

ARTICLE

# Genome-wide association study confirms extant PD risk loci among the Dutch

Javier Simón-Sánchez<sup>\*,1</sup>, Jacobus J van Hilten<sup>2</sup>, Bart van de Warrenburg<sup>3</sup>, Bart Post<sup>3</sup>, Henk W Berendse<sup>4</sup>, Sampath Arepalli<sup>5</sup>, Dena G Hernandez<sup>5</sup>, Rob MA de Bie<sup>6</sup>, Daan Velseboer<sup>6</sup>, Hans Scheffer<sup>7</sup>, Bas Bloem<sup>3</sup>, Karin D van Dijk<sup>4</sup>, Fernando Rivadeneira<sup>8,9</sup>, Albert Hofman<sup>8</sup>, André G Uitterlinden<sup>8,9</sup>, Patrizia Rizzu<sup>1</sup>, Zoltan Bochdanovits<sup>1</sup>, Andrew B Singleton<sup>5</sup> and Peter Heutink<sup>1</sup>

In view of the population-specific heterogeneity in reported genetic risk factors for Parkinson's disease (PD), we conducted a genome-wide association study (GWAS) in a large sample of PD cases and controls from the Netherlands. After quality control (QC), a total of 514 799 SNPs genotyped in 772 PD cases and 2024 controls were included in our analyses. Direct replication of SNPs within *SNCA* and *BST1* confirmed these two genes to be associated with PD in the Netherlands (*SNCA*, rs2736990:  $P=1.63 \times 10^{-5}$ , OR=1.325 and *BST1*, rs12502586:  $P=1.63 \times 10^{-3}$ , OR=1.337). Within *SNCA*, two independent signals in two different linkage disequilibrium (LD) blocks in the 3' and 5' ends of the gene were detected. Besides, *post-hoc* analysis confirmed *GAK/DGKQ*, *HLA* and *MAPT* as PD risk loci among the Dutch (*GAK/DGKQ*, rs2242235:  $P=1.22 \times 10^{-4}$ , OR=1.51; *HLA*, rs4248166:  $P=4.39 \times 10^{-5}$ , OR=1.36; and *MAPT*, rs3785880:  $P=1.9 \times 10^{-3}$ , OR=1.19).

*European Journal of Human Genetics* (2011) 19, 655–661; doi:10.1038/ejhg.2010.254; published online 19 January 2011

**Keywords:** *SNCA*; *BST1*; *GAK/DGKQ*; *HLA*; *MAPT*; PD

## INTRODUCTION

A genetic contribution to Parkinson's disease (PD) is well recognized<sup>1–3</sup> with ~10% of the cases carrying mutations that lead to rare Mendelian forms of the disease.<sup>4</sup> Recently, independent genome-wide association studies (GWAS) established an unequivocal role for common genetic variants in the etiology of PD<sup>5–11</sup> and suggested a population-specific genetic heterogeneity, with *SNCA*, *PARK16*, *BST1* and *LRRK2* as shared risk loci for PD,<sup>6–8,10,11</sup> and *MAPT* as an European-specific risk locus.<sup>6,7</sup> Besides, variation in the *GAK/DGKQ*,<sup>5</sup> the *HLA* region<sup>10</sup> and a locus in chromosome 12q24<sup>11</sup> has been proposed to exert a risk for PD in different European populations.

To determine the role of these loci in the Dutch population and to find new genetic factors exerting a risk for PD, we carried out what is, to our knowledge, the first GWAS for PD in the Dutch population.

## SUBJECTS AND METHODS

### Subjects

As a product of a national collaborative venture, a total of 841 PD patients were recruited from four different centers within the Netherlands (Scales for Outcomes in Parkinson's disease, SCOPA, <http://www.scopa-propark.eu>; the Academic Medical Center Amsterdam, AMC, <http://www.amc.uva.nl>; the Parkinson Centrum Nijmegen, ParC, <http://www.umcn.nl>; and the VU University medical centre, VUmc, <http://www.vumc.nl>). All patients were

self-reported Caucasian individuals from the Netherlands. The assessed samples consisted of 533 males and 308 females with a mean age at onset ranging from 16 to 84 years (mean=57.5 years; standard deviation=12). For more information about these samples, please visit the websites listed above.

Genome-wide genotyping data from 2082 control participants from the Rotterdam study III<sup>12–14</sup> (ERGO Young) genotyped with Human610K beadchips from Illumina (<http://www.illumina.com>) were used as our control population. Of these, 912 were males and 1116 females. The mean age was 53.75 years with a range of 45–95 years.

### Genotyping

All 841 PD cases were genotyped at 592 839 unique positions with Human660W-Quad beadchips from Illumina, a powerful tool for GWAS. For more information about this genotyping platform, please visit <http://www.illumina.com>.

### Quality control (QC) procedures

After extensive QC approaches (see Supplementary Material for details), the final number of fully genotyped samples from the Netherlands was 2796 including 772 cases and 2024 controls. Each of these was genotyped in a total of 514 799 unique autosomal SNPs.

### Statistical analyses

Quanto software was used to estimate power (University of South California, <http://hydra.usc.edu/gxe>). Odds ratios (OR) considering a  $P$ -value of  $9.71 \times 10^{-8}$  (genome-wide significance level after Bonferroni correction) and different allele frequencies were calculated (Supplementary Figure 1).

<sup>1</sup>Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre, Amsterdam, The Netherlands; <sup>2</sup>Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands; <sup>3</sup>Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; <sup>4</sup>Department of Neurology, VU University Medical Centre, Amsterdam, The Netherlands; <sup>5</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; <sup>6</sup>Department of Neurology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; <sup>7</sup>Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; <sup>8</sup>Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands and <sup>9</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

\*Correspondence: Dr J Simón-Sánchez, Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre, Amsterdam, North Holland 1081 BT, The Netherlands. Tel: +31 20 598 3662; Fax: +31 20 598 3596; E-mail: [j.simon Sanchez@vumc.nl](mailto:j.simon Sanchez@vumc.nl)

Received 28 July 2010; revised 30 November 2010; accepted 14 December 2010; published online 19 January 2011

For association analyses, all estimates and tests were carried out with PLINK<sup>15</sup> (<http://pngu.mgh.harvard.edu/~purcell/plink/>).

For each SNP that successfully passed our QC process, a multi-covariate logistic regression analysis following an additive model was applied. Covariates used in our model were gender, age at onset and age at ascertainment (for cases and controls, respectively), and the first two components of the MDS values after calculation of pairwise IBS. Although the genomic inflation factor (based on median  $\chi^2$ ) was low ( $\lambda=1.06$ , Supplementary Figure 3), we chose to adjust for genomic control to correct for possible population substructure.

In a first approach aiming to determine (in the Dutch) the role of previously identified PD genetic risk factors, we decided to look closely at the results for *SNCA* (chromosome 4q21),<sup>5–7</sup> *MAPT locus* (chromosome 17q21.1),<sup>5,7</sup> *LRKK2* (chromosome 12q12),<sup>6,7</sup> *PARK16*<sup>6,7</sup> (chromosome 1q32), *BST1* (chromosome 4p15),<sup>6</sup> the *GAK/DGKQ locus* (chromosome 4p16),<sup>5</sup> the *HLA* region (chromosome 6p21.3)<sup>10</sup> and chromosome 12q24 locus.<sup>11</sup> A total of 30 SNPs from these loci were selected for closer scrutiny.

The LD structure of the associated loci was analyzed using Haploview 4.1<sup>16</sup> (Broad Institute of MIT and Harvard, [www.broad.mit.edu/mpg/haploview/](http://www.broad.mit.edu/mpg/haploview/)) and blocks delimited using the  $D'$ -based confidence interval method developed by Gabriel *et al*<sup>17</sup> as implemented in Haploview. Risk haplotypes were counted and OR values were plotted with R v.2.7.2 (<http://cran.r-project.org/>).

A logistic regression analysis of the most associated SNP in each block conditioned to variation at the most associated SNP in the other block (if applicable) was carried out to test for statistical independence of the signals detected.

For *SNCA*, the population attributable risk (PAR) was calculated using the formula  $PAR = (p(RR-1))/(p(RR-1)+1) \times 100$ , where  $p$  is the prevalence of the risk allele in the population and  $RR$  is the relative risk. As we showed that association from the 3' and 5' LD block is independent (see results), the combined PAR was calculated with the formula  $cPAR = 1 - (1 - PAR_{3' \text{ block}}) \times (1 - PAR_{5' \text{ block}})$ .

For *MAPT* locus, the presence of alleles in the H1 and H2 haplotypes was accomplished using rs1981997 as a haplotype tag SNP because the major (G) and minor (A) alleles of this SNP are fixed in the H1 and H2 haplotypes, respectively.<sup>9,18</sup> To determine which of the associated alleles in the *MAPT* locus were present in H1 (previously associated with PD<sup>19–24</sup>) a two-locus haplotype association analysis of rs1981997 and SNPs in the *MAPT* region was carried out using PLINK.<sup>15</sup>

Although we are aware that the sample size of this cohort has a limited power and a GWAS would probably fail to find any associated locus after correcting for 514 799 independent tests, we decided to carry out this analysis to look for specific PD risk loci in the Dutch population. For this purpose, each genotyped SNP was tested for association using the multi-covariate logistic regression explained above.

## RESULTS

After QC, a total of 772 Dutch PD cases and 2024 controls from the Rotterdam study<sup>12–14</sup> genotyped in a total of 514 799 unique autosomal SNPs, were included in our analyses (Table 1). To assess the homogeneity of our cohort, identity by state (IBS) distances were calculated. This analysis revealed that both cases and controls share common ancestry (Supplementary Figure 1A) consistent with that seen in CEU (CEPH Europeans from Utah) (Supplementary Figure 1B, C). Power calculations showed that our sample has adequate power to detect variants conferring a risk with an OR of 1.4–1.5 depending on the minor allele frequency of such variant (Supplementary Figure 2).

As we are aware that our population is of limited size to detect risk loci with the effect size previously described for PD in Europeans at the significance level requested by GWAS, the first objective of this project was to determine the role, in the Dutch population, of previously identified genetic risk factors. Thus, association results from 30 SNPs in *SNCA*,<sup>5–7</sup> *MAPT locus*,<sup>5,7</sup> *LRKK2*,<sup>6,7</sup> *PARK16*,<sup>6</sup> *BST1*<sup>6</sup> and *GAK/DGKQ*,<sup>5,10</sup> *HLA*<sup>10</sup> and chromosome 12q24 loci,<sup>11</sup> were selected for close scrutiny. Two SNPs in *SNCA* and *BST1* were significantly associated with PD in our population after correcting for 30 independent tests ( $P < 0.00166$  in both instances, Table 2). Although no association was found in any of the other SNPs, a trend toward an association in the *DGKQ/GAK* and *MAPT* loci was detected (Table 2). *Post-hoc* analysis of results at these two locations and the *HLA* region, provided evidence of the role of these three loci in the pathogenesis of PD in the Dutch. No association was found in *LRKK2*, *PARK16*, or the chromosome 12q24 locus. A more detailed description of the results obtained at each locus is given below.

### SNCA

Although association of *SNCA* with PD is well established, it remains unclear whether it is derived from the 3' or the 5' end of the gene.<sup>6–8,25–32</sup> In an effort to delineate the signal at this locus, we examined the LD structure across this region. This analysis revealed two LD blocks delimited at the intron 4 of *SNCA* as previously described<sup>6–8,25</sup> (Supplementary Figure 4). Although the strongest signal detected in *SNCA* was located at the 3' block (rs2736990,  $P = 1.63 \times 10^{-5}$ ), a closer examination of the association signals revealed that SNPs both in the 3' (lowest  $P = 1.63 \times 10^{-5}$ , rs2736990) and the 5' blocks (lowest  $P = 1.78 \times 10^{-3}$ , rs356188) of the gene appeared to be associated with PD in our population (Supplementary Table 1). A logistic regression analysis using the allele dosage of the most associated SNP in each block as a covariate, did not change the results observed at the most associated signal in the other, indicating that the two signals are independent (Supplementary Table 2) and being in disagreement with previous results.<sup>7,25,32</sup> LD analysis between these two SNPs revealed a  $r^2$  value of 0.028 and a  $D'$  value of 0.388, further supporting this hypothesis.

Examination of the risk conferred by each of the haplotypes contained in these blocks confirmed these results and showed that a single haplotype contained in the 3' block and two other in the 5' block exerted the largest risk for PD in our population with ORs of 1.42, 1.41 and 1.48, respectively (Supplementary Figure 5).

### BST1

Rs12502586 showed significant association with PD in our population, with a  $P$ -value of  $1.63 \times 10^{-3}$  and an OR of 1.337. A closer look into the association results at this locus revealed that three SNPs presented with  $P$ -values of  $\sim 10^{-3}$  (range  $1.63 \times 10^{-3}$ – $3.33 \times 10^{-3}$ , Supplementary Table 3). Haplotype analysis of this region showed the existence of two LD blocks containing haplotypes highly associated with PD ( $P$ -values of  $6 \times 10^{-4}$  and  $4 \times 10^{-4}$ , respectively). These blocks span from *LOC285550* to intron 3 of *BST1* and from intron 4 of *BST1* to intron 8 of this same gene. The associated haplotypes exert a risk for

**Table 1** Population characteristics of the Dutch cohort used for these experiments

	Country of origin	Number	Male/female ratio	Age at onset (Mean (range))	Age at sampling (Mean (range))
Cases	The Netherlands	772	1.75	55.3 (16–84)	62.36 (29–89)
RSIII Controls	The Netherlands	2024	0.78	—	53.75 (45–95)

**Table 2** Replication of previous GWAS results in the Dutch cohort

SNP	Chromosome	Position	Locus	Previous GWAS		Dutch GWAS	
				P-value	OR	P-value	OR
rs11931074	4q22	90 858 538	<i>SNCA</i>	$1.62 \times 10^{-14}$	1.46	$2.15 \times 10^{-3}$	1.387
rs3857059	4q22	90 894 261	<i>SNCA</i>	$3.74 \times 10^{-15}$	1.48	$2.13 \times 10^{-3}$	1.387
rs2736990	4q22	90 897 564	<i>SNCA</i>	$2.24 \times 10^{-16}$	1.23	$1.63 \times 10^{-5}$	1.325
rs393152	17q21	41 074 926	<i>MAPT</i>	$1.95 \times 10^{-16}$	0.77	$9.45 \times 10^{-3}$	0.8167
rs17563986	17q21	41 347 100	<i>MAPT</i>	$1.67 \times 10^{-14}$	0.78	$7.65 \times 10^{-3}$	0.8119
rs1981997	17q21	41 412 603	<i>MAPT</i>	$3.55 \times 10^{-14}$	0.78	$7.69 \times 10^{-3}$	0.812
rs8070723	17q21	41 436 901	<i>MAPT</i>	$9.64 \times 10^{-14}$	0.79	$8.64 \times 10^{-3}$	0.8146
rs2532274	17q21	41 602 941	<i>MAPT</i>	$8.24 \times 10^{-11}$	0.83	$9.16 \times 10^{-3}$	0.8161
rs2532269	17q21	41 605 885	<i>MAPT</i>	$5.62 \times 10^{-13}$	0.79	0.014	0.8255
rs2668692	17q21	41 648 797	<i>MAPT</i>	$7.87 \times 10^{-14}$	0.79	0.02	0.8338
rs169201	17q21	421 45 386	<i>MAPT</i>	$4.39 \times 10^{-14}$	0.78	0.056	0.8617
rs199533	17q21	42 184 098	<i>MAPT</i>	$1.09 \times 10^{-14}$	0.78	0.050	0.8581
rs11564162	12q12	38 729 159	<i>LRRK2</i>	$9.52 \times 10^{-5}$	0.87	0.961	0.9962
rs2896905	12q12	38 779 683	<i>LRRK2</i>	$7.81 \times 10^{-3}$	0.93	0.909	0.9926
rs1491923	12q12	38 877 384	<i>LRRK2</i>	$1.55 \times 10^{-5}$	1.14	0.705	1.026
rs947211	1q32	202 484 322	<i>PARK16</i>	0.01	0.93	0.123	0.8892
rs708730	1q32	202 509 437	<i>PARK16</i>	$1.59 \times 10^{-3}$	0.9	0.848	1.017
rs823128	1q32	203 980 001	<i>PARK16</i>	$7.29 \times 10^{-8}$	0.66	0.385	0.8316
rs823156	1q32	204 031 263	<i>PARK16</i>	$7.60 \times 10^{-4}$	0.89	0.838	0.9827
rs11240572	1q32	204 074 636	<i>PARK16</i>	$6.11 \times 10^{-7}$	0.67	0.961	1.01
rs11931532	4p15.3	15 334 864	<i>BST1</i>	0.44	0.94	0.642	1.095
rs12502586	4p15.3	15 335 662	<i>BST1</i>	0.07	1.08	$1.63 \times 10^{-3}$	1.337
rs12645693	4p15.3	15 338 632	<i>BST1</i>	0.44	0.94	0.766	1.06
rs4698412	4p15.3	15 346 446	<i>BST1</i>	0.03	0.94	0.021	0.8616
rs4538475	4p15.3	15 347 035	<i>BST1</i>	0.27	0.98	0.602	0.9563
rs12646913	4p15.3	15 348 374	<i>BST1</i>	0.03	0.92	0.372	0.8931
rs1564282	4p16	842 313	<i>DGKQ/GAK</i>	$6.7 \times 10^{-7}$	1.7	$4.56 \times 10^{-3}$	1.309
rs11248060	4p16	954 359	<i>DGKQ/GAK</i>	$2.5 \times 10^{-7}$	1.69	0.013	1.27
rs3129882	6p21.3	32 517 508	<i>HLA</i>	$2.9 \times 10^{-8}$	1.31	0.312	1.07
rs4964469	12q24	105 474 117	<i>12q24</i>	$2.38 \times 10^{-7}$	1.16	0.381	1.04

Abbreviations: GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

A total of 30 SNPs significantly associated with PD in four GWAS performed in European<sup>5,7,10,11</sup> and Japanese<sup>6</sup> populations were selected for closer scrutiny. *P*-values and odds ratios for these SNPs are compared with the results at those same positions in the Dutch cohort. To note, although variation in rs3129882 was not associated with PD in our population, *post-hoc* analysis revealed 36 SNPs nominally associated with PD in this locus (see *HLA* region results section).

PD with an OR of 1.24 and 1.36, respectively (Supplementary Figure 6). A logistic regression analysis of the most associated SNP in each block (rs4583752 and rs12502586, respectively) conditioned to variation at the most associated SNP in the other, yielded a prominent reduction of the detected association, indicating that both risk haplotypes are tagging the same risk variant (Supplementary Table 4). The linkage disequilibrium between these two had  $r^2$  value of 0.280 and a *D'* of 0.938, further supporting the previous results.

#### **GAK/DGKQ**

Only two out of the three SNPs in the *GAK/DGKQ* locus, which were reported to be associated with PD by Pankratz *et al*<sup>5</sup> were genotyped in our cohort. One of these (rs1564282), which is most associated in Pankratz's study, was associated with PD in our population (Table 1). Although this association did not surpass our conservative Bonferroni correction for 30 independent tests, *post-hoc* analysis revealed the existence of 11 SNPs with *P*-values <0.05 spanning 260.5 kb (lowest  $P=1.22 \times 10^{-4}$ , rs2242235) (Supplementary Table 5, Supplementary Figure 7). Due to this, and because Bonferroni correction is unnecessarily punitive when replicating genes that are known to be associated with a certain trait, we consider these results as positive.

The two most associated SNPs in this region (rs2242235 and rs4690296 with *P*-values of  $1.22 \times 10^{-4}$  and  $2.02 \times 10^{-4}$ , respectively)

are located in *PCGF3*, a gene 78 kb away from *GAK* and 188 kb away from *DGKQ*, which encodes a protein with unknown function containing a C3HC4 type RING finger for protein–protein interaction. These two SNPs are contained in two different haplotype blocks spanning from intron 1 to intron 8 of *PCGF3* and intron 8 to intron 9 of this same gene, respectively. In addition, a predicted transcript, *LOC100118084* is contained in the first block. Each of the two blocks contain a single haplotype associated with PD in our population ( $P=2 \times 10^{-4}$  and  $7.38 \times 10^{-5}$ , respectively) with an OR of 1.47 and 1.52, respectively (Supplementary Figure 8). The most associated SNPs in each of these haplotype blocks are in strong LD with each other ( $r^2=0.818$ ,  $D'=0.916$ ), indicating that they are, most likely, tagging the same risk variant in the region. A logistic regression analysis using the allele dosage at the other marker as a covariate confirmed this hypothesis (Supplementary Table 6).

#### **HLA region**

In the GWAS carried out by Hamza *et al*<sup>10</sup> only one SNP (rs3129882) in this region exceeded genome-wide significance level. After *in silico* replication and meta-analysis of the most significant SNPs, another six SNPs in this region were associated with PD in their population.<sup>10</sup> Of these, only rs3129882 was present in our genotyping platform. Although this SNP was not associated with PD in our population

( $P=0.13$ ), *post-hoc* analysis of this region revealed the existence of 36 SNPs with  $P$ -values below 0.05 (lowest  $P=4.39 \times 10^{-5}$ , rs4248166) spanning a total of 440.1 kb (Supplementary Table 7, Supplementary Figure 9).

The four most associated SNPs in this region ( $P$ -values ranging from  $4.39 \times 10^{-5}$  to  $7.98 \times 10^{-5}$ ) are located in two consecutive LD blocks. Each of these blocks contains a haplotype associated with PD in our population ( $P$ -values of  $6.90 \times 10^{-5}$  and  $6.63 \times 10^{-5}$ , respectively) with ORs of 1.32 and 1.28, respectively (Supplementary Figure 10). The most associated SNPs in each of these haplotype blocks are in strong LD with each other ( $r^2=0.758$ ,  $D'=0.984$ ), indicating that they are tagging the same risk variant. A logistic regression analysis using the allele dosage at the other marker as a covariate confirmed this hypothesis (Supplementary Table 8).

The next association signal in this locus lies on rs17533090. Although this signal is 223.46 kb away from the one explained in the previous paragraph, it is in relatively strong LD with it ( $r^2=0.376$ ,  $D'=0.616$ ), indicating that they are, most likely, tagging the same risk variant. A logistic regression analysis using the allele dosage at rs4248166 as a covariate confirmed this hypothesis (Supplementary Table 9).

### MAPT locus

Again, Bonferroni correction is unnecessarily conservative when replicating genes known to be associated with PD. A total of 20 SNPs in this locus were nominally associated with PD in our population (lowest  $P=1.9 \times 10^{-3}$  at rs3785880) (Supplementary Table 10). Genotypes at rs1981997 allowed us to differentiate between H1 and H2 haplotypes at this locus as the major (G) and the minor (A) alleles of this SNP are fixed in the H1 and H2 haplotypes, respectively.<sup>9,18</sup> This analysis showed that the H1 haplotype frequency in the Dutch population is 77.64%.

Two-locus haplotype analysis of this SNP and the most associated SNPs in our data set ( $P<0.01$ ) revealed medium to high  $r^2$  values in all SNP pairs (Supplementary Figure 11). A two-locus association analysis showed that, in all instances, positive effect sizes were found in haplotypes in which the G allele of rs1981997 was present (Supplementary Table 11).

### Genome-wide analysis

As expected, none of the SNPs tested reached genome-wide significance (Figure 1). The most significant signal was found on rs7995973, with a  $P$ -value of  $5.41 \times 10^{-6}$  and an OR of 0.72. This SNP is part of an

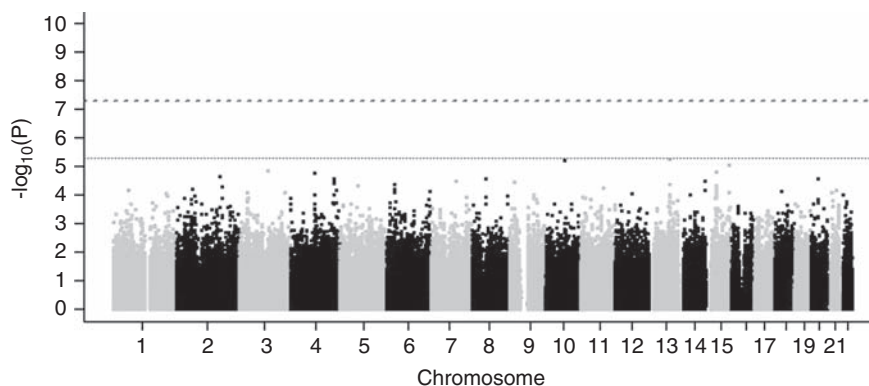
associated locus containing 39 SNPs nominally associated with PD in our population of which eight had a  $P$ -value below  $10^{-3}$  (Supplementary Table 1). This locus is on chromosome 13q31 and contains *LOC729479* and *SPRY2* (Supplementary Figure 12).

In an attempt to replicate these results, logistic regression models were applied to these 39 SNPs from two recently published GWAS datasets in European populations.<sup>5,7</sup> This approach failed to find any association after correcting for 39 independent tests (Supplementary Table 12).

### DISCUSSION

In this study, we present the results of the first GWAS of PD in the Dutch population. Given the population-specific heterogeneity in reported genetic risk factors for PD,<sup>6,7</sup> the first objective of this project was to determine the role, in our population, of the genetic risk factors identified in previous GWAS in European and Japanese populations.<sup>5-7,9</sup> This approach confirmed *SNCA* and *BST1* as PD risk loci. Association of the tested SNPs in *GAK/DGKQ*, *HLA* and *MAPT* loci did not surpass our stringent Bonferroni correction for 30 independent tests. However, given the strong *a priori* for each of these loci, correcting across all the tests carried out is unnecessarily conservative. For this reason and because *post-hoc* analyses revealed the existence of 11, 36 and 20 SNPs nominally associated with PD at these loci (*GAK/DGKQ*, *HLA* and *MAPT*, respectively), we are confident that the signals detected represent a true association with PD.

Although a role of *SNCA* in the risk for PD is well established, it is still unclear whether this association is derived from the 3' or the 5' end of the gene.<sup>6-8,25-32</sup> The data presented here supports the hypothesis that two different signals in these blocks are present in the Dutch population, which is in controversy with previous results.<sup>7,28,30-32</sup> In a report by Mueller *et al*,<sup>25</sup> the authors also detected two association signals in the 3' and 5' blocks of *SNCA*. Interestingly, when stratifying their cohort into males and females, as well as young and old PD groups (according to the median age of 57 years), additional significant association in the intron 4 region of the 5' block was found for the female and the young patients' subgroups. Stratifying our data on the basis of gender and median age at onset/ascertainment did not replicate such differences between either of the subgroups (data not shown). These results do not support gender and age heterogeneity for PD at this locus in the Dutch population, but support the fact that two different variants are exerting a risk for PD in our population.



**Figure 1** Manhattan plot of the results obtained after applying a logistic regression model following an additive mode of association. Age, sex and the two first components of the Multidimensional Scaling (MDS) plot were used as covariates. No SNP reached genome-wide significance level after multiple tests correction.

Collectively, these results indicate that the population attributable risk (PAR) by variation at the *SNCA* locus in the Dutch is larger than previously described in other European populations,<sup>7</sup> as it should be calculated by adding the effect of variation from each independent LD block. When the most associated SNPs at each block are considered, the combined PAR at the *SNCA* locus would be 17.56% (6.39% at the 3' and 11.97% at the 5' block).

Several groups have focused their association studies on polymorphisms in the promoter region of *SNCA*, especially a microsatellite marker ~10 kb upstream of the gene (REP1, D4S3481). Although there are some conflicting results regarding the association of this marker with PD,<sup>33–37</sup> most reports consistently describe a positive association between the longer forms of REP1 and PD.<sup>26,28,38,39</sup> Gene expression differences caused by this repeat have also been described both *in vitro*<sup>35,40</sup> and *in vivo*.<sup>41,42</sup> We agree that examining this microsatellite in our population would help understanding whether the effect detected on the 5' block is a residual effect of variation in REP1 caused by LD between the associated alleles. However, these experiments could not be performed because only genotypic information was available for the control population used in this study.

Results derived from the analyses performed herein, successfully replicate the association at *BST1*. Although association at this locus was originally presented as specific for the Japanese population,<sup>6,7</sup> it has recently been replicated in British<sup>8</sup> and French<sup>11</sup> populations as well. Thus, this is the third study implicating *BST1* on PD risk in a European population.

Although association at this locus was not detected in the GWAS carried out in a combined cohort of German and American samples,<sup>7</sup> separating the data set by country of origin shows that rs12502586 is significantly associated with PD in the Germans, but not in North Americans (*P*-values of  $4.82 \times 10^{-3}$  and 0.179, respectively) (Supplementary Table 13).

These results together point to the hypothesis that variation in *BST1* represents a risk factor for PD in Japanese<sup>6</sup> and European populations<sup>7,8,11</sup> (British, French, German and Dutch populations) and that this effect has been diluted in the North America.<sup>7</sup> A recent report by Hamza *et al*<sup>10</sup> in which a larger cohort of North American samples have been assayed (2000 North American PD cases versus 971 typed in the previous German/North American study), identified two of the six reported SNPs at this locus to be nominally associated with PD. These results point to the hypothesis that the association of *BST1* is, indeed, diluted in North America and that large datasets are needed to detect the effect of variation at this locus.

Given the known European population genetic substructure (specially between northern and southern European populations groups)<sup>43,44</sup> and that American population represents an admixture of individuals with northern and southern European ancestry,<sup>10,45</sup> one would expect allele frequency differences between North American and European populations at the most heterogeneous loci. In case of rs12502586, the allele frequency of the associated allele has been diluted from 11.24 and 10.41% in the Dutch and German controls, respectively, to 9.3% in the North American controls (Supplementary Table 14). Owing to the small size effect of the associated allele, this minor allele frequency changes can lead to differences of statistical power to detect association. Thus, it is not surprising that there is a diminished power to detect association at *BST1* locus in the North American population.

The events that trigger neuronal degeneration in PD remain unknown. Some authors argue that PD may result from inflammatory processes within the nervous system.<sup>46–48</sup> From this perspective, it is

interesting that *BST1* encodes a molecule that facilitates pre-B-cell growth<sup>49</sup> and has an amino acid sequence that is 33% similar to CD38, a key player in inflammatory processes. Although no autoimmune disease has been found to be a convincing risk factor for PD, a possible pathogenic link has been proposed between rheumatoid arthritis and PD.<sup>50–52</sup> It is noteworthy that *BST1* expression is enhanced in bone marrow stromal cell lines derived from patients with rheumatoid arthritis.<sup>49</sup>

Interestingly, another region with genes implicated in neuroinflammation (*HLA* region) has recently been associated with PD.<sup>10</sup> In this report, the authors identified rs3129882 to be associated with PD in their population and chose to name this locus *PARK18*. Although rs3129882 is not associated with PD in our population, *post-hoc* analysis revealed 36 SNPs nominally associated with PD in the Dutch (lowest  $P=4.39 \times 10^{-5}$ ), confirming, for the first time, the role of this locus in PD etiology.

Association between certain *HLA* antigens and PD has largely been suggested since postmortem analysis of the *substantia nigra* (SN) in PD patients revealed the presence of activated microglia.<sup>46</sup> In addition, the frequency of various *HLA* antigens was increased in patients with PD and in patients dying from postencephalitic PD.<sup>53–55</sup> Although PD is clearly not an autoimmune disorder, the occurrence of a localized attack of microglia during the disease course has been demonstrated with epidemiological,<sup>56,57</sup> animal model<sup>58,59</sup> and cell culture<sup>60–62</sup> approaches. Zhang *et al*<sup>63</sup> reported that aggregated  $\alpha$ -synuclein activates microglia which, as a consequence, are toxic toward cultured dopaminergic neurons.

Thus, is it hypothesized that in response to something related to  $\alpha$ -synuclein accumulation, microglial cells are activated to undergo a neuroinflammation process that, ultimately, kills the dopaminergic neurons of the SN.

In a recent GWAS by Pankratz and colleagues, the authors carried out a GWAS focusing only on familial PD cases. With this approach they detected evidence of association to several chromosomal regions including a locus containing *GAK*, *TMEM175* and *DGKQ*, of which *GAK* is of special interest because it was one of the 137 genes shown to be differentially expressed in PD cases in a gene expression profiling approach of parkinsonian SN.<sup>64</sup> In our cohort, only two out of the three SNPs reported to be associated with PD by Pankratz *et al*,<sup>5</sup> were successfully genotyped, one of them (rs1564282) being associated with PD in our population.

Although this association did not surpass our conservative Bonferroni correction, the *a priori* for this hypothesis is strong, as it has been replicated in three independent studies.<sup>8,10,11</sup> Thus, we would at most correct for the SNPs tested within this locus and both of them would be significantly associated with PD in our population ( $P < 0.025$  in both instances, Table 2).

*Post-hoc* analysis revealed the existence of 11 SNPs nominally associated with PD at this locus. Interestingly, the two most associated ones are located in *PCGF3*, which is 78 kb away from *GAK* and 188 kb away from *DGKQ*, and encodes a protein with unknown function containing a C3HC4 type RING finger for protein–protein interaction. Although the function of this transcript is still unknown, it is expressed in the brain. Haplotype analyses also point to this gene as the one carrying the risk variants for the development of PD in the Dutch population, pointing to population-specific heterogeneity for PD in this genomic region with different genes exerting risk in different Caucasian populations. Further genotyping of cases and controls from different populations will help understanding the role of this locus in PD.

No significant association was detected in the *MAPT* locus after multiple tests correction. This is probably because of the relatively

small size of the sample used for these experiments, as this locus is a well-known PD risk locus among Europeans.<sup>5,7,9,23,65,66</sup> If the same hypothesis as that presented for *GAK/DGKQ* is applied, five out of the nine tested SNPs would be borderline significant (in the range of  $10^{-3}$ ) supporting a role of this locus in the development of PD among the Dutch.

In our population, associated alleles at this locus occur in the H1 haplotype. Genotyping of a larger cohort of Dutch PD cases and controls will help confirming these results and defining what H1 subhaplotypes exert the risk for PD in the Dutch.

Last, we failed to find any new PD risk loci in the Dutch population. The most associated locus in our cohort was located in chromosome 13q31 and contained *LOC729479* and *SPRY2*. As the signal detected was not significant after multiple test correction and because we failed to replicate it in two other Caucasian GWAS data sets,<sup>5,7</sup> we refrained from drawing any further conclusions. For the same reason, further conclusions cannot be drawn for other clusters of nominally associated SNPs in regions with interesting biological candidates for PD that did not reach genome-wide significance level. A table showing the top 100 results is displayed in the Supplementary Material (Supplementary Table 10).

In summary, we presented the results of the first GWAS in PD in the Dutch population. These data directly confirmed *SNCA* and *BST1* as PD genes in the Dutch population. Besides, *post-hoc* analysis confirmed a role for *DGKQ/GAK*, *HLA* and *MAPT* loci in modulating the risk for PD in this population.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank the subjects involved in this study for making this work possible. We also want to thank the Hersenstichting Nederland (<http://www.hersenstichting.nl>), the Neuroscience Campus Amsterdam and the section of Medical genomics (MGA) and the Intramural Research Program of the National Institute on Aging (National Institutes of Health, Department of Health and Human services; project number Z01 AG000949-04) for supporting in part the work presented here. Last, we would like to thank the Prinses Beatrix Fonds (<http://www.prinsesbeatrixfonds.nl>) for sponsoring this work.

- Burn DJ, Mark MH, Playford ED *et al*: Parkinson's disease in twins studied with 18F-dopa and positron emission tomography. *Neurology* 1992; **42**: 1894–1900.
- Piccini P, Morrish PK, Turjanski N *et al*: Dopaminergic function in familial Parkinson's disease: a clinical and 18F-dopa positron emission tomography study. *Ann Neurol* 1997; **41**: 222–229.
- Piccini P, Burn DJ, Ceravolo R, Maraganore D, Brooks DJ: The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. *Ann Neurol* 1999; **45**: 577–582.
- Pirkevi C, Lesage S, Brice A, Basak AN: From genes to proteins in mendelian Parkinson's disease: an overview. *Anat Rec (Hoboken)* 2009; **292**: 1893–1901.
- Pankratz N, Wilk JB, Latourelle JC *et al*: Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet* 2009; **124**: 593–605.
- Satake W, Nakabayashi Y, Mizuta I *et al*: Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet* 2009; **41**: 1303–1307.
- Simon-Sanchez J, Schulte C, Bras JM *et al*: Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* 2009; **41**: 1308–1312.
- The UK Parkinson's Disease Consortium and The Wellcome Trust Case Control Consortium 2: Dissection of the genetics of Parkinson's disease identifies an additional association 5' of *SNCA* and multiple associated haplotypes at 7q21. *Hum Mol Genet* 2011; **20**: 345–353.
- Edwards TL, Scott WK, Almonte C *et al*: Genome-wide association study confirms SNPs in *SNCA* and the *MAPT* region as common risk factors for Parkinson disease. *Ann Hum Genet* 2010; **74**: 97–109.
- Hamza TH, Zabetian CP, Tenesa A *et al*: Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat Genet* 2010; **42**: 781–785.
- Saad M, Lesage S, Saint-Pierre A *et al*: Genome-wide association study confirms *BST1* and suggests a locus on 12q24 as risk loci for Parkinson's disease in the European population. *Hum Mol Genet* 2011; **20**: 615–627.
- Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA: Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 1991; **7**: 403–422.
- Hofman A, Breteler MM, van Duijn CM *et al*: The Rotterdam Study: objectives and design update. *Eur J Epidemiol* 2007; **22**: 819–829.
- Hofman A, Breteler MM, van Duijn CM *et al*: The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* 2009; **24**: 553–572.
- Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- Gabriel SB, Schaffner SF, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
- Stefansson H, Helgason A, Thorleifsson G *et al*: A common inversion under selection in Europeans. *Nat Genet* 2005; **37**: 129–137.
- Golbe LI, Lazzarini AM, Spychala JR *et al*: The tau A0 allele in Parkinson's disease. *Mov Disord* 2001; **16**: 442–447.
- Maraganore DM, Hernandez DG, Singleton AB *et al*: Case-control study of the extended tau gene haplotype in Parkinson's disease. *Ann Neurol* 2001; **50**: 658–661.
- Skipper L, Wilkes K, Toft M *et al*: Linkage disequilibrium and association of *MAPT* H1 in Parkinson disease. *Am J Hum Genet* 2004; **75**: 669–677.
- Zhang J, Song Y, Chen H, Fan D: The tau gene haplotype h1 confers a susceptibility to Parkinson's disease. *Eur Neurol* 2005; **53**: 15–21.
- Zabetian CP, Hutter CM, Factor SA *et al*: Association analysis of *MAPT* H1 haplotype and subhaplotypes in Parkinson's disease. *Ann Neurol* 2007; **62**: 137–144.
- Tobin JE, Latourelle JC, Lew MF *et al*: Haplotypes and gene expression implicate the *MAPT* region for Parkinson disease: the GenePD Study. *Neurology* 2008; **71**: 28–34.
- Mueller JC, Fuchs J, Hofer A *et al*: Multiple regions of alpha-synuclein are associated with Parkinson's disease. *Ann Neurol* 2005; **57**: 535–541.
- Maraganore DM, de Andrade M, Elbaz A *et al*: Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *JAMA* 2006; **296**: 661–670.
- Mizuta I, Satake W, Nakabayashi Y *et al*: Multiple candidate gene analysis identifies alpha-synuclein as a susceptibility gene for sporadic Parkinson's disease. *Hum Mol Genet* 2006; **15**: 1151–1158.
- Winkler S, Hagenah J, Lincoln S *et al*: Alpha-synuclein and Parkinson disease susceptibility. *Neurology* 2007; **69**: 1745–1750.
- Kay DM, Factor SA, Samii A *et al*: Genetic association between alpha-synuclein and idiopathic Parkinson's disease. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1222–1230.
- Mylhre R, Toft M, Kachergus J *et al*: Multiple alpha-synuclein gene polymorphisms are associated with Parkinson's disease in a Norwegian population. *Acta Neurol Scand* 2008; **118**: 320–327.
- Rajput A, Vilarino-Guell C, Rajput ML *et al*: Alpha-synuclein polymorphisms are associated with Parkinson's disease in a Saskatchewan population. *Mov Disord* 2009; **24**: 2411–2414.
- Sotiriou S, Gibney G, Baxevas AD, Nussbaum RL: A single nucleotide polymorphism in the 3'UTR of the *SNCA* gene encoding alpha-synuclein is a new potential susceptibility locus for Parkinson disease. *Neurosci Lett* 2009; **461**: 196–201.
- Izumi Y, Morino H, Oda M *et al*: Genetic studies in Parkinson's disease with an alpha-synuclein/NACP gene polymorphism in Japan. *Neurosci Lett* 2001; **300**: 125–127.
- Khan N, Graham E, Dixon P *et al*: Parkinson's disease is not associated with the combined alpha-synuclein/apolipoprotein E susceptibility genotype. *Ann Neurol* 2001; **49**: 665–668.
- Holzmann C, Kruger R, Saecker AM *et al*: Polymorphisms of the alpha-synuclein promoter: expression analyses and association studies in Parkinson's disease. *J Neural Transm* 2003; **110**: 67–76.
- Spadafora P, Annesi G, Pasqua AA *et al*: NACP-REP1 polymorphism is not involved in Parkinson's disease: a case-control study in a population sample from southern Italy. *Neurosci Lett* 2003; **351**: 75–78.
- Tan EK, Chai A, Teo YY *et al*: Alpha-synuclein haplotypes implicated in risk of Parkinson's disease. *Neurology* 2004; **62**: 128–131.
- Farrer M, Maraganore DM, Lockhart P *et al*: alpha-Synuclein gene haplotypes are associated with Parkinson's disease. *Hum Mol Genet* 2001; **10**: 1847–1851.
- Pals P, Lincoln S, Manning J *et al*: alpha-Synuclein promoter confers susceptibility to Parkinson's disease. *Ann Neurol* 2004; **56**: 591–595.
- Chiba-Falek O, Nussbaum RL: Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (*SNCA*) on transcription in a cell culture luciferase reporter system. *Hum Mol Genet* 2001; **10**: 3101–3109.
- Cronin KD, Ge D, Manninger P *et al*: Expansion of the Parkinson disease-associated *SNCA*-Rep1 allele upregulates human alpha-synuclein in transgenic mouse brain. *Hum Mol Genet* 2009; **18**: 3274–3285.
- Linnertz C, Saucier L, Ge D *et al*: Genetic regulation of alpha-synuclein mRNA expression in various human brain tissues. *PLoS One* 2009; **4**: e7480.
- Seldin MF, Shigeta R, Villoslada P *et al*: European population substructure: clustering of northern and southern populations. *PLoS Genet* 2006; **2**: e143.
- Moskvina V, Smith M, Ivanov D *et al*: Genetic differences between five European populations. *Hum Hered* 2010; **70**: 141–149.

- 45 Price AL, Butler J, Patterson N *et al*: Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 2008; **4**: e236.
- 46 McGeer PL, Itagaki S, Boyes BE, McGeer EG: Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains. *Neurology* 1988; **38**: 1285–1291.
- 47 Hirsch EC, Hunot S: Neuroinflammation in Parkinson's disease: a target for neuroprotection? *Lancet Neurol* 2009; **8**: 382–397.
- 48 Røgeberg K, Friis S, Ritz B, Schernhammer ES, Korbo L, Olsen JH: Autoimmune disease and risk for Parkinson disease: a population-based case-control study. *Neurology* 2009; **73**: 1462–1468.
- 49 Kaisho T, Ishikawa J, Oritani K *et al*: BST-1, a surface molecule of bone marrow stromal cell lines that facilitates pre-B-cell growth. *Proc Natl Acad Sci USA* 1994; **91**: 5325–5329.
- 50 Hrycaj P, Korczowska I, Lacki JK: Severe Parkinson's disease in rheumatoid arthritis patient treated with infliximab. *Rheumatology (Oxford)* 2003; **42**: 702–703.
- 51 Melikoglu MA, Sezer I, Kacar C: Rheumatoid-like hand deformities in Parkinson disease. *J Clin Rheumatol* 2007; **13**: 236–237.
- 52 Kogure T, Tatsumi T, Kaneko Y, Okamoto K: Rheumatoid arthritis accompanied by Parkinson disease. *J Clin Rheumatol* 2008; **14**: 192–193.
- 53 Emile J, Truelle JL, Pouplard A, Hurez D: Association of Parkinson's disease with HLA-B17 and B18 antigens. *Nouv Presse Med* 1977; **6**: 4144.
- 54 Eliazan TS, Terasaki PI, Yahr MD: HLA-B14 antigen and postencephalitic Parkinson's disease. Their association in an American-Jewish ethnic group. *Arch Neurol* 1980; **37**: 542–544.
- 55 Lampe JB, Gossrau G, Herting B *et al*: HLA typing and Parkinson's disease. *Eur Neurol* 2003; **50**: 64–68.
- 56 Chen H, Zhang SM, Hernan MA *et al*: Nonsteroidal anti-inflammatory drugs and the risk of Parkinson disease. *Arch Neurol* 2003; **60**: 1059–1064.
- 57 Esposito E, Di Matteo V, Benigno A, Pierucci M, Crescimanno G, Di Giovanni G: Non-steroidal anti-inflammatory drugs in Parkinson's disease. *Exp Neurol* 2007; **205**: 295–312.
- 58 McGeer PL, Schwab C, Parent A, Doudet D: Presence of reactive microglia in monkey substantia nigra years after 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine administration. *Ann Neurol* 2003; **54**: 599–604.
- 59 Barcia C, Sanchez Bahillo A, Fernandez-Villalba E *et al*: Evidence of active microglia in substantia nigra pars compacta of parkinsonian monkeys 1 year after MPTP exposure. *Glia* 2004; **46**: 402–409.
- 60 Gao HM, Hong JS, Zhang W, Liu B: Distinct role for microglia in rotenone-induced degeneration of dopaminergic neurons. *J Neurosci* 2002; **22**: 782–790.
- 61 Lee DY, Oh YJ, Jin BK: Thrombin-activated microglia contribute to death of dopaminergic neurons in rat mesencephalic cultures: dual roles of mitogen-activated protein kinase signaling pathways. *Glia* 2005; **51**: 98–110.
- 62 Wang XJ, Yan ZQ, Lu GQ, Stuart S, Chen SD: Parkinson disease IgG and C5a-induced synergistic dopaminergic neurotoxicity: role of microglia. *Neurochem Int* 2007; **50**: 39–50.
- 63 Zhang W, Wang T, Pei Z *et al*: Aggregated alpha-synuclein activates microglia: a process leading to disease progression in Parkinson's disease. *FASEB J* 2005; **19**: 533–542.
- 64 Grunblatt E, Mandel S, Jacob-Hirsch J *et al*: Gene expression profiling of parkinsonian substantia nigra pars compacta; alterations in ubiquitin-proteasome, heat shock protein, iron and oxidative stress regulated proteins, cell adhesion/cellular matrix and vesicle trafficking genes. *J Neural Transm* 2004; **111**: 1543–1573.
- 65 Fung HC, Xiromerisiou G, Gibbs JR *et al*: Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts. *Neurodegener Dis* 2006; **3**: 327–333.
- 66 Vandrovicova J, Pittman AM, Malzer E *et al*: Association of MAPT haplotype-tagging SNPs with sporadic Parkinson's disease. *Neurobiol Aging* 2009; **30**: 1477–1482.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)