



HAL
open science

SNCA locus duplication carriers: from genetics to Parkinson's disease phenotypes

Eugénie Mutez, Frederic Lepretre, Emilie Le Rhun, Lydie Larvor, Aurélie Dufлот, Vincent Mouroux, Jean-Pierre Kerckaert, Martin Figeac, Kathy Dujardin, Alain Destée, et al.

► **To cite this version:**

Eugénie Mutez, Frederic Lepretre, Emilie Le Rhun, Lydie Larvor, Aurélie Dufлот, et al.. SNCA locus duplication carriers: from genetics to Parkinson's disease phenotypes. *Human Mutation*, 2011, 32 (4), 10.1002/humu.21459 . hal-00613918

HAL Id: hal-00613918

<https://hal.science/hal-00613918>

Submitted on 8 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



SNCA locus duplication carriers: from genetics to Parkinson's disease phenotypes

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0524.R1
Wiley - Manuscript type:	Mutation in Brief
Date Submitted by the Author:	20-Dec-2010
Complete List of Authors:	Mutez, Eugénie; CHRU, Neurology lepretre, frederic; UMR 837 INSERM, Functional Genomic Platform Le Rhun, Emilie; CHU, Neurology Larvor, Lydie; UMR 837 INSERM, IRCL Dufлот, Aurélie; UMR 837 INSERM, IRCL Mouroux, Vincent; UMR 837 INSERM, IRCL kerckaert, jean-pierre; UMR 837 INSERM, Functional Genomic Platform figeac, martin; UMR 837 INSERM, Functional Genomic Platform Dujardin, Kathy; CHU, Neurology Destée, Alain; CHU, Neurology Chartier-Harlin, Marie-Christine; UMR 837 INSERM, IRCL
Key Words:	Parkinson, alpha-synuclein, SNCA, duplication, overexpression

SCHOLARONE™
Manuscripts

SNCA locus duplication carriers: from genetics to Parkinson's disease phenotypes



Eugénie Mutez,^{1,2,3}‡ Frédéric Leprêtre,^{1,2,4}‡ Emilie Le Rhun,³‡ Lydie Larvor,^{1,2}‡ Aurélie Duflot,^{1,2} Vincent Mouroux,^{1,2} Jean-Pierre Kerckaert,^{2,4} Martin Figeac,^{2,4} Kathy Dujardin,^{2,3} Alain Destée,^{1,2,3} and Marie-Christine Chartier-Harlin^{1,2*}

¹UMR 837 INSERM, Team 6, JParc, IRCL, Lille, France; ²Univ Lille Nord de France, Lille, France; ³Movement Disorders Unit, Lille University Hospital, Lille, France; ⁴Functional Genomic Platform, UDSL, IRCL, Lille, France.
‡ These authors contributed equally to the manuscript.

*Correspondence to Dr Marie-Christine Chartier-Harlin, UMR 837 INSERM, Team 6, JParc, IRCL, Place de Verdun, 59045 Lille Cedex, France. E-mail: marie-christine.chartier-harlin@inserm.fr
Contract grant sponsor: Univ Lille Nord de France, INSERM, Lille University Hospital and French Ministry of Research;
Contract grant number: PHRC PARKFANORD (2005/1914).

Communicated by <Please don't enter>

ABSTRACT: Genomic multiplication of the alpha-synuclein gene (*SNCA*) locus is one cause of familial Parkinson's disease (PD). We performed a detailed genomic, *SNCA* expression level, clinical, neuropsychological and functional imaging analysis of a parkinsonian kindred with a known duplication of the *SNCA* locus. We demonstrated that the duplication spanned 4.928 Mb (encompassing 31 known and putative genes) and was the largest to have been described at this locus. The presence of several repetitive long interspersed nuclear elements (LINEs) flanking the potential break area suggested that the duplication resulted from a genomic recombination between LINEs. We sequenced the break junction and confirmed the involvement of L1PA2 and L1PA4 in a non-allelic, homologous recombination. An analysis of mRNA levels in immortalized lymphoblastoid cells and peripheral blood mononuclear cells showed *SNCA* overexpression in subjects with the duplication, as well as overexpression of 13 other genes highlighting the usefulness of such cell models to study this duplication. Interestingly, abnormal tracer uptake in DaTSCAN[®] imaging correlated with the severity of the clinical symptoms. Our detailed genomic analysis and clinical exploration enabled us to specify the genotype-phenotype relationship, identify a case of presymptomatic PD and gain insight into the role of LINEs in *SNCA* locus duplication. ©2010 Wiley-Liss, Inc.

KEY WORDS: Parkinson; alpha-synuclein; *SNCA*; duplication; overexpression

Received <date>; accepted revised manuscript <date>.

© 2010 WILEY-LISS, INC.

2 <Mutez et al.>

INTRODUCTION

Genetic factors have been shown to contribute to the etiology of Parkinson's disease (PD). Missense mutations of the *SNCA* locus (MIM# 163890) were the first to be identified as a cause of familial PD (Polymeropoulos et al., 1997). More recently, multiplications of the *SNCA* locus were reported in 11 families (and 4 isolated patients) with PD (Lesage & Brice, 2009). This new type of mutation has shed light on the importance of the level of expression of the wild-type gene and suggests that overexpression of alpha-synuclein is enough to cause PD. Our comparison of the *SNCA*-duplicated patients in family P59 with the Iowa kindred carrying a *SNCA* triplication prompted us to suggest that the severity of PD correlates with the gene dosage (Singleton et al., 2003; Chartier-Harlin et al., 2004; Ross et al., 2008; Ibáñez et al., 2009). However, the rarity of this multiplication phenomenon makes it difficult to know whether the phenotype is due to *SNCA* alone or to other genes involved in the same genomic rearrangement. For example, the cognitive dysfunction seen in some subjects might be due in part to the presence of *MMRN1* (MIM# 601456) in the multiplied locus (Nishioka et al., 2006). This gene is also present in the duplicated region of the P59 kindred also reported as FPD-131. This duplication was evaluated as the longest described to date and was expected to encompass 33 to 34 genes (Ibáñez et al., 2009). Moreover, it has been suggested that non-allelic, homologous recombination (NAHR) causes the multiplication of this locus (with mediation by low copy repeats (LCRs)) (Ross et al., 2008). However, characterization of the exact breakpoint sequence will be required to confirm or refute this hypothesis.

Hence, in the present study, we sought to (i) expand and deepen our description of the clinical phenotype of all the members of the P59 kindred, (ii) determine the level of expression of *SNCA* in both lymphoblastoid cell lines and peripheral blood mononuclear cells (PBMCs), (iii) identify the exact number of genes present at the genomic *SNCA* locus, (iv) explore the variations of expression of the duplicated genes in PBMCs and (v) define the exact breakpoint sequence and the mechanism underlying this locus multiplication.

MATERIALS AND METHODS

Patients

P59 was the first *SNCA* duplication kindred to be identified within a set of 9 families in the Parkfanord study (Chartier-Harlin et al., 2004). Further screening of 149 families did not reveal any other *SNCA* duplication carriers and confirmed the rarity of this type of event. Moreover, no deleterious punctual mutation in *SNCA*, *LRRK2* (MIM# 609007) or *GBA* (MIM# 606463) genes was found in the affected members of this parkinsonian kindred with autosomal dominant inheritance (data not shown). In order to describe the clinical phenotype in detail, after obtaining IRB approval and informed consent, the ten members of family P59 were examined in a movement disorder clinic by physicians trained in PD and blinded to the genetic status of the subjects. The diagnosis of PD was based on the United Kingdom Parkinson's Disease Society Brain Bank criteria. Extensive neuropsychological investigations (including tests known to be particularly sensitive to cognitive dysfunction in PD) were performed on all participants. Global cognitive efficiency was assessed using the Mattis dementia rating scale (MDRS). Test procedures evaluating working and episodic memory and executive function have been described elsewhere (Dujardin et al., 2000).

DaTSCAN® single-photon emission computed tomography

Single-photon emission computed tomography (SPECT) with the radioligand (¹²³I) FP-CIT (DaTSCAN®) was performed in all subjects. Interpretation was based on semiquantitative analysis of uptake in the caudate nucleus and the putamen (Binding Index calculated by using the occipital cortex as the radioactivity uptake reference).

SNCA gene dosage

The *SNCA* gene dosage was determined in a semi-quantitative PCR with genomic DNA (ABI Prism 7700 sequence detector, Applied Biosystems, Foster City, USA). Exons 1 and 7 of *SNCA* were amplified for each member of the P59 kindred. The ribosomal protein S18 (*RPS18*) gene (MIM# 180473) was amplified as an endogenous reference. DNA sample ND00139 from the Iowa kindred (<http://ccr.coriell.org/>) was used as a positive control of triplication of the *SNCA* locus (Singleton et al., 2003). All the primer sequences used for this study are given in Supp. Table S1.

Fluorescent *in situ* hybridization analysis

We derived Epstein-Barr-virus (EBV)-transformed lymphoblastoid cell lines for each subject and we re-sampled PBMCs from alive subjects. Fluorescent *in situ* hybridization (FISH) was performed using bacterial artificial chromosome clones from the RP11 library according to standard procedures. These clones include RP11-115D19 (encompassing a part of the *SNCA* gene), RP11-77E12 (encompassing parts of the *SNCA* and *MMRN1* genes) and RP11-347P22 (encompassing the *FAM190A* gene).

Determination of the size of the duplication using comparative genomic hybridization (CGH) arrays

Genomic DNA from 2 subjects was labeled with Cy3 and Cy5-dCTP, respectively, then 244k CGH arrays (Agilent, Santa Clara, CA, USA) were scanned according to the manufacturer's instructions. The acquired images were analyzed using Agilent's CGH Analytics software. Spot intensities were corrected for the background and only spots with both Cy5 and Cy3 signal intensities at least 1.7 times greater than the local background were selected. Data were normalized against the median of Cy5/Cy3 ratios from all spots. At least 3 consecutive probes with an abnormal ratio (arbitrarily set between -0.3 and 0.3) indicated a genomic multiplication or deletion event.

Sequence analysis of the duplication junction

We used a PCR walking strategy to pinpoint the exact breakpoints of the duplicated segment. As reported previously (Ross et al., 2008), rearrangements may occur within DNA repeat structures such as *Alu* repeat subtypes or long interspersed nuclear elements (LINEs). Sequence and genomic information on the locus was obtained from the UCSC Human March 2006 Assembly database (www.genome.ucsc.edu). The centromeric and telomeric flanking limits observed in our CGH analysis included two LINE-1 elements. These two LINEs (L1PA2 at 5' and L1PA4 at 3') were selected for further analysis. Primers were designed to amplify the theoretical breakpoint in these regions. The experiment was repeated until a PCR fragment of about 500bp was obtained for sequencing. Both strands were sequenced using the ABI BigDye Terminator Kit v. 3.1 (Applied Biosystems).

SNCA relative expression analysis

RNA from immortalized leukocytes was extracted with the RNeasy® mini kit and was reverse-transcribed using the QuantiTect® Kit (Qiagen, Courtaboeuf, France). Next, the *SNCA* cDNA was PCR-amplified. The method enables the estimation of the relative expression of *SNCA* compared to three reference genes (*RPS18*, *GAPDH* MIM# 138400 and β -actin *ACTB* MIM# 182790) using SYBR green PCR master mix on an 7500 Real-Time PCR System (Applied Biosystems) following the manufacturer's instructions. Standard curves of each amplified gene were created to calculate the PCR efficiency. The relative expression levels were calculated method using the comparative C_T method (Schmittgen & Livak, 2008). All PCR reactions were performed at least in duplicate. We compared *SNCA* gene expression in immortalized cell lines derived from (i) Iowa kindred member carrying a *SNCA* triplication, (ii) two P59 subjects with *SNCA* duplication and (iii) seven controls (three P59 subjects without *SNCA* duplication and four unrelated controls).

Next, we conducted a second *SNCA* expression analysis in PBMCs from subject III4 in duplicate and 15 healthy unrelated controls with the same procedure using *RPS18* as reference gene.

Microarray procedure

To define the expression of the genes present within the duplication, we performed Agilent one-color whole human genome 44K microarrays according to manufacturer's instructions. The experimental design included the subject III4 in triplicate and 20 unrelated controls free of any personal or familial neurological disorders (mean age 58.9 ± 13.4) and the microarray procedure and normalization was described previously (Mutez et al., 2010). Data analyses were performed with GeneSpring GX 7.3.1 software from Agilent. A Volcano plot was used to identify differentially expressed genes by comparing the *SNCA* duplication carrier in triplicate to the controls, using a ≥ 1.2 -fold change and $p < 0.05$ as the significance threshold (Welch's t-test). Furthermore, we compared our results with the significantly deregulated genes ($p < 0.05$ t-test) from (i) an Affymetrix transcriptomic analysis of PBMCs between 50 PD patients and 22 controls from the GEO dataset GDS2519 (www.ncbi.nlm.nih.gov/sites/gds) (Scherzer et al., 2007) and (ii) an Affymetrix transcriptomic analysis of substantia nigra between 16 PD patients and 9 controls from the GEO dataset GDS2821 (Lesnick et al., 2007).

RESULTS

Subjects (clinical evaluation and DaTSCAN results)

4 <Mutez et al.>

Family P59's pedigree is presented in Fig. 1A. The study subjects included 4 PD patients (II1, II2, III1 and III2) and 6 asymptomatic or paucisymptomatic individuals (III3, III4, III5, III6, III7 and III8). Subject II1 had deceased before the study and no DNA sample was available - only historical data were noted. His age at PD onset was 39 and the disease duration was 23 years. Subject II2 presented a classical PD phenotype with rest tremor, bradykinesia, rigidity and asymmetrical onset at the age of 65 and a disease duration of 19 years. Subject III1 began his disease at 48 years old with bradykinesia. He developed a classical PD with Levodopa response. At death, the disease duration was 9 years. Subject III2 was diagnosed with PD at age of 40 and she did not present any sign of dementia. At time of the last evaluation, she was bedridden and met the criteria for PD dementia (Emre et al., 2007) with global efficiency degradation, working memory impairment, episodic memory disorder and dysexecutive syndrome; she deceased a few months later. Her disease duration was 18 years. Subject III3 presented a slight postural tremor and a dysexecutive syndrome with a MDRS score of 131/144. Subject III4 was 52 at evaluation and presented a rigidity of the left arm and also a global cognitive efficiency degradation (MDRS score 130/144), a working memory impairment and a dysexecutive syndrome. Subject III5 was 58 and presented a slight bradykinesia without cognitive complaint and a MDRS score of 136/144. Subject III6 had presented mystic and megalomaniac delirium at the age of 37. He had been treated with neuroleptics for 10 years and the symptoms had never recurred. He was 47 at time of evaluation and showed a postural tremor and bradykinesia related to the neuroleptics' use. There was a cognitive complaint with working memory impairment, episodic memory disorder and dysexecutive syndrome. The MDRS score was 136/144. No parkinsonian clinical or neuropsychological abnormality was observed for subjects III7 and III8 (MDRS scores of 141 and 144/144, respectively). Apart from osteoarthritis, all these subjects did not present another debilitating condition, for example cardiovascular, hematologic disorders or cancer.

DaTSCAN SPECTs were normal for subjects III3, III5, III6, III7 and III8. Conversely, there were a severe defect in putamen and pallidum with right predominance for subject III2 and a hypo-uptake in right putamen for subject III4. DaTSCAN images and semi-quantitative analyses for subjects III2, III3 and III4 are presented in Fig. 1B and 1C.

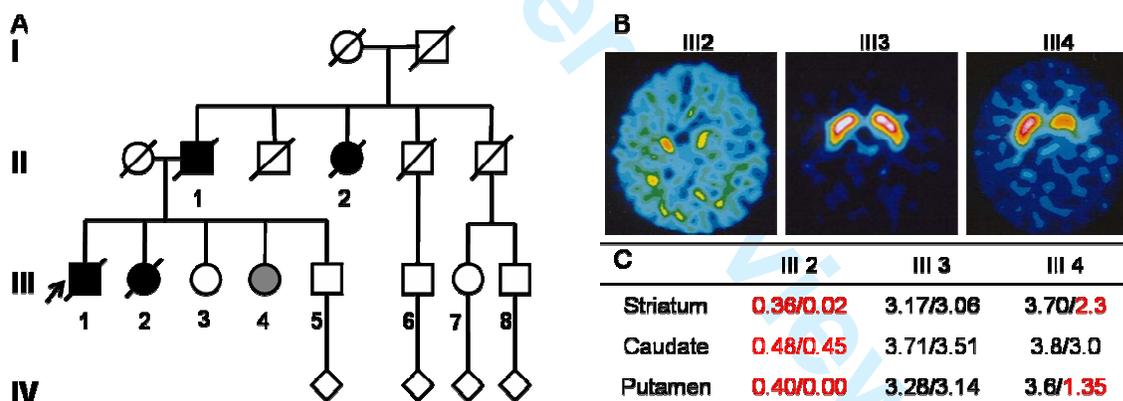


Figure 1. Presentation of the kindred's pedigree, with the DaTSCAN results. (A) Pedigree of the P59 family. Subject III1 was the probant (arrow). Subject III4 was colored in gray, since she carried the *SNCA* duplication at a prodromal stage of the disease. (B) DaTSCAN images for subjects III2, III3 and III4. (C) Semi-quantitative analysis of radioligand uptake in the brain, with details of the Binding Index (left/right values). Abnormal values are given in red type.

SNCA gene dosage and FISH

By using semiquantitative PCR to determine gene dosage, we found 4 subjects (II2, III1, III2 and III4) with a heterozygous *SNCA* duplication of exons 1 and 7 (Fig. 2A). To provide independent confirmation, FISH was performed in lymphoblastoid cell lines from the different family members. Again, a *SNCA* locus duplication was found in subjects II2, III1, III2 and III4 with 3 FISH probe signals (Fig. 2B). Moreover to evaluate if mosaicism could be present, we scored the FISH probe signals in a minimum of 100 interphase nuclei from non immortalized lymphocytes from subject III4 and controls (III3, III5, III6, III7 and III8). The duplication signals were present only in subject III4 in more than 80% of the interphase cells.

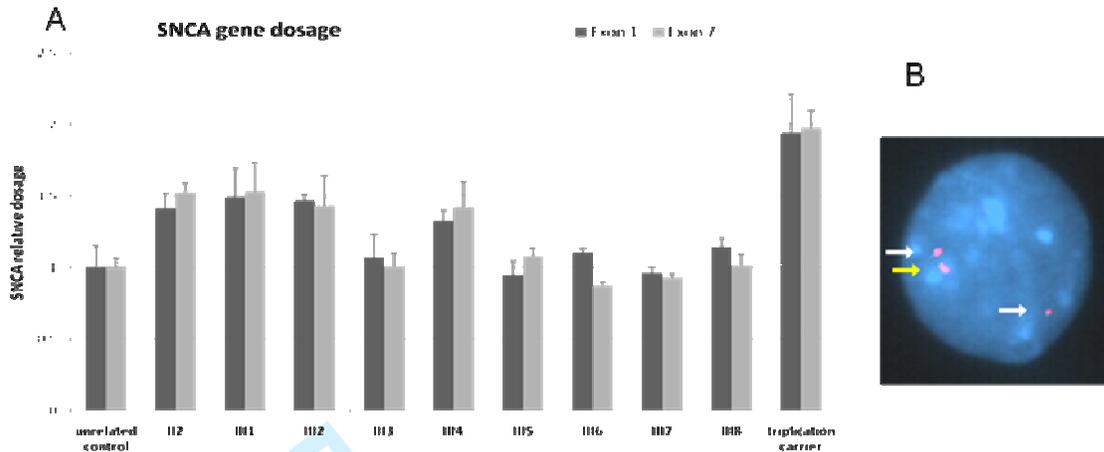


Figure 2. Identification of SNCA duplication by gene dosage and FISH (A) SNCA gene dosage assay for exons 1 and 7. The dosage was normalized to the *RPS18* gene and compared with a triplication carrier from the Iowa kindred. (B) Identification of 3 FISH signals (represented by arrows) in interphase nucleus showing the duplication.

Identification of the duplicated genomic segment

High-density CGH array analysis brought a better definition than the microsatellite markers previously used (Ibáñez et al., 2009) and showed that the duplication spanned 4.928 Mb and encompassed 31 known and putative genes (Fig. 3A, Table 1) instead of the expected 33 to 34 genes. The location of the probes at the extremity of the duplication indicated that the centromeric end of the duplication mapped to a 955 bp region between exons 10 and 11 of the *MAPK10* gene (MIM# 602897), whereas the telomeric end occurred within an 11188 kb region into the putative *FAM190A* gene (*KIAA1680*).

Table 1: List of the 31 known and putative genes present in the duplicated locus.

Symbol	Gene Name	RefSeq	Agilent Probe	FC1	FC2	FC3	Function/Associated disease
<i>ABCG2</i>	ATP-binding cassette, sub-family G, member 2	NM_004827	A_23_P18713	NS	NS	NS	Chemoresistance Colon, breast cancer
<i>AFF1</i>	AF4/FMR2 family, member 1	NM_005935	A_24_P414332	1.44	NS	NS	Acute leukemia
<i>C4ORF36</i>	chromosome 4 open reading frame 36	NM_144645	A_23_P377339	A		A	
<i>DMP1</i>	dentin matrix acidic phosphoprotein 1	NM_004407	A_23_P133153	NS	A	A	Dentinogenesis Ricket, osteomalacia
<i>DSPP</i>	dentin sialophosphoprotein	NM_014208	A_23_P373541	A	NS	NS	Dentinogenesis
<i>FAM13A</i>	family with sequence similarity 13, member A	NM_014883	A_23_P370651 A_23_P7132	2.24 A	0.85	1.51	Chronic obstructive pulmonary disease
<i>FAM13AOS</i>	FAM13A opposite strand	NR_002806	A_24_P614362	1.63		A	
<i>FAM190A</i>	family with sequence similarity 190, member A	NM_001145065	A_32_P306678	A		A	Attention deficit hyperactivity disorder
<i>GPRIN3</i>	GPRIN family member 3	NM_198281	A_24_P76649	1.36		0.65	Parkinson's disease GWAS
<i>HERC3</i>	hect domain and RLD 3	NM_014606	A_23_P121586	A	NS	1.41	Ubiquitin-ligase
<i>HERC5</i>	hect domain and RLD 5	NM_016323	A_23_P110196	NS	NS	1.40	Ubiquitin-ligase Cell cycle
<i>HERC6</i>	hect domain and RLD 6	NM_017912	A_23_P250353	1.38	NS	0.68	Ubiquitin-ligase
<i>HSD17B11</i>	hydroxysteroid (17-beta) dehydrogenase 11	NM_016245	A_23_P21644 A_23_P408271	1.65 1.58	NS	NS	Short-chain alcohol dehydrogenases
<i>HSD17B13</i>	hydroxysteroid (17-beta) dehydrogenase 13	NM_178135	A_24_P362969	A			Short-chain alcohol dehydrogenases
<i>IBSP</i>	integrin-binding sialoprotein	NM_004967	A_23_P144549	A	NS	A	Skeletal mineralization
<i>KLHL8</i>	kelch-like 8	NM_020803	A_23_P328729 A_24_P54808	1.53 1.44		1.38 1.25	Ubiquitin-ligase
<i>MAPK10</i>	mitogen-activated protein	NM_002753	A_23_P45025	A	NS	NS	Epilepsia (Lennox-Gastaut)

6 <Mutez et al.>

	kinase 10						
<i>MEPE</i>	matrix extracellular phosphoglycoprotein	NM_020203	A_23_P155931	A	A	A	Bone mineralization Osteomalacia
<i>MMRN1</i>	multimerin 1	NM_007351	A_23_P18539	2.98	NS	A	Factor V-binding protein Hemostasis Parkinson's disease GWAS
<i>NAPIL5</i>	nucleosome assembly protein 1-like 5	NM_153757	A_23_P144369	2.46		0.54	Hepatocellular carcinoma
			A_24_P166073	1.76			
<i>NUDT9</i>	nudix type motif 9	NM_024047	A_23_P94860	1.57	NS	1.45	ADP-ribose pyrophosphatase
<i>PIGY</i>	phosphatidylinositol glycan anchor biosynthesis, class Y	NM_032906	A_24_P140391	1.63		NS	Glycosylphosphatidylinositol biosynthesis
			A_24_P120380	1.60			
<i>PKD2</i>	polycystic kidney disease 2	NM_000297	A_24_P106112	1.63	0.77	NS	Ca ²⁺ cation channel Polycystic kidney disease 2
			A_23_P167324	NS			
<i>PPMIK</i>	protein phosphatase 1K	NM_152542	A_24_P214598	NS		NS	Mitochondrial permeability transition pore
<i>PTPN13</i>	protein tyrosine phosphatase, non-receptor type 13	NM_080683	A_23_P18493	1.55	A	NS	Carcinomas
			A_24_P586264	1.26			
<i>SLC10A6</i>	solute carrier family 10, member 6	NM_197965					Anion transporter
<i>SNCA</i>	synuclein, alpha	NM_000345	A_23_P29939	2.48	0.70	0.64	Parkinson's disease
			A_32_P109653	3.13		0.57	
<i>SPARCL1</i>	SPARC-like 1 (hevin)	NM_004684	A_23_P113351	A	NS	NS	Astroglial function Leukocyte extravasation
<i>SPPI</i>	secreted phosphoprotein 1	NM_001040058	A_23_P7313	NS	NS	1.42	Osteoblast maturation Elastic fibers
<i>TIGD2</i>	tigger transposable element derived 2	NM_145715	A_32_P20523	NS		NS	DNA transposon
<i>TMSL3</i>	thymosin-like 3	NM_183049	A_24_P143189	NS	NS	NS	Actin cytoskeleton organization

The genes were annotated with their symbol, their description and RefSeq accession number. FC1 corresponds to the fold change (expression ratio) between subject III4 and the 20 controls. The fold change value was indicated for the significantly deregulated genes with the corresponding Agilent probe. FC2 corresponds to the fold change of the significantly deregulated genes between PD patients' PBMCs and controls obtained from the GEO dataset GDS2519; FC3 corresponds to the fold change of the significantly deregulated genes between PD patients' substantia nigra and controls obtained from the GEO dataset GDS2821. NS = no significant difference between the PD patients and controls and A = absent flag in the microarray procedure. The last column corresponds to the functions and associated diseases linked to the genes (<http://www.ncbi.nlm.nih.gov/omim> or <http://www.ncbi.nlm.nih.gov/gene>).

Sequence analysis of the duplication junction

By using a PCR walking strategy, we sequenced PCR amplicons to produce the breakpoint junction between the L1PA4 and L1PA2 elements (Fig. 3B, 3C and 3D). The junction of these two elements results in the deletion of 349 bases (301 from L1PA4 and 48 from L1PA2). The analyses of the L1PA4 and L1PA2 sequences with the L1Base (Penzoker et al., 2005), indicated that these 2 LINES sequences had inactive ORF. We identified 95% sequence similarities for 751 pb between L1PA4 and L1PA2 using BLAST program (<http://blast.ncbi.nlm.nih.gov/>), suggesting that the duplication event was generated by NAHR between LCRs (Inoue & Lupski, 2002).

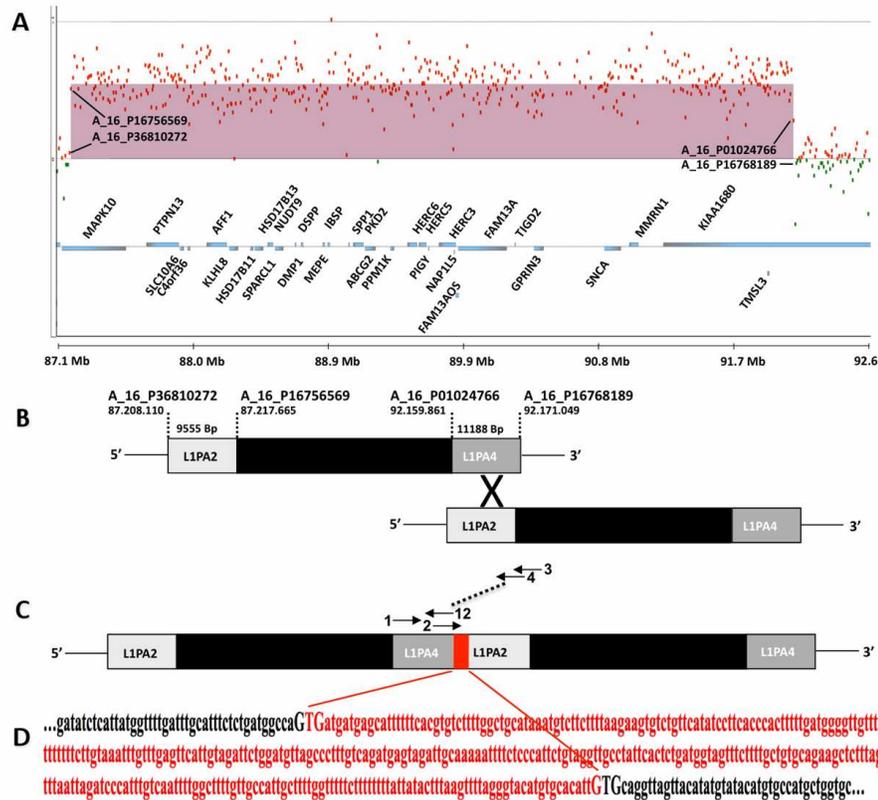


Figure 3. Identification of the duplicated segment (CGH) and the breakpoint junction. (A) CGH analysis between subjects III4 and III5, showing the duplication highlighted in pink (at locus 4q21.3-q22.1) encompassing 31 genes. (B) Agilent CGH probes enclosing L1PA2 and L1PA4. These probes are also represented in A. (C) Graphical representation of the hypothetical rearrangement and duplication, showing the location of each of the 12 probes used in PCR and the sequence analysis. The deleted segment is represented in red. (D) Breakpoint junction sequence for the L1PA4 and L1PA2 elements (in red, the 349 deleted bases). The bases involved at the breakpoint junction are capitalized.

Level of expression of the duplicated genes

SNCA level expression in lymphoblastoid cell lines from controls (2 copies), duplication carriers (3 copies) and triplication carrier (4 copies) revealed an expression gradient which correlated with the number of *SNCA* copies using 3 reference genes (Fig. 4A). Since the expression phenotype of lymphoblastoid cells in culture may change with the EBV transformation (Lee et al., 2010), we decided to confirm these results using fresh leukocytes. We were able to obtain PBMCs from subject III4 since subject III2 had deceased meanwhile. Again, we observed an overexpression of *SNCA* of 2.38 fold (± 0.41) in the III4 PBMCs compared to 15 healthy controls (Fig. 4B). To extend these results to the other duplicated genes, we used a transcriptomic approach with Agilent microarrays. Among the duplicated interval, 14 genes were significantly deregulated between the 3 RNA samples from subject III4 and 20 unrelated controls (Table 1). Again, we observed a *SNCA* overexpression in PBMCs from subject III4. Of interest, all the significantly deregulated genes present in the duplication showed an overexpression. Moreover, we compared our results with the literature data from GEO datasets. We observed a significantly deregulation for 3 genes belonging to the duplicated segment in PBMCs from PD patients (GDS2519) and for 10 genes in the substantia nigra from sporadic PD patients (GDS2821) (Table 1).

8 <Mutez et al.>

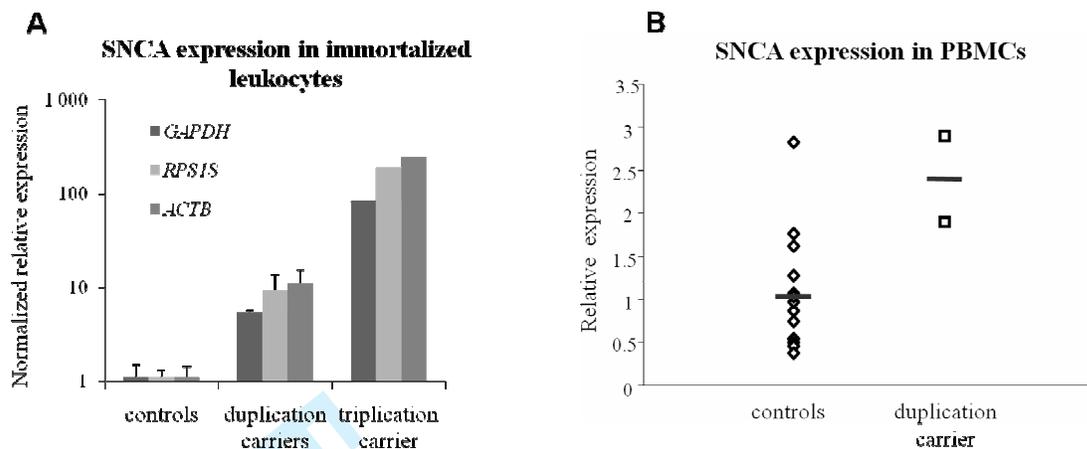


Figure 4. Expression of *SNCA* in lymphoblastoid cell lines and PBMCs. (A) *SNCA* mRNA expression levels in immortalized leukocytes from subjects with *SNCA* locus multiplications. The duplicated group corresponds to lymphoblastoid cell lines from subjects III2 and III4. The triplicated cell line was obtained from the Iowa kindred. Expression levels were normalized against healthy controls. The 3 reference genes were *GAPDH*, *RPS18* and *ACTB*. (B) *SNCA* mRNA expression levels in PBMCs of 15 healthy controls and the III4 subject in duplicate. The bars indicated the average values, arbitrarily set at 1 for the controls. The reference gene was *RPS18*.

DISCUSSION

The French P59 kindred provided us with the opportunity to describe the relationship between the genotype and the clinical phenotype in parkinsonian conditions caused by largest known duplication of the *SNCA* locus. The clinical features were similar to those of classic PD, with mid-life disease onset and late-onset dementia. In addition, we identified a subject (III4) at a paucisymptomatic stage of the disease; the subject carried the *SNCA* locus duplication and displayed an asymmetric dopaminergic defect in DaTSCAN SPECT images. Hypo-uptake has not been observed in the other asymptomatic *SNCA* duplication carriers (Ahn et al., 2008). The disease being age-dependant, this difference might be explained by their younger age (from 44 to 51). Our data confirm that DaTSCAN can demonstrate a striatal dopaminergic deficiency even in a presymptomatic stage of the disease and suggests that subject III4 is at a more advanced stage of the disease and might soon develop the entire PD phenotype. However, the interval between SPECT abnormalities and PD development remain unknown.

In the present study, we showed that the expression levels in immortalized cells from two affected members were significantly elevated. The overexpression correlated with the number of multiplied *SNCA* alleles and the clinical phenotype, since *SNCA* expression was greater in the triplicated cell line than in duplicated cell lines. Furthermore, we confirmed its overexpression with a fold change of 2.38 in fresh leukocytes from the P59 carrier compared to controls. Given that the functional effect of duplication cannot often be confirmed by examination of the substantia nigra from deceased patients, we believe that analysis of *SNCA* expression levels of lymphoblastoid cell lines or PBMCs is a valuable investigational technique. These findings demonstrate a biological dosage effect of *SNCA* locus multiplication in human cells. Our data further emphasize the importance of *SNCA* expression levels on the development of PD. *SNCA* overexpression was also recently implicated in the PD pathogenesis associated with the *SNCA* A53T substitution with an allelic imbalance and an epigenetic regulation (Voutsinas et al., 2010). Such overexpression may also contribute to sporadic cases since an association exists between the PD risk and the REP1 polymorphism within the *SNCA* promoter, which controls the level of alpha-synuclein expression in PD patients (Maraganore et al., 2006).

In addition to *SNCA*, we confirmed that thirty other genes or putative transcripts are present within this 4.928 Mb, which is thus the largest identified *SNCA* duplicated locus leading to PD. To investigate their potential involvement in the PD phenotype, we analyzed their expression in the PBMCs from the P59 carrier by a transcriptomic approach. Nine of them were not expressed at a sufficient level in these cells and seven of them are not dosage-sensitive in those cells (Table 1). Two probes confirmed the overexpression of *SNCA* in the duplication

1
2
3
4 carrier PBMCs. Moreover, we demonstrated that 13 other genes present in the duplicated locus were significantly
5 overexpressed. Especially, *FAM13A* (MIM# 613299), *MMRN1*, *NAP1L5* (MIM# 612203) together with *SNCA*
6 presented a two-fold overexpression in these cells (Table 1). To determine if these genes were overexpressed in
7 other PD conditions, we compared these results with published data. Interestingly, these genes are not
8 overexpressed in PBMCs from other PD patients compared to controls. Thus, their overexpression seems directly
9 linked to the presence of the duplication and may impact the phenotype of this patient. We cannot preclude that
10 RNA overexpression will not lead to a protein dysfunction. Nevertheless, 10 out of the 31 duplicated genes were
11 deregulated in substantia nigra from sporadic PD subjects (Table 1) suggesting that these genes might contribute to
12 the disease development. Among them, expression of *FAM13A* gene was increased both in PBMCs duplication
13 carrier and brain tissue of PD patients compared to controls, but was slightly decreased in other PD PBMCs.
14 Although little is known on its role, this may suggest a particular role of *FAM13A* in the substantia nigra of
15 duplicated patients that has not been found yet. *MMRN1* is also overexpressed in the duplicated carrier PBMCs
16 and since, when multiplied, it was suggested to play a role in the clinical features notably on the cognitive
17 impairment (Nishioka et al., 2006); we also investigated cognitive functions. In our kindred, subject III2 met the
18 criteria for dementia 17 years after PD onset (Emre et al., 2007). Our tests revealed that two subjects (III2 and III6)
19 had both cognitive complaints and cognitive dysfunction. However, only III2 carried the *SNCA* duplication.
20 Likewise, we observed poor cognitive test results for subjects III3, III4 and III5 but only III4 carried the
21 pathogenic mutation. These data support the hypothesis whereby the 30 other genes present within the duplication
22 do not significantly contribute to the cognitive impairment. However, we cannot rule out the possibility that one or
23 more of these genes may exacerbate susceptibility to cognitive dysfunction already present in the kindred members
24 or contribute to another clinical condition. Indeed, osteoarthritis was present in duplication carriers and recent
25 analyses reported a slightly increased risk for PD among patients with osteoarthritis (Rugbjerg et al., 2010).
26 Nevertheless, since arthritis is one of the most common diseases affecting almost every person over age 60, to date
27 it is difficult to draw further conclusion, on the presence of duplicated genes involved in osteoblast maturation or
28 mineralization and the arthritis reported in the patients carrying the duplication (Table 1, last column). This table
29 also indicates that several variants in the vicinity of *SNCA*, *GPRIN3* (MIM# 611241) and *MMRN1* were associated
30 with PD risk by genome wide association studies (GWAS) (Satake et al., 2009; Simón-Sánchez et al., 2009).
31 However, GWAS data alone are not able to define these genes as a molecular cause in the disease. These single
32 nucleotide polymorphisms are indeed localized in the intergenic regions and without further analyses showing the
33 functionality of these variants; they could either reflect the association with other causative polymorphisms
34 affecting the *SNCA* gene or the involvement of another sequence within this locus. Today, the most plausible
35 explanation is the overexpression of *SNCA* in the disease, but other mechanisms such as mRNA stability,
36 preferential splicing or miRNA might also exist. Indeed, *SNCA* gene showed a reduction of expression of 0.6 and
37 0.7-fold in idiopathic PD brain tissue and PBMCs, respectively; while it was overexpressed in P59 *SNCA*
38 duplication carrier (Table 1). The significance of this decrease of total *SNCA* expression in idiopathic PD brain and
39 PBMCs indicates that even if the clinical phenotype between sporadic PD and duplication carrier was similar, the
40 underlying primary mechanisms can differ at least in a number of patients.

41 In the last part of the present study, we sought to define the duplication's boundaries and thus decipher the
42 multiplication mechanism. The genome rearrangement was confirmed by several methods. We provided a better
43 definition of the duplication's boundaries and demonstrated that 31 genes were located within the duplicated
44 region. Next, by using a PCR walking strategy to sequence the breakpoint junction, we discovered that a
45 rearrangement had occurred between two LINE-1 retrotransposons (L1PA2 and L1PA4) and not between *Alu*
46 repeats, as it could have been expected (Ross et al, 2008). A 349 bp deletion was observed at the junction between
47 the LINEs. Of the many transposable elements in the human genome (and which account for almost half of the
48 latter), LINE-1 (L1) sequences are thought to be the major driving force behind genome expansion (Babashok et
49 al., 2007). A transpositionally active L1 element comprises a 5'-untranslated region (UTR), two open reading
50 frames (ORFs) (which notably code for proteins involved in RNA binding and genome integration), a 3'-UTR.
51 However, most copies of L1 are 5'-truncated and often only consist of part of the 3'-UTR. The two L1 elements
52 (L1PA2 and L1PA4) in family P59 had inactive ORFs. Even when these transposable elements are not active, they
53 can nevertheless cause human disease by creating insertional mutations in genes and by contributing to genetic
54 instability through NAHR (leading to rearrangements, such as duplications) (Belancio et al., 2009). In the
55 duplicated locus, we identified a >750 pb length homology sequence between L1PA2 and L1PA4 with several
56 perfect sequence matches; which is a required condition to generate NAHR (Inoue & Lupski, 2002; Conrad et al.,
57 2009). Moreover, this region of chromosome 4q21 has a high DNA flexibility and A/T-content probably which is
58 suspected to promote the occurrence of fragile sites (Rozier et al., 2004). Taken as a whole, these data strengthen
59
60

10 <Mutez et al.>

1
2
3 the idea that homologous recombination has preferred positions and architectural DNA characteristics, since
4 rearrangement breakpoints are associated with LCRs more frequently than would be expected by chance alone
5 (Shaw et al., 2004).

6 In conclusion, for the first time, we identified the exact sequence of the breakpoints from *SNCA* locus
7 multiplication carrier and demonstrated the involvement of LINEs in *SNCA* locus duplication which led to *SNCA*
8 overexpression and abnormal dopamine metabolism in the carriers. We showed that DaTSCAN SPECT,
9 immortalized cell lines and PBMCs are effective tools for demonstrating the preclinical and biological features of
10 PD. While our data highlight the importance of the *SNCA* overexpression as the major cause of the PD phenotype
11 in our kindred, we also demonstrate that the duplication affects the expression of other genes. Therefore their
12 potential role in PD phenotype should deserve further attention.
13

14
15 **ACKNOWLEDGMENTS**

16 The authors thank the subjects for their participation in this research study. We wish to thank Professor Marc
17 Steinling (Nuclear Medicine Department, Lille University Hospital, Lille, France) for the DaTSCAN analyses, Dr
18 Joris Andrieux (Laboratory of Medical Genetics, Lille University Hospital, France) for his help in the FISH
19 experiment. We are particularly grateful to the Institut de Recherches sur le Cancer de Lille for its assistance.
20

21 **REFERENCES**

- 22
23 Ahn TB, Kim SY, Kim JY, Park SS, Lee DS, Min HJ, Kim YK, Kim SE, Kim JM, Kim HJ, Cho J, Jeon BS. 2008. alpha-
24 Synuclein gene duplication is present in sporadic Parkinson disease. *Neurology* 70:43-49.
25
26 Babashok DV, Kazazian HH. 2007. Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat* 28:527-
27 539.
28
29 Belancio VP, Deininger PL, Roy-Engel AM. 2009. LINE dancing in the human genome: transposable elements and disease.
30 *Genome Med* 1:97.
31
32 Chartier-Harlin MC, Kachergus J, Roumier C, Mouroux V, Douay X, Lincoln S, Levecque C, Larvor L, Andrieux J, Hulihan
33 M, Waucquier N, Defebvre L, Amouyel P, Farrer M, Destée A. 2004. Alpha-synuclein locus duplication causes familial
34 Parkinson's disease. *Lancet* 364:1167-1169.
35
36 Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, Turner DJ, Hurles ME. 2010. Mutation spectrum revealed
37 by breakpoint sequencing of human germline CNVs. *Nat Genet* 42:385-391.
38
39 Dujardin K, Krystkowiak P, Defebvre L, Blond S, Destée A. 2000. A case of severe dysexecutive syndrome consecutive to
40 chronic bilateral pallidal stimulation. *Neuropsychologia* 38:1305-1315.
41
42 Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, Broe GA, Cummings J, Dickson DW, Gauthier S,
43 Goldman J, Goetz C, Korczyn A, Lees A, Levy R, Litvan I, McKeith I, Olanow W, Poewe W, Quinn N, Sampaio C, Tolosa
44 E, Dubois B. 2007. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* 22:1689-1707.
45
46 Ibáñez P, Lesage S, Janin S, Lohmann E, Durif F, Destée A, Bonnet AM, Brefel-Courbon C, Heath S, Zelenika D, Agid Y,
47 Dürr A, Brice A; French Parkinson's Disease Genetics Study Group. 2009. Alpha-synuclein gene rearrangements in
48 dominantly inherited parkinsonism: frequency, phenotype, and mechanisms. *Arch Neurol* 66:102-108.
49
50 Inoue K, Lupski JR. 2002. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* 3:199-242.
51
52 Lee JE, Nam HY, Shim SM, Bae GR, Han BG, Jeon JP. 2010. Expression phenotype changes of EBV-transformed
53 lymphoblastoid cell lines during long-term subculture and its clinical significance. *Cell Prolif* 43:378-384.
54
55 Lesage S, Brice A. 2009. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* 18:R48-
56 59.
57
58 Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE,
59 Maraganore DM. 2007. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS*
60 *Genet* 3:e98.

- 1
2
3 Maraganore DM, de Andrade M, Elbaz A, Farrer MJ, Ioannidis JP, Krüger R, Rocca WA, Schneider NK, Lesnick TG, Lincoln
4 SJ, Hulihan MM, Aasly JO, Ashizawa T, Chartier-Harlin MC, Checkoway H, Ferrarese C, Hadjigeorgiou G, Hattori N,
5 Kawakami H, Lambert JC, Lynch T, Mellick GD, Papapetropoulos S, Parsian A, Quattrone A, Riess O, Tan EK, Van
6 Broeckhoven C; Genetic Epidemiology of Parkinson's Disease (GEO-PD) Consortium. 2006. Collaborative analysis of
7 alpha-synuclein gene promoter variability and Parkinson disease. *JAMA* 296:661-670.
- 8
9 Mutez E, Larvor L, Leprêtre F, Mouroux V, Hamalek D, Kerckaert JP, Pérez-Tur J, Waucquier N, Vanbesien-Mailliot C,
10 Dufлот A, Devos D, Defebvre L, Kreisler A, Frigard B, Destée A, Chartier-Harlin MC. Transcriptional profile of Parkinson
11 blood mononuclear cells with LRRK2 mutation. doi: 10.1016/j.neurobiolaging.2009.10.016.
- 12
13 Nishioka K, Hayashi S, Farrer MJ, Singleton AB, Yoshino H, Imai H, Kitami T, Sato K, Kuroda R, Tomiyama H, Mizoguchi
14 K, Murata M, Toda T, Imoto I, Inazawa J, Mizuno Y, Hattori N. 2006. Clinical heterogeneity of α -synuclein gene
15 duplication in Parkinson's disease. *Ann Neurol* 59:298-309.
- 16
17 Nishioka K, Ross OA, Ishii K, Kachergus JM, Ishiwata K, Kitagawa M, Kono S, Obi T, Mizoguchi K, Inoue Y, Imai H,
18 Takashi M, Mizuno Y, Farrer MJ, Hattori N. 2009. Expanding the clinical phenotype of SNCA duplication carriers. *Mov*
19 *Disord* 24:1811-1819.
- 20
21 Penzkofer T, Dandekar T, Zemojtel T. 2005. L1base: from functional annotation to prediction of active LINE-1 elements.
22 *Nucleic Acids Res* 33(Database issue):D498-500.
- 23
24 Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES,
25 Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe
26 LI, Nussbaum RL. 1997. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*
27 276:2045-2047.
- 28
29 Ross OA, Braithwaite AT, Skipper LM, Kachergus J, Hulihan MM, Middleton FA, Nishioka K, Fuchs J, Gasser T, Maraganore
30 DM, Adler CH, Larvor L, Chartier-Harlin MC, Nilsson C, Langston JW, Gwinn K, Hattori N, Farrer MJ. 2008. Genomic
31 investigation of α -synuclein multiplication and Parkinsonism. *Ann Neurol* 63:743-750.
- 32
33 Rozier L, El-Achkar E, Apiou F, Debatisse M. 2004. Characterization of a conserved aphidicolin-sensitive common fragile site
34 at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene* 23:6872-6880.
- 35
36 Rugbjerg K, Friis S, Jørgensen TL, Ritz B, Korbo L, Olsen JH. 2010. Risk for Parkinson's disease among patients with
37 osteoarthritis: a Danish cohort study. *Mov Disord* 25:2355-2360.
- 38
39 Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, Kawaguchi T, Tsunoda T, Watanabe M, Takeda A, Tomiyama
40 H, Nakashima K, Hasegawa K, Obata F, Yoshikawa T, Kawakami H, Sakoda S, Yamamoto M, Hattori N, Murata M,
41 Nakamura Y, Toda T. 2009. Genome-wide association study identifies common variants at four loci as genetic risk factors
42 for Parkinson's disease. *Nat Genet* 41:1303-1307.
- 43
44 Scherzer CR, Eklund AC, Morse LJ, Liao Z, Locascio JJ, Fefer D, Schwarzschild MA, Schlossmacher MG, Hauser MA, Vance
45 JM, Sudarsky LR, Standaert DG, Growdon JH, Jensen RV, Gullans SR. 2007. Molecular markers of early Parkinson's
46 disease based on gene expression in blood. *Proc Natl Acad Sci U S A.* 104:955-960.
- 47
48 Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3:1101-1108.
- 49
50 Shaw CJ, Lupski JR. 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis
51 of disease. *Hum Mol Genet* 13:R57-64.
- 52
53 Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG,
54 Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA, Illig T, Gieger C, Houlden H, Steffens M,
55 Okun MS, Racette BA, Cookson MR, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K,
56 van der Brug M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D,
57 Deuschl G, Chen H, Riess O, Hardy JA, Singleton AB, Gasser T. 2009. Genome-wide association study reveals genetic risk
58 underlying Parkinson's disease. *Nat Genet* 41:1308-1312.
- 59
60 Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R,
Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muentner M, Baptista M, Miller D, Blacato J,
Hardy J, Gwinn-Hardy K. 2003. Alpha-synuclein locus triplication causes Parkinson's disease. *Science* 302:841.

12 <Mutez et al.>

1
2
3 Voutsinas GE, Stavrou EF, Karousos G, Dasoula A, Papachatzopoulou A, Syrrou M, Verkerk AJ, van der Spek P, Patrinos GP,
4 Stöger R, Athanassiadou A. 2010. Allelic imbalance of expression and epigenetic regulation within the alpha-synuclein
5 wild-type and p.Ala53Thr alleles in Parkinson disease. Hum Mutat 31:685-691.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Supp. Table S1: Primer sequences used for gene dosage, *SNCA* expression, PCR walking strategy and breakpoint junction sequencing.

PCR Use	Primer sequences (5'-3')
<i>SNCA</i> exon 1 gene dosage	CCTTCAAGCCTTCTGCCTTTC CGAATGGCCACTCCCAGTT
<i>SNCA</i> exon 7 gene dosage	TCTTTGCTCCCAGTTTCTTGAGA TGGAAGTGAAGCACTTGTACAGGAT
<i>SNCA</i> expression	ATGTAGGCTCCAAAACCAAG TACCCTTCCTCAGAAGGCA
<i>RPS18</i> expression	CCTGGATACCGCAGCTAGGA GCGGCGCAATACGAATGCCCC
<i>ACTB</i> expression	CACGGCATCGTCACCAACT AGCCACACGCAGCTCATTG
<i>GAPDH</i> expression	TACATGTTCCAATATGATTCCACC ACCTTCTTGATGTCATCATATTTG
PCR Walking Strategy and Break junction sequencing	1 TCAGAAGAGGCATATTACAT 2 GTGACAACATTCTAAAAGGG 3 TTCCTTATGATTCCACTTAG 4 GCACAGACAAACAAAAAGAC 5 AAGACCATCGAGACTAGGAA 6 ACACCTCTACGCAAATAAAC 7 CCAAGTCAATCCTAAGCC 8 AATCATGCTGCTATAAAGAC 9 AAAAGTGGGTGAAGGATATG 10 CCCTTTTAGAATGTTGTCAC 11 CAAAACAGAGAAAACAGAAG 12 ATGTGGGTCTGTAGCAAC

Concerning the PCR walking strategy, the numbers correspond to the primers as indicated in Figure 3C.