Article

Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population

SAAD, Mohamad, *et al*.

Abstract

We performed a three-stage genome-wide association study (GWAS) to identify common Parkinson's disease (PD) risk variants in the European population. The initial genome-wide scan was conducted in a French sample of 1039 cases and 1984 controls, using almost 500 000 single nucleotide polymorphisms (SNPs). Two SNPs at SNCA were found to be associated with PD at the genome-wide significance level (P < 3 × 10(-8)). An additional set of promising and new association signals was identified and submitted for immediate replication in two independent case-control studies of subjects of European descent. We first carried out an in silico replication study using GWAS data from the WTCCC2 PD study sample (1705 cases, 5200 WTCCC controls). Nominally replicated SNPs were further genotyped in a third sample of 1527 cases and 1864 controls from France and Australia. We found converging evidence of association with PD on 12q24 (rs4964469, combined P = 2.4 × 10(-7)) and confirmed the association on 4p15/BST1 (rs4698412, combined P = 1.8 × 10(-6)), previously reported in Japanese data. The 12q24 locus includes RFX4, an isoform of which, [...]

Reference

SAAD, Mohamad, *et al*. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Human molecular genetics*, 2011, vol. 20, no. 3, p. 615-27

DOI : 10.1093/hmg/ddq497
PMID : 21084426

Human Molecular Genetics, 2011, Vol. 20, No. 3 **615–627**
doi:10.1093/hmg/ddq497
Advance Access published on November 17, 2010

Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population

Mohamad Saad[1,2,†], Suzanne Lesage[3,4,5,†], Aude Saint-Pierre[1,2], Jean-Christophe Corvol[3,4,5,6], Diana Zelenika[7], Jean-Charles Lambert[8,9,10], Marie Vidailhet[3,4,5], George D. Mellick[11,12], Ebba Lohmann[3,4,5], Franck Durif[13], Pierre Pollak[14], Philippe Damier[15], François Tison[16], Peter A. Silburn[11,12], Christophe Tzourio[17,18], Sylvie Forlani[3,4,5], Marie-Anne Loriot[19,20,21], Maurice Giroud[22], Catherine Helmer[23], Florence Portet[24], Philippe Amouyel[8,9,10,25], Mark Lathrop[7], Alexis Elbaz[17,18], Alexandra Durr[3,4,5,26], Maria Martinez[1,2,*] and Alexis Brice[3,4,5,26,*] for the French Parkinson's Disease Genetics Study Group[‡]

[1]INSERM U563, CPTP, CHU Purpan, 31024 Toulouse, France, [2]Paul Sabatier University, Toulouse, France, [3]Université Pierre et Marie Curie-Paris6, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, UMR-S975, Paris, France, [4]INSERM U975, Paris, France, [5]CNRS, UMR 7225, Paris, France, [6]INSERM CIC-9503, Hôpital Pitié-Salpêtrière, Paris, France, [7]Centre National de Génotypage, Institut Génomique, Commissariat à l'Energie Atomique, Evry, France, [8]INSERM U744, Lille, France, [9]Institut Pasteur de Lille, Lille, France, [10]Université de Lille Nord, Lille, France, [11]National Centre for Adult Stem Cell Research, Eskitis Institute for Cell and Molecular Therapies, Griffith University, Brisbane, Queensland, Australia, [12]Department of Neurology, Princess Alexandra Hospital, Brisbane, Queensland, Australia, [13]Service de Neurologie, Hôpital Gabriel Montpied, Clermont-Ferrand, France, [14]Service de Neurologie, CHU de Grenoble, Grenoble, France, [15]CHU Nantes, CIC0004, Service de Neurologie, Nantes, France, [16]Service de Neurologie, Hôpital Haut-Lévêque, Pessac, France, [17]INSERM U708, Paris, France, [18]Université Pierre et Marie Curie Paris6, Paris, France, [19]UMR-S775, Paris, France, [20]Université Paris Descartes, Paris, France, [21]AP-HP, Hôpital Européen Georges Pompidou, Paris, France, [22]Centre Hospitalier Dijon, Dijon, France, [23]INSERM, CR897, Université Victor Segalen Bordeaux-2, Bordeaux, France, [24]INSERM U888, Montpellier, France, [25]CHRU de Lille, Lille, France, and [26]Department of Genetics and Cytogenetics, AP-HP, Pitié-Salpêtrière Hospital, Paris, France

Received July 16, 2010; Revised and Accepted November 10, 2010

We performed a three-stage genome-wide association study (GWAS) to identify common Parkinson's disease (PD) risk variants in the European population. The initial genome-wide scan was conducted in a French sample of 1039 cases and 1984 controls, using almost 500 000 single nucleotide polymorphisms (SNPs). Two SNPs at SNCA were found to be associated with PD at the genome-wide significance level ($P < 3 \times 10^{-8}$). An additional set of promising and new association signals was identified and submitted for immediate replication in two independent case–control studies of subjects of European descent. We first carried out an *in silico* replication study using GWAS data from the WTCCC2 PD study sample (1705 cases, 5200 WTCCC controls). Nominally replicated SNPs were further genotyped in a third sample of 1527 cases and 1864 controls from France and Australia. We found converging evidence of association with PD on 12q24 (rs4964469,

*To whom correspondence should be addressed. Tel: +33 562744587; Fax: +33 562744558; Email: maria.martinez@inserm.fr (M.M.); Tel: +33 142162189; Fax: +33 144243658; Email: alexis.brice@upmc.fr (A.B.)
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
‡The French Parkinson's Disease Genetics Study Group includes Y. Agid, M. Anheim, A.-M. Bonnet, M. Borg, A. Brice, E. Broussolle, J.-C. Corvol, Ph. Damier, A. Destée, A. Dürr, F. Durif, S. Klebe, E. Lohmann, M. Martinez, C. Penet, P. Pollak, O. Rascol, F. Tison, C. Tranchant, M. Vérin, F. Viallet and M. Vidailhet.

combined $P = 2.4 \times 10^{-7}$) and confirmed the association on 4p15/BST1 (rs4698412, combined $P = 1.8 \times 10^{-6}$), previously reported in Japanese data. The 12q24 locus includes RFX4, an isoform of which, named *RFX4_v3*, encodes brain-specific transcription factors that regulate many genes involved in brain morphogenesis and intracellular calcium homeostasis.

INTRODUCTION

Parkinson's disease (PD) is the second most common degenerative disease, affecting 1–2% of individuals older than 65 years. Clinical features of PD result primarily from the loss of dopaminergic neurons in the substantia nigra. Although the common form of PD is sporadic, six genes have been identified, mainly by linkage analyses of Mendelian forms of the disease. Two genes, SNCA (encoding α-synuclein) and LRRK2, have an autosomal dominant inheritance and four other genes, PARK2 (parkin), PARK6 (PINK1), PARK7 (DJ-I) and PARK13 (ATP13A2), have an autosomal recessive inheritance (1). Frequently, mutations in these genes are found in patients with early-onset PD, particularly those with autosomal recessive inheritance. However, in most populations, Mendelian forms of parkinsonism are rare when compared with the most common form of PD, a frequent and complex disorder probably explained by the interaction between genetic and environmental factors.

The first two genome-wide association studies (GWASs) in PD (2,3) provided evidence of association with several loci but most often not at the genome-wide significant level, and most initial association findings were not confirmed by subsequent replication analyses (4). Two recent GWASs (5,6) reported strong or genome-wide significant associations with one or more of the known PD genes (i.e. SNCA, MAPT and/or LRRK2). So far, only two 'new' loci have been identified, 1q32/PARK16 and 4p15/BST1, in the Japanese data (5). The US/UK/German GWAS (6) replicated positive association with variants at PARK16 but failed to replicate the association at BST1.

To identify additional variants that affect PD risk in the European population, we designed a three-stage GWAS of PD in three case–control samples from France, the UK and Australia (total of >13 300 subjects). A set of 50 top association signals was identified in the scan sample (1039 cases and 1984 controls from France) using the Illumina-610Quad chip. Promising and new signals were followed-up for stepwise replication in two further UK and French/Australian case–control studies (>3200 cases and 7000 controls).

RESULTS

The genome-wide association results from logistic test corrected for genomic inflation (GC) revealed two single nucleotide polymorphisms (SNPs) with $P_{GC} < 10^{-7}$, and a substantial number of SNPs with strong ($P_{GC} < 10^{-4}$) evidence of association (Fig. 1 and Table 1). For practical reasons, we focus our attention on the 50 best-associated SNPs to prioritize for immediate *in silico* replication (Table 1). Secondary logistic analyses, adjusted for the first two principal componenets (PCs) led to similar rank order of SNPs, albeit slightly weaker association signals (Table 1). This suggests that the significant results, revealed by our primary analyses, are not biased by residual population substructure within our French scan sample. The 50 best-associated SNPs spanned 23 distinct genomic loci, and were associated with $P_{GC} < 5.6 \times 10^{-5}$. Sixteen associations were found within two well-known PD genes, i.e. SNCA (4q22, 4 SNPs) and MAPT (17q12–q21, 11 SNPs), or within BST1 (4p15, one SNP), a recently reported PD risk locus established at the genome-wide significance level in a Japanese population (5). The remaining 34 SNPs were located in 20 distinct previously unreported putative PD loci. The two genome-wide significant SNPs were located on 4q22/SNCA [rs356220, $P_{GC} = 2.82 \times 10^{-8}$, OR = 1.37; 95% CI (1.22–1.53) and rs2736990, $P_{GC} = 2.88 \times 10^{-8}$, OR = 1.35 (1.22–1.50)]. The next most significant SNP was on chromosome 12q21/LOC401725 [rs7954761, $P_{GC} = 2.09 \times 10^{-7}$, OR = 1.34 (1.20–1.50)].

The 50 top SNPs were tested for *in silico* replication in the WTCCC2 PD study data (Table 1). For the sake of clarity, OR values are reported as a function of the number of risk alleles as identified in the stage-1 data. Associations for all 15 SNPs in SNCA and MAPT genes were replicated at nominal P-values $< 4 \times 10^{-5}$. Association with the BST1 variant was also replicated but at a weaker significance level (OR = 1.08, $P = 0.025$). For all SNPs at SNCA, MAPT and BST1, the results in the French scan and the UK replication samples were highly congruent in terms of risk alleles and allele frequencies. As expected, ORs estimated in our scan study tend to be higher than those obtained in the replication-stage data, especially for BST1. Of the remaining 20 loci, association signals at three loci (four SNPs) were replicated with nominal $P < 5\%$ and with the same direction of effect. These SNPs were located on chromosomes 2q21.3 (rs621341, OR = 1.08, $P = 0.028$ and rs6723108, OR = 1.11, $P = 0.005$), 12p13.3 (rs11064524, OR = 1.08, $P = 0.045$) and 12q24 (rs4964469, OR = 1.11, $P = 0.0045$). Differences in allele frequencies across the data from France and UK were notable for the 2q21–q22 SNPs. Indeed, the region encompasses the LCT (lactase) gene whose SNPs are known to vary in frequency across Europe, and rs6723108 has been shown to have different allele frequencies in the French and the UK–Irish populations (7).

We further followed-up the five replicated SNPs (from three newly identified loci and from BST1) in the second replication dataset (1527 cases and 1864 controls from France and Australia) (Table 2). In stage 3, evidence of association was assessed with the Mantel–Haenszel test to control for the potential confounding owing to the different geographical origins (France versus Australia). Evidence of association was replicated for two SNPs located on 12q24 (rs4964469, OR = 1.12, $P = 0.0175$) and on 4p15/BST1

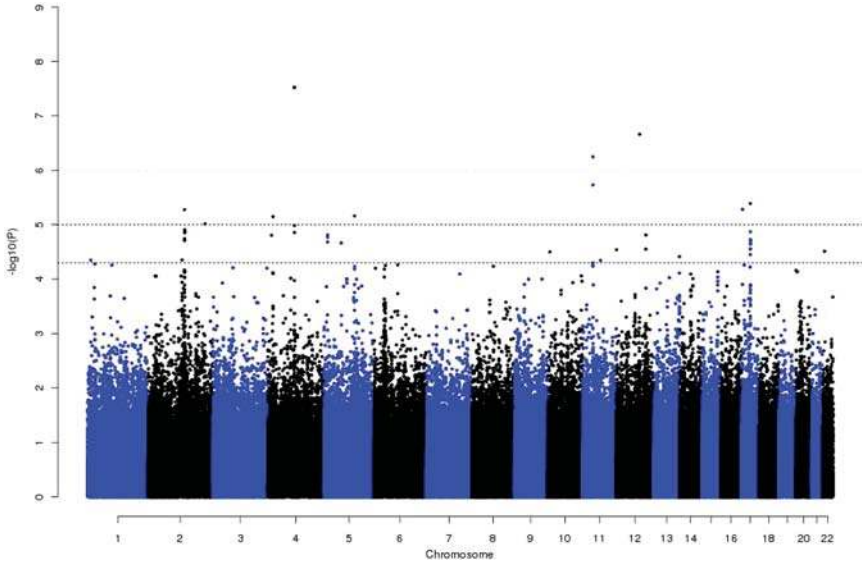

Figure 1. Manhattan plot of the genome-wide association results for 492 929 SNPs. Logistic analysis corrected for genomic inflation (GC results).

(rs4698412, OR = 1.10, $P = 0.029$). Association signals from joint analysis of the two replication datasets were improved for the same two SNPs only: at 4p15/BST1 (stage 2 + stage 3, $P = 0.0033$) and at 12q24 (stage 2 + stage 3, $P = 0.00036$) loci. Notably, joint analysis of the three datasets showed a consistently greater support for association for the newly identified locus on chromosome 12q24 ($P = 2.38 \times 10^{-7}$) than for 4p15/BST1 ($P = 1.79 \times 10^{-6}$). Additional analyses showed that our initial association signals were not confounded by age and they did not appear to be driven either by the subgroup of cases having an early age (<50) of onset of the disease or by those having a positive family history of PD (results not shown). The population-attributable risk (PAR) associated with SNCA, MAPT, BST1 and the 12q24 locus estimated in stage-1 data was 11, 20, 13 and 8%, respectively; in the combined data, PAR was 7% and 4% for BST1 and 12q24, respectively.

Finally, we also examined 18 SNPs from five loci, previously reported to be associated with PD at a genome-wide significance level from two published GWASs of PD (5,6) (Table 3). We added three SNPs (at SNCA and BST1) that were found strongly associated in our stage-1 data. The table also shows the results for a suggestive PD risk locus (GAK) reported by the published GWAS of PD from familial cases (8). As for the previously reported PD loci, two loci only (SNCA and MAPT) have been identified with genome-wide significance at the screen stage: SNCA in both the Japanese and European populations and MAPT in the European population only. The two new PD risk loci (BST1 and PARK16) were identified in the Japanese population: association signals were strong ($P < 10^{-6}$) in the discovery sample, and exceeded $P < 10^{-8}$ in the combined data (5). As already reported here, our GWAS provided genome-wide significance for two SNCA variants and replicated positive associations for variants at MAPT and BST1 loci. It is worth noting that allele frequencies may differ markedly between the Japanese and the European datasets, especially for SNPs at the SNCA and BST1 loci. Saliently, the directions of effects (i.e. risk allele) and effect sizes at SNCA variants are rather congruent across the European and Japanese datasets. For the remaining three PD loci, evidence of association was nominal (PARK16, $P_{GC} = 0.03$; LRKK2, $P_{GC} = 0.04$; GAK $P_{GC} = 0.008$) in the France-GWAS data.

DISCUSSION

Our genome-wide association analyses in the French scan data revealed two SNPs with genome-wide significance ($P_{GC} < 10^{-7}$), and a number of additional SNPs with suggestive evidence ($P_{GC} < 10^{-4}$). Here, we focused on the 50 top associated SNPs for immediate replication in two independent case–control samples. We used a stepwise replication design. To refine the set of most promising results, we first conducted *in silico* replication for the 50 SNPs in the WTCCC2 PD data (1705 cases and 5200 WTCCC controls). Replicated SNPs were genotyped and tested in a third dataset of 1527 cases and 1864 controls from France and Australia. Our scan stage showed genome-wide significance of association with PD for two SNPs at the 4q22/SNCA locus ($P_{GC} < 2.88 \times 10^{-8}$). Indeed, out of the 50 top associated SNPs, 15 are located in genomic regions of two known PD genes (SCNCA, MAPT) and one is located on 4p15/BST1, a risk locus recently reported with genome-wide significance in Japanese samples. SNPs at SNCA and MAPT were all significantly associated with PD in the UK-GWAS data (SNCA, $P < 8 \times 10^{-5}$; MAPT, $P < 2.75 \times 10^{-6}$). Evidence of association with 4p15/BST1 was also replicated in the UK sample but at a lower (rs4698412, $P = 0.025$) significance level. Out of the remaining 34 SNPs, four SNPs (three loci) showed significant ($P < 0.05$) and consistent evidence of association in the UK data. A total of five SNPs (four loci: 2q21.3, 12p13.3, 12q24 and BST1) were followed-up for replication in the third case–control sample. Two of the four tested regions were replicated: 4p15/BST1 ($P = 0.03$) and 12q24

Table 1. Top 50 SNPs in scan stage and *in silico* replication results

Chromosome (gene)	Position (bp)	SNP	Stage-1: scan (France) data					Stage-2: replication (UK) data		
			RA[a]	RAF[b]	OR	P_{GC} (two-tailed)[c]	P_{2PCs} (two-tailed)[d]	RAF	OR	P (one-tailed)[e]
Known PD genes/previously published loci										
4q22 (SNCA)	90858538	rs11931074	T	0.07	1.52	1.35E-05	9.04E-05	0.07	1.33	4.01E-05
	90860363	rs356220	T	0.35	1.37	2.82E-08	6.26E-07	0.36	1.27	2.59E-09
	90894261	rs3857059	G	0.07	1.54	1.00E-05	6.32E-05	0.07	1.33	3.95E-05
	90897564	rs2736990	G	0.44	1.35	2.88E-08	1.32E-07	0.45	1.24	3.98E-08
17q12–21 (MAPT)	41074926	rs393152	A	0.75	1.32	2.68E-05	1.43E-04	0.76	1.31	2.20E-08
	41279463	rs12185268	A	0.76	1.32	3.44E-05	1.62E-04	0.76	1.30	3.59E-08
	41279910	rs12373139	G	0.76	1.33	1.81E-05	7.60E-05	0.76	1.30	2.77E-08
	41281077	rs17690703	C	0.72	1.34	3.94E-06	6.61E-06	0.72	1.24	1.37E-06
	41347100	rs17563986	A	0.75	1.34	1.30E-05	5.65E-05	0.76	1.31	2.58E-08
	41412603	rs1981997	G	0.76	1.33	2.20E-05	8.81E-05	0.76	1.30	4.61E-08
	41436901	rs8070723	A	0.75	1.33	2.19E-05	8.91E-05	0.76	1.30	2.61E-08
	41544850	rs7225002	A	0.59	1.27	2.72E-05	4.08E-05	0.57	1.23	1.14E-07
	41602941	rs2532274	A	0.75	1.33	2.21E-05	1.06E-04	0.75	1.28	2.92E-07
	41605885	rs2532269	T	0.75	1.33	1.90E-05	8.58E-05	0.76	1.29	1.11E-07
	41648797	rs2668692	G	0.76	1.33	1.97E-05	8.20E-05	0.76	1.29	1.22E-07
4p15 (BST1)	15346446	rs4698412	A	0.52	1.28	6.88E-06	1.96E-06	0.55	1.08	0.0247
Newly identified loci										
1p36.22	11880226	rs12724129	T	0.34	1.26	4.35E-05	4.45E-05	0.40	0.99	[f]
1p36.11	26448619	rs10902724	A	0.05	1.55	5.13E-05	5.00E-04	0.07	0.96	[f]
2q14.3	126112282	rs1365783	G	0.42	1.26	4.33E-05	9.10E-05	0.46	0.94	[f]
2q21.3	135011650	rs621341	T	0.28	1.30	5.11E-06	4.37E-04	0.48	1.08	0.0277
	135196450	rs6723108	G	0.31	1.25	6.85E-05	3.20E-03	0.51	1.11	0.0053
	135318626	rs6729702	G	0.46	1.26	1.75E-05	6.02E-04	0.64	1.04	0.15
	135339278	rs6430552	C	0.46	1.26	1.89E-05	6.95E-04	0.64	1.05	0.14
	135367236	rs6714498	T	0.46	1.26	1.84E-05	6.20E-04	0.64	1.05	0.14
2q22	136611978	rs4954564	A	0.52	1.27	1.21E-05	4.07E-04	0.74	1.03	0.27
	136722668	rs6430612	T	0.41	1.27	1.35E-05	1.89E-03	0.65	1.03	0.23
	136730076	rs10221893	T	0.41	1.27	1.31E-05	1.88E-03	0.65	1.03	0.22
2q35	216463846	rs6741233	C	0.87	1.51	9.34E-06	4.46E-04	0.93	1.01	0.46
4p16	11054284	rs368039	A	0.11	1.41	1.52E-05	2.11E-05	0.13	0.86	[f]
5p15.2	10016889	rs1428954	G	0.53	1.27	1.65E-05	3.73E-04	0.57	1.04	0.18
	10026935	rs10072891	G	0.53	1.27	2.02E-05	6.68E-04	0.57	1.04	0.16
	10037418	rs38065	A	0.65	1.29	1.50E-05	2.19E-04	0.69	1.02	0.33
5q12.1	60896208	rs1423326	T	0.60	1.27	2.10E-05	6.54E-04	0.65	1.02	0.36
5q22.2	112814742	rs26990	C	0.13	1.41	6.67E-06	8.67E-06	0.19	0.94	[f]
6q12	70963882	rs9360414	T	0.38	1.25	5.34E-05	4.00E-05	0.38	1.05	0.13
10p14	6933911	rs10905042	C	0.06	1.53	3.08E-05	3.82E-04	0.07	1.02	0.42
11p12	36589978	rs12419750	A	0.89	1.47	4.96E-05	1.65E-05	0.90	0.97	[f]
	36600652	rs1391542	A	0.89	1.47	5.44E-05	1.89E-05	0.90	0.98	[f]
	36613848	rs7128419	A	0.89	1.47	4.91E-05	1.71E-05	0.90	0.98	[f]
	36618299	rs12271660	A	0.90	1.50	5.64E-05	4.56E-05	0.92	0.96	[f]
	36684837	rs12294719	T	0.79	1.44	5.42E-07	6.72E-07	0.81	1.00	0.48
	36687460	rs1533588	A	0.79	1.41	1.79E-06	3.78E-06	0.82	0.97	[f]
11q13.5	75709727	rs12295401	T	0.06	1.53	4.40E-05	5.16E-05	0.06	1.04	0.29
12p13.3	760163	rs11064524	G	0.20	1.32	2.80E-05	1.21E-04	0.24	1.08	0.0447
12q21.31	82691472	rs7954761	A	0.60	1.34	2.09E-07	2.59E-07	0.61	0.99	[f]
12q24	105474117	rs4964469	A	0.33	1.27	2.73E-05	1.30E-04	0.37	1.11	0.0045
	105513235	rs1035767	T	0.11	1.42	1.50E-05	2.15E-05	0.11	0.98	[f]
13q34	113253980	rs2259599	G	0.83	1.39	3.74E-05	5.61E-03	0.88	0.96	[f]
17p13.2	4376339	rs9899558	G	0.73	1.34	5.04E-06	3.82E-04	0.77	0.98	[f]
22q11.23	22917303	rs9608247	A	0.16	1.33	2.99E-05	9.67E-05	0.17	0.95	[f]

[a]Risk allele in stage-1 data.
[b]Risk allele frequency in controls.
[c]Logistic tests corrected for genomic inflation.
[d]Logistic tests including 2PCs as covariates.
[e]*P*-values shown when the direction of effect in stage-1 and stage-2 data are consistent.
[f]One-tailed $P > 0.5$.

($P = 0.018$). Of the four regions, only one (12p13.3) showed no evidence of association from the combined analysis of the two replication datasets. Overall, evidence of association was consistently stronger with the region of the newly identified PD risk locus than with BST1, in each replication sample as well as in the combined (genome-wide and two replication samples) data (12q24, $P = 2.38 \times 10^{-7}$; BST1, $P = 1.79 \times 10^{-6}$).

Table 2. GWAS and replication: loci considered to follow-up

Locus	Stage *n* (case/control) Position (bp)	SNP	Stage-1 (France) (1039/1984) RA[a]	RAF[b]	%OR (95% CI)	*P* (two-tailed)	Stage-2 (UK) (1705/5200) RAF	OR (95% CI)	*P* (one-tailed)[c]	Stage-3 (France/Australia) (1527/1864) RAF	OR (95% CI)	*P* (one-tailed)[c,d]	Combined Stage 2 + 3 (3232/7064) OR (95% CI)	*P* (one-tailed)[c,d]	Stage 1 + 2 + 3 (4271/9048) OR (95% CI)	*P* (two-tailed)[d]
BST1	15346446	rs4698412	A	0.52	1.28 (1.15–1.42)	6.9E-06	0.55	1.08 (1.00–1.17)	0.0247	0.54	1.10 (1.00–1.21)	0.0290	1.09 (1.02–1.16)	0.00333	1.14 (1.08–1.20)	1.79E-06
2q21.3	135011650	rs6721341	T	0.28	1.30 (1.16–1.46)	5.1E-06	0.48	1.08 (1.00–1.17)	0.0277	0.35	1.02 (0.92–1.13)	0.3399	1.06 (1.00–1.13)	0.03500	1.12 (1.04–1.18)	7.50E-05
	135196450	rs6723108	G	0.31	1.25 (1.12–1.40)	6.9E-05	0.51	1.11 (1.02–1.20)	0.0053	0.38	1.03 (0.93–1.14)	0.3047	1.08 (1.01–1.15)	0.00916	1.12 (1.06–1.18)	2.99E-05
12p13.3	760163	rs11064524	G	0.20	1.32 (1.16–1.50)	2.8E-05	0.24	1.08 (0.99–1.18)	0.0447	0.25	0.93 (0.83–1.04)	[e]	–	–	–	–
12q24	105474117	rs4964469	A	0.33	1.27 (1.13–1.41)	2.7E-05	0.37	1.11 (1.03–1.21)	0.0045	0.34	1.12 (1.01–1.24)	0.0175	1.12 (1.05–1.19)	0.00036	1.16 (1.09–1.22)	2.38E-07

[a]Risk allele in stage-1 data.
[b]Risk allele frequency in controls; %odds ratio computed for the stage-1 risk allele.
[c]*P*-values shown when the direction of effect in stage-1 and each replication data are consistent.
[d]*P*-values from stratified association tests.
[e]One-tailed $P > 0.5$.

The evidence of association ($P_{GC} < 1.35 \times 10^{-5}$, Tables 1 and 3) that we detected with several SNPs in the 3′ block of linkage disequilibrium (LD) of the SNCA locus (Fig. 2A), including the two SNPs reaching genome-wide significance in our scan sample, is highly consistent with previous PD GWAS studies (5,6).

MAPT is located in a large block of LD on chromosome 17q12–q22, which contains several additional genes (Fig. 2B). Previous studies (9,10) have identified a large haplotypic block associated with PD, with H1 and H2 being the at-risk and the protective haplotype, respectively. Our two most associated SNPs in the 17q12–q22 region are located within this haplotypic block: rs17690703 ($P_{GC} = 3.9 \times 10^{-6}$) and rs17563986 ($P_{GC} = 1.3 \times 10^{-5}$), the latter being at MAPT. In addition, H2 is tagged by the minor alleles of four of our genotyped SNPs: rs12185268/G, rs12373139/A, rs1981997/A and rs8070723/G. In our scan data, we found the same H1/H2 association signals, with all minor alleles of these four SNPs being significantly associated ($P_{GC} < 3.44 \times 10^{-5}$) with a decreased risk of PD (Table 1).

The BST1 gene has previously been associated with PD in a Japanese GWAS at a genome-wide significance level (5). Strong evidence of association for rs4698412 was found in the Japanese scan ($P = 5.3 \times 10^{-5}$, OR = 1.25) and in the combined (scan + replication) data ($P = 1.8 \times 10^{-8}$, OR = 1.24) (5). A much weaker signal was obtained in the US/UK/German data, in both the scan ($P = 0.09$, OR = 1.07) and the combined ($P = 0.03$, OR = 1.06) data (6). Here, we report strong evidence of association of PD with BST1 (combined $P = 1.79 \times 10^{-6}$, OR = 1.14). The most associated SNP (rs4698412) maps to a 15 kb LD-block (Fig. 2C) and is in high LD ($r^2 = 0.74/0.79$) with the next top two BST1 variants (Table 3). Despite the variation in the allele frequency of the risk allele between the Japanese (RAF = 0.33) and the European (RAF = 0.52–0.56) samples (Tables 2 and 3), we found marked homogeneity in the direction of effects across the groups, but effect sizes seemed to be lower in European than in Japanese samples. BST1 has been proposed to play a role in generating cyclic ADP-ribose that serves as a second messenger for Ca^{2+} mobilization in endoplasmic reticulum and thus Ca homeostasis-related BST1 could be a cause of selective vulnerability of dopaminergic neurons in PD (11).

Our most associated SNP, on 12q24 (combined $P = 2.38 \times 10^{-7}$, OR = 1.16), is 26 kb centromeric of RFX4 (Regulatory factor X4) (Fig. 2D). Two other close genes, POLR3B (Polymerase RNA III polypeptide B) and RIC8B (Resistance to inhibitors of cholinesterase 8 homolog B), are 200 kb centromeric and telomeric of the 12q24 SNP, respectively. The RFX proteins belong to the winged-helix subfamily of helix–turn–helix transcription factors. The *RFX4_v3* transcript variant is the only *RFX4* isoform that is significantly expressed in the fetal and adult brain, and its expression is restricted to the brain. In addition, it has a role in the transcription of many genes involved in brain morphogenesis, such as the signaling components in the wnt, bone morphogenetic protein (BMP) and retinoic acid (RA) pathways. In particular, cx3cl1, a CX3C-type chemokine gene, which is highly expressed in brain in response to injury or infection and regulates intracellular calcium concentration, was downregulated

Table 3. Association results of previously reported PD loci

		Japanese samples							European samples										
		Scan phase					Combined (scan + replication)		Scan phase					Combined (scan + replication)		France-GWAS			
SNP	bp	GWAS	Risk all	RAF	OR	P^a	OR	P	GWAS	Risk all	RAF	OR	P	OR	P	Risk all	RAF	OR	P_{GC}
(A) Genome-wide significant loci																			
4q22 (SNCA)																			
rs11931074	90858538	S1	T	0.58	1.50	*6.2E-13*	1.37	*7.4E-17*	S2	T	0.07	1.49	*4.8E-08*	1.46	1.6E-14	T	0.07	1.52	1.3E-05
rs356220	90860363	NA							NA							T	0.35	1.37	*2.82E-08*
rs3857059	90894261	S1	G	0.59	1.49	*1.2E-12*	1.36	*5.7E-16*	S2	G	0.07	1.49	*3.6E-08*	1.48	3.7E-15	G	0.07	1.54	1.0E-05
rs2736990	90897564	NA							S2	C	0.46	1.27	*5.7E-09*	1.23	2.2E-16	G	0.44	1.35	*2.9E-08*
vrs6532194	90999925	S1	T	0.6	1.44	*7.0E-11*	1.32	*4.2E-13*	NA							T	0.09	1.22	0.028
17q12-q22 (MAPT)																			
rs393152	41074926	NA							S2	A	0.78	1.32	1.4E-07	1.30	2.0E-16	A	0.75	1.32	2.7E-05
rs17563986	41347100	NA							S2	T	0.78	1.30	3.4E-07	1.28	1.7E-14	A	0.75	1.34	1.3E-05
rs199533	42184098	NA							S2	C	0.80	1.33	*5.1E-08*	1.28	1.1E-14	G	0.78	1.28	3.0E-04
4p15 (BST1)																			
rs3213710	15326419	NA							NA							A	0.50	1.24	7.7E-05
rs4698412	15346446	S1	A	0.33	1.25	5.3E-05	1.24	*1.8E-08*	S2	A	0.56	1.07	0.09	1.06	0.03	A	0.52	1.28	6.9E-06
rs4538475	15347035	S1	A	0.36	1.25	4.1E-05	1.24	*3.9E-09*	NA							A	0.83	1.15	0.07
rs12646913	15348374	NA							S2	A	0.92	1.18	0.04	1.09	0.03	A	0.91	1.19	0.08
rs4698120	15352430	NA							NA							C	0.53	1.24	7.4E-05
1q32 (PARK16)																			
rs16856139	203905087	S1	C	0.86	1.50	2.6E-06	1.46	*1.0E-07*	NA							T	0.06	1.10	-
rs947211	204019288	S1	G	0.52	1.23	1.2E-04	1.3	*1.5E-12*	NA							G	0.78	1.10	0.14
rs823156	204031263	S1	A	0.83	1.40	1.2E-05	1.37	*3.6E-09*	S2	T	0.82	1.12	4.3E-03	1.12	7.6E-04	A	0.81	1.17	0.03
rs708730	204044403	S1	A	0.82	1.37	2.6E-05	1.33	*2.4E-08*	NA							A	0.82	1.12	0.13
12q12 (LRRK2)																			
rs11564162	38729159	NA							S2	T	0.81	1.28	4.0E-05	1.15	9.5E-05	G	0.17	1.11	-
rs2708453	38764919	S1	T	0.08	1.41	7.5E-05	1.38	9.7E-08	NA							T	0.16	1.16	0.04
rs2896905	38779683	NA							S2	C	0.60	1.22	5.0E-06	1.07	7.8E-03	G	0.64	1.02	0.75
rs1491923	38877384	NA							S2	C	0.31	1.20	2.2E-04	1.14	1.6E-05	G	0.31	1.10	0.09
(B) Suggestive loci																			
4p16 (GAK)																			
rs1564282	842 313	NA							S3	T	0.09	1.7	6.0E-06			T	0.09	1.27	0.008
rs11248060	954 359	NA							S3	T	0.10	1.69	3.4E-06			T	0.11	1.18	0.05

S1: Table 1 from Satake *et al.*, *Nat. Genet.*, 2009; Japanese data – scan phase (1078 PD/2521 controls). S2: Table 2 from Simon-Sanchez *et al.*, *Nat. Genet.*, 2009; US/GE/UK data – scan phase (1745 PD/4047 controls). S3: Table 2 from Pankratz *et al.*, *Hum. Genet.*, 2009; US (PROGENI + GenePD) data – (857 familial PD cases/867 controls).
[a]Genome-wide significant values ($<10^{-7}$) are italicized.

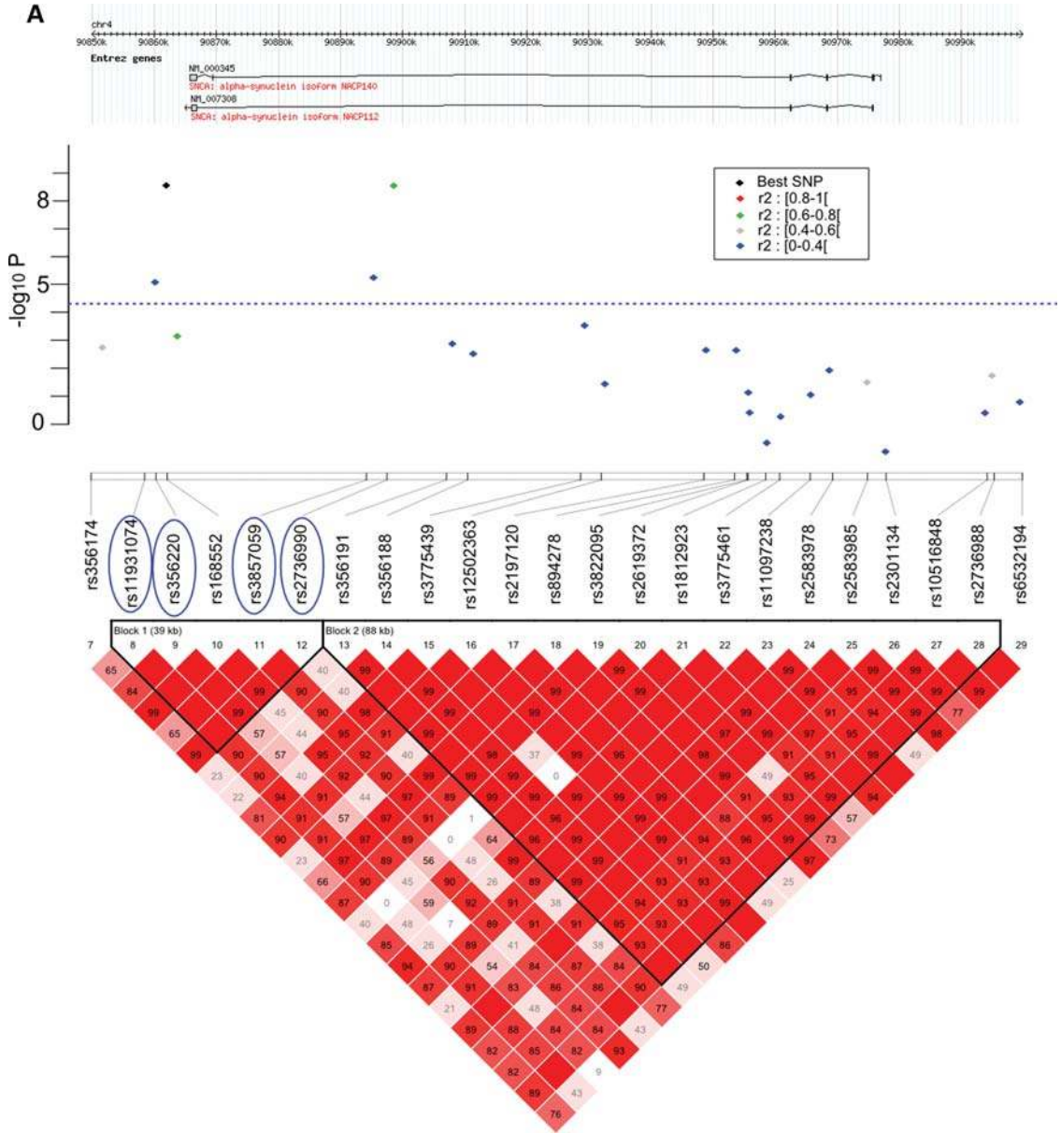

Figure 2. Regional association plots and LD structure for the four PD risk loci (**A**) 4q22/SNCA, (**B**) 17q12–q22/MAPT, (**C**) 4p15/BST1 and (**D**) 12q24/RFX4. The –log10 *P*-values (logistic regression tests corrected for genomic inflation) in the GWAS stage. In each panel, the blue horizontal line indicates a *P*-value of 5×10^{-5}. Pairwise linkage disequilibrium (D') values are displayed and the SNPs with the strongest association signals are circled. SNPs are color-coded for LD relationships (r^2) to the best (colored in black) SNP: red, $0.8 \leq r^2 < 1$; green, $0.6 \leq r^2 < 0.8$; gray, $0.4 \leq r^2 < 0.6$; blue, $0 \leq r^2 < 0.4$. Positions are NCBI build 36 coordinates. Intron and exon structures of genes are taken from the UCSC Genome Browser.

in RFX4_v3-null mice (12). This allows speculation that RFX4 and BST1 are functionally linked and indirectly involved in the regulation of intracellular Ca^{2+} concentrations, which plays an important role in various cellular functions and cell death. Finally, polymorphisms in RFX4 have been shown to be risk factors for the bipolar disorder, manic-depressive illness (13). A recent study showed that a substantial proportion (10–15%) of top GWAS hits, so far identified, are e-quantitative trait loci (eQTLs), i.e. associated with gene expression levels (14). We have initiated eQTL analysis using an existing brain expression database (15), but so far failed to identify any association of the PD-associated rs4964469 SNP with the expression of known genes contained within the 12q24 region.

In conclusion, we have conducted a large GWAS of PD in three case–control samples from France, the UK and Australia. The GWAS stage has 75% and 33% power to detect the loci of the effect sizes observed in stage-1 data for the 12q24 variant (OR = 1.27) at a significance of $P < 5 \times 10^{-5}$ and $P < 10^{-7}$, respectively. In the scan-step, we detected genome-wide significance of association with PD for two SNPs on 4q22, and strong evidence of association with 17q12–q22 SNPs. The two regions encompass previously reported loci: SNCA and MAPT, respectively.

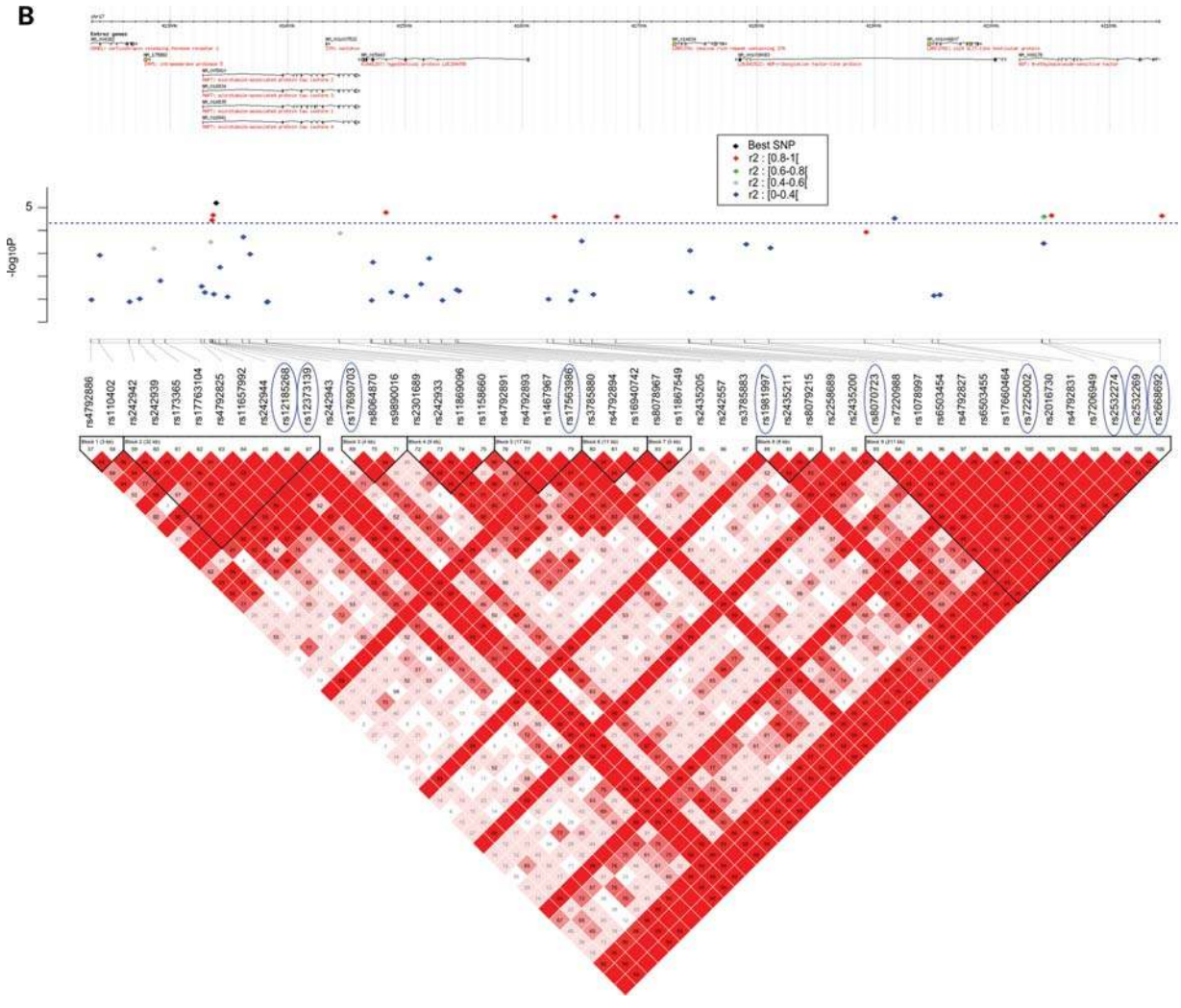

Figure 2. Continued

In addition, we confirmed, for the first time in subjects of European ancestry, the association of PD with 4p15/BST1, recently identified in Japanese samples. Finally, we identified a new locus on 12q24, potentially associated with PD. Further replication studies conducted in large case–control samples are warranted to evaluate the contribution of this locus to PD risk.

MATERIALS AND METHODS

Sample ascertainment and diagnostic criteria

The main characteristics of the three case–control samples are shown in Table 4.

Stage-1 subjects. The total number of cases and controls from France included in stage 1 was 1070 and 2023 controls, respectively.

- **PD subjects:** Patients were recruited through the French network for the study of Parkinson's disease Genetics (PDG) that comprises 15 university hospitals across France. Definite and probable PD was defined according to standard criteria. Definite PD required at least two of three cardinal signs (akinesia and/or rigidity and/or tremor) and absence of exclusion criteria (ophthalmoplegia, pyramidal or cerebellar signs, early dementia, urinary incontinence or postural instability and prior exposure to neuroleptic drugs), and a positive and sustained response to levodopa therapy. Probable PD required at least two of the five following criteria: the parkinsonian triad, a good response to levodopa therapy and asymmetrical onset. Most (>80%) of the PD cases fulfilled the criteria for definite PD. Patients were selected in an effort to enrich for individuals who may have greater genetic predisposition to PD, through selection of cases with a positive family history of PD (Table 4). Cases were of European origin, mostly French ($n = 930$). Subjects diagnosed genetically with known *PARK* mutations (*SNCA*, *LRRK2*, *parkin* and *PINK1*) were excluded.

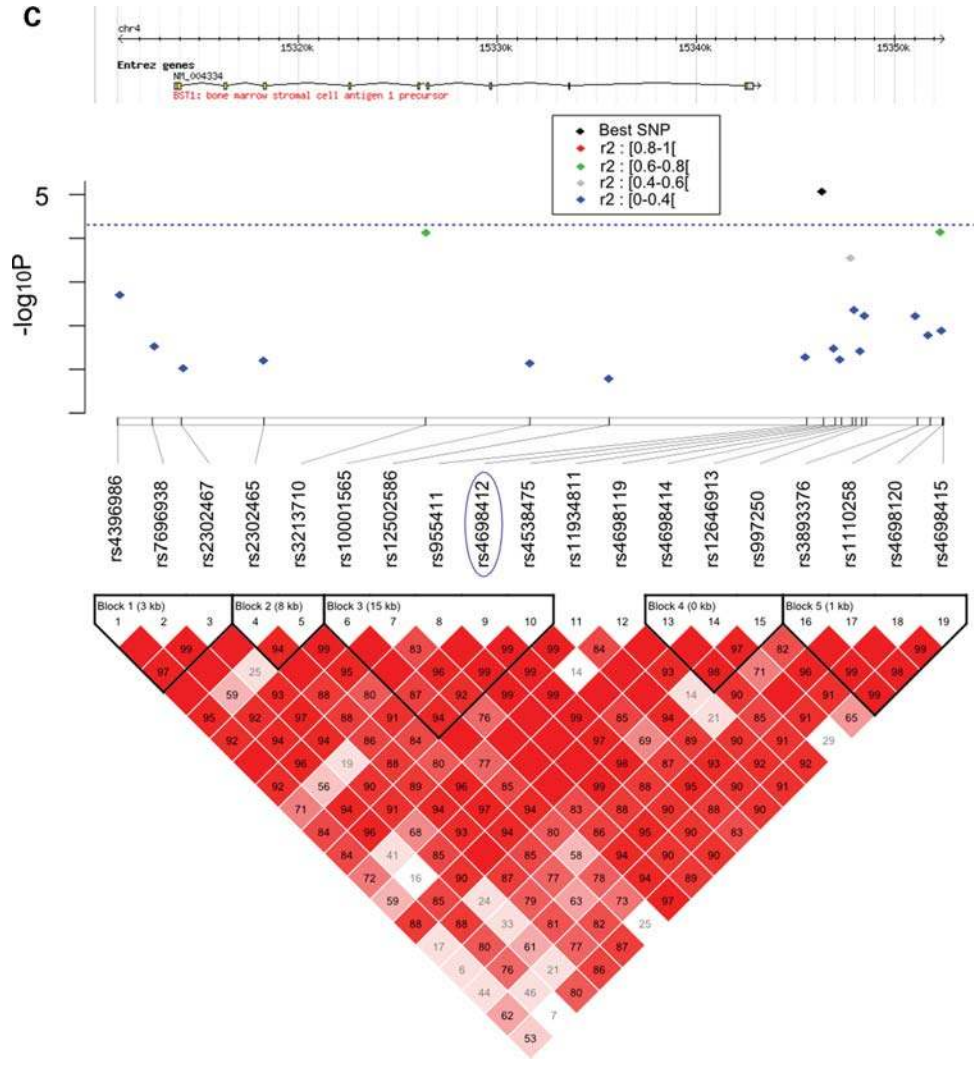

Figure 2. Continued

- 3C neurologically normal controls: The French Three-City (3C) cohort is a population-based, prospective (4-year follow-up) study of the relationship between vascular factors and dementia, carried out in three French cities: Bordeaux (Southwest France), Dijon (central eastern France) and Montpellier (Southeast France) (16). Participants (>9000) are non-institutionalized subjects, over 65 years of age, randomly selected from the electoral rolls of each city. Patients with Alzheimer's disease or other types of dementia, and individuals for whom information on their dementia status during the 4-year follow-up was missing were further excluded. Here, we used a sample of 2023 neurologically normal subjects matched on gender with PD cases, randomly selected from all the participants.

Stage-2 subjects. *In silico* replication sample: we exchanged genome-wide association data with the WTCCC2 PD study group (Spencer *et al.*, submitted). This case–control study consisted of 1705 PD cases and 5200 controls from the 1958 Birth Cohort and from the OK Blood Services Controls (17).

Stage-3 subjects. *De novo* genotyping was conducted in two independent case–control datasets from France (872 PD, 1440 controls) and Australia (655 PD, 424 controls). The subjects from France were combined from three French studies: TERRE (207 cases, 468 controls), PARTAGE (313 cases, 593 controls) and an extension of PDG (352 cases, 378 controls). The extension PDG study includes patients who were not available at the time of the stage-1 genotyping execution and neurologically normal spouses of PDG patients. In cases, the mean age at examination and the mean age of onset of PD is 59 (30–86) and 50 (20–84) years, respectively. The mean age of controls is 60 (31–85) years. In PARTAGE, patients and controls were identified among affiliates of the Mutualité Sociale Agricole (MSA) from five French districts. Parkinsonism was defined as the presence of at least two cardinal signs (rest tremor, bradykinesia, rigidity, impaired

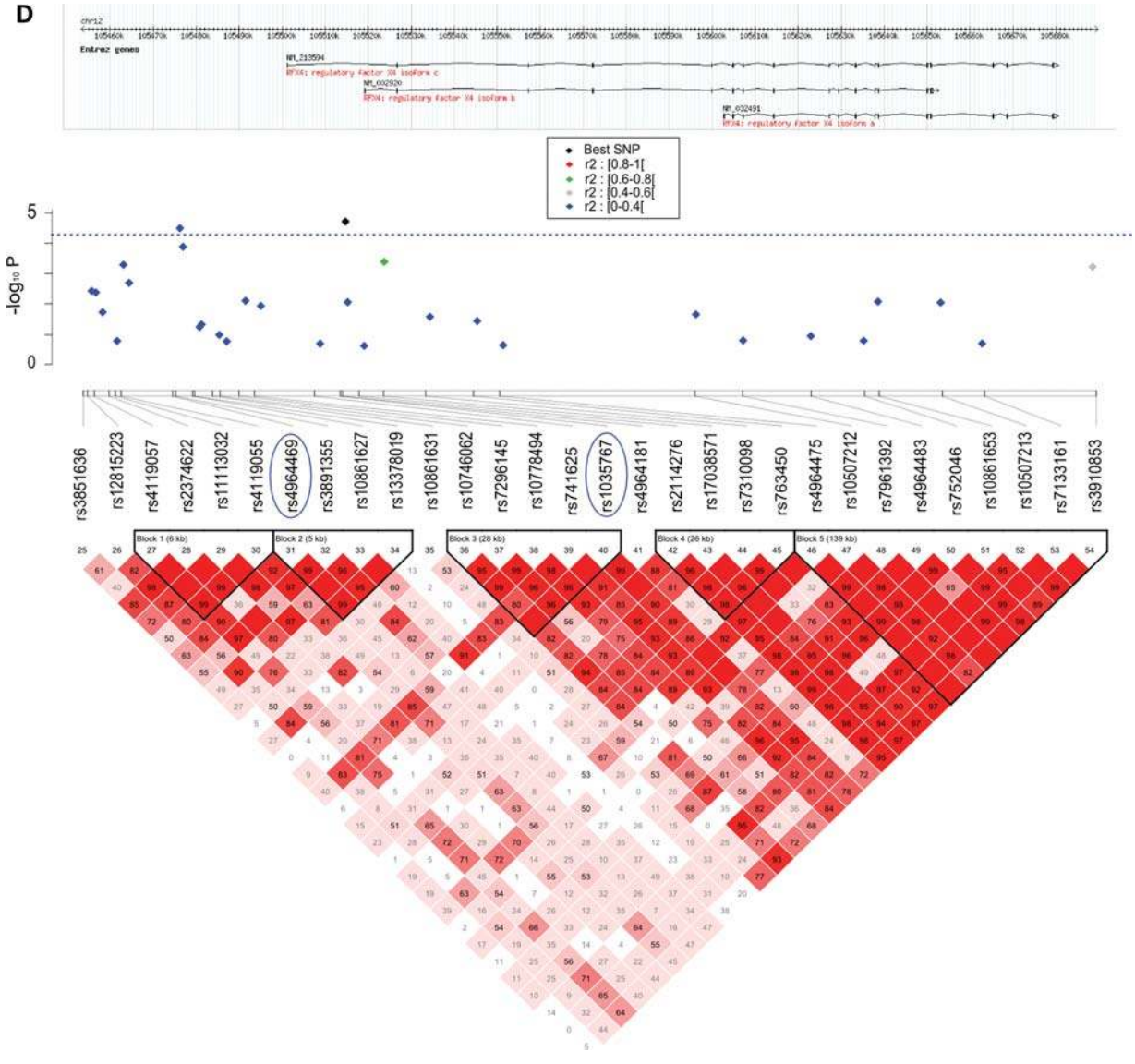

Figure 2. Continued

postural reflexes); PD was defined as the presence of parkinsonism after exclusion of other causes of parkinsonism. Controls were randomly selected from all MSA affiliates in the same districts and matched for sex and age (±2 years). DNA was collected from saliva (Oragene kit). Cases and controls have a mean age of 67 (37–79) years, and the mean age of onset of disease is 63 (35–75) years. TERRE is based on a similar protocol (18), but DNA was collected from blood; the mean age in cases and controls is 73 (46–82) years, and the mean age of onset is 66 (39–80) years in cases.

Australian study: Subjects with PD were recruited from one private and two public movement disorder clinics in Brisbane. Controls were electoral roll volunteers and patient spouses, excluding the subjects demonstrating signs of parkinsonism (19). The mean age is 72 (34–105) and 74 (33–107) years in controls and cases, respectively; the mean age of onset is 59 (23–96) years in cases. Only Caucasian subjects were included in stage 3; in the Australian study, analyses were restricted to participants who reported having four European grandparents (>85% British). There was no overlap between the subjects used in the replication datasets and those included in the stage-1 data. Written informed consent was obtained for all participating subjects and research protocols were approved by local ethics committees.

Genotyping

Stage-1 genotyping. DNA samples of PDG cases and 3C controls were transferred to the French Centre National de Génotypage. First-stage samples that passed DNA quality control (QC) (1064 PD cases and 2023 controls) were genotyped with Illumina Human610-Quad BeadChip and subjected to standard QC procedures.

Table 4. Samples used (post-QC) in this study

Center Genotyping platform	Stage-1 Scan French Illumina 610-Quad	Stage-2 Replication UK Illumina 650Y	Stage-3 Replication French–Australian Illumina GoldenGate	Total
Cases	1039	1705	1527	4271
Sex ratio: M/F	1.42	1.37	1.42	
Age: mean ± SD (*n*)[a]	57.5 ± 16.6 (1003)	NA	69.0 ± 12.7 (1365)	
AOO: mean ± SD (*n*)[a]	48.9 ± 12.8 (970)	65.2 ± 11.3 (1109)	61.3 ± 12.4 (1351)	
FH+ (%)	47	0	17	
Controls	1984	5200	1864	9048
Sex ratio (M/F)	1.33	1.02	1.05	
Age: mean ± SD (*n*)[a]	73.7 ± 5.4 (1984)	51	68.1 ± 10.0	
Total	3023	6905	3391	13 319

[a]Number of subjects for which age/age of onset of disease is known.

Stage-2 genotyping. This WTCCC2 PD study sample was genotyped by the Welcome Trust Case–Control Consortium using the Illumina 650Y genotyping array (Spencer *et al.*, submitted).

Stage-3 genotyping. Genotyping in the extended PDG sample was carried out in the UMR/S 975 laboratory, using predesigned TaqMan probes (C_537709_10/ rs621341; C_29330880_10/ rs6723108; C_12096605_10/ rs11064524; C_2775670_10/ rs4964469; C_1216796_10/ rs4698412) on an ABI 7500 Real-Time PCR system Applied Biosystems, Foster City, CA, USA), according to the manufacturer's instructions. Data were then analyzed using the 7500 software v.2.0.1. The TERRE/PARTAGE and Australian samples were genotyped using the Sequenom MassARRAY platform, with the iPLEX protocol (Genoscreen, France). The basic protocol involves a multiplex primer extension followed by matrix-assisted laser desorption ionization-time of flight mass spectroscopy detection. In order to avoid any genotyping bias, cases and controls were randomly mixed when genotyping and, laboratory personnel were blinded to case–control status.

Quality control of France GWAS scan data

Various stringent QC filters were applied to remove poorly performing SNPs and samples using tools implemented in PLINK version 1.7 (20).

SNP QC: Markers were removed if they had a genotype-missing rate >0.03 or a minor allele frequency (MAF) < 0.05 or a Hardy–Weinberg $P \leq 10^{-5}$. This SNP QC step led to the removal of 74 660 autosomal SNPs. Thus, subsequent analyses were based on 492 929 SNPs.

Individual QC: Samples were removed based on standard exclusion criteria: call rate of $<96\%$ (22 subjects), inconsistencies between reported gender and genotype-determined gender (11 subjects) and genetic relatedness (identity-by-descent estimate >0.14; 6 subjects). Applying these QC filters led to the removal of 39 subjects (14 cases, 25 controls).

Population stratification and principal component analysis: To detect individuals of non-European ancestry, we thinned the SNPs to reduce LD to a set of 55 193 SNPs. To this end, we removed SNPs from the extensive regions of LD (CHR2, CHR5, CHR6, CHR8, CHR11) (21), and excluded SNPs if any pair within a 1000-SNP window had $r^2 > 0.2$. Our stage-1 genotype data were then merged with genotypes at the same SNPs from 381 unrelated European (CEU), Yoruban (YRI) and Asian (CHB and JPT) samples from the HapMap project. Principal component analysis was applied using EIGENSTRAT (22). The two PCs clearly separated the HapMap data into three distinct clusters according to ancestry, and most of our stage-1 samples were clustered with the HapMap European samples (Fig. 3). Thirty-two samples appeared to be ethnic outliers (including one subject clearly sharing African ancestry) from the European cluster and were excluded from further analysis. The final post-QC scan sample comprised 1039 PD cases and 1984 controls.

Statistical analysis

Association analysis of the genotype data was conducted with PLINK (20).

Stage-1 association analyses. Logistic regression was used to study the allelic association between each SNP and PD assuming an additive genetic model. Our analysis was based on 492 929 SNPs, and on a conservative genome-wide significance threshold of $0.05/492\,929 = 10^{-7}$. The distribution of the association results was found to be marginally inflated (median $\chi^2 = 0.521$); genomic inflation factor $\lambda = 1.14$ ($\lambda_{1000} = 1.10$). Logistic regression analysis adjusted for the two first PCs of the EIGENSTRAT analysis revealed a genomic inflation of 1.03 (median $\chi^2 = 0.472$). As for our primary analyses, we applied the genomic inflation correction method (23); the median of the GC-corrected χ^2 value was 0.447.

Sensitivity analyses: Two further analyses were conducted to assist in the interpretation of results of the identified GWAS SNPs. We performed age-adjusted regression analysis and conducted subgroup analyses of two subtypes of cases against all controls. Cases with a disease onset before 50 years ($n = 428$) were classified as 'early AOO', and cases

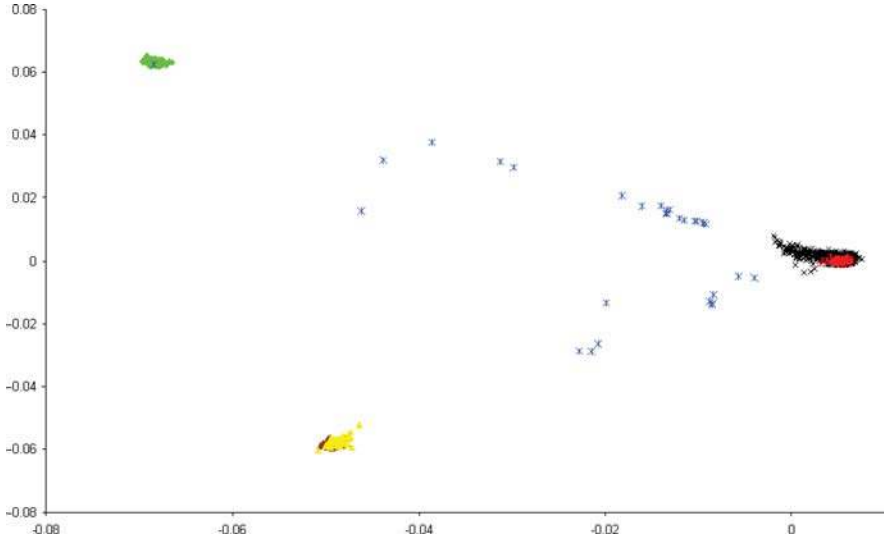

Figure 3. Principal components for our genome-wide stage 1. Plot of the first two principal components from the analysis of our stage-1 (post-QC) data combined with HapMap data. Ethnicity of HapMap samples indicated by color: Africa (YRI) in green, Japan (JPT) in brown, Chinese (6) in yellow and Europe (CEU) in red. Study samples identified to be non-European or not clustering with European samples (outliers) are colored in blue and the remaining study samples assumed to be of European origin are colored in black.

having at least one first-degree relative with PD ($n = 452$) were classified as 'FH+'.

Stage-2 in silico *association analyses*. Statistical data (ORs, effective sample sizes and nominal *P*-values for each of the 50 top SNPs) in the UK sample were obtained from the WTCC2 PD study group that used similar analytical methods (Spencer *et al.*, submitted).

Stage-3 association analyses. For the *de novo* replication stage, we computed association statistics with the Mantel–Haenszel test to control for the potential confounding owing to the geographical center (France versus Australia) for the five SNPs replicated at stage-2. Using raw genotypes from all the study samples, we computed similar stratified (France versus Australia versus UK) association statistics in the combined (stage-2 + stage-3 and stage-1 + stage-2 + stage-3) data.

The PAR associated with the detected variants was estimated with the following formula: $PAR = p\ (OR-1)/[p(OR-1) + 1]$, where p is the frequency of the risk allele in controls, and OR is the odds ratio associated with the risk allele.

AUTHOR CONTRIBUTIONS

S.L. supervised DNA sampling; J.C.C., M.V., E.B., F.D., P.P., P.D., F.T., A.D. and A.B. recruited patients; J.C.C., A.D. and AB supervised clinical work; D.Z. and M.L. supervised PD and 3C GWAS genotyping and DNA QC work; S.L. and J.C.L. supervised genotyping of stage-3 samples. A.E., J.C.L., M.A.L., C.T., G.D.M. and P.A.S. contributed to stage-3 replication; M.S., A.S.P. and M.M. executed QC analyses and performed statistical association analyses; A.E., A.B. and M.M. were involved in obtaining funding; M.M. drafted the manuscript and S.L., A.B. and A.E. contributed to the writing of the final version; A.B. and M.M. conceived and oversaw the design and execution of the GWAS.

ACKNOWLEDGEMENTS

The authors are grateful to the patients and their families. They thank the DNA and Cell Bank of UMR_S975 for sample preparation. We thank the members of the 3C consortium: Drs Annick Alpérovitch, Claudine Berr and Jean-Francois Dartigues for giving us the possibility to use part of the 3C cohort. This study makes use of data generated by the Wellcome Trust Case–Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

Conflict of Interest statement. The authors declare no competing financial interests.

FUNDING

This work was supported by the French National Agency of Research (ANR-08-MNP-012).

REFERENCES

1. Lesage, S. and Brice, A. (2009) Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum. Mol. Genet.*, **18**, R48–R59.
2. Maraganore, D.M., De Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V., Frazer, K.A., Cox, D.R. and Ballinger, D.G. (2005) High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.*, **77**, 685–693.
3. Fung, H.C., Scholz, S., Matarin, M., Simon-Sanchez, J., Hernandez, D., Britton, A., Gibbs, J.R., Langefeld, C., Stiegert, M.L., Schymick, J. *et al.* (2006) Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.*, **5**, 911–916.

4. Myers, R.H. (2006) Considerations for genomewide association studies in Parkinson disease. *Am. J. Hum. Genet.*, **78**, 1081–1082.
5. Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., Tsunoda, T., Watanabe, M., Takeda, A. *et al.* (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.*, **41**, 1303–1307.
6. Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G. *et al.* (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.*, **41**, 1308–1312.
7. Tian, C., Kosoy, R., Nassir, R., Lee, A., Villoslada, P., Klareskog, L., Hammarstrom, L., Garchon, H.J., Pulver, A.E., Ransom, M. *et al.* (2009) European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol. Med.*, **15**, 371–383.
8. Pankratz, N., Wilk, J.B., Latourelle, J.C., Destefano, A.L., Halter, C., Pugh, E.W., Doheny, K.F., Gusella, J.F., Nichols, W.C., Foroud, T. *et al.* (2009) Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum. Genet.*, **124**, 593–605.
9. Healy, D.G., Abou-Sleiman, P.M., Lees, A.J., Casas, J.P., Quinn, N., Bhatia, K., Hingorani, A.D. and Wood, N.W. (2004) Tau gene and Parkinson's disease: a case-control study and meta-analysis. *J. Neurol. Neurosurg. Psychiatry*, **75**, 962–965.
10. Zhang, J., Song, Y., Chen, H. and Fan, D. (2005) The tau gene haplotype h1 confers a susceptibility to Parkinson's disease. *Eur. Neurol.*, **53**, 15–21.
11. Zhang, D., Stumpo, D.J., Graves, J.P., Degraff, L.M., Grissom, S.F., Collins, J.B., Li, L., Zeldin, D.C. and Blackshear, P.J. (2006) Identification of potential target genes for RFX4_v3, a transcription factor critical for brain development. *J. Neurochem.*, **98**, 860–875.
12. Chan, C.S., Gertler, T.S. and Surmeier, D.J. (2009) Calcium homeostasis, selective vulnerability and Parkinson's disease. *Trends Neurosci.*, **32**, 249–256.
13. Glaser, B., Kirov, G., Bray, N.J., Green, E., O'donovan, M.C., Craddock, N. and Owen, M.J. (2005) Identification of a potential bipolar risk haplotype in the gene encoding the winged-helix transcription factor RFX4. *Mol. Psychiatry.*, **10**, 920–927.
14. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. and Lathrop, M. (2009) Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.*, **10**, 184–194.
15. Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P. *et al.* (2007) A survey of genetic human cortical gene expression. *Nat. Genet.*, **39**, 1494–1499.
16. Alperovitch, A. (2003) Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology*, **22**, 316–325.
17. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature*, **447**, 661–678.
18. Elbaz, A., Clavel, J., Rathouz, P.J., Moisan, F., Galanaud, J.P., Delemotte, B., Alperovitch, A. and Tzourio, C. (2009) Professional exposure to pesticides and Parkinson disease. *Ann. Neurol.*, **66**, 494–504.
19. Sutherland, G.T., Halliday, G.M., Silburn, P.A., Mastaglia, F.L., Rowe, D.B., Boyle, R.S., O'sullivan, J.D., Ly, T., Wilton, S.D. and Mellick, G.D. (2009) Do polymorphisms in the familial Parkinsonism genes contribute to risk for sporadic Parkinson's disease? *Mov. Disord.*, **24**, 833–838.
20. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
21. Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D. *et al.* (2008) Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.*, **83**, 132–135. author reply 135–139.
22. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
23. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.