Human Molecular Genetics, 2004, Vol. 13, No. 12 **1267–1274**
DOI: 10.1093/hmg/ddh138
Advance Access published on April 28, 2004

The structure of the tau haplotype in controls and in progressive supranuclear palsy

Alan M. Pittman[1,2,†], Amanda J. Myers[3,†], Jaime Duckworth[3], Leslie Bryden[3], Melissa Hanson[3], Patrick Abou-Sleiman[2], Nicholas W. Wood[2], John Hardy[1,3], Andrew Lees[1,2] and Rohan de Silva[1,2,*]

[1]Reta Lila Weston Institute of Neurological Studies, University College London, Windeyer Building, 46 Cleveland Street, London W1T 4JF, UK, [2]Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK and [3]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Building 10, Bethesda, MD 20892, USA

Received February 18, 2004; Revised and Accepted April 7, 2004

The group of neurodegenerative diseases collectively known as tauopathies are characterized by hallmark lesions consisting of fibrillar aggregates of the microtubule-associated protein, tau (MAPT). Mutations of the tau gene (*MAPT*) are the cause of frontotemporal dementia with parkinsonism linked to chromosome 17, giving tau a central role in the pathogenic process. The chromosomal region containing *MAPT* has been shown to evolve into two major haplotypes, H1 and H2, which are defined by linkage disequilibrium (LD) between several polymorphisms over the entire *MAPT* gene. Studies to date suggest a complete absence of recombination between these two haplotypes. The more common haplotype H1 is over-represented in patients with progressive supranuclear palsy (PSP) and corticobasal degeneration. Using single nucleotide polymorphisms, we mapped LD in the regions flanking *MAPT* and have established the maximum extent of the haplotype block on chromosome 17q21.31 as a region covering ~2 Mb. This gene-rich region extends centromerically beyond the corticotrophin releasing hormone receptor 1 gene (*CRHR1*) to a region of ~400 kb, where there is a complete loss of LD. The telomeric end is defined by an ~150 kb region just beyond the *WNT3* gene. We show that the entire, fully extended H1 haplotype is associated with PSP, which implicates several other genes in addition to *MAPT*, as candidate pathogenic loci.

INTRODUCTION

The tauopathies are a group of neurodegenerative disorders that are characterized by fibrillar aggregates of the microtubule-associated protein, tau (MAPT). These disorders include Alzheimer's disease (AD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), Pick's disease (PiD) and frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17), with a clinical spectrum ranging from dementia to parkinsonian phenotypes (1). The recent identification of missense and splice-site mutations in the tau gene, *MAPT* (MIM 157140), causing FTDP-17 (MIM 600274) affirmed a central role of tau dysfunction in neurodegenerative aetiology (2,3). Although the other related tauopathies, including AD, PSP and CBD are defined by fibrillar tau pathology, *MAPT* is not mutated in these diseases.

PSP (MIM 601104; Steele–Richardson–Olszewski syndrome) (4) is a sporadic disorder of late adult life. However, robust genetic association of PSP with *MAPT* and rare reports of families with more than one affected member (5,6) indicate that genetic factors could play a role in PSP. Conrad *et al.* (7) were the first of many groups that proposed that variation in *MAPT* itself could be an important genetic influence in sporadic PSP, by demonstrating the association of the more common allele (a_0) of a dinucleotide marker in *MAPT* intron 9 with PSP. Over-representation of this allele was subsequently reported in the related, but rarer, four repeat primary tauopathy CBD (8). The implication is that this polymorphism itself contributes to increased risk or that it is in linkage disequilibrium (LD) with the actual causative variant. Although some *MAPT* mutations in FTDP-17 cause a clinical PSP phenotype (9–11), no obvious causative

*To whom correspondence should be addressed. Tel: +44 2076799266; Fax: +44 2076799429; Email: rsilva@ion.ucl.ac.uk
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

variations of *MAPT* have been identified in clinically and pathologically diagnosed sporadic and familial PSP (12).

The allelic association of *MAPT* with PSP and CBD was subsequently extended to a series of polymorphisms extending over the entire *MAPT* coding region spanning nearly 62 kb (13). In ~200 unrelated Caucasians, these polymorphisms were shown to be in complete LD, thereby forming two extended haplotypes, H1 and H2 (13). The authors suggested that the establishment of these two haplotypes was an ancient event and that either recombination was suppressed in this region or recombinants were selected against. Baker *et al.* (13) also demonstrated that the more common haplotype H1, with which the a_0 allele segregated, was significantly over-represented in PSP. The *MAPT* haplotype was subsequently extended a further 68 kb to the three single nucleotide polymorphisms (SNPs) in the promoter region of *MAPT* (14,15).

Identifying the functional basis of the *MAPT* H1 haplotype association provides the most promising prospect for understanding the aetiopathogenesis of PSP and CBD. Pathological evidence is compelling in implicating the *MAPT* locus itself, with hitherto unidentified polymorphism(s) that could affect tau expression or splicing. It is therefore important to analyze the underlying LD within the H1 haplotype. Any neighbouring locus that is in LD with the *MAPT* haplotype could also be a candidate gene that harbours a pathogenic variant. The recent identification of the saitohin gene (*STH*) illustrates this possibility (16). We have shown that a coding polymorphism (Q7R) in this gene is in complete LD with the *MAPT* haplotype and the Q (H1) variant is associated with PSP (17). The function of saitohin and the effect of the Q7R polymorphism need to be determined so as to assess its importance in disease predisposition.

The pool of candidate genes has increased considerably with recent findings that show that the H1/H2 haplotype block consisting of the region of almost complete LD extends far beyond *MAPT* (18). The possible presence of pathogenic loci other than *MAPT* is supported by variants of FTDP-17 that are characterized by the absence of tau pathology and *MAPT* mutations (19–21).

The aim of this work was to define the complete *MAPT* haplotype block by delineating the outer edges of this extended region of LD. Several recent studies have suggested the concept of haplotype blocks, discrete block-like regions with near complete LD, typically <100 kb (22,23). By virtue of this, haplotype blocks correspond to regions of low recombination compared with recombination hotpots in intervening regions (22,23). Evidence from previous studies strongly suggests that the H1/H2 haplotype covering the *MAPT* gene has undergone little recombination. Haplotypic diversity within the H1 haplotype may suggest that one or more unidentified H1 haplotypes could harbour functional polymorphic variants that are responsible for the well-documented increased risk of PSP and CBD. Knowing the extent of the *MAPT* haplotype block would enable us to analyze the underlying variation in candidate genes within the H1 haplotype and establish H1-specific subtypes that could flag up potential pathogenic variants. These could also account for cases of FTDP-17 that lack tau pathology and are without any functional mutations in *MAPT* (19–21).

For this LD analysis, we selected a series of SNPs from the SNP Consortium database (http://snp.cshl.org), covering regions of 1 Mb from each edge of the tau gene (5′ end of the *MAPT* promoter and 3′ end of *MAPT* exon 14). We analyzed the LD of the SNPs with the tau HI/H2-defining intron 9 deletion polymorphism (*del-In9*) (13) as the point of reference. We also analyzed the haplotype block in a PSP case–control group (63 controls and 60 PSP cases) and showed that the entire extended *MAPT* H1 haplotype block is associated with PSP. These data define the extent of the candidate region, which harbours a genetic variant(s) that is an important contributory factor in the pathogenesis of PSP.

RESULTS

LD

To analyze the LD across the *MAPT* region, SNPs were initially selected from the SNP Consortium (http://snp.cshl.org) at intervals of ~50 kb, covering regions of 1 Mb on either side of *MAPT* and also on the basis of having available frequency information (minor allele frequencies >0.1). SNPs (characterized and uncharacterized) were also selected on the basis of their position and type. The SNPs were genotyped using pyrosequencing or restriction enzyme analysis [restriction fragment length polymorphism (RFLP)] in 63 normal individuals and 60 PSP cases, all cases and controls were unrelated and of Caucasian, western European origin. The LD was calculated pair-wise, separately in the cases and in the controls, using the statistical LD calculations for D' and r^2 from the expectation–maximization (EM) derived haplotypes. There was no significant deviation from Hardy–Weinberg equilibrium (HWE; significance level was set at $P < 0.05$) in the controls and PSP cases at any of the loci (Table 1). Previous studies show that the polymorphisms scattered throughout the coding and promoter regions of *MAPT* are in complete LD (13–15,17). Working outwards from *MAPT*, we defined the maximum size of the haplotype block and proceeded to locate the edges of this extended LD block. In both the control and the PSP cohorts, there was tight LD extending beyond *MAPT* in both directions as was reflected by the match of genotypes at these loci.

Centromeric from *MAPT*, LD extends at least 0.39 Mb beyond the *CRHR1* gene and ends within a ~0.4 Mb region that includes several known genes such as mitogen activated kinase kinase kinase 14 (*MAP3K14*) and phospholipase C, delta 3 (*PLCD3*) (Figs 1 and 2). The SNP rs894685 and the glial fibrillary acidic protein gene (*GFAP*), which is ~53 kb centromeric from *CRF*, clearly lie outside the haplotype block. In the telomeric direction, LD extends ~0.8 Mb from the 3′ end of *MAPT* to the *N*-ethylmaleimide sensitive factor gene (*NSF*). Within the region of 65 kb between *NSF* and the *WNT3* genes (wingless-type MMTV integration site family, member 3), there is a progressive decay in LD (Figs 1 and 2). With this, we have defined the maximum extent of the *MAPT* haplotype block as a region of ~2 Mb that is in LD. The centromeric end of the haplotype block could only be resolved to a minimum region of ~0.4 Mb within which there is complete loss of LD (Figs 1 and 3). This was due to the lack of informative SNPs in this region.

Table 1. SNPs used in the analysis of the *MAPT* haplotype block

SNP	Position	Assay	Minor allele frequency		Hardy–Weinberg equilibrium		Case–control association	
			Controls	Cases	Controls (*P*)	Cases (*P*)	Genotypic (*P*)	Allelic (*P*)
rs894685	43530204	RFLP	0.185	0.310	0.734	0.386	0.070	0.030
rs732589	43946182	RFLP	0.295	0.353	0.672	0.888	0.586	0.336
rs2668643	44146521	RFLP	0.192	0.070	0.315	0.133	0.005	0.005
rs1880748	44264218	RFLP	0.193	0.200	0.917	0.197	0.662	0.893
rs110402	44355457	PYRO	0.492	0.377	0.379	0.749	0.103	0.045
rs1396862	44378417	PYRO	0.194	0.070	0.282	0.139	0.004	0.005
rs916793	44430116	PYRO	0.198	0.070	0.240	0.128	0.004	0.002
del-In9	44562589	PCR	0.198	0.070	0.240	0.128	0.002	0.002
rs1468241	44671571	PYRO	0.189	0.070	0.328	0.139	0.007	0.006
rs1528072	44712141	PYRO	0.200	0.070	0.282	0.144	0.004	0.004
rs2240756	44735106	PYRO	0.143	0.313	0.769	0.463	0.004	0.001
rs142167	45270057	PYRO	0.186	0.036	0.078	0.780	0.002	0.002
rs199528	45317944	RFLP	0.196	0.083	0.121	0.324	0.007	0.017
rs70602	45334515	PYRO	0.173	0.085	0.112	0.333	0.024	0.048
rs1662577	45471397	RFLP	0.460	0.375	0.396	0.810	0.392	0.175
rs758391	45491879	RFLP	0.500	0.466	0.370	0.670	0.567	0.599

Approximate positions of the SNPs are given according to chromosome 17 co-ordinates from the July 2003 assembly (http://genome.ucsc.edu). Comparisons of minor allele frequencies between PSP cases and controls and *P*-values for HWE and for case–control associations for each SNP are indicated.

We are currently working to refine the definition of this 'edge'. At the telomeric end, which is marked by the *WNT3* gene, we resolved a minimal region of ~150 kb within which there is a complete loss in LD (Figs 1 and 3). Two outlying SNPs, rs758391 and rs1662577, that are not in LD with *del-In9*, are in LD with each other, suggesting that they are part of an adjacent haplotype block. We calculated both D' and r^2 as statistical values for LD pair-wise between each SNP. Although the two methods showed differences in the LD pattern, i.e. D' appears less discriminatory in assigning significant LD between pairs of loci, they were both in agreement on the maximum size of the *MAPT* haplotype block. Furthermore, comparing the LD patterns between the controls and PSP cases did not reveal any major differences; the SNP rs2240756 appeared less strongly in LD by D' in the PSP group, as did SNP rs732589 in the control group. In this study, we encountered three types of SNPs. Firstly, those SNPs that had strong LD with the H1/H2 *MAPT* intron 9 deletion polymorphism (*del-In9*), by both $r^2$and D' and thereby define the extended haplotype block (rs70602, rs199528, rs142167, rs1528072, rs1468241, H1/H2 del, rs916793, rs1396862, rs110402 and rs2668643). SNPs rs2668643, rs1396862, rs916793, H1/H2 del, rs1468241 and rs1528072 had an r^2 value >0.8 and SNPs rs142167, rs199528 and rs70602 showed a decline in r^2, and to a lesser extent in D', which defines the decay in LD at the edge of the extended haplotype.

Secondly, those SNPs that have only modest to weak LD with *del-In9* but were contained within the extended haplotype (rs2240756, rs110402, rs1880748 and rs732589). The degree of LD of these SNPs was low when determined by the r^2 statistics but, depending on the SNP, showed modest to high LD according to the D' values. These differences in LD measure are because of the fact that r^2 has taken into account the differences in the allele frequencies of these SNPs (Table 1). D' can reach a high value of LD when the frequencies of the SNPs vary widely, high values imply that copies of the rarer allele occur exclusively with one of the possible alleles of the other marker. Recent work suggests that r^2 is viewed more favourably as a measure of association for the causal variant than that of D' (24). D' is good at identifying the 'block' structure of LD and r^2 is better for defining the 'associated interval' and identifying potential causal variants. By D', the haplotype block extends from rs732589 to rs70602. Using an EM algorithm, we predicted the haplotypes that could be formed by these SNPs and the *del-In9*, and noted that multiple haplotypes segregate with the H1 variant of *del-In9*, whereas only one major form segregates with the H2 variant of *del-In9*. This prediction suggests that these SNPs are specific to the H1 haplotype in that variation occurs only within the H1 background (H1-specific SNPs) and that there is negligible variation at these SNPs within the H2 haplotype in our case–control cohorts. It is interesting to note that the H1-specific SNPs (some with relatively high minor allele frequencies) have extremely low LD when compared with one-another and measured by r^2, but have medium LD when measured with D'. These data suggest that there is considerable diversity and DNA recombination across the extended H1 haplotype.

In the third group of SNPs (rs758391, rs1662577 and rs894685), there was no detectable LD by r^2 ($r^2 < 0.1$) and very low D' with *del-In9* as they lie outside the limits of the haplotype block. They show random variation relative to the H1 and H2 haplotypes confirming that they are not in LD with *del-In9*.

Association of the extended haplotype block with PSP

We performed a case–control, locus-by-locus association study of all the SNPs used in this study and, not surprisingly, all of the SNPs that were in LD with *del-In9* forming the extended haplotype block were associated with PSP (Table 1 and Fig. 2). Interestingly, some of the H1-specific SNPs

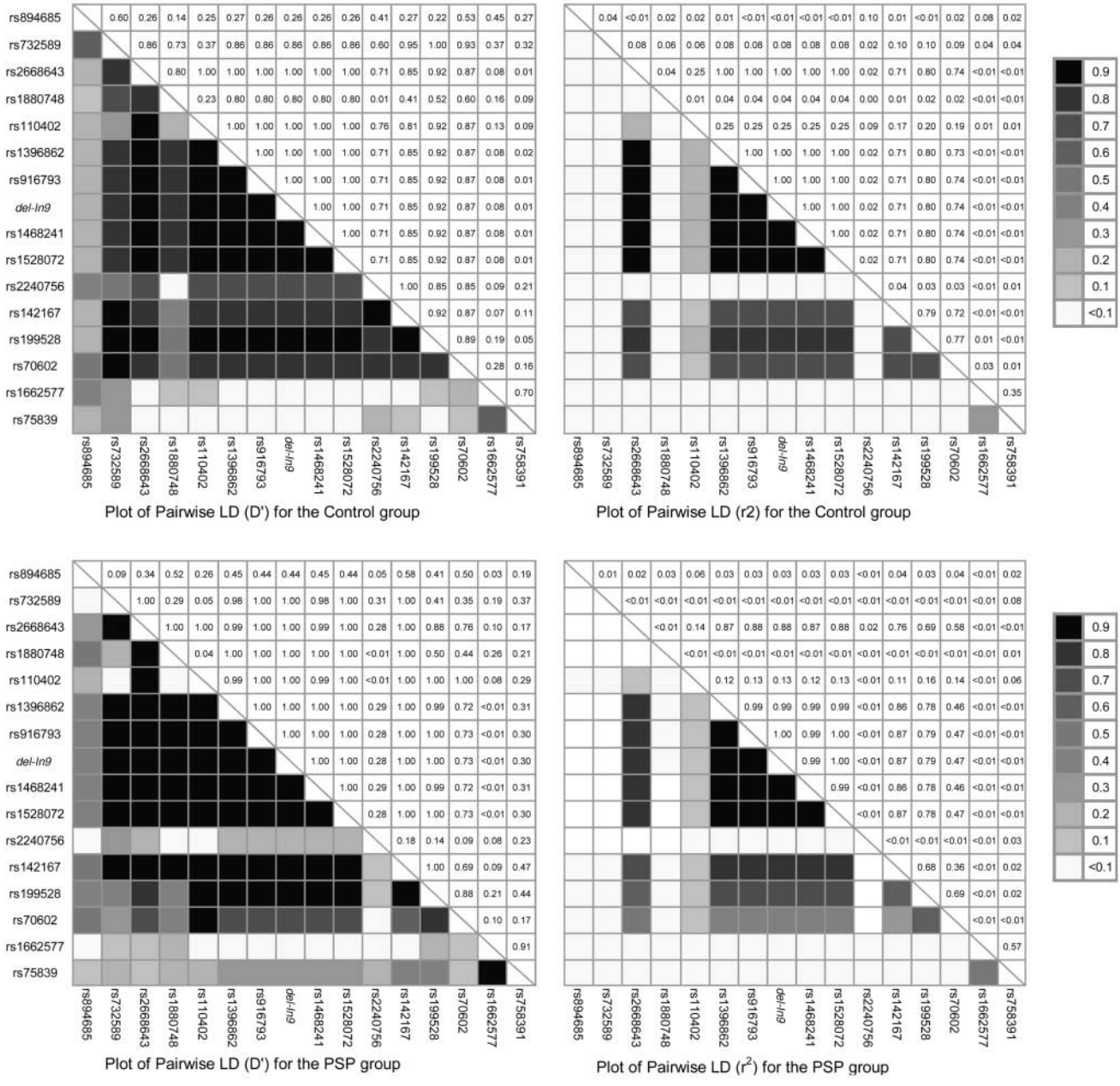

Figure 1. Pair-wise r^2 and D' LD analysis of all of the SNPs for the control and the PSP populations. The blocks are shaded corresponding to the values which were obtained from the LD analysis programme TagIT.

(rs2240756 and rs110402), but not others (rs732589 and rs1880748), showed significant allelic and genotypic association with PSP (Table 1). The associated H1-specific SNPs flank *MAPT* (Fig. 3) and could be part of a sub-class of the H1 haplotype that includes the pathogenic variant(s). We deem these SNPs important for future study of the underlying architecture of the H1 haplotype in order to progressively refine the candidate regions responsible for PSP, CBD and some cases of Parkinson's disease (PD).

At the telomeric end of the haplotype, the SNPs rs142167, rs199528 and rs70602 show a progressive decay in r^2 LD with increased distance from *MAPT*. This is reflected by a corresponding decrease in association with PSP, with a complete loss of LD and association with the outlying SNPs rs758391 and rs1662577 (Table 1 and Figs 1 and 2).

At the centromeric end of the haplotype, the outlying SNP rs894685 also shows evidence of allelic and genotypic association with PSP (Table 1 and Fig. 2). This is surprising because it is completely out of LD ($r^2 < 0.1$) with the extended *MAPT* haplotype block, and does not follow the expected correspondence of decrease in LD and association observed at the centromeric end. This could be

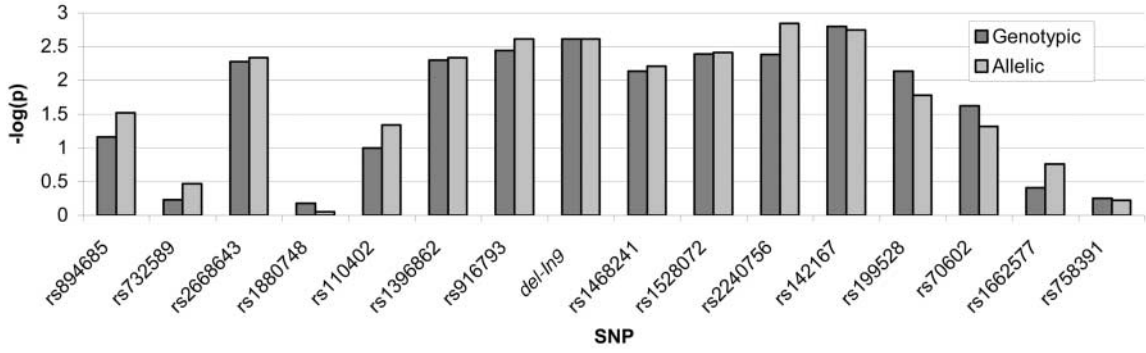

Figure 2. Significant association of *MAPT* haplotype SNPs with PSP. Plot of $-\log(P)$ for case–control test for both allele and genotype association of the SNPs with PSP.

due to variability in the extent of LD in the region. However, we typed a further SNP (rs911) that lies more centromeric from rs894685 and showed that it also was completely out of LD and not associated with PSP (data not shown). We are currently mapping this region in question with a higher density of informative SNPs and in larger study cohorts and in multiple ethnic groups.

DISCUSSION

With the *MAPT* haplotype block, we have defined a region of ~2 Mb on chromosome 17q21 that is in complete LD (Fig. 1) both pair-wise and relative to the H1/H2 defining *MAPT* intron 9 deletion (*del-In9*). A shorter block of LD extending from the end of the *CRHR1* gene to the beginning of the *NSF* locus was found in a cohort of idiopathic late-onset PD cases and age-, region- and gender-matched controls (25). The region that we have defined provides us with the candidate region for more detailed analysis of the underlying architecture of the H1 haplotype and by more detailed LD analysis pinpoints the locus involved in PSP and CBD pathogenesis. The region is relatively gene rich and, other than *MAPT*, contains genes, both confirmed and predicted, that could be important for neuronal function or even be involved directly with tau function. *MAPT* lies in the centre of the haplotype block, with the presenilin homologue (*IMP5*/*PSH2*) and corticotrophin releasing hormone receptor (*CRHR1*) genes between *MAPT* and the centromeric edge of the haplotype block. *IMP5/PSH2*, which is at 47 kb from *MAPT* belongs to a family of proteins with sequence and structural homology to the presenilins (PS), PS1 and PS2, that are mutated in early-onset familial AD (26). Nine coding SNPs in *IMP5/PSH2* were shown to be in complete LD with the *MAPT* haplotype (26) but no pathological mutations were identified in *MAPT* mutation negative frontotemporal dementia families ('tau-negative' FTDP-17) (26). The telomeric end of the *MAPT* haplotype block is defined by the *NSF* and *WNT3* genes. The NSF protein is a synaptic protein that is important for the regulation of neurotransmitter release and acts as part of a protein complex involved in the docking and fusion of synaptic vesicles with target membranes (reviewed in 27). It has been suggested that the *NSF* gene may be involved in schizophrenia (28) and epilepsy (29). The WNT3 protein is involved in Central nervous system development (30,31).

There are possibly up to seven more genes between *MAPT* and *NSF*. Confirmed genes include LOC284058 and *ARF*. The former is a large gene just 5 kb telomeric to *MAPT* and its expression in brain has been confirmed separately by the cDNA clone KIAA1267 (32). In the most recent assembly of the chromosome 17 sequence (July 2003 from http://genome.ucsc.edu), the region of the *ARF* and *NSF* genes is duplicated, as indicated by the perfect tandem duplication of a region covering the entire *ARF* gene and exons 1–13 of the *NSF* gene (Fig. 2). Nested within each duplicated *ARF* gene there could be a gene resembling the *NBR2* gene (neighbour of *BRCA1* gene 2) (33) as suggested by the high degree of homology (93.6%). *NBR2* is also on chromosome 17q21, outside the *MAPT* haplotype block, about 2.66 Mb centromeric to the *CRHR1* and lies head-to-head with the breast cancer gene, *BRCA1* (33). In fact, the *NBR2* and *BRCA1* region is also duplicated tandemly with a region containing *NBR1* and the *LBRCA1* pseudogene (34) which has an arrangement similar to the duplicated *ARF* and *NSF* gene region. The *BRCA1/NBR2* gene region is also characterized by extended LD forming two major, highly conserved haplotypes of frequencies 0.64 and 0.33 that similarly show a lack of recombination (34).

We analyzed each SNP for association with PSP in cases and controls. As expected, all the SNPs in LD with *del-In9* showed evidence of association ($P \leq 0.05$) (Fig. 2). At the telomeric end of the haplotype block, the association of the SNPs tended to decrease when the LD decreased. This was not apparent at the centromeric end of the haplotype. The association with PSP may be explained by the SNP perhaps having LD with other markers inside the haplotype block that are also associated with PSP.

In a previous study, it was shown that the markers D17S810 and rs1816 had significant allelic association with PSP (18). Both markers lie within the extended haplotype block that we have defined (Fig. 3), thus supporting the validity of our findings. Future studies would require a denser, higher resolution map of SNPs in this region to comprehensively map association and LD at the edges of the *MAPT* haplotype block.

Haplotype-based association studies have been touted as powerful and comprehensive method for identifying the genetic cause of complex disorders (22,23). Both the LD

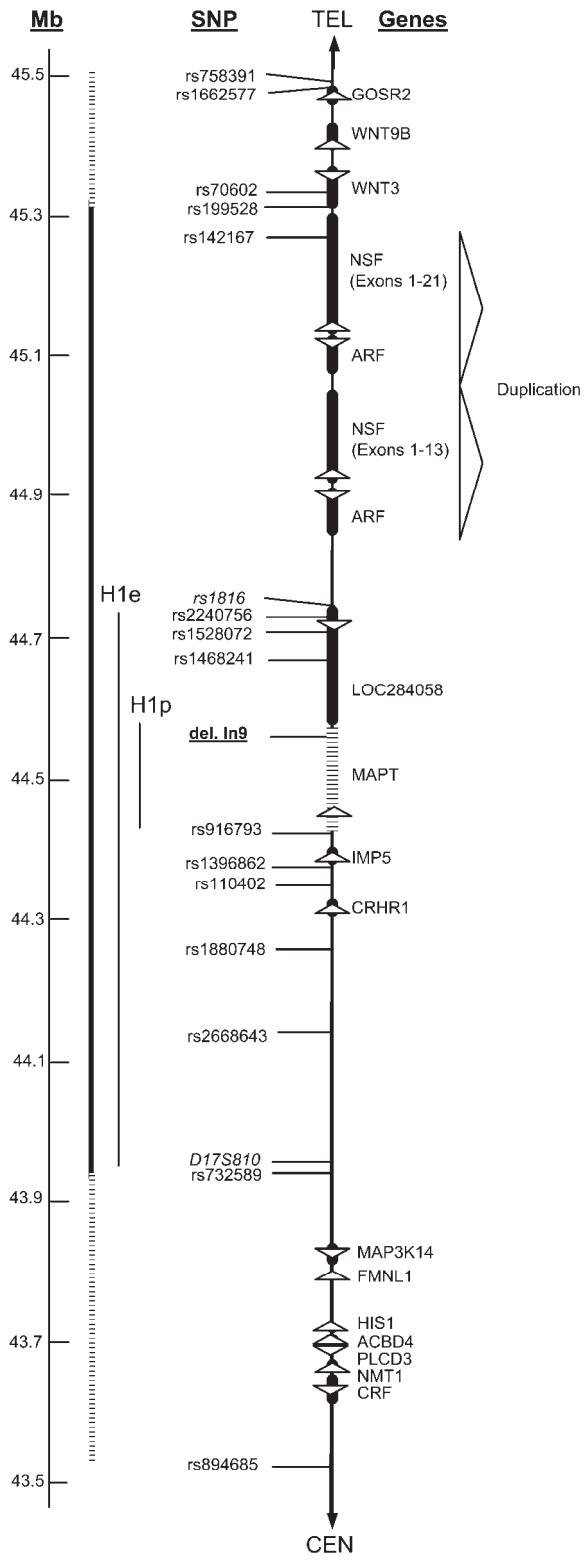

Figure 3. Region of chromosome 17q21.31 containing the extended *MAPT* haplotype block. The chromosomal co-ordinates (Mb; million base pairs) are indicated on the left-hand axis. They are based on the July 2003 draft of the human genome sequence (http://genome.ucsc.edu). Relative positions of SNPs and confirmed genes are indicated. Arrowheads on genes indicate the direction of transcription. The extents of the previously reported *MAPT* haplotypes H1p and H1e are indicated by bars on the left. CEN, centromeric; TEL, telomeric.

and the association studies give further strong evidence of a gene responsible for PSP in the region of the *MAPT* haplotype block. The extension and delineation of the outer limits of the *MAPT* haplotype block significantly increases the candidate region in both PSP and CBD, classifying the genes within this block as potential pathogenic candidates. However, based on the considerable evidence from tau related pathology of both PSP and FTDP-17, the *MAPT* locus should remain as the main focus of attention in the quest for a pathogenic polymorphism or H1 haplotype. This should include investigation of regions within *MAPT* that could influence gene expression, splicing or stability of message. However, this does not exclude the possibility that any other gene within this haplotype is directly or indirectly involved with *MAPT*, or that defects in a gene other than *MAPT* lead to tau pathology such as in the case of mutations in the *APP* and *PS* genes and the *APOE-ε4* allele in AD. For example, the *STH* gene nested within intron 9 of *MAPT* (7) could be involved directly in tau function. This is based on parallels with the vesicular acetylcholine transporter (*VAChT*) gene which is nested within the choline acetyltransferase (*ChAT*) gene and both genes, which could be co-regulated, are needed for expression of the cholinergic phenotype (35). At this stage it would be premature to speculate on the importance of the association of any single gene or polymorphism with PSP. It is clear that the entire *MAPT* haplotype block and the genes it contains are associated with PSP. In future studies, it would be important to study the underlying architecture of the extended *MAPT* H1 haplotype block so as to hone in on candidate H1 haplotype clades and their genes that are more tightly associated with PSP.

This study provides new data in defining the outer extent of LD of the *MAPT* haplotype block, providing the minimal chromosomal region for more detailed studies of haplotype architecture and association with PSP and CBD with the final goal of pinpointing the ancestral pathogenic locus/loci. Future investigations should therefore include: (1) investigation of intra-H1 variation across the entire *MAPT* haplotype block in order to establish H1 haplotypes within *MAPT* and elsewhere that, by virtue of being more tightly associated with PSP, could harbour any pathogenic variants; (2) re-sequence obvious candidate genes within the *MAPT* haplotype block in order to find PSP specific polymorphisms; and (3) investigate the LD pattern in the *MAPT* haplotype block in ethnic diversity panels. In a recent study of recombination across the MHC class II region, it was shown that very similar patterns of LD exist between divergent populations studied (North Europeans, Saami and Zimbabweans), with recombination hotspots playing the major role in shaping LD (36). This supports other studies that show that LD blocks seem to be common and are often shared between populations (23,37) indicating that haplotype structure in humans is ancient, predating the recent diversification (36).

MATERIALS AND METHODS

Subjects

The control population ($n = 63$; average age of death 72 years) were of Caucasian origin, from the European Brain Bank series and had neither clinical evidence of neurodegenerative disease

nor abnormal histopathology. The PSP cases ($n = 60$; average age at death 77 years) also of Caucasian, western European origin were all pathologically confirmed. Many of the above samples have been used in previous studies (12,15,17,38) and additional pathologically confirmed cases have been added to this study. Pathological confirmation of the diagnosis of PSP was made following standardized criteria (38). All patients and controls were collected under approved protocols followed by informed consent and his work was approved by the Joint Medical Ethics Committee of the National Hospital of Neurology and Neurosurgery, London.

PCR

The 238 bp *MAPT* H2 deletion in intron 9 (*del-In9*) was used to determine unambiguously the tau haplotype in the individuals as previously described (13). Oligonucleotide primer pairs were designed to specifically amplify by PCR, the *MAPT* haplotype SNPs of interest (rs895685, rs732589, rs2668643, rs1880748, rs110402, rs1396862, rs916793, rs1468241, rs1528072, rs2240756, rs142167, rs199528, rs70602, rs1662577 and rs758391). The reactions contained 25 ng of template DNA, 1 U of *Taq* DNA polymerase, 10× PCR buffer, 10 mM dNTP mix, 10 pmol of each oligonucleotide primer pair and 5× Q solution (Qiagen, Crawley, UK). Thermocycling conditions were carried out using the 'touchdown' method. Genotyping of the SNPs was carried out either by pyrosequencing (Pyrosequencing Inc., Westbourough, MA, USA) or by restriction digests (RFLP). PCR and pyrosequencing primer sequences and thermocycling conditions are available on request.

Genotyping

The SNPs rs110402, rs1396862, rs916793, rs1468241, rs1528072, rs2240756, rs142167 and rs70602 were analyzed by pyrosequencing. Twelve microlitres of the biotinylated PCR product was immobilized on streptavidin–sepharose™ HP (Amersham Pharmacia Biotech, Piscataway, NJ, USA): The gel slurry (4 μl) was resuspended in binding buffer (10 mM Tris–HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween-20). Template and beads were mixed continuously for >5 min at room temperature. The immobilized DNA template was transferred to a 96-well filter plate attached to a vacuum manifold (Millipore Inc., Bedford, MA, USA), then immersed for 10 s in ethanol followed by denaturing buffer (0.2 M NaOH) and finally wash buffer (10 mM Tris–acetate, pH 7.6). The sequencing primer (15 pmol) was then annealed to the single-stranded template in 12 μl of annealing buffer (20 mM Tris–acetate, 2 mM magnesium acetate, pH 7.6) at 80°C for 2 min before cooling to room temperature. Samples were analyzed using a PSQ 96 system together with SNP software and SNP reagent kits (Pyrosequencing Inc.) following the manufacturer's instructions.

For SNP analysis by RFLP, 15 μl of PCR product—SNPs: rs1880748 [*Bsp* HI (C)]; rs199528 [*Pvu* II (C)]; rs758391 [*Hph* I (A)]; rs1662577 [*Bsr*GI (C)]; rs2668643 [*Apo* I (A)] and rs894685 [*Acc* I (T)]—was digested by 1 U of the corresponding restriction endonuclease [New England Biolabs, Hertfordshire, UK (*Bsp* HI, *Dra* III, *Hph* I, *BsrG* I, *Apo* I and *Acc* I) Promega, Southampton, UK (*Pvu* II)] in a reaction volume of 20 μl for 4 h. The PCR products are all cleaved by the corresponding enzyme once at the indicated (N) allele. Digests were run on a 4% agarose gel for analysis. Genotype scoring was carried out blindly by two individuals. Any discrepancies between the two were resolved by repeating the assay.

LD and statistical analysis

For each SNP, the allele and genotype distributions in the group of PSP patients were compared with those in the control group. Statistical assessments for the allele and genotype frequencies and HWE were made using the genetics software program TagIt (http://popgen.biol.ucl.ac.uk/software.html) (24). The square of the correlation coefficient (r^2) (39) and D' for LD was calculated pair-wise between each. Both measures of LD are based on D, the basic pair-wise disequilibrium coefficient, the difference between the probabilities of observing them independently in the population: $D = f(A_1B_1) - f(A_1)f(B_1)$ (40), where A and B refer to two genetic markers and f is their frequency. D' is obtained from D/D_{max} and a value of 0.0 suggests independent assortment, whereas 1.0 means that all copies of the rarer allele occur exclusively with one of the possible alleles at the other marker. r^2 has a more strict interpretation than that of D', $r^2 = 1.0$ only when the marker loci have identical allele frequencies; the allele at the one locus can always be predicted by the allele at the other locus. In contrast, D' can reach a value of 1.0 when the allele frequencies vary. An r^2 value >0.5 represents a high level of LD, r^2 values <0.1 represent essentially no LD. Haplotype predictions were made using an EM algorithm using TagIT. Case–control locus-by-locus association was calculated statistically using a χ^2 distribution and the significance was calculated using a Monte–Carlo approach as implemented by CLUMP software (41).

ACKNOWLEDGEMENTS

The authors would like to thank the patients and their families, without whose generous support none of this research would have been possible. This work was supported by the Reta Lila Weston Trust, the PSP (Europe) Association (http://www.pspeur.org), the Parkinson's Disease Society, UK, the Brain research trust (P.A.S.) and by the NIA/NIH. A.P. would like to thank the UCL Bogue Fellowship. J.H. and A.J.M. would like to thank the Verum Foundation. A.J.M. is a resident research associate of the National Academy of Sciences.

REFERENCES

1. Lee, V.M., Goedert, M. and Trojanowski, J.Q. (2001) Neurodegenerative tauopathies. *Annu. Rev. Neurosci.*, **24**, 1121–1159.
2. Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A. *et al.* (1998) Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*, **393**, 702–705.
3. Spillantini, M.G., Murrell, J.R., Goedert, M., Farlow, M.R., Klug, A. and Ghetti, B. (1998) Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc. Natl Acad. Sci. USA*, **95**, 7737–7741.

4. Steele, J.C., Richardson, J.C. and Olszewski, J. (1964) Progressive supranuclear palsy. A heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. *Arch. Neurol.*, **10**, 333–359.
5. de Yebenes, J.G., Sarasa, J.L., Daniel, S.E. and Lees, A.J. (1995) Familial progressive supranuclear palsy. Description of a pedigree and review of the literature. *Brain*, **118**, 1095–1103.
6. Rojo, A., Pernaute, R.S., Fontan, A., Ruiz, P.G., Honnorat, J., Lynch, T., Chin, S., Gonzalo, I., Rabano, A., Martinez, A. *et al.* (1999) Clinical genetics of familial progressive supranuclear palsy. *Brain*, **122**, 1233–1245.
7. Conrad, C., Andreadis, A., Trojanowski, J.Q., Dickson, D.W., Kang, D., Chen, X., Wiederholt, W., Hansen, L., Masliah, E., Thal, L.J. *et al.* (1997) Genetic evidence for the involvement of tau in progressive supranuclear palsy. *Ann. Neurol.*, **41**, 277–281.
8. Houlden, H., Baker, M., Morris, H.R., MacDonald, N., Pickering-Brown, S., Adamson, J., Lees, A.J., Rossor, M.N., Quinn, N.P., Kertesz, A. *et al.* (2001) Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. *Neurology*, **56**, 1702–1706.
9. Poorkaj, P., Muma, N.A., Zhukareva, V., Cochran, E.J., Shannon, K.M., Hurtig, H., Koller, W.C., Bird, T.D., Trojanowski, J.Q., Lee, V.M. *et al.* (2002) An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype. *Ann. Neurol.*, **52**, 511–516.
10. Wszolek, Z.K., Tsuboi, Y., Uitti, R.J., Reed, L., Hutton, M.L. and Dickson, D.W. (2001) Progressive supranuclear palsy as a disease phenotype caused by the S305S tau gene mutation. *Brain*, **124**, 1666–1670.
11. Morris, H.R., Osaki, Y., Holton, J., Lees, A.J., Wood, N.W., Revesz, T. and Quinn, N. (2003) Tau exon 10 + 16 mutation FTDP-17 presenting clinically as sporadic young onset PSP. *Neurology*, **61**, 102–104.
12. Morris, H.R., Katzenschlager, R., Janssen, J.C., Brown, J.M., Ozansoy, M., Quinn, N., Revesz, T., Rossor, M.N., Daniel, S.E., Wood, N.W. *et al.* (2002) Sequence analysis of tau in familial and sporadic progressive supranuclear palsy. *J. Neurol. Neurosurg. Psych.*, **72**, 388–390.
13. Baker, M., Litvan, I., Houlden, H., Adamson, J., Dickson, D., Perez-Tur, J., Hardy, J., Lynch, T., Bigio, E. and Hutton, M. (1999) Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum. Mol. Genet.*, **8**, 711–715.
14. Ezquerra, M., Pastor, P., Valldeoriola, F., Molinuevo, J.L., Blesa, R., Tolosa, E. and Oliva, R. (1999) Identification of a novel polymorphism in the promoter region of the tau gene highly associated to progressive supranuclear palsy in humans. *Neurosci. Lett.*, **275**, 183–186.
15. de Silva, R., Weiler, M., Morris, H.R., Martin, E.R., Wood, N.W. and Lees, A.J. (2001) Strong association of a novel tau promoter haplotype in progressive supranuclear palsy. *Neurosci. Lett.*, **311**, 145–148.
16. Conrad, C., Vianna, C., Freeman, M. and Davies, P. (2002) A polymorphic gene nested within an intron of the tau gene: implications for Alzheimer's disease. *Proc. Natl Acad. Sci. USA*, **99**, 7751–7756.
17. de Silva, R., Hope, A., Pittman, A., Weale, M.E., Morris, H.R., Wood, N.W. and Lees, A.J. (2003) Strong association of the Saitohin gene Q7 variant with progressive supranuclear palsy. *Neurology*, **61**, 407–409.
18. Pastor, P., Ezquerra, M., Tolosa, E., Munoz, E., Marti, M.J., Valldeoriola, F., Molinuevo, J.L., Calopa, M. and Oliva, R. (2002) Further extension of the H1 haplotype associated with progressive supranuclear palsy. *Mov. Disord.*, **17**, 550–556.
19. Bird, T.D., Wijsman, E.M., Nochlin, D., Leehey, M., Sumi, S.M., Payami, H., Poorkaj, P., Nemens, E., Rafkind, M. and Schellenberg, G.D. (1997) Chromosome 17 and hereditary dementia: linkage studies in three non-Alzheimer families and kindreds with late-onset FAD. *Neurology*, **48**, 949–954.
20. Froelich, S., Basun, H., Forsell, C., Lilius, L., Axelman, K., Andreadis, A. and Lannfelt, L. (1997) Mapping of a disease locus for familial rapidly progressive frontotemporal dementia to chromosome 17q12–21. *Am. J. Med. Genet.*, **74**, 380–385.
21. Lendon, C.L., Lynch, T., Norton, J., McKeel, D.W., Jr, Busfield, F., Craddock, N., Chakraverty, S., Gopalakrishnan, G., Shears, S.D., Grimmett, W. *et al.* (1998) Hereditary dysphasic disinhibition dementia: a frontotemporal dementia linked to 17q21–22. *Neurology*, **50**, 1546–1555.
22. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. and Lander, E.S. (2001) High-resolution haplotype structure in the human genome. *Nat. Genet.*, **29**, 229–232.
23. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
24. 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.
25. Kwok, J.B.J., Teber, E.T., Loy, C., Hallupp, M., Nicholson, G., Mellick, G.D., Buchanan, D.D., Silburn, P.A. and Schofield, P.A. (2003) Tau promoter haplotypes regulate transcription and are associated with Parkinson's disease. *Ann. Neurol.*, **55**, 329–334.
26. Ponting, C.P., Hutton, M., Nyborg, A., Baker, M., Jansen, K. and Golde, T.E. (2002) Identification of a novel family of presenilin homologues. *Hum. Mol. Genet.*, **11**, 1037–1044.
27. Südhof, T.C. (1995) The synaptic vesicle cycle: a cascade of protein–protein interactions. *Nature*, **375**, 645–653.
28. Mirnics, K., Middleton, F.A., Marquez, A., Lewis, D.A. and Levitt, P. (2000) Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron*, **28**, 53–67.
29. Yu, F., Guan, Z., Zhuo, M., Sun, L., Zou, W., Zheng, Z. and Liu, X. (2002) Further identification of *NSF** as an epilepsy related gene. *Brain. Res. Mol. Brain. Res.*, **99**, 141–144.
30. Katoh, M. (2002) Regulation of WNT signaling molecules by retinoic acid during neuronal differentiation in NT2 cells: threshold model of WNT action. *Int. J. Mol. Med.*, **10**, 683–687.
31. Braun, M.M., Etheridge, A., Bernard, A., Robertson, C.P. and Roelink, H. (2003) Wnt signaling is required at distinct stages of development for the induction of the posterior forebrain. *Development*, **130**, 5579–5587.
32. Nagase, T., Ishikawa, K., Kikuno, R., Hirosawa, M., Nomura, N. and Ohara, O. (1999) Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.*, **6**, 337–345.
33. Xu, C.F., Brown, M.A., Nicolai, H., Chambers, J.A., Griffiths, B.L. and Solomon, E. (1997) Isolation and characterisation of the *NBR2* gene which lies head to head with the human *BRCA1* gene. *Hum. Mol. Genet.*, **6**, 1057–1062.
34. Liu, X. and Barker, D.F. (1999) Evidence for effective suppression of recombination in the chromosome 17q21 segment spanning RNU2-BRCA1. *Am. J. Hum. Genet.*, **64**, 1427–1439.
35. Mallet, J., Houhou, L., Pajak, F., Oda, Y., Cervini, R., Bejanin, S. and Berrard, S. (1998) The cholinergic locus: ChAT and VAChT genes. *J. Physiol. Paris.*, **92**, 145–147.
36. Kauppi, L., Sajantila, A. and Jeffreys, A.J. (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum. Mol. Genet.*, **12**, 33–40.
37. Reich, D.E., Schaffner, S.F., Daly, M.J., McVean, G., Mullikin, J.C., Higgins, J.M., Richter, D.J., Lander, E.S. and Altshuler, D. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat. Genet.*, **32**, 135–142.
38. Morris, H.R., Janssen, J.C., Bandmann, O., Daniel, S.E., Rossor, M.N., Lees, A.J. and Wood, N.W. (1999) The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases. *J. Neurol. Neurosurg. Psych.*, **66**, 665–667.
39. Weir, B.S. (1979) Inferences about linkage disequilibrium. *Biometrics*, **35**, 235–254.
40. Lewontin, R.C. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49–67.
41. Sham, P.C. and Curtis, D. (1995) Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann. Hum. Genet.*, **59**, 97–105.