Human Molecular Genetics, 2005, Vol. 14, No. 23 ***3549–3556***
doi:10.1093/hmg/ddi376
Advance Access published on November 3, 2005

Comprehensive evaluation of common genetic variation within *LRRK2* reveals evidence for association with sporadic Parkinson's disease

Lisa Skipper[1], Yi Li[2], Carine Bonnard[1], Ratnagopal Pavanni[3], Yuen Yih[3], Eva Chua[3,4], Wing-Kin Sung[2], Louis Tan[5], Meng-Cheong Wong[5], Eng-King Tan[3,4] and Jianjun Liu[1,*]

[1]Population Genetics and [2]Information and Mathematical Sciences, Genome Institute of Singapore, Singapore 138672, [3]Department of Neurology and [4]Division of Research, SingHealth, Singapore General Hospital, Singapore 169611 and [5]National Neuroscience Institute, Singapore, 308433

Received August 15, 2005; Revised and Accepted September 30, 2005 GenBank accession no. AY792511

Parkinson's disease (PD) is a complex neurodegenerative disorder whose aetiologies are largely unknown. To date, mutations in six genes have been found causal for some rare familial forms of the disease and common variation within at least three of these is associated with the more common sporadic forms of PD. *LRRK2* is the most recently identified familial PD gene, although its role in sporadic disease is unknown. In this study, we have performed the first comprehensive evaluation of common genetic variation within *LRRK2* and investigated its contribution to risk of sporadic PD. We first characterized the linkage disequilibrium within *LRRK2* using a panel of densely spaced SNPs across the gene. We then identified a subset of tagging-SNPs (tSNP) that capture the majority of common variation within *LRRK2*. Both single tSNP and tSNP haplotype analyses, using a large epidemiologically matched sporadic case–control series comprising 932 individuals, yielded significant evidence for disease association. We identified a haplotype that dramatically increases disease risk when present in two copies (OR = 5.5, 95%CI = 2.1–14.0, *P* = 0.0001). Thus, we provide the first evidence that common genetic variation within *LRRK2* contributes to the risk of sporadic PD in the Chinese population.

INTRODUCTION

Parkinson's disease (PD) is a complex neurodegenerative disorder currently thought to affect almost one million in the USA and as many as 1.7 million in China (1). Clinically, PD is characterized by resting tremor, muscle rigidity, bradykinesia and postural instability (2). Dopaminergic neuronal death within the substantia nigra pars compacta of the midbrain and the presence of proteinaceous intracellular aggregates (Lewy bodies) in surviving cells are defining pathological features (3), although the pathogenic mechanisms are poorly understood at present. The majority of cases are sporadic, although mutations in a handful of genes are responsible for a small proportion of rare familial forms of PD (4). Study of these familial PD genes has offered useful insight and is assisting efforts in understanding aetiologies in common forms of the disease.

The first familial PD gene identified was alpha-synuclein (*SNCA*). Mis-sense mutations are responsible for a small number of autosomal dominant PD cases (5–7). A central role for *SNCA* in PD stems from the finding that fibrillar forms of the protein are a major structural component of LBs in both familial and sporadic cases (8). Genomic multiplications leading to increased expression of *SNCA* found in other autosomal dominant PD families [PARK4 (MIM 605543)] (9,10) suggest that protein levels may be important in pathogenesis. Polymorphic variability within the *SNCA* promoter is also associated with increased risk for sporadic PD (11).

Mutations in the parkin gene (*PRKN*) are responsible for about 50% of autosomal recessive (AR) PD cases [PARK2 (MIM 600166)] (12). The genetic aberrations in this gene are numerous ranging from single base-pair substitutions and deletions to multiple exonic deletions and duplications. Although most parkin-linked disease is inherited in AR

*To whom correspondence should be addressed at: Population Genetics, Genome, Genome Institute of Singapore, #02-01 60 Biopolis Street, Singapore 138672, Singapore. Tel: +65 64788088; Fax: +65 64789059; Email: liuj3@gis.a-star.edu.sg

fashion, mutations have been described in sporadic cases and haploinsufficiency may also predispose to common PD (13). A common promoter variant that decreases *PRKN* transcription is also associated with sporadic PD (14). A function of the parkin protein is to act as an E3 ubiquitin protein ligase within the ubiquitin proteasomal pathway (UPP) (15); one of the principle mechanisms by which damaged/misfolded proteins are cleared in the cell. The protein product of another gene initially implicated in familial PD, *UCH-L1* (16), also plays a role within the UPP. Impairment of UPP function is observed in sporadic cases (17) and a common polymorphism that alters UCH-L1 enzymatic activity is inversely associated with risk of sporadic PD (18).

The leucine-rich repeat kinase 2 gene (*LRRK2*) is the most recently identified causative gene for autosomal dominant PD (19). Although its biological function is unknown at present, peptide sequence analogy suggests the LRRK2 protein is a member of the ROCO family (20) that contains five conserved domains within the C-terminal half of the protein. Most *LRRK2* mutations thus far identified are within exons encoding these putative functional domains and they account for up to 6% of autosomal dominant cases (21). The most common of these, G2019S, may also account up to 2% of sporadic cases (22). As variation in genes initially implicated in familial PD is becoming increasingly relevant to PD in the wider community, we sought to investigate whether common variation within *LRRK2* may also predispose risk to sporadic PD. Here, we report our findings in the first study to address this question, using a large ($n = 932$) epidemiologically matched sporadic case–control series.

RESULTS

LRRK2 linkage disequilibrium and tagging-SNPs

Linkage disequilibrium (LD) is defined as the non-random association of alleles at adjacent loci. Loci and thus their allelic variants are expected to segregate independently in a randomly mating population. However, if a particular allele at one locus is found together on the same chromosome with a specific allele at a second locus, more often than expected by chance, then the loci are in linkage disequilibrium (23). Common measures of LD include D' and r^2. D' values range from -1 to 1; 1 indicates 'complete LD' and is indicative of minimal recombination between loci (three out of a possible four haplotypes exist) and r^2 ranges from 0 to 1; 1 indicates 'perfect LD' (only two of a possible four haplotypes exist). Perfect LD between two loci means that information for one of these is redundant. Extensive regions showing strong LD interspersed with (usually smaller) regions in equilibrium are typically manifested as regions of limited haplotype diversity ('haplotype block'). The identification of haplotype blocks has allowed the use of haplotype 'tagging-SNPs' (tSNPs) in association studies, under the assumption that much of the variation can be identified with a smaller subset of SNPs, i.e. those that 'tag' a haplotype (24).

We characterized the LD and haplotype structure of *LRRK2* (including 10 kb of sequence upstream from coding start, based on Genbank accession number AY792511) in 92 control individuals. Twenty-five SNPs (Table 1) with minor allele frequencies (MAF) ranging from 0.03 to 0.49 were genotyped, yielding an average density of 1 SNP/6 kb (Fig. 1B). Seventy-eight unique multilocus haplotypes were constructed and used to calculate pair-wise D' and r^2 values (Fig. 1C and D).

Complete LD exists throughout much of the gene overall, suggestive of minimal recombination in this region. In addition, r^2 values suggest that information for some SNPs is redundant. We therefore, sought to identify tSNPs that would efficiently capture common variation within *LRRK2*. As there is no unique definition of what constitutes a haplotype block and the number of tSNPs can be further reduced if inter-block disequilibrium is also considered (25), we used block-independent methods to identify six tSNPs (tSNPset) sufficient to uniquely distinguish 95% of all *LRRK2* haplotypes with a frequency >0.01. We also evaluated the performance of our tSNPset to predict observed haplotypic variation (measured by R^2_h) and unobserved genotypic variation (measured by r^2) within *LRRK2*. R^2_h values were found to be >0.8 between the tSNPset and all common (frequency >0.05) full haplotypes (comprising the 25 SNPs). Average r^2 was 0.94 between the tSNPset and all remaining common (MAF > 0.05) genotypes. These performance levels exceed the criteria proposed by Ahmadi *et al*. (26) and indicate that our tSNPset affords only a modest reduction in power relative to genotyping all common SNPs within *LRRK2*.

Association analyses

The tSNPset was genotyped in all cases ($n = 466$) and remaining control individuals ($n = 374$). We initially carried out single marker tests of association for each *LRRK2* tSNP, by calculating odds ratios under recessive, additive and dominant models (Table 2). There was a marginally significant association between rs10506151 and disease under a dominant model (OR = 1.3, 95%CI = 1.0–1.7, $P = 0.03$). However, carrying two copies greatly increased disease risk under both recessive (OR = 3.7, 95%CI = 1.8–7.8, $P = 0.0001$) and additive models (OR = 4.0, 95%CI = 1.9–8.3, $P = 0.0001$). This association remains highly significant after a conservative Bonferroni correction for multiple testing ($P = 0.002$).

We next performed tSNP haplotype association analysis, as haplotype analysis may provide more power to detect association than single marker analyses alone (27). tSNP haplotypes were constructed for all cases and controls and odds ratios for tSNP haplotypes (frequency >0.01) were calculated by comparing frequencies of each against all others present within the sample group (Table 3). Haplotype 2 is over-represented in the patient group (OR = 1.4, 95%CI = 1.1–1.8, $P = 0.005$; corrected $P = 0.08$); an effect partly because of signals emanating from the minor allele of rs10506151 (A) carried on this haplotype. Odds ratios for Haplotype 2 were calculated under recessive, additive and dominant models (Table 4). We observed a striking association between carrying two copies and disease status, under both recessive (OR = 5.2, 95%CI = 2.0–13.3, $P = 0.0002$; corrected $P = 0.0008$) and additive (OR = 5.5, 95%CI = 2.1–14.0, $P = 0.0001$; corrected $P = 0.0004$) models. We also investigated a possible age of onset effect related to Haplotype 2 dosage, using

Table 1. SNP assay oligomer sequences

SNP	Alleles	Oligomer Sequence (5′–3′)		
	Major (1)/ Minor (2)	Forward	Reverse	Extension
rs11564131	A/G	ACGTTGGATGTAAGGCACAGATGATAAAGG	ACGTTGGATGGGTGAGTTTCATTAAACCAG	CAGATGATAAAGGATAAGTGC
rs1388598	G/A	ACGTTGGATGAGGTTTTGAAAAGGAAATTC	ACGTTGGATGGTGCATCACTCATGTTTAGG	AGGAAATTCATTTTGAATGAC
rs1491941	A/G	ACGTTGGATGGCTCAGGCTTGGGCAATTTC	ACGTTGGATGAGGGTCTAGATAAGGCAGTG	GTATGTTCCAGGCTCCC
rs11564115	T/G	ACGTTGGATGCCATGTTGTATGGCATGCAC	ACGTTGGATGCCATCGCCAAATGGAGAATG	GAGTCTTTAACCATGATTTAATT
rs10506148	G/A	ACGTTGGATGTGATTTATGTGGGCCCTGTG	ACGTTGGATGCTATCCCATTCTGGGAACTG	CCCTGTGTCAGTTATCA
rs1388596	T/C	ACGTTGGATGCTGGTTTGTTGAATGCAGCG	ACGTTGGATGGCTGAAACAAGAGCTGTGAG	CAGGATTCATGCTTTCCCAC
rs1491938	G/A	ACGTTGGATGTGGAATCAAGGGTCATTTGG	ACGTTGGATGCTCTGCAATTTCTATCACCC	TCTTTCTATTTTCTGGCTGATC
rs4293189	G/A	ACGTTGGATGGCTACGATTATTCCAGCTAC	ACGTTGGATGGCAACAATGGACAACATGAG	CTTTTTATGTTGTGGTCCAA
rs10784470	G/T	ACGTTGGATGGTACGAATGTATTTCCCAGG	ACGTTGGATGTGAGGGAGAGGACATAAAGC	TATTCGTTTTTTCCTTCCT
rs10506151	C/A	ACGTTGGATGTTAAGATGATCCTTGGTGGC	ACGTTGGATGGTCTGTCCTCTTCACTCATG	GATCCTTGGTGGCTTTACC
rs10784486	C/A	ACGTTGGATGGGTGCCACAAAAACTAGAAC	ACGTTGGATGATAGCTCTTCCTGAGTAGGG	ACTAGAACAACTCAGCAA
rs4310676	A/G	ACGTTGGATGTCCGATGGAGAGGAATTGTG	ACGTTGGATGACTCCAGACAGAACAAAGAG	GCCCATTGGAGAGGGAC
rs7966550	T/C	ACGTTGGATGCTTACTGAGTGAATCATCTG	ACGTTGGATGGGCCCATTTTTGATCATGAAG	GAGTGAATCATCTGAAGATA
rs721709	T/A	ACGTTGGATGTTTACCCCAAGGATTTCCAC	ACGTTGGATGTGCTGCTTTACTAGATCCTG	CCCAAGGATTTCCACAAAGCC
rs1427263	G/T	ACGTTGGATGTTTTTCCACATCTCTACGCG	ACGTTGGATGGTGAAAGTGGAAGGTTGTCC	CACATCTCTACGCGAAATAAT
rs11176013	A/G	ACGTTGGATGTATTTCGCGTAGAGATGTGG	ACGTTGGATGCTATTGGCAAAGCAATCTGG	ATGTGGAAAAATTTCTTTCAAA
rs11564148	T/A	ACGTTGGATGCTATTGGCAAAGCAATCTGG	ACGTTGGATGTATTTCGCGTAGAGATGTGG	CTAGGAGCTTAAAATACTGTG
rs10878371	T/C	ACGTTGGATGCCTCTGCTTTCTTCATCAAG	ACGTTGGATGACTCTGTTGAAGAAATGGGC	AGTAAGATTTTTTGATGTTCTTC
rs963243	C/T	ACGTTGGATGGCTTCACGAGTACTTATTCC	ACGTTGGATGTATGTAGGCCATTTGATTCC	GTACTTATTCCAAGTACAGAAG
rs6581667	G/C	ACGTTGGATGAGTTTGGATGCACTTTGCAG	ACGTTGGATGTTGGTTCAGTATGTCCAGGG	ACTGTCAAGGAATTGCAAG
rs1365763	C/T	ACGTTGGATGAATTTGATTTGCCTCACAAGTG	ACGTTGGATGCTTTAGATACTCTAGAGGTC	CCTCACAAGTGCCAACA
rs1365765	G/A	ACGTTGGATGTCAGTTCTGTGTTCTGCATC	ACGTTGGATGTCCTCACAGATGGTTTGTGC	GCTACGCCTCTTATAAAAA
rs4768233	T/C	ACGTTGGATGCACAAGCTTGTTGTCAATAC	ACGTTGGATGAATGCAGTTCATGGCTTGAG	CAATTCATAGACCAGATTTAC
rs7306545	A/T	ACGTTGGATGTCTGATGAGAGACAGGGAAG	ACGTTGGATGCAATCTATTCTTTTTCCCATG	AGTATAACCCTTAGGTTATTTCTT
rs3761863	A/G	ACGTTGGATGTGAAGGCAGAGGGTTTTCAC	ACGTTGGATGGTATACTTTATGGTTCTAGGG	TTTGATTCCTTGTTTTCTTTTACC

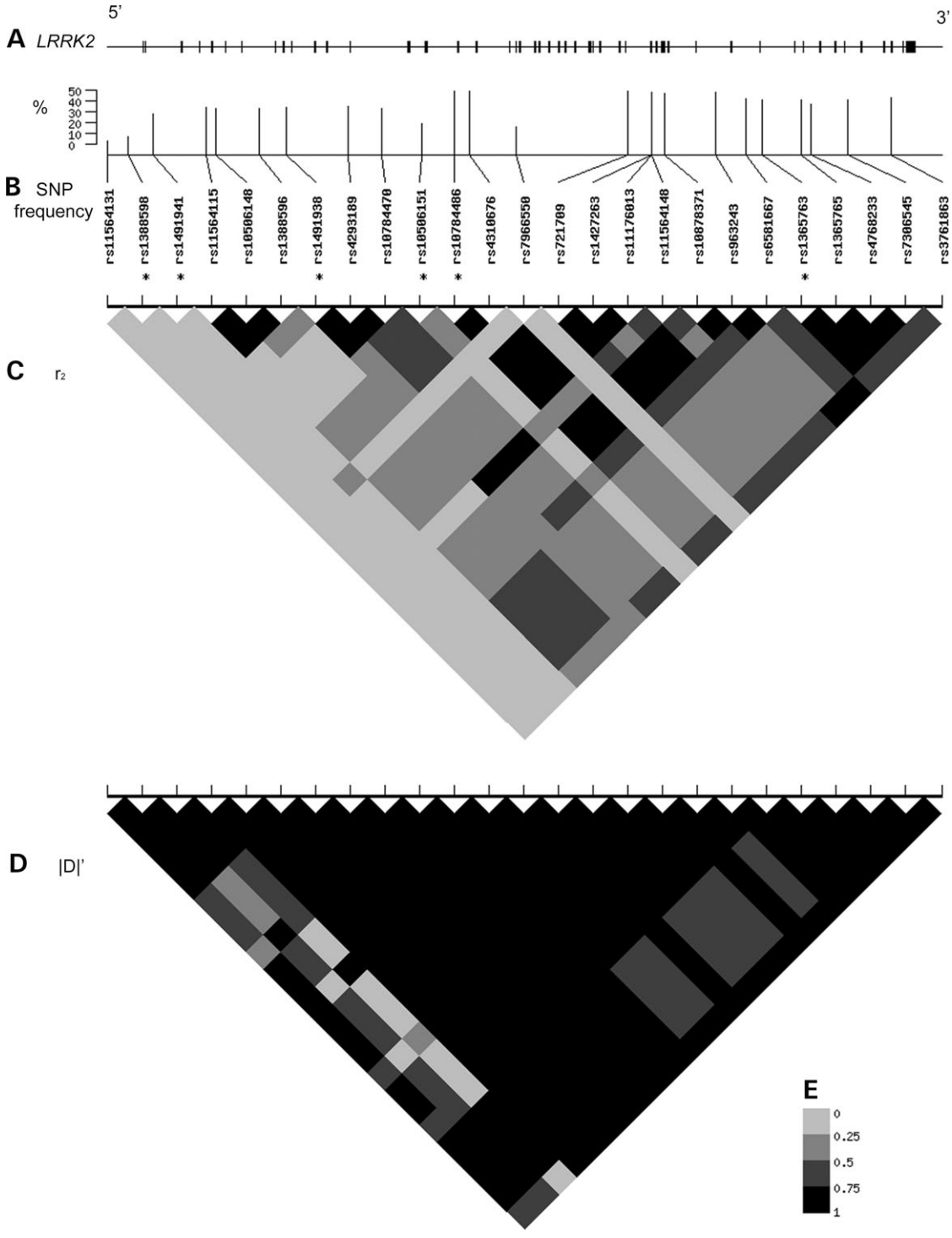

Figure 1. Genomic organisation and LD within *LRRK2*. (**A**) Genomic organisation of *LRRK2*. Exons are shown as vertical lines. (**B**) Relative positions and frequencies of 25 SNPs used to determine tSNPs. Vertical bars represent the position within *LRRK2* (*x*-axis) and the frequency in 94 Chinese control individuals (*y*-axis). tSNPs are denoted with an asterisk. (**C**) Pair-wise r^2 and (**D**) pair-wise D' values between 25 SNPs. Values are shaded according to the key (**E**) such that lighter shading indicates low values (low LD) and darker shading indicates higher values (stronger LD).

Kruskal–Wallis tests, although results were insignificant (data not shown).

In addition to the identification of a risk haplotype, our results also raise the possibility that protective variants exist within *LRRK2*. Haplotype 3 (OR = 0.8, 95%CI = 0.6–0.9, $P = 0.04$) and to a greater extent Haplotype 11 (OR = 0.2, 95%CI = 0.1–0.7, $P = 0.01$) are both under-represented in the case group (Table 3).

Table 2. *LRRK2* tSNP association tests

tSNP	Group	No. of genotypes (frequency)			No. of alleles (frequency)		OR (95%CI; *P*)			
		11	12	22	1	2	Recessive[a]	Additive 1 copy[b]	2 copies[c]	Dominant[d]
rs1388598	Control	368 (0.83)	73 (0.16)	5 (0.01)	809 (0.91)	83 (0.09)	1.0 (0.3–3.2; 1.0)	0.7 (0.5–1.1; 0.10)	0.8 (0.2–2.6; 0.93)	0.7 (0.5–1.1; 0.09)
	Case	393 (0.87)	57 (0.13)	4 (0.01)	843 (0.93)	65 (0.07)				
rs1491941	Control	209 (0.46)	193 (0.43)	50 (0.11)	611 (0.68)	293 (0.32)	0.8 (0.5–1.2; 0.26)	0.9 (0.7–1.2; 0.36)	0.7 (0.5–1.2; 0.18)	0.9 (0.7–1.1; 0.22)
	Case	228 (0.50)	185 (0.41)	40 (0.09)	641 (0.71)	265 (0.29)				
rs1491938	Control	173 (0.39)	211 (0.47)	63 (0.14)	557 (0.62)	337 (0.38)	1.3 (0.9–1.9; 0.10)	0.9 (0.7–1.2; 0.50)	1.3 (0.9–1.9; 0.20)	1.0 (0.8–1.3; 0.95)
	Case	172 (0.39)	190 (0.43)	80 (0.18)	534 (0.60)	350 (0.40)				
rs10506151	Control	285 (0.64)	152 (0.34)	9 (0.02)	722 (0.81)	170 (0.19)	3.7 (1.8–7.8; 0.0001)	1.2 (0.9–1.6; 0.20)	4.0 (1.9–8.3; 0.0001)	1.3 (1.0–1.7; 0.03)
	Case	248 (0.57)	158 (0.36)	31 (0.07)	654 (0.75)	220 (0.25)				
rs10784486	Control	112 (0.25)	243 (0.54)	97 (0.21)	467 (0.52)	437 (0.48)	1.0 (0.3–1.3; 0.9)	0.9 (0.7–1.2; 0.38)	0.9 (0.6–1.3; 0.54)	0.9 (0.6–1.2; 0.87)
	Case	121 (0.27)	228 (0.52)	93 (0.21)	470 (0.53)	414 (0.47)				
rs1365763	Control	426 (0.97)	15 (0.03)	0 (0.00)	598 (0.67)	298 (0.33)	1.0 (0.9–1.1; 0.81)	1.0 (0.7–1.3; 0.70)	1.0 (0.6–1.4; 0.74)	1.0 (0.8–1.3; 0.03)
	Case	433 (0.93)	30 (0.06)	1 (0.00)	596 (0.68)	284 (0.32)				

[a]Minor allele homozygotes versus all other individuals.
[b]Heterozygotes versus major allele homozygotes.
[c]Minor allele homozygotes versus major allele homozygotes.
[d]Major allele homozygotes versus all other individuals.

Table 3. *LRRK2* tSNPset haplotype frequencies and ORs

Haplotype (Allelic conformation)	No. of haplotypes (frequency)		
	Control	Case	OR (95%CI; P)
1 (CTCCCT)	258 (0.28)	268 (0.29)	1.1 (0.9–1.3; 0.61)
2 (CTTAAC)	157 (0.17)	205 (0.22)	1.4 (1.1–1.8; 0.005)
3 (CTTCAC)	152 (0.16)	121 (0.13)	0.8 (0.6–0.9; 0.04)
4 (CCCCCC)	144 (0.15)	138 (0.15)	1.0 (0.7–1.2; 0.70)
5 (TCCCAC)	38 (0.04)	34 (0.04)	0.9 (0.6–1.4; 0.63)
6 (CCCCAC)	36 (0.04)	40 (0.04)	1.1 (0.7–1.8; 0.64)
7 (CTCCAC)	25 (0.03)	26 (0.03)	1.0 (0.6–1.8; 0.88)
8 (CTCCCC)	19 (0.02)	18 (0.02)	1.0 (0.5–1.8; 0.87)
9 (CTCCAT)	16 (0.02)	17 (0.02)	1.1 (0.5–2.1; 0.86)
10 (TCTAAC)	15 (0.02)	13 (0.01)	0.9 (0.4–1.8; 0.70)
11 (TCCCAT)	14 (0.02)	3 (0.003)	0.2 (0.1–0.7; 0.007)
12 (TCCCCC)	13 (0.01)	9 (0.01)	0.7 (0.3–1.6; 0.39)
13 (CCTAAC)	12 (0.01)	11 (0.01)	0.9 (0.4–2.0; 0.83)
14 (CTTCCT)	9 (0.01)	6 (0.01)	0.7 (0.2–1.8; 0.44)
15 (CCCCCT)	9 (0.01)	4 (0.004)	0.4 (0.1–1.4; 0.27)
16 (CCCCAT)	5 (0.01)	7 (0.01)	1.4 (0.5–4.2; 0.56)
Rare (frequency <0.01)	10 (0.01)	12 (0.01)	—

In silico analysis of rs10506151

SNP rs10506151 is positioned within the intron following exon 16, 2203 bp downstream of exon 16 and 692 bp upstream of exon 17. *In silico* analysis shows that the wild-type allele (C) is conserved in vertebrates through to the rat. In addition, the SNP is within a putative 6-mer hnRNP K binding site (ACC**C**AA; SNP position shown in bold).

DISCUSSION

This study is the first to comprehensively evaluate common genetic variation within *LRRK2* and assess its potential role in sporadic PD. By using a powerful tSNP approach, we have identified a common haplotype that is highly over-represented within cases ($P = 0.005$) and, when present in two copies, significantly increases disease risk (OR = 5.5, 95%CI = 2.1–14.0, $P = 0.0001$). Interestingly, this haplotype carries a SNP (rs10506151) that is within a putative hnRNP K binding site and whose wild-type allele is conserved in vertebrates. hnRNP K is a component of the heterogeneous ribonucleoprotein (hnRNP) particle that is implicated in a variety of cellular processes including transcription and translation (28). It is possible that this SNP is a functional variant that alters the processing of *LRRK2* and directly increases disease risk. However, given the level of LD, we observed within this gene, we postulate the association is indirect, emanating from a true risk allele in LD with rs10506151. This yet-to-be-identified variant could lie anywhere between rs10506151 and all 21 SNPs from rs11564115 through rs3761863 as LD is complete ($D' = 1$) between these SNPs (Fig. 1D). However, the identification of a risk haplotype has enabled us to prioritize 25 individuals (cases carrying two copies of Haplotype 2) out of 189 cases (cases carrying at least one minor allele of rs10506151), who are likely to harbour a variant responsible for our remarkable observations. As none of the currently known *LRRK2* mutations are present, detailed

Table 4. Haplotype 2 frequencies and ORs

I			II	
No. of copies of Haplotype 2	No of individuals		Model	OR (95%CI; *P*)
	Control	Case		
None	314	286	Recessive[a]	5.2 (2.0–13.3; 0.0002)
1	147	155	Additive[b] (1 copy)	1.2 (0.9–1.5; 0.30)
2	5	25	Additive[c] (2 copies)	5.5 (2.1–14.0; 0.0001)
1 or 2	152	180	Dominant[d]	1.3 (1.0–1.7; 0.06)

I. Breakdown of the number of copies of Haplotype 2 carried.
II. ORs for Haplotype 2 under different disease assumptions.
[a]Individuals carrying 2 copies versus all other individuals.
[b]Individuals carrying 1 copy versus individuals carrying no copies.
[c]Individuals carrying 2 copies versus individuals carrying no copies.
[d]Individuals carrying no copies versus all other individuals.

analysis of coding and regulatory sequences in these individuals is now justified. Our haplotype association analyses also suggest that protective variants may exist within *LRRK2*. However, these observations are based on marginal significance (for Haplotype 3) or small numbers (for Haplotype 11) and must be interpreted with caution.

We acknowledge the possibility that our findings may represent 'spurious' associations—a common pitfall of many association studies. However, we have taken care to minimize potential confounders, in particular by using a large case–control series well matched in terms of age, gender and ethnicity. This strategy is also believed sufficient to attenuate the possible, and debatable effect of population stratification (27).

Paisan-Ruiz *et al*. (29) recently reported negative association of four *LRRK2* SNPs (two coding non-synonymous and two coding synonymous) in a smaller and unmatched Caucasian case–control series. However, patients in this study were selected on the basis of early onset and/or family history of PD and the level of *LRRK2* variation assessed was extremely limited. It is also possible that sporadic disease risk associated with common variation in *LRRK2* is influenced by other genetic and non-genetic factors, some of which may be population specific. It will therefore now be of value to investigate our findings in other populations.

In conclusion, our results strongly suggest that, in addition to the established role in familial disease, *LRRK2* genetic variation may indeed make a significant contribution to common forms of PD in the general community. Isolation of the responsible factors will provide further clues to understanding the pathogenesis of this complex disorder and further study is warranted.

MATERIALS AND METHODS

Samples

From the movement disorder clinics of two major tertiary institutions in Singapore (Singapore General Hospital and National Neuroscience Institute) 466 PD patients were included in this study. PD diagnosis was made in accordance with the UK PD Society Brain Bank Clinical Diagnostic Criteria of PD, by two movement disorders specialists (EK Tan and L Tan). The mean age of disease onset (AOO) was 60 ± 12 years (range 24–81), mean current age was 65 ± 11 years (range 30–91) and 56% were male. All patients were considered sporadic, as no family history of PD was evident. All were of Chinese ethnicity. None carry the common G2019S mutation (30) or any of the other 13 *LRRK2* mutations reported to date (R793M, R1067Q, L1114L, I1122V, I1371V, R1441C, R1441G, R1441H, IVS31+3 A>G, IVS33+6 T>A, Y1699C, M1869T, I2020T); (EK Tan *et al*., manuscript in preparation). Each patient was matched with an unrelated control individual (mean age 60 ± 11 years, range 26–93), free of neurological disease ($n = 466$). Matching criteria included a similar age (maximum age difference was ± 10 years), same sex and ethnicity. Institutional ethics committees approved this study and informed consent was obtained from all participants.

SNP selection and genotyping

Aiming for an average marker density of one SNP/5 kb, 30 SNPs with a MAF >0.02 in the Chinese population (according to HapMap Project population data, URL http://www.hapmap.org/) were selected for genotyping in 92 control samples. SNPs were genotyped using MALDI-TOF mass spectrometry using the Sequenom MassARRAY™ system (San Diego, CA, USA) as previously described (19). Briefly, multiplex genotyping assays were designed using the Sequenom DESIGNER software. Initial PCR (5 ng of genomic DNA) and primer extension reactions were carried out according to the Sequenom MassEXTEND protocol. Primers and allele codes are shown in Table 1. After purification, 15 nl of primer extension product was analysed with a MassARRAY Sequenom-Bruker Spectrometer (Bruker Biosciences). SNPs with overall call rates <0.85 (percentage of samples giving a reliable genotype call) and/or those that significantly deviated from Hardy–Weinberg equilibrium (at $\alpha = 0.05$) were excluded from further analyses ($n = 5$).

LD characterization, tSNP selection and performance evaluation

Pair-wise D' and r^2 values were calculated using multilocus haplotypes inferred from phase-unknown genotype data using PHASE version 2.1 (31). tSNPs were identified using the TSSA algorithm. This is a block-independent approach that finds a minimum set of tSNPs which uniquely distinguish a certain percentage of all possible haplotypes (25). α was set at 0.95 in order that the tSNPset returned was able to uniquely distinguish 95% of all possible haplotypes.

Performance of the selected tSNPset to capture haplotypic variation within *LRRK2* was assessed by calculating R^2_h between tSNP haplotypes and common (frequency >0.05) full haplotypes (comprising 25 SNPs) (32). This value can range from 0 to 1, with 1 indicating that tSNP haplotypes can perfectly predict all common 25 SNP haplotypes. To evaluate tSNP performance in capturing unobserved genotypic variation within *LRRK2* we performed the 'leave-one-out' analysis (33). In brief, each of the 25 genotyped SNPs (k) was dropped in turn and tSNPs were selected from the

remaining SNPs ($k - 1$). Haplotype r^2 (possible value range 0–1) was then calculated between each tSNP set and remaining SNPs. This evaluation can provide an unbiased and accurate estimate of tSNP performance in predicting genotypes that are not directly assessed (34).

Statistical analysis

Odds ratios, 95%CIs and *P*-values were calculated using web-based simple interactive statistical analysis (SISA) tools (URL http://home.clara.net/sisa/index.htm). Kruskal–Wallis tests were carried out using Microsoft® Excel 2002.

In silico analysis

Sequence conservation around rs10506151 was analysed using conservation tracks available at the University of California at Santa Cruz Human Genome Browser Gateway (May 2004 assembly) (URL: http://genome.ucsc.edu/cgi-bin/hg). Intronic regulatory sequence analysis was performed using the web-based Alternative Splicing Workbench (URL: http://www.ebi.ac.uk/asd-srv/wb.cgi).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge all participants in this study. They also would like to thank Jason Ong and Seok Shin Ng for genotyping and Marie Wong for construction of Figure 1. This study was supported by funding from the Agency for Science and Technology and Research of Singapore.

Conflict of Interest statement: The authors declare no conflicts of interest.

REFERENCES

1. Rocca, W.A. (2005) Prevalence of Parkinson's disease in China. *Lancet Neurol.*, **4**, 328–329.
2. Gelb, D.J., Oliver, E. and Gilman, S. (1999) Diagnostic criteria for Parkinson disease. *Arch. Neurol.*, **56**, 33–39.
3. Forno, L.S. (1996) Neuropathology of Parkinson's disease. *J. Neuropathol. Exp. Neurol.*, **55**, 259–272.
4. Vila, M. and Przedborski, S. (2004) Genetic clues to the pathogenesis of Parkinson's disease. *Nat. Med.*, **10** (suppl.), S58–S62.
5. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R. *et al.* (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, **276**, 2045–2047.
6. Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Graeber, M., Kosel, S., Przuntek, H., Epplen, J.T., Schols, L. and Riess, O. (1998) Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat. Genet.*, **18**, 106–108.
7. Zarranz, J.J., Alegre, J., Gomez-Esteban, J.C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., Atares, B. *et al.* (2004) The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.*, **55**, 164–173.
8. Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M. and Goedert, M. (1998) alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc. Natl. Acad. Sci. USA*, **95**, 6469–6473.
9. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R. *et al.* (2003) alpha-Synuclein locus triplication causes Parkinson's disease. *Science*, **302**, 841.
10. Chartier-Harlin, M.C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M. *et al.* (2004) alpha-Synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet*, **364**, 1167–1169.
11. Pals, P., Lincoln, S., Manning, J., Heckman, M., Skipper, L., Hulihan, M., Van Den Broeck, M., De Pooter, T., Cras, P., Crook, J. *et al.* (2004) alpha-Synuclein promoter confers susceptibility to Parkinson's disease. *Ann. Neurol.*, **56**, 591–595.
12. Lucking, C.B., Durr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B.S., Meco, G., Denefle, P., Wood, N.W. *et al.* (2000) Association between early-onset Parkinson's disease and mutations in the parkin gene. French Parkinson's Disease Genetics Study Group. *N. Engl. J. Med.*, **342**, 1560–1567.
13. Mata, I.F., Lockhart, P.J. and Farrer, M.J. (2004) Parkin genetics: one model for Parkinson's disease. *Hum. Mol. Genet.*, **13** (Spec no. 1), R127–R133.
14. West, A.B., Maraganore, D., Crook, J., Lesnick, T., Lockhart, P.J., Wilkes, K.M., Kapatos, G., Hardy, J.A. and Farrer, M.J. (2002) Functional association of the parkin gene promoter with idiopathic Parkinson's disease. *Hum. Mol. Genet.*, **11**, 2787–2792.
15. Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K. *et al.* (2000) Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat. Genet.*, **25**, 302–305.
16. Leroy, E., Boyer, R., Auburger, G., Leube, B., Ulm, G., Mezey, E., Harta, G., Brownstein, M.J., Jonnalagada, S., Chernova, T. *et al.* (1998) The ubiquitin pathway in Parkinson's disease. *Nature*, **395**, 451–452.
17. Mcnaught, K.S., Belizaire, R., Isacson, O., Jenner, P. and Olanow, C.W. (2003) Altered proteasomal function in sporadic Parkinson's disease. *Exp. Neurol.*, **179**, 38–46.
18. Liu, Y., Fallon, L., Lashuel, H.A., Liu, Z. and Lansbury, P.T., Jr. (2002) The UCH-L1 gene encodes two opposing enzymatic activities that affect alpha-synuclein degradation and Parkinson's disease susceptibility. *Cell*, **111**, 209–218.
19. Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B. *et al.* (2004) Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*, **44**, 601–607.
20. Bosgraaf, L. and Van Haastert, P.J. (2003) Roc, a Ras/GTPase domain in complex proteins. *Biochim. Biophys. Acta*, **1643**, 5–10.
21. Pardo, L.M. and Van Duijn, C.M. (2005) In search of genes involved in neurodegenerative disorders. *Mutat. Res.*, in press.
22. Gilks, W.P., Abou-Sleiman, P.M., Gandhi, S., Jain, S., Singleton, A., Lees, A.J., Shaw, K., Bhatia, K.P., Bonifati, V., Quinn, N.P. *et al.* (2005) A common *LRRK2* mutation in idiopathic Parkinson's disease. *Lancet*, **365**, 415–416.
23. Cardon, L.R. and Abecasis, G.R. (2003) Using haplotype blocks to map human complex trait loci. *Trends Genet.*, **19**, 135–140.
24. Johnson, G.C., Esposito, L., Barratt, B.J., Smith, A.N., Heward, J., Di Genova, G., Ueda, H., Cordell, H.J., Eaves, I.A., Dudbridge, F. *et al.* (2001) Haplotype tagging for the identification of common disease genes. *Nat. Genet.*, **29**, 233–237.
25. Liu, T., Sung, W.-K., Li, Y., Liu, J., Mittal, A. and Mao, P. (2005) Effective algorithms for Tag SNP selection. *J. Bioinf. Comp. Biol.*, in press.
26. Ahmadi, K.R., Weale, M.E., Xue, Z.Y., Soranzo, N., Yarnall, D.P., Briley, J.D., Maruyama, Y., Kobayashi, M., Wood, N.W., Spurr, N.K. *et al.* (2005) A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat. Genet.*, **37**, 84–89.
27. Newton-Cheh, C. and Hirschhorn, J.N. (2005) Genetic association studies of complex traits: design and analysis issues. *Mutat. Res.*, **573**, 54–69.
28. Bomsztyk, K., Denisenko, O. and Ostrowski, J. (2004) hnRNP K: one protein multiple processes. *Bioessays*, **26**, 629–638.
29. Paisan-Ruiz, C., Lang, A.E., Kawarai, T., Sato, C., Salehi-Rad, S., Fisman, G.K., Al-Khairallah, T., St George-Hyslop, P., Singleton, A. and Rogaeva, E. (2005) *LRRK2* gene in Parkinson disease. *Neurology*, **44**, 595–600.

30. Tan, E.K., Shen, H., Tan, L.C., Farrer, M., Yew, K., Chua, E., Jamora, R.D., Puvan, K., Puong, K.Y., Zhao, Y. *et al.* (2005) The G2019S *LRRK2* mutation is uncommon in an Asian cohort of Parkinson's disease patients. *Neurosci. Lett.*, **384**, 327–329.
31. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
32. Stram, D.O., Haiman, C.A., Hirschhorn, J.N., Altshuler, D., Kolonel, L.N., Henderson, B.E. and Pike, M.C. (2003) Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum. Hered.*, **55**, 27–36.
33. Weale, M.E., Depondt, C., Macdonald, S.J., Smith, A., Lai, P.S., Shorvon, S.D., Wood, N.W. and Goldstein, D.B. (2003) Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene *SCN1A*: implications for linkage-disequilibrium gene mapping. *Am. J. Hum. Genet.*, **73**, 551–565.
34. Iles, M.M. (2005) Obtaining unbiased estimates of tagging SNP performance. *Ann. Hum. Genet.*, in press.