

Spillantini, M. G., M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert. Alpha-synuclein in Lewy bodies. *Nature*, v.388, n.6645, Aug 28, p.839-40. 1997.

Staropoli, J. F., C. Mcdermott, C. Martinat, B. Schulman, E. Demireva and A. Abeliovich. Parkin is a component of an SCF-like ubiquitin ligase complex and protects postmitotic neurons from kainate excitotoxicity. *Neuron*, v.37, n.5, Mar 6, p.735-49. 2003.

Strauss, K. M., L. M. Martins, H. Plun-Favreau, F. P. Marx, S. Kautzmann, D. Berg, T. Gasser, Z. Wszolek, T. Muller, A. Bornemann, H. Wolburg, J. Downward, O. Riess, J. B. Schulz and R. Kruger. Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Hum Mol Genet*, v.14, n.15, Aug 1, p.2099-111. 2005.

Summerfield, C., C. Junque, E. Tolosa, P. Salgado-Pineda, B. Gomez-Anson, M. J. Marti, P. Pastor, B. Ramirez-Ruiz and J. Mercader. Structural brain changes in Parkinson disease with dementia: a voxel-based morphometry study. *Arch Neurol*, v.62, n.2, Feb, p.281-5. 2005.

Sun, M., J. C. Latourelle, G. F. Wooten, M. F. Lew, C. Klein, H. A. Shill, L. I. Golbe, M. H. Mark, B. A. Racette, J. S. Perlmutter, A. Parsian, M. Guttman, G. Nicholson, G. Xu, J. B. Wilk, M. H. Saint-Hilaire, A. L. Destefano, R. Prakash, S. Williamson, O. Suchowersky, N. Labelle, J. H. Growdon, C. Singer, R. L. Watts, S. Goldwurm, G. Pezzoli, K. B. Baker, P. P. Pramstaller, D. J. Burn, P. F. Chinnery, S. Sherman, P. Vieregge, I. Litvan, T. Gillis, M. E. Macdonald, R. H. Myers and J. F. Gusella. Influence of heterozygosity for parkin mutation on onset age in familial Parkinson disease: the GenePD study. *Arch Neurol*, v.63, n.6, Jun, p.826-32. 2006.

Sun, S. G. [Juvenile familial Gaucher's disease]. *Zhonghua Shen Jing Jing Shen Ke Za Zhi*, v.19, n.6, Dec, p.352-3. 1986.

Tabrizi, S. J., M. Orth, J. M. Wilkinson, J. W. Taanman, T. T. Warner, J. M. Cooper and A. H. Schapira. Expression of mutant alpha-synuclein causes increased susceptibility to dopamine toxicity. *Hum Mol Genet*, v.9, n.18, Nov 1, p.2683-9. 2000.

Taira, T., Y. Saito, T. Niki, S. M. Iguchi-Arigo, K. Takahashi and H. Ariga. DJ-1 has a role in antioxidative stress to prevent cell death. *EMBO Rep*, v.5, n.2, Feb, p.213-8. 2004.

Takahashi, H., E. Ohama, S. Suzuki, Y. Horikawa, A. Ishikawa, T. Morita, S. Tsuji and F. Ikuta. Familial juvenile parkinsonism: clinical and pathologic study in a family. *Neurology*, v.44, n.3 Pt 1, Mar, p.437-41. 1994.

Takeda, S., K. Yamazaki, T. Miyakawa and H. Arai. Parkinson's disease with involvement of the parasympathetic ganglia. *Acta Neuropathol*, v.86, n.4, p.397-8. 1993.

Tan, E. K., A. Chai, Y. Y. Teo, Y. Zhao, C. Tan, H. Shen, V. R. Chandran, M. L. Teoh, Y. Yih, R. Pavanni, M. C. Wong, K. Puvan, Y. L. Lo and E. Yap. Alpha-synuclein haplotypes implicated in risk of Parkinson's disease. *Neurology*, v.62, n.1, Jan 13, p.128-31. 2004.

Tan, E. K., H. Chung, Y. Zhao, H. Shen, V. R. Chandran, C. Tan, M. L. Teoh, Y. Yih, R. Pavanni and M. C. Wong. Genetic analysis of Nurr1 haplotypes in Parkinson's disease. *Neurosci Lett*, v.347, n.3, Aug 28, p.139-42. 2003.

Tan, E. K., T. Matsuura, S. Nagamitsu, M. Khajavi, J. Jankovic and T. Ashizawa. Polymorphism of NACP-Rep1 in Parkinson's disease: an etiologic link with essential tremor? *Neurology*, v.54, n.5, Mar 14, p.1195-8. 2000.

Tan, E. K., K. Y. Puong, S. Fook-Chong, E. Chua, H. Shen, Y. Yuen, R. Pavanni, M. C. Wong, K. Puvan and Y. Zhao. Case-control study of UCHL1 S18Y variant in Parkinson's disease. *Mov Disord*, v.21, n.10, Oct, p.1765-8. 2006.

Tan, E. K., H. Shen, L. C. Tan, M. Farrer, K. Yew, E. Chua, R. D. Jamora, K. Puvan, K. Y. Puong, Y. Zhao, R. Pavanni, M. C. Wong, Y. Yih, L. Skipper and J. J. Liu. The G2019S LRRK2 mutation is uncommon in an Asian cohort of Parkinson's disease patients. *Neurosci Lett*, v.384, n.3, Aug 26, p.327-9. 2005.

Tan, E. K., C. Tan, H. Shen, A. Chai, S. Y. Lum, M. L. Teoh, Y. Yih, M. C. Wong and Y. Zhao. Alpha synuclein promoter and risk of Parkinson's disease: microsatellite and allelic size variability. *Neurosci Lett*, v.336, n.1, Jan 9, p.70-2. 2003.

Tan, E. K., C. Tan, Y. Zhao, K. Yew, H. Shen, V. R. Chandran, M. L. Teoh, Y. Yih, R. Pavanni and M. C. Wong. Genetic analysis of DJ-1 in a cohort Parkinson's disease patients of different ethnicity. *Neurosci Lett*, v.367, n.1, Aug 26, p.109-12. 2004.

Tan, E. K., K. Yew, E. Chua, K. Puvan, H. Shen, E. Lee, K. Y. Puong, Y. Zhao, R. Pavanni, M. C. Wong, D. Jamora, D. De Silva, K. T. Moe, F. P. Woon, Y. Yuen and L. Tan. PINK1 mutations in sporadic early-onset Parkinson's disease. *Mov Disord*, v.21, n.6, Jun, p.789-93. 2006.

Tan, E. K., K. Yew, E. Chua, H. Shen, R. D. Jamora, E. Lee, K. Y. Puong, Y. Zhao, R. Pavanni, M. C. Wong, K. Puvan, Y. Yih and L. C. Tan. Analysis of PINK1 in Asian patients with familial parkinsonism. *Clin Genet*, v.68, n.5, Nov, p.468-70. 2005.

Tanner, C. M., R. Ottman, S. M. Goldman, J. Ellenberg, P. Chan, R. Mayeux and J. W. Langston. Parkinson disease in twins: an etiologic study. *Jama*, v.281, n.4, Jan 27, p.341-6. 1999.

Tassin, J., A. Durr, T. De Broucker, N. Abbas, V. Bonifati, G. De Michele, A. M. Bonnet, E. Broussolle, P. Pollak, M. Vidailhet, M. De Mari, R. Marconi, S. Medjbeur, A. Filla, G. Meco, Y. Agid and A. Brice. Chromosome 6-linked autosomal recessive early-onset Parkinsonism: linkage in European and Algerian families, extension of the clinical spectrum, and evidence of a small homozygous deletion in one family. The French Parkinson's Disease Genetics Study Group, and the European Consortium on Genetic Susceptibility in Parkinson's Disease. *Am J Hum Genet*, v.63, n.1, Jul, p.88-94. 1998.

Taymans, J. M., C. Van Den Haute and V. Baekelandt. Distribution of PINK1 and LRRK2 in rat and mouse brain. *J Neurochem*, v.98, n.3, Aug, p.951-61. 2006.

Tecott, L. H., J. L. Rubenstein, G. Paxinos, C. J. Evans, J. H. Eberwine and K. L. Valentino. Developmental expression of proenkephalin mRNA and peptides in rat striatum. *Brain Res Dev Brain Res*, v.49, n.1, Sep 1, p.75-86. 1989.

Tretiakoff, M. Contribution a L'étude de L'anatomie Pathologique de Locus Niger de Soemmerling. Université de Paris. 1919.

Trottier, Y., Y. Lutz, G. Stevanin, G. Imbert, D. Devys, G. Cancel, F. Saudou, C. Weber, G. David, L. Tora and Et Al. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature*, v.378, n.6555, Nov 23, p.403-6. 1995.

Tuite, P. J., E. A. Rogaeva, P. H. St George-Hyslop and A. E. Lang. Dopa-responsive parkinsonism phenotype of Machado-Joseph disease: confirmation of 14q CAG expansion. *Ann Neurol*, v.38, n.4, Oct, p.684-7. 1995.

Tumilowicz, J. J., G. E. Gallick, J. L. East, S. Pathak, J. J. Trentin and R. B. Arlinghaus. Presence of retrovirus in the B95-8 Epstein-Barr virus-producing cell line from different sources. *In Vitro*, v.20, n.6, Jun, p.486-92. 1984.

Turpin Jc, D. G., Brice a, Masson M, Nadaud Mc, and S. A. Boutry Jm, Tager Jm, Baumann N. Parkinsonian symptomatology in a patient with type I (adult) Gaucher's disease. In *Lipid Storage Disorders, Biological and Medical Aspects*. New York: Plenum Press, p.103-104. 1987.

Twelves, D., K. S. Perkins and C. Counsell. Systematic review of incidence studies of Parkinson's disease. *Mov Disord*, v.18, n.1, Jan, p.19-31. 2003.

Tylki-Szymanska, A., G. Millat, I. Maire and B. Czartoryska. Types I and III Gaucher disease in Poland: incidence of the most common mutations and phenotypic manifestations. *Eur J Hum Genet*, v.4, n.6, p.334-7. 1996.

Urtasun, M., A. Saenz, C. Roudaut, J. J. Poza, J. A. Urtizberea, A. M. Cobo, I. Richard, F. Garcia Bragado, F. Leturcq, J. C. Kaplan, J. F. Marti Masso, J. S. Beckmann and A. Lopez De Munain. Limb-girdle muscular dystrophy in Guipuzcoa (Basque Country, Spain). *Brain*, v.121 ( Pt 9), Sep, p.1735-47. 1998.

Uversky, V. N., J. Li and A. L. Fink. Pesticides directly accelerate the rate of alpha-synuclein fibril formation: a possible factor in Parkinson's disease. *FEBS Lett*, v.500, n.3, Jul 6, p.105-8. 2001.

Valente, E. M., P. M. Abou-Sleiman, V. Caputo, M. M. Muqit, K. Harvey, S. Gispert, Z. Ali, D. Del Turco, A. R. Bentivoglio, D. G. Healy, A. Albanese, R. Nussbaum, R. Gonzalez-Maldonado, T. Deller, S. Salvi, P. Cortelli, W. P. Gilks, D. S. Latchman, R. J. Harvey, B. Dallapiccola, G. Auburger and N. W. Wood. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science*, v.304, n.5674, May 21, p.1158-60. 2004.

Valente, E. M., A. R. Bentivoglio, P. H. Dixon, A. Ferraris, T. Ialongo, M. Frontali, A. Albanese and N. W. Wood. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36. *Am J Hum Genet*, v.68, n.4, Apr, p.895-900. 2001.

Valente, E. M., F. Brancati, A. Ferraris, E. A. Graham, M. B. Davis, M. M. Breteler, T. Gasser, V. Bonifati, A. R. Bentivoglio, G. De Michele, A. Durr, P. Cortelli, D. Wassilowsky, B. S. Harhangi, N. Rawal, V. Caputo, A. Filla, G. Meco, B. A. Oostra, A. Brice, A. Albanese, B. Dallapiccola and N. W. Wood. PARK6-linked parkinsonism occurs in several European families. *Ann Neurol*, v.51, n.1, Jan, p.14-8. 2002.

Valente, E. M., S. Salvi, T. Ialongo, R. Marongiu, A. E. Elia, V. Caputo, L. Romito, A. Albanese, B. Dallapiccola and A. R. Bentivoglio. PINK1 mutations are associated with sporadic early-onset parkinsonism. *Ann Neurol*, v.56, n.3, Sep, p.336-41. 2004.

Van Den Eeden, S. K., C. M. Tanner, A. L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch and L. M. Nelson. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am J Epidemiol*, v.157, n.11, Jun 1, p.1015-22. 2003.

Van Duijn, C. M., M. C. Dekker, V. Bonifati, R. J. Galjaard, J. J. Houwing-Duistermaat, P. J. Snijders, L. Testers, G. J. Breedveld, M. Horstink, L. A. Sandkuijl, J. C. Van Swieten, B. A. Oostra and P. Heutink. Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. *Am J Hum Genet*, v.69, n.3, Sep, p.629-34. 2001.

Vanbogaert L, F. A. Un cas de maladie de Gaucher de l'adulte avec syndrome de Raynaud, pigmentation, et rigidite du type extra-pyramidal aux membres inferieurs. *Ann Med*, v.45, p.57-70. 1939.

Vines, J. J., R. Larumbe, I. Gaminde and M. T. Artazcoz. [Incidence of idiopathic and secondary Parkinson disease in Navarre. Population-based case registry]. *Neurologia*, v.14, n.1, Jan, p.16-22. 1999.

Volles, M. J., S. J. Lee, J. C. Rochet, M. D. Shtilerman, T. T. Ding, J. C. Kessler and P. T. Lansbury, Jr. Vesicle permeabilization by protofibrillar alpha-synuclein: implications for the pathogenesis and treatment of Parkinson's disease. *Biochemistry*, v.40, n.26, Jul 3, p.7812-9. 2001.

Von Campenhausen, S., B. Bornschein, R. Wick, K. Botzel, C. Sampaio, W. Poewe, W. Oertel, U. Siebert, K. Berger and R. Dodel. Prevalence and incidence of Parkinson's disease in Europe. *Eur Neuropsychopharmacol*, v.15, n.4, Aug, p.473-90. 2005.

Wakabayashi, K., L. A. Hansen and E. Masliah. Cortical Lewy body-containing neurons are pyramidal cells: laser confocal imaging of double-immunolabeled sections with anti-ubiquitin and SMI32. *Acta Neuropathol*, v.89, n.5, p.404-8. 1995.

Wakabayashi, K., S. Hayashi, A. Kakita, M. Yamada, Y. Toyoshima, M. Yoshimoto and H. Takahashi. Accumulation of alpha-synuclein/NACP is a cytopathological feature common to Lewy body disease and multiple system atrophy. *Acta Neuropathol*, v.96, n.5, Nov, p.445-52. 1998.

Wakabayashi, K. and H. Takahashi. The intermediolateral nucleus and Clarke's column in Parkinson's disease. *Acta Neuropathol*, v.94, n.3, Sep, p.287-9. 1997a.

Wakabayashi, K. and H. Takahashi. Neuropathology of autonomic nervous system in Parkinson's disease. *Eur Neurol*, v.38 Suppl 2, p.2-7. 1997b.

Wakabayashi, K., H. Takahashi, S. Takeda, E. Ohama and F. Ikuta. Parkinson's disease: the presence of Lewy bodies in Auerbach's and Meissner's plexuses. *Acta Neuropathol*, v.76, n.3, p.217-221. 1988.

Wakabayashi, K., K. Tanji, F. Mori and H. Takahashi. The Lewy body in Parkinson's disease: molecules implicated in the formation and degradation of alpha-synuclein aggregates. *Neuropathology*, v.27, n.5, Oct, p.494-506. 2007.

Wang, J., C. Y. Zhao, Y. M. Si, Z. L. Liu, B. Chen and L. Yu. ACT and UCH-L1 polymorphisms in Parkinson's disease and age of onset. *Mov Disord*, v.17, n.4, Jul, p.767-71. 2002.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. Grant, H. Hakonarson and M. Bucan. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*, v.17, n.11, Nov, p.1665-74. 2007.

Wang, Y. S., Y. M. Shi, Z. Y. Wu, Y. X. He and B. Z. Zhang. Parkinson's disease in China. Coordinational Group of Neuroepidemiology, PLA. *Chin Med J (Engl)*, v.104, n.11, Nov, p.960-4. 1991.

Ward, C. D., R. C. Duvoisin, S. E. Ince, J. D. Nutt, R. Eldridge and D. B. Calne. Parkinson's disease in 65 pairs of twins and in a set of quadruplets. *Neurology*, v.33, n.7, Jul, p.815-24. 1983.

Ward, C. D., R. C. Duvoisin, S. E. Ince, J. D. Nutt, R. Eldridge, D. B. Calne and J. Dambrosia. Parkinson's disease in twins. *Adv Neurol*, v.40, p.341-4. 1984.

Waters, C. H. and C. A. Miller. Autosomal dominant Lewy body parkinsonism in a four-generation family. *Ann Neurol*, v.35, n.1, Jan, p.59-64. 1994.

Wellenbrock, C., K. Hedrich, N. Schafer, M. Kasten, H. Jacobs, E. Schwinger, J. Hagenah, P. P. Pramstaller, P. Vieregge and C. Klein. NR4A2 mutations are rare among European patients with familial Parkinson's disease. *Ann Neurol*, v.54, n.3, Sep, p.415. 2003.

West, A., M. Periquet, S. Lincoln, C. B. Lucking, D. Nicholl, V. Bonifati, N. Rawal, T. Gasser, E. Lohmann, J. F. Deleuze, D. Maraganore, A. Levey, N. Wood, A. Durr, J. Hardy, A. Brice and M. Farrer. Complex relationship between Parkin mutations and Parkinson disease. *Am J Med Genet*, v.114, n.5, Jul 8, p.584-91. 2002.

West, A. B., A. Zimprich, P. J. Lockhart, M. Farrer, A. Singleton, B. Holtom, S. Lincoln, A. Hofer, L. Hill, B. Muller-Myhsok, Z. K. Wszolek, J. Hardy and T. Gasser. Refinement of the PARK3 locus on chromosome 2p13 and the analysis of 14 candidate genes. *Eur J Hum Genet*, v.9, n.9, Sep, p.659-66. 2001.

Whitehead, A. S., S. Bertrand, F. Finnan, A. Butler, G. D. Smith and Y. Ben-Shlomo. Frequency of the apolipoprotein E epsilon 4 allele in a case-control study of early onset Parkinson's disease. *J Neurol Neurosurg Psychiatry*, v.61, n.4, Oct, p.347-51. 1996.

Wichmann, B. A. A. H., I.D. An efficient and portable pseudo-random number generator. *Applied Statistics* v.31, p.188-190. 1982.

Wilhelmsen, K. C., D. E. Weeks, T. G. Nygaard, C. B. Moskowitz, R. L. Rosales, D. C. Dela Paz, E. E. Sobrevega, S. Fahn and T. C. Gilliam. Genetic mapping of "Lubag" (X-linked dystonia-parkinsonism) in a Filipino kindred to the pericentromeric region of the X chromosome. *Ann Neurol*, v.29, n.2, Feb, p.124-31. 1991.

Wilkins, A., J. M. Brown and R. A. Barker. SCA2 presenting as levodopa-responsive parkinsonism in a young patient from the United Kingdom: a case report. *Mov Disord*, v.19, n.5, May, p.593-5. 2004.

Wintermeyer, P., R. Kruger, W. Kuhn, T. Muller, D. Voitalla, D. Berg, G. Becker, E. Leroy, M. Polymeropoulos, K. Berger, H. Przuntek, L. Schols, J. T. Epplen and O. Riess. Mutation analysis and association studies of the UCHL1 gene in German Parkinson's disease patients. *Neuroreport*, v.11, n.10, Jul 14, p.2079-82. 2000.

Wszolek, Z. K., B. Pfeiffer, J. R. Fulgham, J. E. Parisi, B. M. Thompson, R. J. Uitti, D. B. Calne and R. F. Pfeiffer. Western Nebraska family (family D) with autosomal dominant parkinsonism. *Neurology*, v.45, n.3 Pt 1, Mar, p.502-5. 1995.

Wszolek, Z. K., R. F. Pfeiffer, Y. Tsuboi, R. J. Uitti, R. D. McComb, A. J. Stoessl, A. J. Strongosky, A. Zimprich, B. Muller-Myhsok, M. J. Farrer, T. Gasser, D. B. Calne and D. W. Dickson. Autosomal dominant parkinsonism associated with variable synuclein and tau pathology. *Neurology*, v.62, n.9, May 11, p.1619-22. 2004.

Wu, Y. R., H. Y. Lin, C. M. Chen, K. Gwinn-Hardy, L. S. Ro, Y. C. Wang, S. H. Li, J. C. Hwang, K. Fang, H. M. Hsieh-Li, M. L. Li, L. C. Tung, M. T. Su, K. T. Lu and G. J. Lee-Chen. Genetic testing in spinocerebellar ataxia in Taiwan: expansions of trinucleotide repeats in SCA8 and SCA17 are associated with typical Parkinson's disease. *Clin Genet*, v.65, n.3, Mar, p.209-14. 2004.

Xiromerisiou, G., G. M. Hadjigeorgiou, V. Gourbali, J. Johnson, I. Papakonstantinou, A. Papadimitriou and A. B. Singleton. Screening for SNCA and LRRK2 mutations in Greek sporadic and autosomal dominant Parkinson's disease: identification of two novel LRRK2 variants. *Eur J Neurol*, v.14, n.1, Jan, p.7-11. 2007.

Xu, P. Y., R. Liang, J. Jankovic, C. Hunter, Y. X. Zeng, T. Ashizawa, D. Lai and W. D. Le. Association of homozygous 7048G7049 variant in the intron six of Nurr1 gene with Parkinson's disease. *Neurology*, v.58, n.6, Mar 26, p.881-4. 2002.

Yamamura, Y., K. Arihiro, T. Kohriyama and S. Nakamura. [Early-onset parkinsonism with diurnal fluctuation--clinical and pathological studies]. *Rinsho Shinkeigaku*, v.33, n.5, May, p.491-6. 1993.

Yamamura, Y., I. Sobue, K. Ando, M. Iida and T. Yanagi. Paralysis agitans of early onset with marked diurnal fluctuation of symptoms. *Neurology*, v.23, n.3, Mar, p.239-44. 1973.

Zarepari, S., R. Camicioli, G. Sexton, T. Bird, P. Swanson, J. Kaye, J. Nutt and H. Payami. Age at onset of Parkinson disease and apolipoprotein E genotypes. *Am J Med Genet*, v.107, n.2, Jan 15, p.156-61. 2002.

Zarranz, J. J., J. Alegre, J. C. Gomez-Esteban, E. Lezcano, R. Ros, I. Ampuero, L. Vidal, J. Hoenicka, O. Rodriguez, B. Atares, V. Llorens, E. Gomez Tortosa, T. Del Ser, D. G. Munoz and J. G. De Yebenes. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol*, v.55, n.2, Feb, p.164-73. 2004.

Zeggini, E., M. N. Weedon, C. M. Lindgren, T. M. Frayling, K. S. Elliott, H. Lango, N. J. Timpson, J. R. Perry, N. W. Rayner, R. M. Freathy, J. C. Barrett, B. Shields, A. P. Morris, S. Ellard, C. J. Groves, L. W. Harries, J. L. Marchini, K. R. Owen, B. Knight, L. R. Cardon, M. Walker, G. A. Hitman, A. D. Morris, A. S. Doney, M. I. McCarthy and A. T. Hattersley. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, v.316, n.5829, Jun 1, p.1336-41. 2007.

Zhang, J., N. Hattori, N. Giladi and Y. Mizuno. Failure to find mutations in ubiquitin carboxy-terminal hydrolase L1 gene in familial Parkinson's disease. *Parkinsonism Relat Disord*, v.6, n.4, Oct 1, p.199-200. 2000.

Zhang, J., Y. Song, H. Chen and D. Fan. The tau gene haplotype h1 confers a susceptibility to Parkinson's disease. *Eur Neurol*, v.53, n.1, p.15-21. 2005.

Zhang, L., M. Shimoji, B. Thomas, D. J. Moore, S. W. Yu, N. I. Marupudi, R. Torp, I. A. Torgner, O. P. Ottersen, T. M. Dawson and V. L. Dawson. Mitochondrial localization of the Parkinson's disease related protein DJ-1: implications for pathogenesis. *Hum Mol Genet*, v.14, n.14, Jul 15, p.2063-73. 2005.

Zhang, Y., J. Gao, K. K. Chung, H. Huang, V. L. Dawson and T. M. Dawson. Parkin functions as an E2-dependent ubiquitin- protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. *Proc Natl Acad Sci U S A*, v.97, n.24, Nov 21, p.13354-9. 2000.

Zheng, K., B. Heydari and D. K. Simon. A common NURR1 polymorphism associated with Parkinson disease and diffuse Lewy body disease. *Arch Neurol*, v.60, n.5, May, p.722-5. 2003.

Zhu, X., A. Babar, S. L. Siedlak, Q. Yang, G. Ito, T. Iwatsubo, M. A. Smith, G. Perry and S. G. Chen. LRRK2 in Parkinson's disease and dementia with Lewy bodies. *Mol Neurodegener*, v.1, p.17. 2006.

Zhukareva, V., D. Mann, S. Pickering-Brown, K. Uryu, T. Shuck, K. Shah, M. Grossman, B. L. Miller, C. M. Hulette, S. C. Feinstein, J. Q. Trojanowski and V. M. Lee. Sporadic Pick's disease: a tauopathy characterized by a spectrum of pathological tau isoforms in gray and white matter. *Ann Neurol*, v.51, n.6, Jun, p.730-9. 2002.

Zimprich, A., F. Asmus, P. Leitner, M. Castro, B. Bereznai, N. Homann, E. Ott, A. W. Rutgers, G. Wieditz, C. Trenkwalder and T. Gasser. Point mutations in exon 1 of the NR4A2 gene are not a major cause of familial Parkinson's disease. *Neurogenetics*, v.4, n.4, Aug, p.219-20. 2003.

Zimprich, A., S. Biskup, P. Leitner, P. Lichtner, M. Farrer, S. Lincoln, J. Kachergus, M. Hulihan, R. J. Uitti, D. B. Calne, A. J. Stoessl, R. F. Pfeiffer, N. Patenge, I. C. Carbajal, P. Vieregge, F. Asmus, B. Muller-Myhsok, D. W. Dickson, T. Meitinger, T. M. Strom, Z.

K. Wszolek and T. Gasser. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*, v.44, n.4, Nov 18, p.601-7. 2004.

Zimprich, A., B. Muller-Myhsok, M. Farrer, P. Leitner, M. Sharma, M. Hulihan, P. Lockhart, A. Strongosky, J. Kachergus, D. B. Calne, J. Stoessl, R. J. Uitti, R. F. Pfeiffer, C. Trenkwalder, N. Homann, E. Ott, K. Wenzel, F. Asmus, J. Hardy, Z. Wszolek and T. Gasser. The PARK8 locus in autosomal dominant parkinsonism: confirmation of linkage and further delineation of the disease-containing interval. *Am J Hum Genet*, v.74, n.1, Jan, p.11-9. 2004.

Zivelin, A., F. Bauduer, L. Ducout, H. Peretz, N. Rosenberg, R. Yatuva and U. Seligsohn. Factor XI deficiency in French Basques is caused predominantly by an ancestral Cys38Arg mutation in the factor XI gene. *Blood*, v.99, n.7, Apr 1, p.2448-54. 2002.