ORIGINAL ARTICLE

Linkage disequilibrium fine mapping and haplotype association analysis of the *tau* gene in progressive supranuclear palsy and corticobasal degeneration

A M Pittman, A J Myers, P Abou-Sleiman, H C Fung, M Kaleem, L Marlowe, J Duckworth, D Leung, D Williams, L Kilford, N Thomas, C M Morris, D Dickson, N W Wood, J Hardy, A J Lees, R de Silva

J Med Genet 2005;**42**:837–846. doi: 10.1136/jmg.2005.031377

Background: The haplotype H1 of the *tau* gene, *MAPT*, is highly associated with progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD).
Objective: To investigate the pathogenic basis of this association.
Methods: Detailed linkage disequilibrium and common haplotype structure of *MAPT* were examined in 27 CEPH trios using validated HapMap genotype data for 24 single nucleotide polymorphisms (SNPs) spanning *MAPT*.
Results: Multiple variants of the H1 haplotype were resolved, reflecting a far greater diversity of *MAPT* than can be explained by the H1 and H2 clades alone. Based on this, six haplotype tagging SNPs (htSNPs) that capture 95% of the common haplotype diversity were used to genotype well characterised PSP and CBD case–control cohorts. In addition to strong association with PSP and CBD of individual SNPs, two common haplotypes derived from these htSNPs were identified that are highly associated with PSP: the sole H2 derived haplotype was underrepresented and one of the common H1 derived haplotypes was highly associated, with a similar trend observed in CBD. There were powerful and highly significant associations with PSP and CBD of haplotypes formed by three H1 specific SNPs. This made it possible to define a candidate region of at least ~56 kb, spanning sequences from upstream of *MAPT* exon 1 to intron 9. On the H1 haplotype background, these could harbour the pathogenic variants.
Conclusions: The findings support the pathological evidence that underlying variations in *MAPT* could contribute to disease pathogenesis by subtle effects on gene expression and/or splicing. They also form the basis for the investigation of the possible genetic role of *MAPT* in Parkinson's disease and other tauopathies, including Alzheimer's disease.

See end of article for authors' affiliations

Correspondence to:
Professor Andrew Lees,
Reta Lila Weston Institute
of Neurological Studies,
University College London,
London W1T 4JF, UK;
alees@ion.ucl.ac.uk

Received 26 January 2005
Revised version received
15 March 2005
Accepted for publication
17 March 2005

The tauopathies are a group of neurodegenerative disorders that are characterised pathologically by fibrillar aggregates of the microtubule associated protein, tau. These disorders include Alzheimer's disease, progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), Pick's disease, and frontotemporal dementia with parkinsonism with tau pathology linked to chromosome 17 (FTDP-17T), with a clinical spectrum ranging from dementia to parkinsonian phenotypes.[1] The identification of missense and splice site mutations in the *tau* gene, *MAPT* (MIM 157140), causing FTDP-17T (MIM 600274) affirmed a central role for tau dysfunction in some neurodegenerative diseases.[2 3] Although the other related tauopathies—including Alzheimer's disease, 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 usually a sporadic disorder of late adult life. It is the second most common form of degenerative parkinsonism and is characterised clinically by an akinetic-rigid syndrome, supranuclear gaze palsy, pseudobulbar signs, and cognitive decline of frontal lobe type.[5–7] CBD is an atypical parkinsonian condition occurring much less commonly than PSP and it classically presents with unilateral cortical sensory loss, alien hand, jerky dystonia, rigidity, bradykinesia, and dementia. PSP is sporadic, with no familial history or *MAPT* mutations in the large majority of cases. However, robust genetic association of PSP with *MAPT* and reports of the rare families with more than one affected member[8 9] indicated that genetic factors could play a role. Conrad and colleagues were the first of many groups to show that variation at the *MAPT* locus could be an important genetic influence in sporadic PSP by demonstrating allelic association with PSP of a dinucleotide polymorphism in *MAPT* intron 9.[10] The overrepresentation of the commoner allele (a_0) in PSP and also later in CBD was then confirmed by other groups.[11 12] This suggests either that this polymorphism itself could contribute to increased risk or that it is in linkage disequilibrium (LD) with the actual causative variant. Although some *MAPT* mutations in FTDP-17T cause a clinical picture closely resembling PSP,[13–15] no pathogenic variations

Abbreviations: CBD, corticobasal degeneration; CEPH, Centre d'Etude du Polymorphisme Humain; EM, expectation maximisation; FTD, frontotemporal dementia; FTDP-17T, frontotemporal dementia with parkinsonism with tau pathology linked to chromosome 17; htSNP, haplotype tagging single nucleotide polymorphism; LD, linkage disequilibrium; LRT, likelihood ratio test; MAPT, microtubule associated protein, tau; MIM, mendelian inheritance in man; PSP, progressive supranuclear palsy; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism of *MAPT* have yet been identified in clinically and pathologically diagnosed sporadic and familial PSP.[16]

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 kilobases (kb).[17] In approximately 200 unrelated white subjects, these polymorphisms were in complete LD, forming two extended haplotypes, H1 and H2.[17] The study 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. It also showed that the more common haplotype, H1, with which the a_0 allele segregated, was significantly overrepresented in PSP.[17] Follow up studies[18 19] extended the *MAPT* haplotype a further 68 kb to the promoter region of *MAPT* where three SNPs, highly associated with PSP, were in complete LD with the rest of the *MAPT* haplotype.[19] We have further extended the *MAPT* haplotype to cover a maximal region of ~2 million bases (Mb) which is in near complete LD,[20] and using high density HapMap genotype data for LD analysis we subsequently revised the size of the region to 1.8 Mb (unpublished work). This region associated with PSP includes several other genes in addition to *MAPT*, including *Saitohin*[21 22] (situated within intron 9 of *MAPT*), *NSF* (N-ethylmaleimide sensitive factor), *IMP5* (a presenilin homologue),[23] *CRHR1* (corticotrophin releasing hormone receptor), and *LOC284058*, an unknown gene just adjacent to *MAPT*.

Identifying the functional basis of the H1 haplotype association will be important in providing an insight into the aetiopathogenesis of PSP and CBD. Although all the genes within this multigene haplotype block are associated with PSP and CBD, the hallmark tau pathology of these disorders strongly implicates *MAPT* itself. The aim of our study was therefore to analyse exhaustively the *MAPT* haplotype association with PSP and CBD in order to identify non-coding variants that could affect *tau* gene expression, splicing, or processing, leading to tau pathology and selective neuronal loss. More controversially, recent work shown weak association of the H1 haplotype with sporadic Parkinson's disease[24] and association with Norwegian Parkinson's disease cases of a haplotype within the extended H1 clade, spanning the 5′ half of *MAPT*.[25] This is surprising as Parkinson's disease is traditionally not associated with tau dysfunction or pathology.

In this work, we employed a systematic framework of genetic analyses to investigate the common haplotype structure of *MAPT* in order to refine the association of the *MAPT* haplotype with PSP and CBD. By using the validated high density genotype data available from the International HapMap Project (www.hapmap.org) we analysed the *MAPT* gene in 27 defined CEPH (Centre d'Etude du Polymorphisme Humain) trios (father, mother, and offspring). We analysed LD and haplotype structure with 24 SNPs in relation to the H1 and H2 haplotypes, as defined by the *MAPT* biallelic intron 9 deletion-insertion (*del-In9*),[17] using the software suite TagIT (www.popgen.biol.ucl.ac.uk/software.html), which contains routines specifically tailored for the inference of haplotypes from the CEPH trio data.[26] With this analysis, we identified far greater haplotypic variation of *MAPT* than can be explained by the description of the extended H1 and H2 haplotypes alone. Based on the data for this common haplotypic diversity of *MAPT* in the CEPH trios, we identified a set of six haplotype tagging SNPs (htSNPs): five SNPs that represent intra-H1 variation and *del-In9*.[17] The htSNPs function as a minimal set of highly informative single nucleotide polymorphism (SNP) markers that capture 95% of the common haplotype diversity of *MAPT*.[26] We genotyped the *MAPT* htSNPs in our target populations, namely well characterised PSP case–control cohorts of both British and north American (US) origins and CBD cases of US origin.

Table 1 The 24 single nucleotide polymorphisms and *del-In9* used for the linkage disequilibrium and haplotype structure analysis of *MAPT* in the CEPH trios

SNP	Position*	dbSNP ID	Alleles	Ancestral	*F*1†	*F*2†	p Value‡
1	41291420	rs962885	C/T	T	0.639	0.361	0.572
2	41301910	rs1078830	C/T	C	0.189	0.811	0.426
3	41307507	rs2055794	A/G	A	0.185	0.815	0.442
4	41324209	rs7210728	A/G	A	0.259	0.741	0.248
5	41333623	rs1864325	C/T	C	0.811	0.189	0.426
6	41334330	rs1560310	A/G	G	0.185	0.815	0.442
7	41336326	rs3885796	G/T	C	0.189	0.811	0.426
8	41342006	rs1467967	A/G	A	0.648	0.352	0.851
9	41349204	rs3785880	G/T	T	0.462	0.538	0.709
10	41354402	rs1467970	G/T	T	0.185	0.815	0.442
11	41354620	rs767058	A/G	C	0.815	0.185	0.442
12	41361649	rs1001945	C/G	G	0.546	0.454	0.301
13	41374593	rs2435205	A/G	A	0.593	0.407	0.251
14	41375548	rs242557	A/G	G	0.396	0.604	0.854
15	41382599	rs242562	A/G	G	0.375	0.625	0.684
16	41409284	rs2217394	A/G	G	0.815	0.185	0.442
17	41410268	rs3785883	A/G	G	0.204	0.796	0.524
18	41411483	rs754512	A/T	T	0.185	0.815	0.442
19	41419081	rs2435211	C/T	C	0.632	0.368	0.061
20	41429726	rs1052553	A/G	G	0.815	0.185	0.442
21	41431900	rs2471738	C/T	C	0.713	0.287	0.335
22	41442488	*del-In9*	+/−	+	0.823	0.177	0.617
23	41445400	rs733966	C/T	C	0.815	0.185	0.442
24	41457408	rs9468	C/T	C	0.185	0.815	0.442
25	41461242	rs7521	A/G	G	0.434	0.566	0.569

The analysis was carried out on the available genotype data for these single nucleotide polymorphisms (SNP) from HapMap (http://www.hapmap.org/). In addition, we genotyped the *del-In9* in the same CEPH trios. Allele and genotype frequencies and p values for test to fit Hardy–Weinberg equilibrium were calculated in the program TagIt. The ancestral allele (Chimpanzee) is also indicated. Position on chromosome (in bp) is based on May 2004 build of Human Genome Sequence (http://genome.ucsc.edu).
*SNP position on chromosome.
†Allelic frequencies in the CEPH trios.
‡p Values for test to fit Hardy–Weinberg equilibrium.

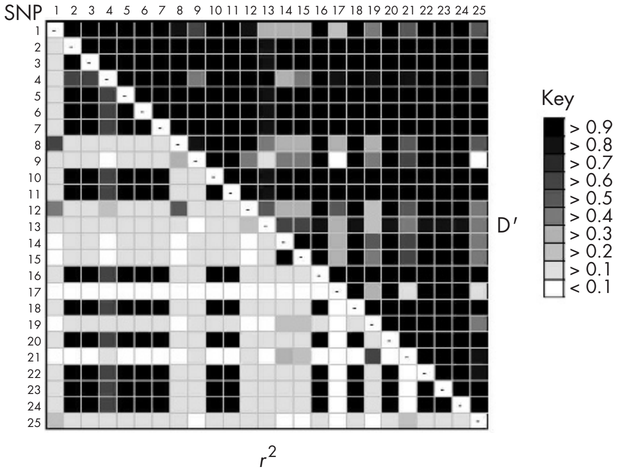

Figure 1 Linkage disequilibrium (LD) across the *MAPT* gene. Numerical LD is presented by grey scale, pairwise between each single nucleotide polymorphism (SNP) by both D′ (upper right) and the more stringent measure r^2 (bottom left). The darker the shading indicates a greater extent of LD between the SNPs.

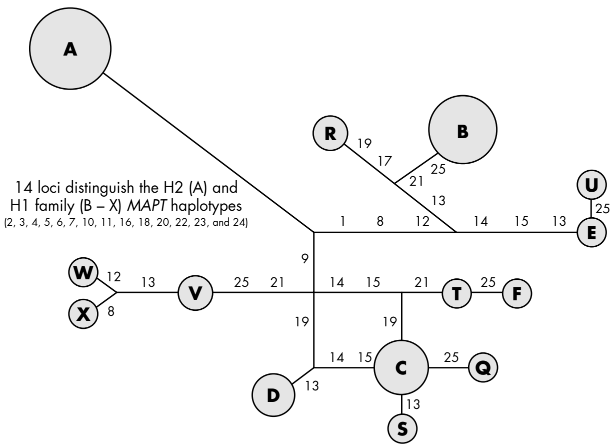
A
B
C
D
E
F
Q
R
S
T
U
V
W
X
14 loci distinguish the H2 (A) and H1 family (B – X) MAPT haplotypes
(2, 3, 4, 5, 6, 7, 10, 11, 16, 18, 20, 22, 23, and 24)
1 8 12 14 15 13
19 17 21 25 13
25
9
12 13 25 21 14 15 21 25
8
19 19
14 15 25
13
13

Figure 2 Reduced median network of *MAPT* haplotypes (frequencies that exceed 1% by expectation maximisation, from table 2) in the CEPH trios. Node size is proportional to haplotype frequency. Any two given haplotypes differ by the single nucleotide polymorphism(s) (as numbered from table 1) along the lines that connect them.

METHODS

Analysis of the linkage disequilibrium and haplotype structure

SNP data for the region of the *MAPT* locus in 27 CEPH trios (Corriell Institute for Medical Research; http://locus.umdnj.edu/nigms/) from the International HapMap project (HapMap) web site (http://www.hapmap.org/) were downloaded for genetic analysis of the *MAPT*. The raw SNP genotype data were analysed in TagIT, a software package for identifying and evaluating tagging SNPs applied to haplotype data, which also contains routines for inferring haplotypes from trio material and LD analysis (http://popgen.biol.ucl.ac.uk/software).[26]

We initially removed from the HapMap data any SNPs that had a minor allele frequency of less than 5%. We also checked for any inconsistencies in the data through the parent–offspring relationship in the CEPH trios. We used a resulting set of 24 SNPs and the *del-In9* (table 1) which covers the entire *MAPT* gene from upstream of the promoter to beyond exon 13, to infer haplotypes and their respective frequencies by an expectation–maximisation (EM) algorithm ($\epsilon = 1\times10^{-6}$) specifically for CEPH trio material (EM trio).[26] For convenience, we designated the biallelic (+/−) intron 9 deletion-insertion polymorphism (*del-In9*) as an SNP. In all, 34 haplotypes were resolved from parental chromosomes. The pairwise LD across *MAPT* for each SNP was then evaluated by both the measures of D′ and the square of the correlation coefficient (r^2). Both measures were calculated, first by estimating pairwise haplotype frequencies through EM trio, then by assessing the statistical strength of association through a likelihood ratio test (LRT), by comparing the EM frequencies with haplotype frequencies estimated assuming no LD. Both measures of LD are based upon D, the basic pairwise disequilibrium coefficient, the difference between the probabilities of observing the alleles independently in the population: $D = f(A_1B_1) - f(A_1)f(B_1)$.[27] 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. The measure of r^2 has a more strict interpretation than that of D′; $r^2 = 1.0$ only when the marker loci also have identical allele frequencies. The allele at the one locus can always be predicted by the allele at the second locus. Recent work suggests that r^2 is the preferred measure of LD for association based studies.[26]

Allelic and genotype frequencies followed by statistical assessment of Hardy–Weinberg equilibrium were made at each locus in the CEPH trios as implemented by TagIT.

From the LD and haplotype structure of *MAPT*, htSNPs were selected to capture the diversity of known *MAPT* HapMap SNPs in the CEPH trios. We selected six tagging SNPs (*del-In9*, SNPs 8, 14, 17, 21, and 25); using TagIT, we then assessed their performance on the CEPH trios. Our tagging approach focused on the coefficient of determination (that is, haplotype r^2) in a linear regression, which uses the haplotypes defined by the htSNPs to predict the state of the tagged SNPs.[26] The basis of this design is that even when individual haplotypes defined by the htSNPs do not correlate perfectly with tagged SNPs, haplotype combinations might do so, and these combinations are identified by selection of the appropriate coefficients in the linear regression. Haplotype r^2 is the coefficient of determination from an analysis of variance of locus i (coding alleles at locus i as "0" or "1") among the G groups (number of haplotypes, or groups, defined in the dataset in question by the htSNP set): $r^2_{[\mathrm{hap}]i} = 1 - R'_i/D_i$, where $R'_i = 2\Sigma p'_{ig}(1-p'_{ig})/x_g$, which can be interpreted as the sum of the within group variances weighted by their frequency.

The PSP cases and control subjects

The unrelated PSP cases (n = 83), from the Queen Square brain bank for neurological disorders, were all white and of western European origin and were all pathologically confirmed. Most of these cases have been used in previous studies.[16 19 20 22 28] Pathological confirmation of the diagnosis of PSP was made following standardised criteria.[28] The unrelated British control population (n = 169), all white, were taken from brain bank tissue with no clinical evidence of neurodegenerative disease and no abnormal histopathology, from the MRC Building, Newcastle, UK. The samples were age matched, where the average age at death was 73.5 years for the PSP cases (63% male) and 76 years for the controls (51% male). All patients and controls were collected under approved protocols followed by informed consent, and this work was approved by the joint research ethics committee of the Institute of Neurology and the National Hospital for Neurology and Neurosurgery.

Table 2 The haplotype structure of the *MAPT* gene in CEPH-trios based upon the 25 markers in Table 1

																											Frequency (%)	
ID*	Haplotype†																										EM‡	R§
A	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	1	1	0	1	:	18.1	17.6
B	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0	:	17.2	23.5
C	0	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1	1	:	14.3	23.5
D	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	:	3.8	…
E	0	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0	1	0	0	1	1	:	1.9	2.9
F	0	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	:	1.9	2.9
G	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	:	…	2.9
H	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	:	…	2.9
I	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	:	…	2.9
J	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	1	:	…	2.9
K	0	1	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	1	1	0	1	0	0	1	1	:	…	2.9
L	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	:	…	2.9
M	1	1	1	1	0	1	1	0	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	1	0	:	…	2.9
N	1	1	1	1	0	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	:	…	2.9
O	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1	:	…	2.9
P	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0	:	…	2.9
Q	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0	1	1	0	1	0	0	1	1	:	1.9	…
R	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	:	1.9	…
S	0	1	1	1	0	1	1	0	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	1	0	:	1.9	…
T	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0	:	1.9	…
U	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	:	1.9	…
V	0	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	0	:	1.9	…
W	0	1	1	1	0	1	1	0	0	1	0	0	1	0	0	0	1	1	1	0	1	0	0	1	1	:	1.9	…
X	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	:	1.9	…
	1	0	0	0	0	1	–	0	1	1	–	1	0	1	1	1	1	1	0	1	0	0	0	0	1		Ancestral	

Alleles represented in binary (1 = highest letter in alphabet of SNP allele). Haplotypes shown if observed in resolved chromosomes (parental chromosomes, n = 34) or if the expectation-maximisation (EM trio) inferred haplotype frequency exceeded 1%. Also presented is the build of the ancestral haplotype (Chimpanzee).
*Haplotype identity.
†Binary representation.
‡Inferred frequency by expectation-maximisation (all data).
§Resolved haplotype frequency.

The unrelated US control population consisted of individuals (n = 131; 50% male) free of abnormal histopathology and with an average age at death of 79.9 years. The unrelated PSP cases (n = 238; 50% male) were pathologically confirmed by standard criteria and had an average age at death of 75.3 years. The unrelated CBD cases (n = 44; 50% males) were pathologically confirmed following standard criteria and had an average age at death of 71.3 years.

Genotyping

The htSNPs (dbSNP numbers: rs1467967, rs242557, rs3785883, rs2471738, and rs7521, and the *del-In9*; table 1) were genotyped in the PSP case–control cohorts as follows. The 238 bp *MAPT del-In9* was genotyped as previously described.[17] Polymerase chain reaction (PCR) primer pairs (available on request) were designed by the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and used to amplify each SNP of interest. PCR reactions were as follows: 10 µl reactions, which contained one unit of DNA polymerase (Qiagen, Crawley, West Sussex), 10×PCR reaction buffer, 5×Q solution (Qiagen), 10 pmol of each oligonucleotide primer pair, and 25 ng of sample template genomic DNA.

Genotyping of the SNPs rs1467967, rs242557, rs3785883, rs2471738, and rs7521 was conducted by Pyrosequencing (Biotage AB, Uppsala, Sweden) (details available on request) or by restriction fragment length polymorphism (RFLP) digest. The following restriction endonucleases cut the PCR product once at the (N) allele: *Dra* I (A), *Apa*L I (A), *Bsa*H I (G), *Bst*E II (T), and *Pst* I (A) (New England Biolabs, Hitchin, Herts, UK). PCR products were incubated overnight with 2 units of the corresponding restriction enzyme at the recommended temperature. Digests were separated on 4% agarose gels and visualised with ethidium bromide staining.

We assessed genotyping accuracy by retyping 20% of all genotypes, whole sets of htSNPs, genotyping by alternative methods and by direct automated DNA sequencing of random samples.

The ancestral allele at each locus was determined by direct sequence comparison of the 24 SNP loci in human and chimpanzee *MAPT* and in addition by searching for the ancestral allele in NCBI (http://www.ncbi.nlm.nih.gov/).

Statistical analysis

For each htSNP, the allele and genotype distribution in the PSP cases were compared with those in the control group. Statistical assessments for the allele and genotype frequencies and Hardy–Weinberg were made using TagIT. Case–control single locus htSNP allelic and genotypic association was calculated statistically in CLUMP software.[29] The p values were derived by standard Pearson's χ^2 tests except in cases where cell counts in the contingency tables were less than 5. When cell counts were less than 5, p values were determined empirically by 100 000 simulations; the program uses a Monte-Carlo approach that performs repeated simulations to generate random tables having the same marginal totals as the one under consideration and counting the number of times that a χ^2 value associated with the actual table is achieved by the randomly generated tables. We tested for heterogeneity between the H1H1 homozygote populations versus the whole population using a standard Pearson χ^2 test.

Distributions of haplotypes defined by the htSNPs were compared in the PSP cases and controls using WHAP software (http://www.broad.mit.edu/personal/shaun/whap/). This is an SNP haplotype analysis suite that performs a regression based haplotype association test through an LRT, which is a χ^2 test with n−1 degrees of freedom to derive the associated p value, where n is the number of haplotypes observed for the data. We used this test to give an initial assessment of haplotype association (an omnibus test) and then carried out individual haplotype tests (haplotype specific tests) of association, again through an LRT (df = 1)

Table 3 Allele frequencies (*F*1) and *p*-values of single-locus association in the three studies

		Frequency (*F* 1%)		Association (p)		Odds ratio (MA)	
	dbSNP ID	Cases	Controls	Allelic	Genotypic	OR	95% CI
US PSP							
8	rs1467967	62.8	62.6	0.963	1.000	0.965	0.703 to 1.325
14	rs242557	54.4	31.0	**2.91ex^{-9}**	**2.29ex^{-8}**	**2.356**	**1.706 to 3.255**
17	rs3785883	17.0	22.4	0.072	*0.168	0.713	0.487 to 1.044
21	rs2471738	67.0	81.5	**1.87ex^{-5}**	***1.15ex^{-4}**	**2.224**	**1.535 to 3.222**
del-In9	....	91.6	77.1	**4.02ex^{-8}**	***1.00ex^{-5}**	**0.298**	**0.193 to 0.462**
25	rs7521	43.2	44.5	0.456	0.671	1.124	0.827 to 1.526
UK PSP							
8	rs1467967	67.9	64.6	0.993	0.770	0.998	0.639 to 1.560
14	rs242557	47.9	35.7	**0.012**	**0.016**	**1.815**	**1.209 to 2.726**
17	rs3785883	25.5	20.6	0.365	0.680	1.227	0.762 to 1.974
21	rs2471738	66.0	80.1	**0.001**	**0.005**	**2.142**	**1.368 to 3.355**
del-In9	....	93.2	76.6	**1.14ex^{-5}**	**5.31ex^{-5}**	**0.215**	**0.099 to 0.466**
25	rs7521	51.2	45.7	0.546	0.814	0.773	0.505 to 1.183
US CBD							
8	rs1467967	61.9	62.6	0.909	*0.870	1.030	0.619 to 1.713
14	rs242557	50.0	31.0	**0.002**	**0.010**	**2.231**	**1.322 to 3.764**
17	rs3785883	33.3	22.4	**0.019**	**0.022**	1.047	0.586 to 1.872
21	rs2471738	67.0	81.5	**0.005**	**0.011**	**2.165**	**1.254 to 3.736**
del-In9	....	86.4	77.1	0.063	†	0.532	0.271 to 1.043
25	rs7521	43.2	44.5	0.826	0.464	0.807	0.494 to 1.320

Significant single locus association of htSNPs are indicated in bold. Odds ratios and their 95% confidence interval are presented for the minor allele versus the major allele for all htSNPs. The p values were derived by standard Pearson's χ^2 tests except in cases where cell counts in the contingency tables were less than 5. When cell counts were less than 5 (*), p values were determined empirically by 100 000 simulations (CLUMP software).

†A genotypic test was not carried out for the *del-In9* in intron 9 in the CBD series, as there were no rare homozygotes in the CBD cases, thus preventing us from performing a valid test.

CI, confidence interval; MA, minor allele; OR, odds ratio.

and by also obtaining empirical p values by Monte-Carlo methods (20 000 simulations used). To test the effect of the H1 specific htSNPs while controlling for the extended H1/H2 haplotype we imposed a set of equality constraints under the null across the haplotypes identical at the *del-In9* and undertook single locus and haplotype analysis as outlined above. We corrected the p values in tables 4 and 5 according to the number of tests performed where appropriate by the Bonferroni correction, the significance of which is discussed throughout the text.

RESULTS

Linkage disequilibrium and haplotype structure of *MAPT*

For the haplotype analysis of the *MAPT* gene, we downloaded genotype data for 27 CEPH trios (mother, father, and offspring) of European descent (CEPH Utah collection) for SNPs spanning the *MAPT* region, from the International HapMap Project web site (www.hapmap.org). The raw SNP data from HapMap were analysed using the software package TagIT (http://popgen.biol.ucl.ac.uk/software). We discarded SNPs that had a minor allele frequency of less than 5%. No inconsistencies in Mendelian inheritance in the parent–offspring relationship were found. We genotyped the *del-In9* marker that defines the extended H1 and H2 clades.[17] The average density of the markers is one SNP every 6.7 kb. None of the polymorphisms deviated from Hardy–Weinberg equilibrium. See table 1 for details of all SNPs analysed in the CEPH trios.

We evaluated pairwise LD across *MAPT* for all 24 selected SNPs and *del-In9* in the 27 CEPH trios both by D prime (D′) and the square of the correlation coefficient (r^2), calculated from the expectation-maximisation trio (EM trio) inferred haplotypes. By pairwise LD analysis of the 25 SNPs in CEPH trios, we identified a greater diversity than reflected by the description of the two extended H1 and H2 haplotypes alone (fig 1). The entire *MAPT* gene is featured by significant LD as is particularly evident by the measure of D′ (fig 1). However, when LD was assessed by the more stringent measure of r^2 (which accounts for differences in allele frequencies), it appeared more fragmented, with SNPs that were in high r^2 LD with each another, but in moderate to low r^2 LD with the extended H1 and H2 haplotype (defined by the *del-In9* and other SNP loci), suggesting that they are correlated with either the H1 or H2 haplotypes, but with differing frequency. This supports evidence of variability on the background of these extended haplotypes. In fact, our analyses in the CEPH trios show that these underlying blocks of LD were variable exclusively on the background of the extended H1 haplotype and therefore defined haplotypes within the H1 clade. LD correlation by D′ between many of the described H1 specific SNPs is relatively low (fig 1), suggesting a degree of linkage equilibrium between them; this indicates that, unlike the H1 and H2 haplotypes, there are no constraints to recombination between variants of the extended H1 haplotypes. This pattern of LD across the extended H1 haplotype is essentially similar in the Taiwanese population, in which the extended H2 haplotype is absent (unpublished data).

We obtained the EM inferred *MAPT* haplotypes and their respective frequencies by using the EM estimation algorithm specifically tailored to deal with trio data (EM trio) as structured in the CEPH trios.[26] We also obtained phased haplotypes (n = 34, representing 42% of the total number of haplotypes in the CEPH trios) by resolving parental chromosomes in the CEPH trios. EM predictions depict a total of 14 different *MAPT* haplotypes of frequency greater than 1% (table 2). Three of these haplotypes are common, having a frequency greater than 10%, with the remaining 21 haplotypes having frequencies of less than 5%. Only one of the common predicted haplotypes (haplotype A, frequency = 18.1%) is representative of H2 (table 2). The other two common variants (B and C; frequencies = 17.2% and 14.2%, respectively) are based upon the H1 haplotype and differ from one another at multiple SNP loci, as shown in fig 2. A further 11 rare variants of the H1 haplotype (frequency less than 1%) were predicted.

Table 4 Association of common *MAPT* haplotypes with progressive supranuclear palsy and corticobasal degeneration

htSNP haplotypes							UK PSP			US PSP			US CBD		
							Frequency (%)		Association (*LRT*)	Frequency (%)		Association (*LRT*)	Frequency (%)		Association (*LRT*)
ID	rs1467967	rs242557	rs3785883	rs2471738	del-In9	rs7521	Control	PSP	p (p corrected)	Control	PSP	p (p corrected)	Control	CBD	p (p corrected)
A	A	G	G	C	H2	G	20.7	6.3	**1.46ex-5 (2.77ex-4)**	22.0	6.3	**9.55ex-9 (2.01ex-7)**	22.0	8.2	**0.020** (0.367)
B	G	G	G	C	H1	A	16.5	13.9	0.378 (1.000)	12.2	15.8	0.562 (1.000)	12.2	15.4	0.914 (1.000)
C	A	A	G	T	H1	G	11.3	24.3	**0.001 (0.022)**	7.8	24.0	**6.42ex-9 (1.35ex-7)**	7.8	17.7	0.066 (1.000)
D	A	A	G	C	H1	A	8.9	3.7	0.110 (1.000)	4.0	7.9	0.077 (1.000)	4.0	7.5	0.489 (1.000)
E	A	G	G	C	H1	A	6.4	8.4	0.949 (1.000)	15.7	6.5	**0.014** (0.294)	15.7	4.6	0.148 (1.000)
F	G	G	A	C	H1	A	4.0	1.0	0.291 (1.000)	1.4	0.0	…	1.4	4.6	0.588 (1.000)
G	G	A	A	C	H1	A	3.9	5.1	0.691 (1.000)	2.6	3.5	0.937 (1.000)	2.6	3.4	0.834 (1.000)
H	A	G	A	C	H1	A	2.6	6.5	**0.010** (0.173)	0.0	3.8	0.404 (1.000)	0.0	0.0	…
I	G	A	G	C	H1	A	2.6	3.8	0.960 (1.000)	4.4	5.2	0.376 (1.000)	4.4	3.3	0.610 (1.000)
J	A	G	G	C	H1	G	2.4	0.0	**0.033** (0.621)	0.0	3.0	0.055 (1.000)	0.0	3.4	0.237 (1.000)
K	A	A	A	C	H1	G	2.2	0.9	0.378 (1.000)	0.0	0.0	…	0.0	0.0	…
L	A	G	A	C	H1	G	2.2	4.1	0.496 (1.000)	3.8	3.4	0.338 (1.000)	3.8	0.0	0.759 (1.000)
M	G	A	G	C	H1	G	2.0	2.6	0.744 (1.000)	3.5	3.4	0.930 (1.000)	3.5	5.0	0.319 (1.000)
N	G	G	A	C	H1	G	0.9	3.7	0.331 (1.000)	4.3	0.6	**0.005** (0.105)	4.3	0.0	**0.018** (0.322)
O	A	A	A	C	H1	A	0.0	3.6	0.070 (1.000)	3.4	1.3	0.350 (1.000)	3.4	5.0	0.386 (1.000)
P	G	G	G	T	H1	G	1.2	3.4	0.509 (1.000)	0.4	1.4	0.628 (1.000)	0.4	0.0	…
Q	A	A	G	T	H1	A	0.7	2.8	**0.040** (0.760)	0.0	1.6	**0.003** (0.073)	0.0	1.2	…
R	A	G	G	T	H1	G	0.7	2.7	0.114 (1.000)	2.4	1.6	0.386 (1.000)	2.4	1.5	0.493 (1.000)
S	G	G	G	C	H1	G	1.4	2.4	0.599 (1.000)	2.6	2.0	0.920 (1.000)	2.6	0.0	0.621 (1.000)
T	A	G	A	T	H1	G	0.3	0.0	…	1.1	0.0	…	1.1	7.0	0.713 (1.000)
U	A	A	G	C	H1	G	1.1	0.0	…	1.1	1.7	0.270 (1.000)	1.1	3.5	0.170 (1.000)
v	G	G	A	T	H1	G	1.3	0.0	…	1.9	1.0	0.207 (1.000)	1.9	2.8	0.699 (1.000)
w	G	G	G	C	H2	G	0.0	0.0	…	0.0	0.0	…	0.0	2.9	0.326 (1.000)
x	G	A	A	T	H1	G	0.0	0.0	…	2.7	0.5	0.205 (1.000)	2.7	0	0.174 (1.000)

The above analysis was based on the output of all haplotypes (>90%), but only those with a frequency >2% were tested for association through the likelihood ratio test (LRT). After adjustment of p values, in parentheses, for correction of multiple testing, only haplotypes **A** and **C** in both PSP studies remain significant. No haplotype is significantly associated with CBD after correction for multiple testing.
CBD, corticobasal degeneration; PSP, progressive supranuclear palsy.

It is noteworthy that in addition to the resolved H2 haplotype A, a single resolved haplotype (haplotype G; frequency 2.9% in resolved), based on variation of H2 haplotype A, was resolved which differed from haplotype A by SNP 13 (table 2). However, this haplotype was not predicted by EM trio for output as a significant frequency in the population and represented only ~5% (estimated by EM prediction) of all H2 haplotypes in the CEPH trios. It is thought that haplotype prediction through EM is a more accurate representation of the relative haplotype frequencies in a population than simply resolving "known" haplotypes because of a far greater utilisation of the data. We also constructed the ancestral (chimpanzee) haplotype based upon the alleles of the 24 SNPs and the *del-In9* (table 2). This appears not to resemble any haplotype present in the CEPH trios, though its closest relative (but different by 10 loci) would appear to be that of the extended H2 (CEPH trio haplotype A, from table 2). The other ancestral SNP loci are either consistent with the H1 haplotype family (SNPs 1, 5, 6, 10, 12, 18, and 23), including the presence of the 238 bp insertion sequence (*del-In9*, or SNP 22 in table 1), or the allele is not observed in *Homo sapiens* (SNPs 7 and 11).

Selection, performance assessment, and association analysis of *MAPT* haplotype tagging SNPs

We used an association based criterion (criterion 5 in TagIT, haplotype r^2) in order to select the haplotype tagging SNPs (htSNPs).[26] Six htSNPs (SNPs 8, 14, 17, 21, 22 (*del-In9*), and 25; table 1) are sufficient to represent all the HapMap SNPs in the 27 CEPH trios with a high coefficient of determination. Five of these htSNPs are H1 specific—that is, they vary only on the H1 background. In addition the bi-allelic *del-In9* marker is used to unambiguously distinguish the extended H1 and H2 haplotypes.[17] In CEPH trios[26] the performance value for the 6 htSNPs and *del-In9* in the CEPH trios was interpreted at an average haplotype r^2 value of 0.95 (95%) and a minimum r^2, interpreted as the minimum locus value of 0.68. Excluding the *del-In9* from the set of htSNPs results in a loss of performance of only of 3%, with performance down to 92% with the five remaining H1 specific htSNPs. This is because a particular allelic combination of these five H1 specific SNPs is representative of the extended H2 haplotype. The performance value of just the *del-In9* against the known SNPs in the CEPH trios is just 50%.

Table 5 Association of the subset of htSNP haplotypes with progressive supranuclear palsy and corticobasal degeneration

Haplotype				Frequency (%) and association (LRT) of haplotype								
				UK PSP			US PSP			US CBD		
ID	rs242557	rs3785883	rs2471738	Control	PSP	p (p corrected)	Control	PSP	p (p corrected)	Control	CBD	p (p corrected)
I	G	G	C	50.0	30.7	**3.14ex-4 (2.51ex-3)**	51.3	34.7	**1.65ex-5 (1.32ex-4)**	51.3	32.6	**0.002 (0.019)**
II	A	G	T	12.0	28.3	**2.16ex-4 (1.73ex-3)**	8.3	27.6	**2.31ex-9 (1.85ex-8)**	8.3	17.8	**0.009** (0.070)
III	A	G	C	13.2	10.2	0.349 (1.000)	13.9	17.7	0.091(0.730)	13.9	22.1	0.145 (1.000)
IV	G	A	C	10.0	16.6	0.316 (1.000)	10.2	7.1	**0.008** (0.064)	10.2	0.0	**0.034** (0.275)
V	A	A	C	6.9	9.0	0.454 (1.000)	6.1	6.8	0.728 (1.000)	6.1	12.4	0.619 (1.000)
VI	G	G	T	2.2	5.2	0.087 (0.700)	4.0	3.0	0.611 (1.000)	4.0	4.4	0.603 (1.000)
VII	A	A	T	3.2	0.0	0.907 (1.000)	2.9	1.6	0.751 (1.000)	2.9	0.0	0.321 (1.000)
VIII	G	A	T	2.4	0.0	**0.045** (0.356)	3.4	1.4	0.103 (1.000)	3.4	10.7	0.186 (1.000)

This haplotype analysis was based on a subset of H1 specific htSNP defined haplotypes that show evidence of association after consideration of the *del-In9*. After correction of p values for multiple testing (bracketed p values), haplotypes **I** and **II** in both PSP studies and haplotype **I** in the CBD study are significant.
CBD, corticobasal degeneration; LRT, likelihood ratio test; PSP, progressive supranuclear palsy.

We genotyped the *MAPT* htSNPs in two separate PSP case–control cohorts from the UK and USA and CBD cases from USA. Single locus association results are summarised in table 3. In none of the groups were there any significant deviations from Hardy–Weinberg equilibrium at any of the htSNPs. The strong association of the *del-In9* with PSP was again verified in both the UK and US cohorts ($p = 1.14\times10^{-5}$, 4.021×10^{-8}, respectively; table 3). The same trend was observed in CBD but the difference was not significant, possibly because of the small sample size. No evidence of association was found for htSNPs 8, 17, and 25 in the studies, except in the US CBD study where htSNP 17 is moderately associated ($p = 0.019$, allelic) (table 3). We calculated the odds ratios (OR) and their 95% confidence intervals and present values for all six htSNPs (table 3) by comparison of each minor allele verses each major allele. The H2 haplotype as defined by *del-In9* is a significant protective factor. The H1 specific SNPs rs242557 and rs2471738 are highly associated with these diseases and are arguably as important for risk as the association of the extended H1 haplotype. This could particularly be the case in CBD in the light of the lack of association of *del-In9* in this particular study.

There is potentially a greater power to detect the contribution to association of causal variants by undertaking tests of association for the htSNP defined haplotypes rather than individual htSNPs themselves. The six htSNPs we identified capture 95% of the common haplotypic diversity of *MAPT* and we carried out an omnibus test of haplotype frequency differences estimated by EM between cases and controls in both the UK and US PSP groups. We found the haplotype distribution (all haplotypes >1.0%) was highly significant in the UK PSP cohort ($p = 9.75\times10^{-5}$, df = 19) and in the US PSP cohort ($p = 7.40\times10^{-12}$, df = 20) but not in CBD ($p = 0.120$, df = 17). In addition to the global significance of the haplotype-wide comparison, we undertook individual haplotype tests (df = 1) for significance through LRT, and derived empirical p values through Monte-Carlo methods (20 000 simulations, data not shown); we identified two common haplotypes, A and C, which were strongly associated with both UK and US PSP (table 4). Haplotype A, which derives from the *del-In9* defined H2 haplotype, was the most common type in the controls and was significantly underrepresented in both PSP groups. Haplotype C, a variant of the H1 clade, was highly overrepresented in PSP. It was the commonest haplotype in PSP but not in the control groups. The most common H1 derived haplotype in the control population was not associated with either PSP or CBD. These trends were observed in CBD (table 4), though on correction for multiple comparisons no haplotype was significantly associated. In both PSP cohorts, after strict correction according to the number of tests performed, only associations of haplotypes A and C remained significant. Associated haplotypes A and C, derived from the H2 and H1 haplotypes respectively, differ by only two H1 specific htSNPs, 14 and 21, which, in addition to *del-In9*, also show powerful single locus effects. Haplotypes A and C do not differ by htSNPs 8 and 25, and these SNPs are not associated. The reduction in haplotype A (H2) appears almost entirely accounted for by the increase in the H1 haplotype C.

Common variation in *MAPT* is associated with PSP and CBD

To assess whether the significant association with PSP of any of the H1 specific htSNPs is independent of that of *del-In9*, we incorporated each htSNP as an additional explanatory factor to the logistic regression model of the *del-In9* that serves to define the extended H1 and H2 haplotype status. We found significant association of single locus htSNPs 14, 17, and 21 ($p = 9.00\times10^{-6}$, 2.87×10^{-3} and 2.73×10^{-3} respectively) for

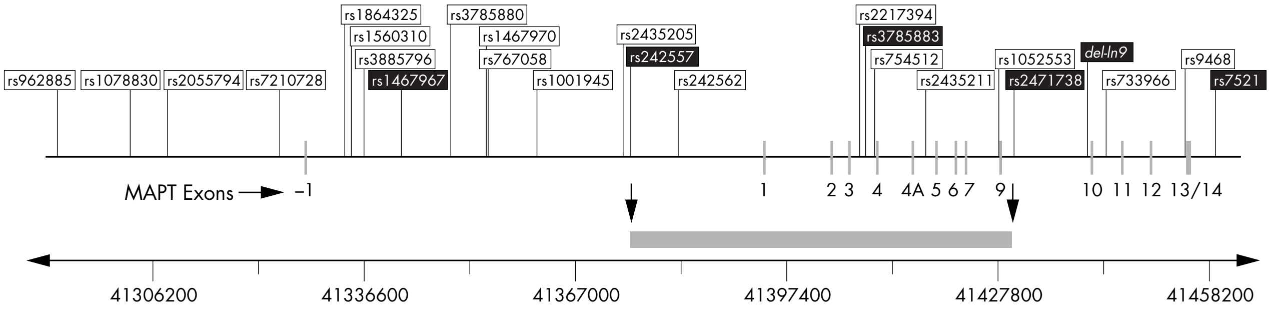

Figure 3 Genomic organisation of *MAPT* and distribution of the 25 markers used in the linkage disequilibrium and haplotype analysis in the CEPH trios. Relative positions of the promoter, coding exons, and genetic markers in the HapMap CEPH trios are to scale. Chromosome coordinates (base pairs) are according to the March 2004 build of the Human Genome Sequence (http://genome.ucsc.edu). Haplotype tagging SNPs selected for this study are indicated by the black boxes. The minimum candidate region identified in this study, in which potential causal variants may lie on the H1 background, is indicated by the grey bar.

the US PSP cases, htSNP 21 ($p = 0.0421$) for the UK PSP cases, and htSNPs 14 and 21 ($p = 0.0183$ and 0.0436, respectively) for the CBD cases. We probed for effects of haplotypes on subsets of htSNPs, again entering the extended haplotype (H1 and H2 status, defined by the *del-In9*) as an explanatory factor. We found highly significant differences in the distribution of haplotypes defined by three htSNPs 14, 17, and 21 in the UK and US PSP, and to a lesser extent in the CBD cases ($p = 9.34\times10^{-4}$, $p = 9.31\times10^{-5}$, and $p = 0.0292$, respectively). This was significant ($p = 2.49\times10^{-5}$, $p = 1.44\times10^{-8}$, and $p = 0.006$) in UK PSP, US PSP, and CBD, respectively, when the extended haplotype was excluded as an explanatory factor (table 5). The haplotypes they define are associated with PSP and CBD after consideration of the *del-in9*, suggesting that variability of *MAPT* within the extended H1 clade is a risk factor in PSP and CBD. Haplotype II (A-G-T) was greatly overrepresented in each group, and the haplotype I (G-G-C) underrepresented (table 5). The SNPs 14, 17, and 21 (rs242557, rs3785883, and rs2471738, respectively) are H1 specific SNPs in *MAPT*—that is, variable only on the H1 background, though the haplotype I allelic combination is fixed and representative of H2 in addition to H1 derived variants.

We also attempted to reanalyse the htSNP data, after removing all individuals with an H2 chromosome, thus leaving us with a biased H1H1 homozygote population. We found significant heterogeneity ($p<0.05$) in both the control groups after the removal of the H2 chromosomes, namely at rs1467967 and rs7521 in the US group and at rs242557, rs2471738, and rs7521 in the UK controls. Removal of the H2 chromosomes would therefore prevent us from performing valid "H1-only" haplotype analyses in our white cohorts. For this purpose, it would be important to extend this study in an H1-only population such as the Japanese and Taiwanese.[30]

DISCUSSION

To date, genetic association studies have involved the study of one or a few random polymorphisms in a gene, an approach that bears the risk of missing adjacent regions of LD within the gene that harbour variants associated with phenotype. It is therefore important that the haplotype architecture of the *entire* gene is considered in order to determine its association with a particular complex phenotype. In our attempt to provide insight into the basis of the well established association of *MAPT* with PSP and CBD, we applied the haplotype tagging approach. This protocol, which uses a minimal set of tagging SNPs to study the LD and common haplotypic diversity of the entire gene or locus, is substantially more stream lined and economical.

We first assessed the underlying LD and haplotype structure of *MAPT* using a high density map of genotype data from the HapMap project (http://www.hapmap.org). This involved LD analysis using genotype data for 24 SNPs that had been validated in CEPH trios. In addition, we included the *del-In9* status, defining the H1 and H2 haplotypes.[17] This revealed multiple distinct haplotypes based upon the H1 and H2, as defined by *del-In9*, with no evidence of recombination between the multiple H1 haplotypes and the H2 in the CEPH trios. The presence of multiple H1 haplotypes, inferred both by EM and resolved to phase, shows a considerable diversity within this extended haplotype. This H1 haplotype specific diversity was first suggested by Golbe and colleagues, based on microsatellite variability.[31] The strict H1/H2 dichotomy and H1 diversity across *MAPT* and beyond has also been demonstrated in other studies.[25 32] In a more recent study,[33] the lack of recombination between H1 and H2 has been shown to be caused by inversion of the chromosomal region on 17q21.31 corresponding to the extended MAPT H1/H2 haplotype block that we had previously described.[20]

We then used association based criteria to assign a set of five haplotype tagging SNPs (htSNPs) which, together with *del-In9* as a sixth biallelic tagging polymorphism, capture 95% of the common haplotype diversity in *MAPT*. We genotyped the six htSNPs in two PSP and one CBD case–control cohorts in order to determine if any particular haplotype had greater association with disease with the extended H1. In PSP we showed clearly that there were very strong associations of two common haplotypes—first, the significant underrepresentation of the "classical" H2 (haplotype A, table 4), and second, strong overrepresentation of an H1 derived haplotype (haplotype C, table 4). The other htSNP derived common H1 haplotype (haplotype B) showed no association in any of the groups. Some weaker associations of rare haplotypes were detected but were not consistent in both the British and American cohorts in PSP, and the significance did not remain after correction for multiple comparisons. Furthermore, it is difficult to assess the association of such low frequency haplotypes in populations of our sample size. Similar trends were observed in the small number of CBD cases ($n = 44$), with underrepresentation of H2 (Haplotype A; table 4) and overrepresentation of the H1 derived haplotype C (table 4). However, they were not significant, possibly because of the smaller number of CBD cases. Assuming that these findings can be confirmed in a larger CBD cohort, they suggest that causative variants in PSP and CBD may affect the same region of *MAPT* or perhaps even be the same variant.

Pastor and colleagues defined an extended region in LD of 1.14 Mb around *MAPT* that is associated with PSP and CBD.[34]

Within this haplotype, they similarly defined a ''protective'' H2 haplotype that has a significant negative association with PSP and CBD, and an H1 derived haplotype that is associated with PSP and CBD.[34] Our work refines the analysis of LD, haplotype structure, and associations of the *MAPT* gene alone and we have demonstrated that a particular H1 derived haplotype in *MAPT* is highly associated with PSP.

In an attempt to further minimise the candidate pathogenic domain of *MAPT*, we also identified particularly strong association with PSP and CBD of three-locus haplotypes based on the subset of H1 specific htSNPs, 14, 17, and 21 (table 3). These associations are independent of the extended H1 and H2 haplotypes, defined by *del-In9*. As indicated in fig 3, haplotypes derived from these SNPs span a minimum region from SNPs 14 (rs242557) to 21 (rs2471738) on the H1 haplotype background in *MAPT*. This minimum region incorporates ~56.3 kb of sequence, from upstream of exon 1 downstream to intron 9, that could harbour potential causal variants that are in LD with these SNPs. Skipper and colleagues defined a similar associated candidate region in the 5′-half of *MAPT* in Norwegian Parkinson's disease cases, thereby proposing genetic variability that could influence the alternative splicing of *MAPT* exons 2 and 3, or expression levels of *MAPT*.[25] However, they carried out their analysis only on H1 homozygous individuals, having removed all H2 carriers.[25] For this reason, we cannot compare findings from both studies. As explained above in Results, unbiased inclusion of the entire study cohort, irrespective of H1/H2 status, is essential in order to obtain an accurate representation of haplotype diversity in the population in question. Another study implicated an *MAPT* promoter haplotype in Parkinson's disease, based not only on allelic association of the previously defined extended H1 haplotype but also on differences in transcriptional activity.[35] In future studies, it would be important to compare LD and association of the *MAPT* locus in PSP, CBD, and Parkinson's disease using standardised procedures, in order to determine if they share the same risk variants of the *MAPT* locus that contribute to disease.

The haplotypes we identified that confer protection, risk, or are neutral in PSP and CBD pathogenesis provide us with the basis for targeted direct sequencing strategies for *MAPT*. It is now clear that there are no obvious pathogenic missense or splice site mutations in *MAPT* in the large majority of sporadic PSP cases.[17] It is more plausible that the associated SNPs in our study that confer greatest risk (SNPs 14 (rs242557) and 21(rs2471738); table 1 and fig 3) or protection (*del-In9* and associated SNPs through LD; table 1 and fig 1) are in LD with variants that could cause subtle changes either in the alternative splicing or in overall expression levels. It is possible that each neuronal subgroup is dependent on a particular tau isoform profile and expression level. Aberrations in this homeostasis could affect one neuronal subgroup more than another and lead to the selective and disease specific neuronal death and tau pathology.[36] Investigating correlations between candidate polymorphisms and *MAPT* splicing and allele specific expression—combined with the association studies described in this work—and the resulting identification of candidate variations by stringently targeted resequencing strategies in individuals carrying the haplotypes described here, could help us gain further insight into the precise nature of the role of *MAPT* in the molecular pathogenesis of PSP, CBD, Parkinson's disease, and the tauopathies.

ACKNOWLEDGEMENTS

We 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 for Medical Research, the PSP (Europe) Association (http://www.pspeur.org), the Society for PSP, USA, the Parkinson's Disease Society, UK, the Brain Research Trust (PAS), Medical Research Council, NIH grant p50-NS40256-06 (DD, NT), The Society for PSP Brain Bank, and by the NIA/NIH Intramural Research Program. AJM is a resident research associate of the National Academy of Sciences.

Many data and biomaterials were collected from several NIA-NACC funded sites. The directors, pathologist and technicians involved include: *National Institute on Aging*: Marcelle Morrison-Bogorad PhD, Tony Phelps PhD, Ruth Seemann; *Johns Hopkins Alzheimer's Disease Research Center* (NIA grant No AG 05146): Juan C Troncoso MD, Dr Olga Pletnikova; *University of California, Los Angeles* (NIA grant No P50 AG16570):Harry Vinters MD, Justine Pomakian; *The Kathleen Price Bryan Brain Bank, Duke University Medical Center* (NIA grant No AG05128, NINDS grant No NS39764, NIMH MH60451 also funded by Glaxo Smith Kline): Christine Hulette MD, Director; *Stanford University*: Dikran Horoupian MD, Ahmad Salehi MD, PhD; *New York Brain Bank, Taub Institute, Columbia University* (NYBB): Jean Paul Vonsattel MD; *Massachusetts General Hospital*: E Tessa Hedley-Whyte MD, Karlotta Fitch; *University of Michigan* (NIH grant P50-AG08671): Dr Roger Albin, Lisa Bain, Eszter Gombosi; *University of Kentucky*: William Markesbery MD, Sonya Anderson; *University Southern California*: Caroll A Miller MD, Jenny Tang MS, Dimitri Diaz; *Washington University, St Louis Alzheimer's Disease Research Center*: Dan McKeel MD, John C Morris MD, Eugene Johnson Jr PhD, Virginia Buckles PhD, Deborah Carter; *University of Washington, Seattle*: Thomas Montine MD, PhD, Aimee Schantz MEd.

Authors' affiliations

A M Pittman, H C Fung, A J Lees, R de Silva, Reta Lila Weston Institute of Neurological Studies, University College London, London, UK
P Abou-Sleiman, N W Wood, Department of Molecular Neuroscience, Institute of Neurology, London, UK
A J Myers, M Kaleem, L Marlowe, J Duckworth, D Leung, J Hardy, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA
D Williams, L Kilford, Sara Koe PSP Research Centre, Institute of Neurology, London, UK
C M Morris, Institute for Ageing and Health, MRC Building, Newcastle General Hospital, Westgate Road, Newcastle-upon-Tyne, UK
N Thomas, D Dickson, Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, Florida, USA

Competing interests: none declared

REFERENCES

1 **Rademakers R**, Cruts M, van Broeckhoven C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum Mutat* 2004;**24**:277–95.

2 **Hutton M**, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, Hackett J, Adamson J, Lincoln S, Dickson D, Davies P, Petersen RC, Stevens M, de Graaff E, Wauters E, van Baren J, Hillebrand M, Joosse M, Kwon JM, Nowotny P, Heutink P, Che LK, Norton J, Morris JC, Reed LA, Trojanowski J, Basun H, Lannfelt L, Neystat M, Fahn S, Dark F, Tannenberg T, Dodd PR, Hayward N, Kwok JBJ, Schofield P, Andreadis A, Snowden J, Craufurd D, Neary D, Owen F, Oostra BA, Hardy J, Goate A, van Swieten J, Mann D, Lynch T, Heutink P. Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* 1998;**393**:702–5.

3 **Spillantini MG**, Murrell JR, Goedert M, Farlow MR, Klug A, Ghetti B. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc Natl Acad Sci USA* 1998;**95**:7737–41.

4 **Steele JC**, Richardson JC, Olszewski J. 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* 1964;**10**:333–59.

5 **Maher ER**, Lees AJ. The clinical features and natural history of the Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). *Neurology* 1986;**36**:1005–8.

6 **Daniel SE**, de Bruin VM, Lees AJ. The clinical and pathological spectrum of Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy): a reappraisal. *Brain* 1995;**118**:759–70.

7 **Litvan I**, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, Goetz CG, Golbe LI, Grafman J, Growdon JH, Hallett M, Jankovic J, Quinn NP, Tolosa E, Zee DS. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology* 1996;**47**:1–9.

8 **de Yebenes JG**, Sarasa JL, Daniel SE, Lees AJ. Familial progressive supranuclear palsy. Description of a pedigree and review of the literature. *Brain* 1995;**118**:1095–103.

9 **Rojo A**, Pernaute RS, Fontan A, Ruiz PG, Honnorat J, Lynch T, Chin S, Gonzalo I, Rabano A, Martinez A, Daniel S, Pramstaller P, Morris H, Wood N, Lees A, Tabernero C, Nyggard T, Jackson AC, Hanson A, de Yebenes JG,

Pramsteller P. Clinical genetics of familial progressive supranuclear palsy. *Brain* 1999;**122**:1233–45.

10 **Conrad C**, Andreadis A, Trojanowski JQ, Dickson DW, Kang D, Chen X, Wiederholt W, Hansen L, Masliah E, Thal LJ, Katzman R, Xia Y, Saitoh T. Genetic evidence for the involvement of tau in progressive supranuclear palsy. *Ann Neurol* 1997;**41**:277–81.

11 **Di Maria E**, Tabaton M, Vigo T, Abbruzzese G, Bellone E, Donati C, Frasson E, Marchese R, Montagna P, Munoz DG, Pramstaller PP, Zanusso G, Ajmar F, Mandich P. Corticobasal degeneration shares a common genetic background with progressive supranuclear palsy. *Ann Neurol* 2000;**47**:374–7.

12 **Houlden H**, Baker M, Morris HR, MacDonald N, Pickering-Brown S, Adamson J, Lees AJ, Rossor MN, Quinn NP, Kertesz A, Khan MN, Hardy J, Lantos PL, St George-Hyslop P, Munoz DG, Mann D, Lang AE, Bergeron C, Bigio EH, Litvan I, Bhatia KP, Dickson D, Wood NW, Hutton M. Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. *Neurology* 2001;**56**:1702–6.

13 **Poorkaj P**, Muma NA, Zhukareva V, Cochran EJ, Shannon KM, Hurtig H, Koller WC, Bird TD, Trojanowski JQ, Lee VM, Schellenberg GD. An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype. *Ann Neurol* 2002;**52**:511–16.

14 **Wszolek ZK**, Tsuboi Y, Uitti RJ, Reed L, Hutton ML, Dickson DW. Progressive supranuclear palsy as a disease phenotype caused by the S305S tau gene mutation. *Brain* 2001;**124**:1666–70.

15 **Morris HR**, Osaki Y, Holton J, Lees AJ, Wood NW, Revesz T, Quinn N. Tau exon 10+16 mutation FTDP-17 presenting clinically as sporadic young onset PSP. *Neurology* 2003;**61**:102–4.

16 **Morris HR**, Katzenschlager R, Janssen JC, Brown JM, Ozansoy M, Quinn N, Revesz T, Rossor MN, Daniel SE, Wood NW, Lees AJ. Sequence analysis of tau in familial and sporadic progressive supranuclear palsy. *J Neurol Neurosurg Psychiatry* 2002;**72**:388–90.

17 **Baker M**, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, Hardy J, Lynch T, Bigio E, Hutton M. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet* 1999;**8**:711–15.

18 **Ezquerra M**, Pastor P, Valldeoriola F, Molinuevo JL, Blesa R, Tolosa E, Oliva R. Identification of a novel polymorphism in the promoter region of the tau gene highly associated to progressive supranuclear palsy in humans. *Neurosci Lett* 1999;**275**:183–6.

19 **de Silva R**, Weiler M, Morris HR, Martin ER, Wood NW, Lees AJ. Strong association of a novel Tau promoter haplotype in progressive supranuclear palsy. *Neurosci Lett* 2001;**311**:145–8.

20 **Pittman AM**, Myers AJ, Duckworth J, Bryden L, Hanson M, Abou-Sleiman P, Wood NW, Hardy J, Lees A, de Silva R. The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum Mol Genet* 2004;**13**:1267–74.

21 **Conrad C**, Vianna C, Freeman M, Davies P. A polymorphic gene nested within an intron of the tau gene: implications for Alzheimer's disease. *Proc Natl Acad Sci U S A* 2002;**99**:7751–6.

22 **de Silva R**, Hope A, Pittman A, Weale ME, Morris HR, Wood NW, Lees AJ. Strong association of the Saitohin gene Q7 variant with progressive supranuclear palsy. *Neurology* 2003;**61**:407–9.

23 **Ponting CP**, Hutton M, Nyborg A, Baker M, Jansen K, Golde TE. Identification of a novel family of presenilin homologues. *Hum Mol Genet* 2002;**11**:1037–44.

24 **Healy DG**, Abou-Sleiman PM, Lees AJ, Casas JP, Quinn N, Bhatia K, Hingorani AD, Wood NW. Tau gene and Parkinson's disease: a case-control study and meta-analysis. *J Neurol Neurosurg Psychiatry* 2004;**75**:962–5.

25 **Skipper L**, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M, Ross OA, Hutton M, Aasly J, Farrer M. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet* 2004;**75**:669–77.

26 **Weale ME**, Depondt C, Macdonald SJ, Smith A, Lai PS, Shorvon SD, Wood NW, Goldstein DB. Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping. *Am J Hum Genet* 2003;**73**:551–65.

27 **Lewontin RC**. The interaction of selection and linkage. I General Considerations; heterotic models. *Genetics* 1964;**49**:49–67.

28 **Morris HR**, Janssen JC, Bandmann O, Daniel SE, Rossor MN, Lees AJ, Wood NW. The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases. *J Neurol Neurosurg Psychiatry* 1999;**66**:665–7.

29 **Sham PC**, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 1995;**59**:97–105.

30 **Evans W**, Fung HC, Steele J, Eerola J, Tienari P, Pittman A, de Silva R, Myers A, Vrieze FW, Singleton A, Hardy J. The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci Lett* 2004;**369**:183–5.

31 **Golbe LI**, Lazzarini AM, Spychala JR, Johnson WG, Stenroos ES, Mark MH, Sage JI. The tau A0 allele in Parkinson's disease. *Mov Disord* 2001;**16**:442–7.

32 **Oliveira SA**, Scott WK, Zhang F, Stajich JM, Fujiwara K, Hauser M, Scott BL, Pericak-Vance MA, Vance JM, Martin ER. Linkage disequilibrium and haplotype tagging polymorphisms in the Tau H1 haplotype. *Neurogenetics* 2004;**5**:147–55.

33 **Stefansson H**, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier JB, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K. A common inversion under selection in Europeans. *Nat Genet* 2005;**37**:129–37.

34 **Pastor P**, Ezquerra M, Perez JC, Chakraverty S, Norton J, Racette BA, McKeel D, Perlmutter JS, Tolosa E, Goate AM. Novel haplotypes in 17q21 are associated with progressive supranuclear palsy. *Ann Neurol* 2004;**56**:249–58.

35 **Kwok JB**, Teber ET, Loy C, Hallupp M, Nicholson G, Mellick GD, Buchanan DD, Silburn PA, Schofield PR. Tau haplotypes regulate transcription and are associated with Parkinson's disease. *Ann Neurol* 2004;**55**:329–34.

36 **Buee L**, Delacourte A. Comparative biochemistry of tau in progressive supranuclear palsy, corticobasal degeneration, FTDP-17 and Pick's disease. *Brain Pathol* 1999;**9**:681–93.