A MOLECULAR GENETIC STUDY OF THE TAU LOCUS IN NEURODEGENERATION

Thesis submitted in fulfilment of the degree of
Doctor of Philosophy

Reta Lila Weston Institute of Neurological Studies
Institute of Neurology
University College London
University of London

October 2006

Alan Michael Pittman

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Tau is a microtubule-associated protein and is deposited as neurofibrillary tangles in the group of neurodegenerative diseases collectively known as the tauopathies including Alzheimer's disease (AD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and frontotemporal dementia with parkinsonism linked to chromosome 17 with tau pathology (FTDP-17T). The pathological tau is hyperphosphorylated and has a reduced microtubule binding capacity. The identification of missense and splice site mutations in the tau gene (*MAPT*) causing FTDP-17T affirmed a central role for tau dysfunction in this neurodegenerative disease. PSP is usually a sporadic disorder of late adult life, with no family history or *MAPT* mutations in the majority of cases. However, it has been shown that common genetic variation at the *MAPT* locus is an important genetic risk factor for sporadic PSP. This finding has been consistently and independently replicated. There are two major *MAPT* haplotypes at 17q21.31 designated H1 and H2. It is the over-representation of H1 that is associated with PSP and in this work the extent of the H1/H2 *MAPT* non recombining haplotypes was mapped to cover a region of approximately 2 million base pairs of perfect linkage disequilibrium.

In order to explore the functional and pathogenic basis of the H1 haplotype, a systematic framework of genetic analysis was employed for a high resolution association study of the *MAPT* gene with PSP. Multiple common variants of the H1 haplotype were identified and one common haplotype, H1c, showed preferential association with PSP above all others. Candidate causal variants were identified on this haplotype background and were tested for their effects on *MAPT* promoter driven expression using luciferase reporter assays. The functional SNP loci that were identified reside in evolutionarily conserved islands and the SNP variants that give rise to higher *MAPT* promoter driven expression were significantly over-represented in PSP. Thus, increased *MAPT* expression could at least in part explain the association of the locus with PSP.

The region of 17q21.31 also shows complex genomic architecture containing low-copy repeats (LCRs) that are responsible for an ancient parcentric inversion that

corresponds to the H1 and H2 lineages. A *de novo* microdeletion encompassing the *MAPT* gene was identified in three individuals with moderate to severe learning disabilities. In this work, a in depth haplotype analysis in these triad families with respect to H1 and H2 haplotypes unambiguously revealed that in two of the cases the parental origins of the deletion were heterozygous carriers for the H1/H2 inversion. A mechanism of inversion mediated non-allelic homologous recombination between the *MAPT* 17q21.31 LCRs was proposed to explain the generation of the deletion.

Acknowledgements

Firstly I would like to thank my supervisors: Andrew Lees, for providing a stimulating and enjoyable environment for research at the Reta Lila Weston Institute of Neurological studies, Nick Wood, Head of the Department of Molecular Neuroscience for guidance and special thanks go to my principal supervisor and friend, Rohan de Silva, for encouragement and for providing an exciting research project. I also give thanks to Amanda Myers for help with genetics in the early days and to John Hardy for his inspiration and contagious enthusiasm for science.

Thanks are given to the Reta Lila Weston Trust and the Parkinson's Disease Society for financial support. I am grateful to all the staff and lab colleagues at the Reta Lila Weston Institute for guidance and assistance: Rina Bandopadhyay, Yvonne Mwelwa, Drew Hope, Andy Reid, Ravindran Kumaran, Connie Luk, Laura Moriyama and Sean O'Sullivan. In particular I would like to thank David Williams for assistance with understanding the clinical aspects of PSP. I would like to thank all the staff at the Queen Square Brain Bank for Neurological Disorders and the Sarah Koe PSP Research Centre: Susan Stoneham, Karen Shaw, Tammaryn Lashley and Tamas Ravesz and I would like to thank Linda Kilford and Ann Kingsbury for providing tissue from PSP samples for DNA extraction. Thanks are given to members of the Laboratory of Neurogenetics, NIA at the NIH for our tau genetics 'collective': Philip Fung and Jaime Duckworth. Thanks are also given to members of Addenbrookes Hospital, Cambridge: Lionel Willatt and in particular, special thanks go to Charles Shaw-Smith for the exciting collaboration in the molecular characterisation of a new microdeletion syndrome.

I am grateful to all the patients and their families, without whom none of this research would have been possible.

Lastly, this thesis is dedicated with love to all of my family and friends.

Collaborations

This thesis has involved a number of important collaborations including the supply of samples, data and the exchange of scientific thoughts and ideas:

Chapter 3, 4 and 5

The pathologically confirmed PSP cases from the UK in this study were all obtained from the Queen Square Brain Bank, Institute of Neurology, University College London, 1 Wakefield Street, London WC1N 1PJ. The pathologically confirmed control samples from the UK were obtained from Dr. Chris Morris, Institute for Ageing and Health, MRC Building, Newcastle General Hospital, Newcastle-upon-Tyne.

The pathologically confirmed PSP cases from the US were obtained from the Mayo Clinic, Jacksonville, Florida from Dr. Dennis W. Dickson and Ms. Natalie Thomas. The pathologically confirmed control samples from the US were obtained from the Laboratory of Neurogenetics NIA, NIH, Bethesda from Dr. Amanda Myers.

The Japanese control DNA samples were from the RIKEN Brain Science Institute, Saitama, Japan with thanks to Drs Masanari Itokawa and Takeo Yoshikawa.

Chapter 6

DNA samples from affected individuals numbers 1 and 2, and their parents were obtained by Dr. Charles Shaw-Smith, Addenbrooke's Hospital, Cambridge, CB2 2QQ. DNA from affected individual number 3 and the parents were collected by Dr. Carla Rosenberg, Department of Genetics and Evolutionary Biology, University of Sao Paolo, Brazil. The fosmid array-CGH results for the three cases were generated by Dr. Charles Shaw-Smith and Dr. Susan Gribble at the Wellcome Trust Sanger Institute, Cambridge, CB10 1SA. The FISH results were generated by Lionel Willatt at the Regional Cytogenetics Laboratory, Addenbrooke's Hospital, Cambridge, CB2 2QQ.

Chapter 7

The FTDP-17T case positive for the *MAPT* mutation was obtained from Dr. Gabor Kovacs, National Institute of Psychiatry and Neurology, Department of Neuropathology, National Reference Centre for Human Prion Diseases, Budapest, Hungary. Thanks are given to Tamas Ravesz for help in understanding the clinical and pathological phenotype of the patient.

Publications

Published papers either as a direct result from (Appendix 10.6) or through collaborative work during this thesis:

Pittman, A.M., Fung, H.C., de Silva, R. (2006). Untangling the tau gene association with neurodegenerative disorders. Human Molecular Genetics 15, R188-R195.

Pittman, A.M.,Shaw-Smith, C., Willatt, L., Martin, H., Rickman, L., Gribble, S., Curley, R., Cumming, S., Dunn, C., Kalaitzopoulos, D., Porter, K., Prigmore, E., Krepischi-Santos, A.C., Varela, M.C., Koiffmann, C.P., Lees, A.J., Rosenberg, C., Firth, H.V., de Silva, R., Carter, N.P. (2006). Microdeletion encompassing *MAPT* at chromosome 17q21.3 is associated with developmental delay and learning disability. Nature Genetics 38(9), 1032-1037.

Hardy, J., **Pittman, A.**, Myers, A., Fung, H.C., de Silva, R., Duckworth, J. (2006). Tangle Diseases and the tau haplotypes. Alzheimer's Disease and Associated Disorders 20(1), 60-62

Bandopadhyay, R., Miller, D.W., Kingsbury, A.E., Jowett, T.P., Kaleem, M.M., **Pittman, A.M.**, de, Silva R., Cookson, M.R., Lees, A.J. (2005). Development, characterisation and epitope mapping of novel monoclonal antibodies for DJ-1 (PARK7) protein. Neuroscience Letters 383(3), 225-230.

Fung,H.C., Evans,J., Evans,W., Duckworth,J., **Pittman,A.**, de Silva,R., Myers,A., Hardy,J. (2005). The architecture of the tau haplotype block in different ethnicities. Neuroscience Letters 377(2), 81-84.

Hardy, J., **Pittman, A.**, Myers, A., Gwinn-Hardy, K., Fung, H.C., de Silva, R., Hutton, M., Duckworth, J. (2005). Evidence suggesting that Homo neanderthalensis contributed the H2 *MAPT* haplotype to Homo sapiens. Biochemical Society Transactions 33(4), 582-585.

Myers, A.J., Kaleem, M., Marlowe, L., **Pittman, A.M.**, Lees, A.J., Fung, H.C., Duckworth, J., Leung, D., Gibson, A., Morris, C.M., de Silva, R., Hardy, J. (2005). The H1c haplotype at the *MAPT* locus is associated with Alzheimer's disease. Human Molecular Genetics 14(16), 2399-2404.

Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L., Thomas, N., Morris, C.M., Dickson, D., Wood, N.W., Hardy, J., Lees, A.J., de Silva, R. (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the *tau* gene in progressive supranuclear palsy and corticobasal degeneration. Journal of Medical Genetics 42(11), 837-846.

Williams, D.R., de Silva, R., Paviour, D.C., **Pittman, A.**, Watt, H.C., Kilford, L., Holton, J.L., Revesz, T., Lees, A.J. (2005). Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. Brain 128(6), 1247-1258

Bandopadhyay, R., Kingsbury, A.E., Cookson, M.R., Reid, A.R., Evans, I.M., Hope, A.D., **Pittman, A.M.**, Lashley, T., Canet-Aviles ,R., Miller, D.W., McLendon, C., Strand, C., Leonard, A.J., Abou-Sleiman, P.M., Healy, D.G., Ariga, H., Wood, N.W., de Silva, R., Revesz, T., Hardy, J.A., Lees, A.J. (2004). The expression of DJ-1 (PARK7) in normal human CNS and idiopathic Parkinson's disease. Brain 127(2), 420-430.

Evans, W., Fung, H.C., Steele, J., Eerola, J., Tienari, P., **Pittman, A.**, de Silva, R., Myers, A., Vrieze, F.W., Singleton, A., Hardy, J. (2004). The tau H2 haplotype is almost exclusively Caucasian in origin. Neuroscience Letters 369(3), 183-185.

Pittman, A., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., de Silva, R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics 13(12), 1267-1274.

de Silva, R., Hope, A., **Pittman, A.**, Weale, M.E., Morris, H.R., Wood, N.W., Lees, A.J. (2003). Strong association of the Saitohin gene Q7 variant with progressive supranuclear palsy. Neurology 61(3), 407-409.

Abbreviations

AD	Alzheimer's disease
AGD	argyrophilic grain disease
ALS	amyotrophic lateral sclerosis
array-CGH	array-comparative genomic hybridisation
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
CBD	corticobasal degeneration
CEPH	Centre d'Etude du Polymorphisme Humain
CNPs	copy-number polymorphisms
CNVs	copy-number variants
CI	confidence interval
CSF	cerebrospinal fluid
CVCD	common variant-common disease hypothesis
DNA	deoxyribosenucleic acid
dNTPs	deoxynucleosides
del-In9	238 bp H1/H2 insertion deletion polymorphism
EM	expectation-maximization algorithm
FISH	fluorescence *in situ* hybridization
FTD	frontotemporal dementia
FTD3	autosomal dominant FTD linked to chromosome 3
FTD9	autosomal dominant ALS with FTD linked chromosome 9
FTLD	frontotemporal lobar degeneration
FTDP-17T	frontotemporal dementia with parkinsonism linked to chromosome 17 with tau pathology
HapMap	The International HapMap project
htSNP	tagging SNP
HWE	Hardy-Weinberg equilibrium
LCR	low-copy repeat
LD	linkage disequilibrium
LOD	logarithm of the odds
LRT	likelihood ratio test

MAP	microtubule associated protein
NAHR	non-allelic homologous recombination
NCBI	National Centre for Biotechnology Information
NFTs	neurofibrillary tangles
NPC	Niemann-Pick disease type C
NTs	Neuropil threads
OR	odds ratio
PAC	P1 artificial chromosome
PL-EM	partition ligation-expextation maximization algorithm
PCR	Polymerase chain reaction
PD	Parkinson's disease
PET	positron emission tomography
PiD	Pick's disease
PPi	pyrophosphate
PSP	progressive supranuclear palsy
PSP-P	PSP-parkinsonism
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
qPCR	quantitative PCR
RS	Richardson's syndrome
SNP	single nucleotide polymorphism
SSPE	subacute sclerosing panencephalitis
tau	microtubule associated protein tau
UCSC	University of California Santa Cruz
UTR	untranslated region
WBS	Williams-Beuren syndrome
3R	three repeat tau
4R	four repeat tau

Gene abbreviations

ACBD4	acyl-coenzyme A binding domain containing 4
CRF	C1q-related factor
CRHR1	corticotrophin releasing hormone receptor 1
FMNL1	formin-like 1
GFAP	glial fibrillary acidic protein gene
HIS1	hexamethylene bisacetamide-inducible
MAP3K14	mitogen activated kinase kinase kinase 14
MAPT	microtubule associated protein, tau
NMT1	N-myristoyltransferase
NSF	N-ethlymaleimide sensitive factor
PLCD3	phospholipase C, delta 3
IMP5	presenlilin homologue 2
STH	saitohin
WNT3	wingless-type MMTV integration site family, member 3

Contents

Tables

Figures

1 Introduction

1.1 Overview

There is continuing evidence to suggest that common genetic variation at the population level can contribute to common, complex diseases as suggested by the common variant-common disease hypothesis (CDCV). For example, the *APOE* locus and the risk of Alzheimer's disease (See section 1.2.3.7). Recent advances in the field of genetics, including the completed sequence of the human genome and the International HapMap project are now fuelling the rapid study of the effect of common genetic variation on human disease.

There are approximately 1 million genetic differences or polymorphisms between any two individuals. The vast majority of these differences are accounted for by single nucleotide polymorphisms (SNPs) but other differences include sub-microscopic structural polymorphisms including DNA copy number changes such as duplications, triplications and deletions and structural polymorphisms such as inversions and translocations where there is no such DNA copy number change.

Neurodegenerative diseases are a group of progressive heterogeneous disorders with a wide clinocopathological spectrum. Recent advances in classification of pathology, biochemistry and genetics are beginning to shed light on their molecular pathogenesis. One such group of diseases, the tauopathies, are characterised pathologically by intra-neuronal inclusions of insoluble aggregates of the microtubule associated protein, tau. These disorders present clinically with a wide spectrum of symptoms ranging from dementia to parkinsonism and include Alzheimer's disease, frontotemporal dementia, progressive supranuclear palsy, Picks disease and corticobasal degeneration.

Tau dysfunction has been linked to the pathogenesis of the tauopathies not only through their hallmark pathology but also with the identification of highly penetrant dominant mutations in the tau gene, *MAPT*, causing frontotemporal dementia with

parkinsonism linked to chromosome 17. This 'smoking gun' highlighted a direct link of tau dysfunction to neurodegeneration and much attention has now been given to investigate tau dynamics in the sporadic tauopathies. Progressive supranuclear palsy is one such sporadic tauopathy and common genetic variation at the *MAPT* locus has been shown to be a genetic risk factor. Understanding the functional basis of this association provides a most promising means to understand the pathogenesis of the tauopathies and was the aim of the work presented in this thesis. Through this work, and work by collaborators and other groups, we now have a better understanding of the functional basis of the association.

1.2 Genetic approaches to studying human disease

1.2.1 Genetic epidemiology

Genetic epidemiology is a discipline closely related to traditional epidemiology that focuses on the familial and the population, toward identification of genetic determinants of disease and the joint effects of genes and non-genetic determinants such as the environment [1]. Importantly, genetic epidemiology is a fusion of traditional epidemiological principles with the biology of genes and their mode of inheritance.

The vast majority of success so far in genetic epidemiology has been related to the identification of disease causing genes in monogenic disorders, relying heavily on linkage studies and positional cloning, where familial recurrence appears to obey the laws of Mendelian inheritance. However, genetic epidemiology today is increasingly focused on complex diseases such as neurodegenerative disease, diabetes mellitus and cancer. These diseases are thought to be caused by several interacting genetic and environmental determinants [2] and require quite different genetic epidemiological study design and interpretation compared to traditional genetic linkage studies in monogenic Mendelian disorders [1].

In the following section, the traditional methods used for identifying Mendelian disease causing genes are briefly described. This is followed by a detailed

description of the rationale and principles behind genetic mapping of complex disease genes by indirect association that were used for much of the study design behind the research detailed in this thesis.

1.2.2 Genetic mapping of Mendelian disease causing genes

1.2.2.2 Parametric linkage analysis

Genetic linkage analysis is used to identify regions of the genome that contain genes predisposing to Mendelian diseases. Linkage extends over long regions of the genome; two loci are linked if, during meiosis, recombination occurs between the two loci with a probability of less than 50%, that is, they are transmitted together from parent to offspring more often than expected under independent inheritance [3]. Linkage analysis is often the first stage in the genetic investigation of a trait, since it can be used to identify broad genomic regions that might contain the disease gene.

Parametric (model-based) linkage analysis is the analysis routinely used to examine the co-segregation of genetic loci in pedigrees. Loci that are in close proximity on the same chromosome segregate together more often than do loci on different chromosomes, which segregate together purely by chance alone. Each genotype for an individual for any genetic marker is made up of two alleles; one inherited from each parent. Specific single alleles are in haploid in gametic phase when they are inherited from the same parent; they are transmitted together in the single gamete originating from that parent. The further apart the two loci are on the same chromosome, the more likely it is that a recombination event at meiosis will break up the co-segregation. In parametric linkage analysis the main quantity of interest is the recombination fraction θ (the probability of recombination between two loci at meiosis). By genotyping polymorphic markers (usually microsatellites) and studying their segregation through pedigrees, it is possible to infer their relative position to each other on the genome and this is used to map the position of disease trait loci.

Linkage is reported as a logarithm of the odds (LOD) score [4] and is a function of the recombination fraction (θ). Large positive scores are evidence for linkage and negative scores are evidence against it. To calculate a LOD score in parametric linkage analysis, a model for disease expression must be specified. This accompanying model includes the frequency of the disease causing allele, the mode of its inheritance (dominant or recessive), analysed marker allele frequencies, the penetrance (expressed as a percentage) and the full marker make up of each chromosome. The ultimate aim of the analysis is to estimate the recombination fraction between individual typed markers and the unknown disease locus (two-point) or position of the disease locus relative to a fixed map of markers (multipoint). Traditionally, a LOD score of 3 or above is regarded as significant evidence of linkage, equivalent to 1:1000 = p, 0.0001 [5].

Once a region of the genome has been linked to a particular Mendelian trait, crossover events and fine haplotype analysis in the pedigree can be used to narrow down the region sufficiently to define a small interval for close scrutiny such that the underlying locus associated with the trait or disease can be identified. For example, dominant, missense and splice-site mutations in *MAPT* were identified in familial frontotemporal dementia with linkage to chromosome 17 (17q21) [6].

1.2.3 Genetic mapping of complex disease genes

1.2.3.1 Perspective

The genetic epidemiology of common disorders has proven much more challenging to study, since it is thought that they are caused by a combined effect of many different susceptibility DNA variants of low penetrance interacting with environmental factors [1]. Studies of common diseases fall into two categories: linkage studies across the entire genome and population-based association studies of candidate genes. Although there has been some notable success, progress has been slow as both methods have limitations; linkage analysis has low power when a trait locus contributes only a small fraction to the disease and in association studies of

single candidate genes, usually only a small fraction of the genetic variation at the gene locus is tested for association [7].

A popular hypothesis about genetic variation and risk of common, complex diseases is that they are caused by disease loci where there is one common variant, or a small collection of several predisposing to the disease; the 'common disease-common variant' (CDCV) hypothesis. If the CDCV hypothesis holds true then association mapping should be a powerful tool for detecting complex disease loci of small effect [8]. Recent work bears out that this hypothesis and at least in some cases holds true, for example the *APOE*-ε4 allele in Alzheimer's disease [9, 10] and *PPARG* (peroxisome proliferative activated receptor, gamma) in type 2 diabetes [11] among several well documented examples.

1.2.3.2 Linkage analysis in complex diseases

For complex diseases in which several genes might contribute to the disease risk, and where there is no clear mode of inheritance (either dominant or recessive), methods to identify linkage in complex diseases have been developed such that they do not require any prior specification of a disease model. These methods are referred to as non-parametric, or model free. The rationale behind this analysis is that, between affected relatives, an excess of haplotype sharing identical by descent (IBD) in the region of a disease causing gene would be expected, irrespective of the unknown mode of inheritance. Thus it is simply a test to identify if there is haplotype sharing IBD greater than expected under the null hypothesis of no linkage.

An example and routinely used non-parametric linkage analysis is to study sibling pairs. In this instance, both individuals are affected with the disease or trait of interest. Linkage would be suggested if the pairs of siblings, both of whom are affected by the disease, share significantly more haplotypes IBD than expected by chance alone. Linkage analysis of complex disease can only identify large regions (typically tens of cM) and although a very strong candidate gene might exist within the linkage region, such regions often contain hundreds of genes, many of which could be biologically plausible candidates. Several whole genome screens in a wide

range of complex diseases have been done using non-parametric linkage analysis. These include type 1 diabetes that successfully identified linkage to the HLA region [12] and for high-density lipoprotein-cholesterol, with evidence for linkage to a locus on chromosome 9 [13]. One study of 1,292 sibling pairs with late onset Alzheimer's disease identified a linkage peak on chromosome 10 of 16 cM in size [14]. However, to this day the underlying genetic lesion has not been identified.

Gene discovery in complex human disease by linkage has been complicated by substantial aetiological heterogeneity, the possibility of genes with small effect and the concomitant requirement for large sample cohorts. The mapping of human susceptibility loci for such diseases is also made difficult by incomplete penetrance, phenocopies and possible epistasis (a combined effect of one or more loci) and as a consequence, linkage approaches to complex diseases without Mendelian inheritance has had very limited success thus far [15].

1.2.3.3 Population-based genetic association studies

The rationale of genetic association studies is to detect association between one or more genetic polymorphism(s) and a trait, which could either be a quantitative characteristic or a discrete attribute or disease. Association differs from linkage in that the same allele(s) is associated with the trait in the same manner across the whole population, whereas linkage allows different alleles to be associated with the trait in different families. Association studies identify polymorphisms in which an allele occurring in the general population occurs at a different frequency in the disease group. In these instances, the disease associated allele does not cause the disease in the same way that a Mendelian mutation does but increases susceptibility to the disease as a genetic risk factor, most likely in conjunction with other genetic and environmental risk factors. Such identified variants have relatively low penetrance compared to variants causing monogenic Mendelian disease.

Association studies can either be direct or indirect. In direct association studies target polymorphisms which are themselves putative functional variants (for example a SNP variant in a gene at a codon that changes an amino acid) are

genotyped in both the general (control) and also trait (disease) population. A statistically different frequency of the alleles and/or genotypes in the control population versus the disease group would suggest that the polymorphism in question has a direct effect on disease pathogenesis. However, it is likely that many causal variants contributing to complex disorders will be non-coding. These variants could include those that affect gene regulation, expression or alternative splicing and such functional variants are difficult to predict. For this reason, most association studies are indirect; where the polymorphisms genotyped in the control and trait populations are surrogates for the unknown causal locus.

Identifying susceptibility genes for complex disorders by the indirect method depends on the existence of an association between the causal variants and surrounding polymorphisms nearby. This association is termed linkage disequilibrium (LD) and is defined as the non-random association of alleles at two or more loci and describes a situation in which some combinations of alleles of genetic markers occur more or less frequently in a population than would be expected from a random formation of haplotypes from the alleles.

Various methods of marker pairwise LD measures have been proposed [16] that are usually based upon Lewontin's D' [17]; this is the association probability. A probability D' value of 0.0 between two markers suggests independent allele assortment, whereas 1.0 means that all copies of the rarer allele occurs exclusively with one of the alleles at the other marker. D' is an important measure for the identification of regions of the genome in which there has been little recombination thus having the potential for mapping causal loci by indirect association studies. This LD measure, however, cannot determine the power of tests for indirect association studies. The latter depends on the LD measure of r^2, the square of the correlation coefficient. Even when loci are in complete linkage disequilibrium (D'=1.0), the pairwise r^2 values can vary widely because the allele frequency of each locus is also taken to account. For perfect r^2 LD (r^2=1.0), the allele frequencies at each loci must be the same.

The nature between r^2 and the power to detect association is such that, if locus A is causal then a proportional sample size increase of $1/ r^2$ would be required to detect the genetic association of locus A by the indirect association of locus B, with r^2 being the pairwise LD value between locus A and B [18].

1.2.3.4 Population genetic association study design

The first step in a case-control association study is to find a plausible candidate gene or genomic interval to test for variants associated with the trait of interest. Good candidate genes can be identified when prior genetic data exists, for example genes residing in proximity to a region of a chromosome that has been previously identified thorough linkage studies. Alternatively a link between a trait and gene can be established through biological data, for example the genes encoding ion channels may influence sporadic epilepsy because ion-channel mutations cause familial epilepsy and antiepileptic drugs target such ion channels [19] or a link between a pathological trait and a gene.

The second step in the study design is to select appropriate case and control samples to test for association variants in the gene or genomic interval of interest. The control samples should consist of random, unrelated individuals representative of the population under study. The controls should be drawn from the same population as the cases with the particular biological trait or disease and the two groups should be age and gender matched as closely as possible [20]. In terms of sample size, the more the better; larger sample sizes provide greater statistical power. The key determinant of quality in an association study is sample size [21]. Sample sizes can vary widely from study to study depending on the availability of samples but typically range from upwards of 50 samples per study group to more than a thousand.

An important measure of sample size in any association study is power. The power of a study is the statistical probability that the study can detect a true association if one is present. Power calculations are based upon the variables of sample size, the prevalence and effect of the risk variant and the threshold of significance. For

example, 500 cases and 500 controls would be required to detect an effect of an odds ratio of 1.5 of a susceptibility variant at a frequency of 0.2 (in the control population) at 80% power. Susceptibility variants of low frequency (<10%) and that also have low relative risks are the most difficult to identify because sample sizes in the thousands are required for sufficient study power and as such rare variants with low relative risks are largely beyond the reach of genetic epidemiology (**Table 1.1**). Susceptibility variants that are most easy to find with a modest number of cases and controls are those with a frequency in the general population close to 0.5 and that have a high relative risk.

Allelic odds ratio	1%	10%	40%
1.1	221,927	24,626	9,505
1.2	58,177	6,509	2,581
1.3	27,055	3,051	1,240
1.5	10,603	1,213	516
2.0	3,193	377	177
4.0	598	78	47

Table 1.1 Power and sample size

Approximate sample sizes necessary to detect significant association (power=90%, two sided significance threshold α=0.001) by effect size and allele frequency for predisposing allele; adapted from Hattersley and McCarthy, 2005 [20].

The third step is to genotype markers (typically SNPs) from the gene or region of interest in the case and control samples and strategies for selection of markers based on LD mapping are described in Chapter 4. Statistical methods for analyzing the population data are described in detail in Chapter 2, and the relevant results chapters. Briefly, this involves statistical tests (usually in a chi-squared distribution) of association by comparing the allele/genotype/haplotype frequency between the case and control populations.

1.2.3.5 Population stratification

Population stratification bias occurs when a case-control study sample contains subgroups with allele frequency differences. This is most commonly due to a mix of different ethnic groups in one of the sample sets and can result in false positive study results. The false positive (or indeed false negative) claims could arise if one

particular ethnic group is over represented in the disease group and has a higher incidence of the variant. Thus the variant could be found to be associated with the disease even if it does not influence it [22] and so care should be taken to select ethnically matched samples to protect against population stratification. There are formal methods to measure covert population stratification, one such method is to genotype multiple unlinked marker polymorphisms across the genome under the presumption that these are independent of the disease state and therefore can detect and correct for potential differences in the genetic make-up of the case and control groups [23].

1.2.3.6 The International HapMap project

This large and ambitions project (http://www.hapmap.org/) aims to construct genome-wide maps of LD patterns in samples collected in the USA (Caucasians of western European origin), Nigeria, China and Japan at a density of at least 1 SNP per kb for public release. The main aims of this project are to facilitate genetic mapping studies across a broad array of complex phenotypes for use in candidate gene case-control studies, and to identify sets of SNPs that take advantage of the LD patterns of the genome and allow more economical genotyping though indirect association studies [24]. For the purposes of candidate gene study association studies, HapMap project data can be analysed to identify haplotype-tagging SNPs for more efficient and economical genotyping in case-control cohorts.

1.2.3.7 The *APOE* genotype and the risk of Alzheimer's disease

The association of *APOE* with AD serves as a paradigm for the potential for identifying risk variants that contribute to common, complex diseases by association study. In 1993 a series of papers were published suggesting that the ε4 allele of the gene encoding apolipoprotein E (*APOE*) was a major risk factor for AD [10, 25-27]. The three major human isoforms, apoE2, apoE3 and apoE4 encoded by different alleles (ε2, ε3 and ε4) regulate lipid metabolism and redistribution. ApoE contains a 22 kDa N-terminal domain (residues 1–191) that is recognized by LDL (low density lipoprotein)- receptors and a 10 kDa C-terminal domain (residues 222–299) that has

high affinity for lipid and is responsible for the association of apoE with lipoproteins [28]. The three major isoforms of apoE that exist in the population differ by cysteine and arginine at residues 112 and 158. The most common isoform, apoE3, contains cysteine and arginine at these positions, respectively, whereas apoE2 contains cysteine at both positions and apoE4 contains arginine at both positions [29] and apoE3, the most common allele represents ~78% of all chromosomes; apoE4, 15% and apoE2, 7% in Caucasians. In AD the frequency of the apoE4 allele rises to ~40%.

It was shown that the estimated onset of disease is shifted to considerably younger ages in AD cases with ε4/ε4 (50% by age of onset at 66 y/o) compared to ε3/ε4 (50% by 73 y/o) or the ε3/ε3 (50% by 86 y/o), and those carries without an ε4 allele and with an ε2 allele have an even later age of onset [25]. In a complementary paper, clinical data was presented regarding the association of the ε4 and suggested that 50% of AD is associated with the ε4 allele. Furthermore, if the relative rarity of the disease among ε2 homozygotes is also considered, then the apoE genotype may be seen as being responsible for as much as 95% of AD and these findings have also been consistently replicated in multiple ethnic groups [30]. The precise mechanism of its contribution to the molecular pathogenesis is not well understood, however, apoe4 disrupts memory function in rodents [31] and further studies have indicated that fragments of apoE may contribute to both plaque and tangle formation [32], the pathology associated with the AD brain.

1.3 Structural variation in the human genome

1.3.1 Copy number polymorphisms and large scale structural variation

Genetic variation from individual to individual in the genome is present in many forms, including SNPs, small insertion-deletion polymorphisms, variable lengths of repetitive sequences and genomic structural alterations. Decades before the availability of DNA sequencing technology, the first observed structural alterations were rare and usually (though not always) associated with a disease. These included aneuploidies (the presence of an abnormal number of chromosomes),

rearrangements, heteromorphisms and fragile sites. These are termed microscopic structural variants since they are typically greater than 3 Mb in size and are large enough to be seen under the microscope [33]. Subsequently, with the advent of DNA sequencing technology, smaller and more abundant alterations were observed that include SNPs, micro- and minisatellites, and small (typically less that 1 kb) insertions and deletions. Significant progress has been made in characterising human genetic variation at the karyotype level [34] and the nucleotide level [24] but knowledge about genetic variation between these two extremes has, until recently, been less extensive because of previous limitations in the ability to determine these changes at the required resolution.

The completion of the sequence of the human genome has led to the creation of new methods for the study of the structural variation of the genome. In particular, genome-wide scanning array technologies [35, 36] and comparative DNA-sequence analysis [37, 38] are beginning to reveal variations that are larger than those recognised by conventional DNA sequencing and that are submicroscopic. Typically these structural variants range from 1 kb to 3 Mb in size [39] and are comprised of deletions and duplications collectively known as copy-number variants (CNVs) or polymorphisms (CNPs). Other structural variants include inversions (a segment of DNA that is reversed in orientation with respect to the rest of the chromosome) and translocations (a change in the position of a chromosome segment that involves no change to the total DNA content), see **Figure 1.1**.

Currently, array-based comparative genomic hybridisation (array-CGH) approaches [40, 41] provide the most promising methods for carrying out genome wide-scans to identify new CNVs. This method uses labelled fragments of the genomic DNA samples of interest which are hybridised alongside a second differentially labelled reference genomic DNA sample to arrays that are spotted with DNA clone fragments (for example, BACs, fosmids or oligonucleotides). The labels are typically the fluorescent dyes Cy5 and Cy3, green and red, respectively. The use of array-CGH is particularly useful, owing to the extensive high-resolution coverage of the genome it provides and it cross references with genome assemblies for ease of confirmatory FISH experiments. The use of arrays that comprise oligonucleotides

(typically 60-100 bp) can improve the detection resolution over that of BACs and is implemented in as assay format known as representational oligonucleotide microarray analysis [42], or ROMA. The principle of ROMA is similar to the applied use of array-CGH, but with an increased signal-to-noise ratio using a method called representation or whole genome sampling [43]. Here the target DNA is first treated by restriction digestion, ligated to adapters and then PCR-based amplification ensures labelled fragments are in a specific size range (<1.2 kb). The DNA fragments make up a fraction of the entire genomic sequence (~2.5%), which leads to the reduction in background noise.

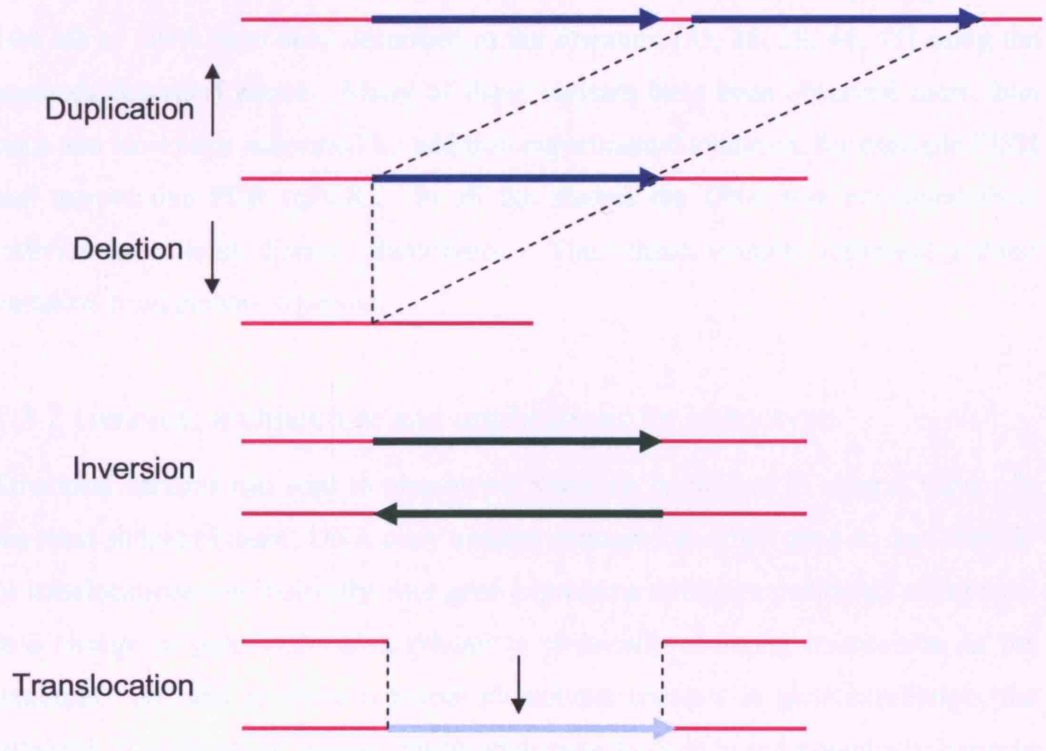

Figure 1.1 Structural polymorphisms in the human genome

Shown is a representation of a duplication (dark blue arrows) and a deletion that both lead to DNA copy number changes. Also shown are two balanced variants; an inversion and a translocation. The inverted segment, represented by red and green arrows, is identical but the orientation is back-to-front. The translocation is represented by a light blue arrow. Simply, translocations are genomic segments that are transferred to another position of the genome. Balanced variants do not lead to DNA copy number changes and are not readily detected by the either array-CGH or ROMA.

Structural variants can also be identified *in silico* by comparing DNA sequences from different sources, for example, aligning two unique human DNA assemblies. One advantage of this method over the array methods is that all types of variant can be detected, in particular balanced variants (for example, inversions and translocations). Balanced variants are not readily identified by array-CGH or ROMA since these array methods rely on the detection of DNA copy number changes; in balanced variants there is no net loss or gain of genetic material.

Before 2004, only a relatively few defined and non-disease-associated submicroscopic structural variants had been documented in the human genome. Since then, more than 600 submicroscopic structural variants, comprising at least 104 Mb of DNA have been described in the literature [35, 36, 38, 44, 45] using the methods described above. Many of these variants have been observed more than once and have been supported by addition experimental evidence, for example FISH and quantitative PCR (qPCR). In all the studies the DNA was examined from individuals without disease phenotypes. Thus these variants represent natural variation from person to person.

1.3.2 Genomic architecture and implications for phenotype

Structural variants can lead to phenotypic variation or disease in several ways. In the most simple of cases, DNA copy number changes can affect gene dosage directly or translocations can indirectly alter gene expression through a positional effect (that is a change in gene expression related to physically changing its position in the genome). As well as these potential phenotypic changes in gene expression, the presence of a structural variant might predispose to further and potentially harmful structural changes [46].

DNA sequence analysis from the human genome project has revealed that about 5-10% of the human genome is comprised of region-specific highly homologous (> 1 kb in size; >95% homology) low-copy repeats (LCRs, [47]). These LCRs appear to have arisen recently during primate speciation via paralogous segmental duplication, thus making the human species particularly susceptible to genomic rearrangements

[48]. Typically, LCRs range in size from a few to several 100 kb in size and can be present in either tandem (direct) or inverted orientation. Molecular studies of recombination breakpoints in genomic disorders have revealed the presence of LCRs flanking the recurrently recombined genomic segments [47]. These findings suggest that the rearrangement might be mediated by non-allelic homologous recombination (NAHR) between these highly homologous LCR substrates. In NAHR, an unequal crossover between two homologous syntenic but non-allelic LCRs results in a loss, gain or inversion of the unique intervening genomic segment that is flanked by the LCRs [49].

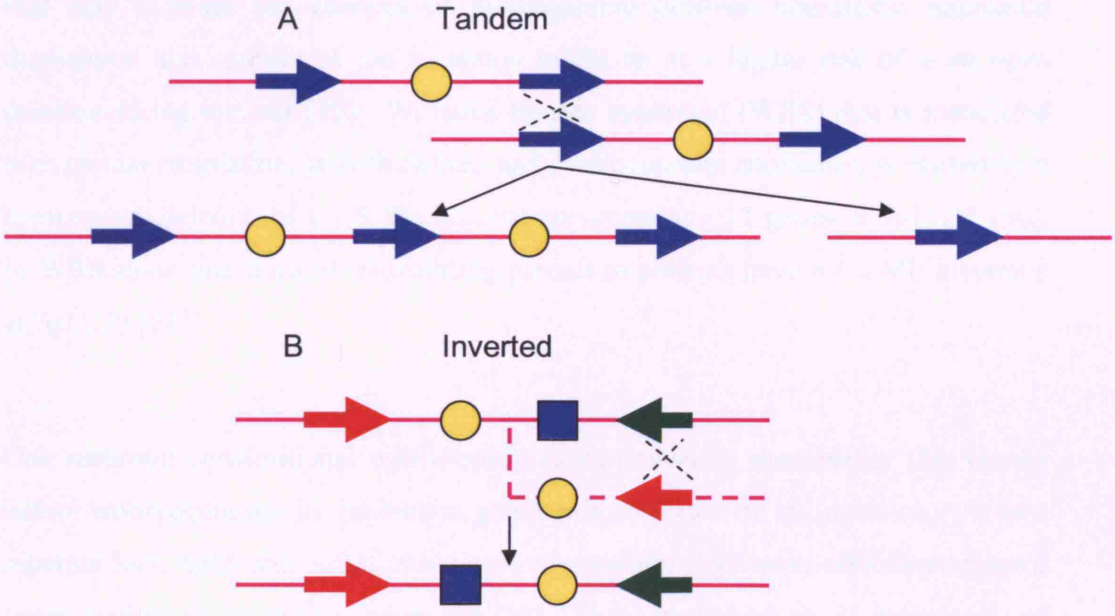

Figure 1.2 Mitotic non-allelic homologous recombination

Schematic representation of molecular mechanisms for meiotic NAHR between LCRs. (A) Crossover between two paralogous LCRs in direct orientation results in reciprocal duplication or a deletion of a unique genomic segment (yellow circle). (B) Recombination between inverted LCRs (red and green arrows) results in an inversion of unique genomic segment (yellow circle and blue square), but no loss or gain of the genomic segment is observed. Figure adapted from Inoue [49].

Each orientation of LCRs yields distinct recombination products. Utilising directly orientated LCRs as substrates in unequal crossing over either between two chromosomes, within a chromosome (intrachromosomal) or within a chromatid

(intrachromatid) results in deletion and duplication of the genomic segment that is flanked by the recombined LCR units. In these instances, the genes within the region are not disrupted but instead, the rearrangement either increases (duplication, triplication) or decreases (deletion, monosomy) the DNA copy number (**Figure 1.2**). Two inversely orientated LCR units aligned together as recombination substrates between two chromosomes results in an inversion of the genomic fragment (**Figure 1.2**). The inversion rearrangement does not alter DNA copy number but it could disrupt a gene that spans the recombination breakpoints.

In the case of polymorphic inversions, there is usually a suppression of recombination resulting from the different orientation of the chromosome segments that may increase the chances of misalignment between non-allelic segmental duplication and carriers of the inversion might be at a higher risk of a *de novo* deletion during meiosis [49]. Williams-Beuren syndrome (WBS) that is associated with mental retardation, growth defects and cardiovascular anomalies, is caused by a hemizygous deletion of a 1.5 Mb interval encompassing 17 genes at 7q11.32 [50]. In WBS about one third of transmitting parents to patients have a 1.5 Mb inversion at 7q11.23 [51].

One recurrent constitutional translocation (a chromosome abnormality that occurs before embryogenesis) in the human genome is mediated by an inversion. At two separate loci, 4p16 and 8p23, clusters of segmentally duplicated olfactory-receptor genes mediate polymorphic inversions [52]. These inversions are at frequencies of 12.5% and 26%, respectively, and are involved in the recurrent t(4;8)(p16;p23) translocation. The transmitting parents of the individuals with the translocations were shown to be heterozygous carriers of both the inversions. The parents show no associated phenotype but the offspring who carry the translocations show phenotypes ranging from Wolf-Hirschhorn syndrome [53] to a milder phenotype of dysmorphic features. These examples demonstrate the importance of identifying the inversion variants that occur in the general population, as they may lead to an increased risk for further disease-causing structural variation to occur.

1.4 Microtubule associated protein tau

The tau protein was first identified as a "factor essential for microtubule assembly"; a heat stable protein that induced the assembly of microtubules from purified tubulin [54]. Tau is a member of the microtubule associated protein family (MAPs) and the main function of these proteins is to stabilise microtubules and to promote microtubule assembly [55]. MAPs are essential for the formation of cytoskeletal architecture. However, they have additional roles on the regulation of organelle transport by their interaction with motor proteins on the microtubules [56]. Tau is abundantly expressed in both the peripheral and central nervous system, where it is enriched in the axons of mature and growing neurones. Low levels of tau are also present in oligodendrocytes and astrocytes [57, 58].

The human microtubule associated protein tau gene (*MAPT*) is located on the long arm of chromosome 17 at band 21 (17q21.31) and spans approximately 150,000 base-pairs (bp) of nucleotide sequence. *MAPT* consists of one non coding, exon -1, followed by 14 coding exons [59]. Intron 13 is retained in human *MAPT* mRNAs, with exon 13 (the final coding exon), intron 13 and exon 14 together constituting a single, last exon comprising the 3' untranslated region (UTR). In the adult human brain, the tau protein exists as six major isoforms produced by interplay of alternatively splicing exons 2, 3 and 10 [60]. The six tau isoforms are named on the basis of the combinations of the number of N-terminal inserts and by the number of C-terminal microtubule binding repeats: 0N3R, 1N3R, 2N3R, 0N4R, 1N4R, 2N4R (**Figure 1.3**). Of the alternative splicing of exons 2 and 3, exclusion of both, inclusion of only exon 3 and the inclusion of both exons 2 and 3 are only possible as inclusion of only exon 2 never occurs. Alternative splicing of exons 2 and 3 result in three isoforms with zero (0N), one (1N) or two (2N) N-terminal inserts, respectively. These N-terminal inserts are part of the projection domain mediating the interaction of microtubules with the plasma membrane [61]. Tau interacts with microtubules through imperfect repeat domains located in the C-terminus portion of the protein, encoded by exons 9-13 [62-64]. The alternative splicing of exon 10 results in either three repeats (3R tau) or four repeat (4R tau) C-terminal domains.

The alternative RNA splicing of *MAPT* is developmentally regulated. In foetal brain, only the transcript encoding the shortest tau isoform with three repeats is expressed in contrast to the adult human brain where six isoforms are expressed [65].

In addition to the six major tau brain isoforms, more *MAPT* transcripts have been identified due to the alternative splicing of exons 4a, 6 and 8. Exon 4a is a large exon and is present in "big tau", so nicknamed, and is found in the PNS [66]. The expression patterns of tau associated with exons 6 and 8 are not well understood. *MAPT* transcripts containing exon 8 have not been observed in humans but have been identified in rhesus monkey and the cow [67, 68].

Tau is a phosphoprotein and phosphorylation is a major post-translational modification of tau at specific serine and threonine residues. At least 25 different phosphorylation sites have been identified. Tau is also developmentally regulated post-translationally because foetal tau is more phosphorylated than adult brain tau [69]. Other post-translational modifications of tau include ubiquitination, glycation and transglutamination. The biological activity of tau in promoting assembly and stability of microtubules is regulated by its degree of phosphorylation. Normal tau contains 2-3 mol phosphate/mol of the protein [70], the level of phosphorylation for its optimal activity. Hyperphosphorylation of tau depresses its microtubule assembly activity and it's binding to microtubules [71, 72].

Hyperphosphorylated tau in insoluble filaments contained in the brain is the pathological hallmark of AD and related neurodegenerative disorders classed as tauopathies [73].

1.5 The tauopathies

The tauopathies are a group of neurodegenerative disorders that are characterised pathologically by the presence of fibrillar aggregates of tau in the brain [62], **Table 1.2**. The most common tauopathy, AD, is characterised clinically by a progressive loss of verbal and visual memory and intellectual function, resulting in severe

dementia. The cognitive decline in AD has been correlated with various bio-markers that include the loss of choline acetyl transferase and synaptophysin reactivity. In addition to abundant extracellular amyloid (Aβ) deposits, the senile plaques' (SP) hyperphosphorylated tau neurofibrillary tangles (NFTs) constitute the pathological lesions [74]. Aβ and NFTs also coexist in some other tauopathies like Down's syndrome [75, 76] however NFTs occur alone in argyrophilic grain disease (AGD) [77], Pick's disease (PiD) [78], corticobasal degeneration (CBD), progressive supranuclear palsy (PSP) [79], familial frontotemporal dementia with parkinsonism linked to chromosome 17 with tau pathology (FTDP-17T) [80], amyotrophic lateral sclerosis (ALS) [81], Niemann-Pick disease type C (NPC) [82] and subacute sclerosing panencephalitis (SSPE). These disorders are classified as primary tauopathies, since pathological aggregates of neurofibrillary tau are their main defining characteristic, (**Table 1.2**). AD is a secondary tauopathy since it is defined not only by aggregates of tau but also by extracellular amyloid deposits.

The Tauopathies
Alzheimer's disease
ALS/parkinsonism-dementia complex
Argyrophilic grain disease
Corticobasal degeneration
Creutzfeld-Jakob disease
Dementia pugilistica
Diffuse neurofibrillary tangles with calcification
Down's syndrome
FTDP-17T
Gerstmann-Staussler-Scheinker disease
Hallervorden-Spatz disease
Myotonic dystrophy
Niemann-Pick disease
Non-Guamanian motor neuron disease with tangles
Pick's disease
Postencephalitic parkinsonism
Prion Protein cerebral amyloid angiopathy
Progreessive subcortical gliosis
Progressive supranuclear palsy
Supacute sclerosing panencephalitis
Tangle only dementia

Table 1.2 The tauopathies

Primary tauopathies are shaded blue; secondary tauopathies are not.

AD and other tauopathies like AGD and PiD are clinically characterised by dementia, while CBD, PSP and post-encephalitic parkinsonism present with motor handicaps. However, CBD can also present with cognitive deficits or aphasia (speech impairment) and in PSP patients behavioural changes and a dysexecution of syndrome may be the most prominent symptoms. Owing to the substantial clinical overlap among various neurodegenerative disorders with tau pathology definite diagnosis still requires neuropathological examination.

Neurofibrillary lesions of filamentous tau form within nerve cells that eventually degenerate and it appears that they die. These lesions are found in nerve cell bodies and apical dendrites as neurofibrillary tangles (NFTs) and in distal dendrites as neuropil threads (NTs). Ultrastructurally, these lesions consist of paired helical filaments (PHFs) and straight filaments [62]. The tau inclusions in the different tauopathies have characteristic morphology and distributions.

The pathological tau filaments are insoluble but can be isolated for biochemical analysis from fractions of brain homogenates that are insoluble in the detergent sarkosyl [83]. Thus in addition to distinct distribution and morphology, tauopathies can also be classified according to the biochemical composition of tau in the respective inclusions. The electrophoretic analysis of the insoluble tau from the different tauopathies shows a banding pattern reflecting the composition of the hyperphosphorylated tau isoforms present in the inclusions. This banding pattern can be divided into three general categories depending on the presence of four bands at 60, 64, 68 and 72 kDa that represent hyperphosphorylated tau isoforms [84], these being predominantly 4R tau pathology (e.g. PSP), mixed 3R/4R tau pathology (e.g. AD) and predominantly 3R tau pathology (e.g PiD).

1.6 Frontotemporal lobar degeneration

1.6.1 Epidemiology

Frontotemporal lobar degeneration (FTLD), more commonly referred to as frontotemporal dementia (FTD), describes a clinically and pathologically

heterogeneous group of dementias that have an onset of illness usually between the ages of 35 and 75 [85]. FTD is the second most common cause of presenile dementia (after AD), accounting for up to 20% of dementia cases [86]. FTD results from severe neuronal degeneration of frontal and temporal brain regions and results in abnormalities in behaviour, personality disturbances including disinhibition and may also be associated with motor dysfunction, including both motor neuron disease and parkinsonism [87]. Survival of pathologically proven FTD is, on average, only 4 years from symptom onset [88]. The molecular pathology of FTD has been divided into at least three groups; a tauopathy, with neurofibrillary tangles; ALS-like, with ubiquitinated and neurofilamentous inclusions and the remainder lacking distinctive histopathological features [89]. In approximately 40% of FTD cases, there is a family history of dementia, indicating a significant genetic contribution [90]. FTD is genetically heterogeneous, with associated loci identified on chromosome 17 [91-93], chromosome 9 [94] and chromosome 3 [95, 96].

1.6.2 FTDP-17T

1.6.2.1 Epidemiology

Frontotemporal dementia and parkinsonism linked to chromosome 17 with tau pathology (FTDP-17T), [97] is inherited as an autosomal dominant condition characterised by behavioural, cognitive, and motor disturbance with an onset usually between 45 to 65 years of age. At autopsy, patients with FTDP-17T display frontotemporal atrophy and intraneuronal tau inclusions are present on histological examination. The morphology and isoform composition of the tau filaments that compose the inclusions is widely variable. In 1998 missense and splice site mutations were identified in *MAPT* associated with FTDP-17T (**Figure 1.3**) thus establishing a firm link between tau dysfunction and this tauopathy [91, 98, 99].

1.6.2.2 The *MAPT* mutation spectrum

After the first identification of mutations in *MAPT*, more than 35 mutations have been identified in more than 100 families worldwide. The mutations affecting

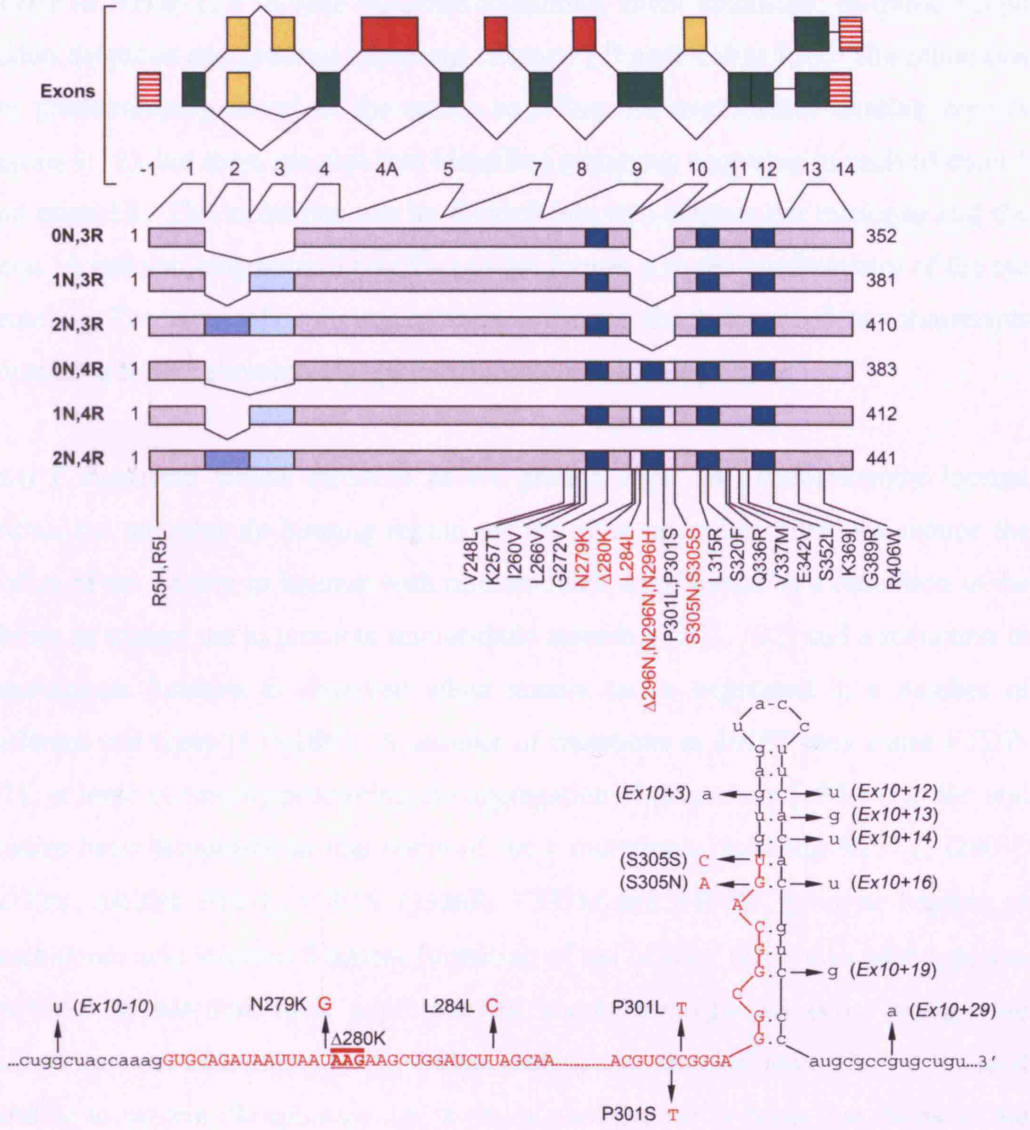

Figure 1.3 *MAPT* structure and the FTDP-17T mutation spectrum

(Upper) Tau in the central nervous system (CNS) exists as six isoforms due to the alternative splicing of exons 2, 3 and 10 (yellow boxes). Exons 4A, 7 and 8 (red boxes) are absent in the CNS, exon 4A is included in peripheral nervous system 'big' tau. Exons 2 and 3 code for amino-terminal inserts, alternative splicing leads to tau isoforms with 2, 1 or no amino-terminal inserts (2N, 1N or 0N). Exon 10 codes for one of four microtubule binding domains – alternative splicing results in tau with 3 or 4 microtubule binding repeat domains (3R, 4R). FTDP1-7 missense and silent mutations and deletions are indicated with numbering relating to the longest 441 residue 2N,4R isoform. Mutations in red affect the alternative splicing of exon 10. Modified from Goedert [100]. (Lower) FTDP-17 mutations affecting the 3' splice donor site of exon 10. The majority of these mutations disrupt a predicted pre-mRNA stem-loop structure, inducing increased incorporation of exon 10. Partial sequence of 3'-end of exon 10 in red. Intronic sequence in black.

MAPT in FTDP-17T include missense mutations, silent mutations, in-frame single codon deletions and intronic mutations (**Figure 1.3 and Table 1.3**). The mutations are predominantly found in the exons encoding the microtubule binding repeats (exons 9-12), but there are also two identified mutations occurring in each of exon 1 and exon 13. The mutations can be divided into two classes: the missense and the exon 10 splicing mutations. The effect of the former is in the biochemistry of the tau protein. The exon 10 splicing mutations disrupt the balance of tau transcripts containing the 4^{th} alternatively spliced microtubule binding repeat.

MAPT mutations whose effect is at the protein level are predominantly located within the microtubule-binding region of tau, most missense mutations reduce the ability of tau protein to interact with microtubules, as reflected by a reduction in the ability of mutant tau to promote microtubule assembly [101, 102] and a reduction in microtubule function is observed when mutant tau is expressed in a number of different cell types [103-106]. A number of mutations in *MAPT* may cause FTDP-17T, at lease in part, by promoting the aggregation of tau protein [107-111]. Several studies have demonstrated that some of these mutations, including K257T, I260V, G272V, ΔK280, P301L, P301S, Q336R, V337M and R406W, promote heparin or arachidonic acid induced filament formation of tau *in vitro* relative to wild-type tau protein. In addition, after expression in human neuroglioma cells, mutant tau assembles into filaments [112]. Some mutant tau proteins have shown reduced binding to protein phosphatase 2A, a major phosphatase in brain that binds to the repeat region of tau [113].

Approximately half of the known *MAPT* mutations have their primary effect at the RNA level (**Table 1.3**). The majority of the coding mutations in exon 10 (N279K, L284L, ΔN296, N296N, N296H, S305N and S305S) alter the splicing of exon 10, thus changing the ratio between three and four repeat tau isoforms [6, 98, 114-118]. For example, the two mutations N279K and ΔK280 have opposite effects on a poly-purine element that is a recognition sequence for a positive regulatory splicing factor. The N279K mutation strengthens this element (AA*U*AAGAAG to AAGAAGAAG), which results in an increase in the proportion of exon 10+ RNA

and soluble 4R tau [98]. The ΔK280 mutation eliminates the poly-purine tract (AAU*AA*GAAG to AAUAAG), resulting in the loss of exon 10 splicing in exon trapping assays [98]. The ΔK280 mutation is interesting because in addition to repressing exon 10 splicing *in vitro* it also disrupts tau-microtubule binding and is highly fibrillogenic [98]. However, since it eliminates exon 10 splicing, its disruption of microtubule binding will be irrelevant as ΔK280 tau is sparsely produced. On the other hand, since it is highly fibrillogenic, even small amounts of mutant tau produced could set a seed for tau aggregation. Another mutation in exon 10 that alters alternative splicing is a silent change in codon L284 [98]. This mutation eliminates a negative cis-acting splice element (U*U*AG to U*C*AG) that lies adjacent to the poly-purine element altered by N279K and ΔK280, and causes an increase in the level of 4R tau [98].

The pathogenic mutations that alter exon 10 splicing also occur in intron 10 at positions +3, +12, +13, +14, +16, +19 and +29 (**Table 1.3 and Figure 1.3**). Studies of the mechanisms of these splice site mutations by reverse transcriptase PCR (RT-PCR) in FTDP-17T brains revealed that these are associated with two- to six-fold increase in the ratio of exon 10+ to 10- tau mRNA [91, 117]. Exon trapping experiments were also used to demonstrate *in vitro* that the splice site mutations increased the incorporation of exon 10 into artificial transcripts [91]. Analysis of soluble tau from the brains of splice site mutation cases revealed an increase of 4 repeat isoforms [116]. Examination of the intronic sequence downstream of exon 10 revealed that each mutation (up to +19) was predicted to disrupt a potential stem-loop structure that was likely involved in regulating exon 10 alternative splicing by competing with the U1 snRNP for binding to the 5' splice site [91, 116], see **Figure 1.3**. Analysis of the secondary structure of *in vitro* transcribed RNA has demonstrated that a stable stem-loop can form at this position and is disrupted by the 5' splice site mutations [119]. The mutation S305N alters the -2 residue relative to the exon 10 5' splice site [98]. This residue is predicted to be part of the stem-loop that is disrupted by other 5' splice site mutations and, in addition, this mutation is predicted to increase directly U1 snRNP binding to the 5' splice site, and increases the splicing-in of exon 10 and the level of 4R-tau. A recent mutation in intron 9 was

identified occurring 10 bases 3' of exon 10 [120]. Exon trapping experiments revealed this mutation to increase the proportion of exon 10 containing transcripts and this suggests that sequences contained within intron 9, as well as intron 10, can influence the alternative splicing of exon 10. Thus these splice site mutations of exon 10 and flanking introns alter the proportion of exon 10+ RNA and 4R tau.

The identified FTDP-17T mutations have multiple effects on tau biology. About half of the known mutations have their primary effect at the protein level; they reduce the ability of tau protein to interact with microtubules and increase its ability to assemble into abnormal filaments. The other mutations have their primary effect on the RNA, perturbing the normal ratio of 3R to 4R tau isoforms by disrupting the alternative splicing of exon 10 and have revealed the presence of different *cis*-acting splice regulatory elements.

MAPT Mutation	Gene Location	RNA/Protein Location
RNA affecting		
Ex10+3 [116]	Intron 10	5' RNA Intron 10 splice site
Ex10+12 [121]	Intron 10	5' RNA Intron 10 splice site
Ex10+13 [91]	Intron 10	5' RNA Intron 10 splice site
Ex10+14 [91]	Intron 10	5' RNA Intron 10 splice site
Ex10+16 [91]	Intron 10	5' RNA Intron 10 splice site
Ex10+29 [122]	Intron 10	5' RNA Intron 10 splice site
S305S [123]	Exon 10	5' RNA Exon 10 splice site
S305N [124]	Exon 10	5' RNA Exon 10 splice site
L284L [98]	Exon 10	5' RNA Exon 10 splice site
Δ280K [125]	Exon 10	5' RNA Exon 10 splice site
N279K [126]	Exon 10	5' RNA Exon 10 splice site
Δ296N [127]	Exon 10	5' RNA Exon 10 splice site
N296N [128]	Exon 10	5' RNA Exon 10 splice site
N296H [118]	Exon 10	5' RNA Exon 10 splice site
Ex10-10 [120]	Intron 9	3' RNA Exon 10 splice site
Protein affecting		
R5H [129]	Exon 1	N-terminal
R5L [130]	Exon 1	N-terminal
K257T [131]	Exon 9	Pre-repeat 1
I260V [132]	Exon 9	Pre-repeat 1
L266V [133]	Exon 9	Repeat 1
G272V [91]	Exon 9	Repeat 1
P301L [91]	Exon 10	Repeat 2
P301S [134]	Exon 10	Repeat 2
L315R [135]	Exon 11	Repeat 3
S320F [90]	Exon 11	Repeat 3
Q336R [136]	Exon 12	Inter-repeat 3-4
V337M [137]	Exon 12	Inter-repeat 3-4
E342V [138]	Exon 12	Inter-repeat 3-4
S352L [139]	Exon 12	Inter-repeat 3-4
K3691 [140]	Exon 12	Repeat 4
G389R [141]	Exon 13	C-terminal
R406W [91]	Exon 13	C-terminal

Table 1.3 FTDP-17T MAPT mutations

Mutation numbering of the amino acids affected is according to the longest human isoform (2N4R)

1.6.2.3 Pathogenesis of FTDP-17T

FTDP-17T had provided great insight into the functional and phenotypic consequences of the various types of tau mutation and their pathological phenotype. Both increased 4R tau expression as a result of exon 10 splice site mutations and reduced microtubule binding capacity as a result of missense mutations affecting the repeat domains and surrounding regions could have the effect of abnormally increasing the availability of unbound tau. This in turn could promote aggregation of excess tau and the formation of insoluble tau fibrils [142]. However, the pathway leading from a mutation in *MAPT* to neurodegeneration is not well understood, it is currently not clear if tau accumulation directly causes neuronal death or if specific defects in aspects of tau function, e.g. microtubule binding, lead to the selective vulnerability of the neuronal populations in FTDP-17T. In AD, PSP, CBD and PiD where pathogenic tau mutations are unknown the underlying pathological lesions may be related to disruption in tau protein dynamics and metabolism.

1.6.3 FTDU-17

Since the identification of *MAPT* mutations in 1998, nine FTD families were conclusively linked to chromosome 17q21, a region that includes *MAPT*, but yet no *MAPT* mutations could be identified (tau negative-FTD-17), [92, 143]. In each family, affected patients lack tau inclusions but develop ubiquitin-immunoreactive pathology typical of frontotemporal dementia with ubiquitin inclusions (FTDU), [144]. The absence of obvious *MAPT* mutations in tau-negative FTDU-17 families suggested that this form of FTD is caused either by a novel *MAPT* mutation that has eluded detection or by mutations in a different gene in the same region. Extensive sequencing of *MAPT* intronic regions in two linked FTDU-17 families and analysis of soluble tau from patient brain extracts failed to identify any evidence of pathogenic *MAPT* mutations [143, 145]. Other candidate genes within the critical region defined by haplotype analysis in the families were examined and mutations were identified in the progranulin gene (*PGRN*), located 1.7 Mb telomeric of *MAPT* [146, 147]. *PGRN* is a growth factor involved in multiple physiological and pathological processes including tumorigenesis [148]. So far all of the mutations

identified result in the expression of null alleles and an overall reduction of progranulin protein [146, 147].

1.6.4 FTD9

Dominant ALS with FTD has been linked chromosome 9 [149]. The region is relatively large (12 cM), from 9p13.2-21.3 and is predicted to contain 103 known genes. The underlying genetic lesion has not yet been identified.

1.6.5 FTD3

In a large Danish pedigree with autosomal dominant FTD, significant linkage was identified on chromosome 3 (FTD3) [150, 151]. The brains of individuals with FTD3 have no distinctive neuropathological features. Onset of dementia is typically in the late 50s and is initially characterised by behavioural and personality changes including apathy, restlessness, disinhibition and hyperorality (the inappropriate placing of objects in the mouth), progressing to stereotyped behaviours, mutism and dystonia, with a marked lack of insight. In subsequent genetic analysis a missense mutation was identified in *CHMP2B* (Charged multivesicular body protein 2B), which encodes a componenet of the ESCRTIII complex [96]. FTD3 is a rare cause of FTD [152, 153].

1.7 Progressive supranuclear palsy

1.7.1 Clinicopathological features and epidemiology

Progressive supranuclear palsy (PSP; MIM 601104; Steele-Richardson-Olszewski syndrome) first described more than 40 years ago [154, 155], is a sporadic late life neurodegenerative disorder. It is the second most common form of degenerative parkinsonism and is characterised by early postural instability and backward falls, slowness of movements an akinetic-rigid syndrome (impairments of movement), vertical supranuclear-gaze palsy, pseudobulbar signs (involuntary problems relating to speech and swallowing) and cognitive decline of frontal lobe type [156].

Epidemiological studies show the prevalence of PSP in Europe, North America and Japan are comparable, in the range of 1.39 to 6.6 per 100,000 [157-159], about as common as motor neurone disease. However, there is widespread agreement that the prevalence of PSP is an underestimate through misdiagnosis.

PSP is pathologically characterised by abnormal accumulation of tau NFTs and NTs in a number of brain areas but particularly the midbrain and basal ganglia [160]. NFT pathology is greatest in the subcortical structures (structures below the cerebral cortex) but PSP is also associated with varying degrees of cortical tau pathology and the NFTs are found in both neurones and glia. Structurally, tau filaments that constitute tangles in PSP are straight. In contrast, the paired helical filament is the most abundant species associated with the tangles in Alzheimer's disease. In normal human brain, 3R and 4R tau occurs in approximately equal levels as determined by the alternative splicing of exon 10. In PSP this ratio is at least 3:1 in favour of 4R tau [161], thus PSP is a 4R tauopathy (**Figure 1.4**). In contrast, other tauopathies such as PiD, 3R tau predominates and so disordered regulation of exon 10 splicing could be the reason for NFT formation in PSP and other tauopathies.

Although tau dysregulation is implicated, the exact mechanism leading to cell death in PSP in unknown but is likely to be multifactorial, with both environmental and genetics influences having roles [162].

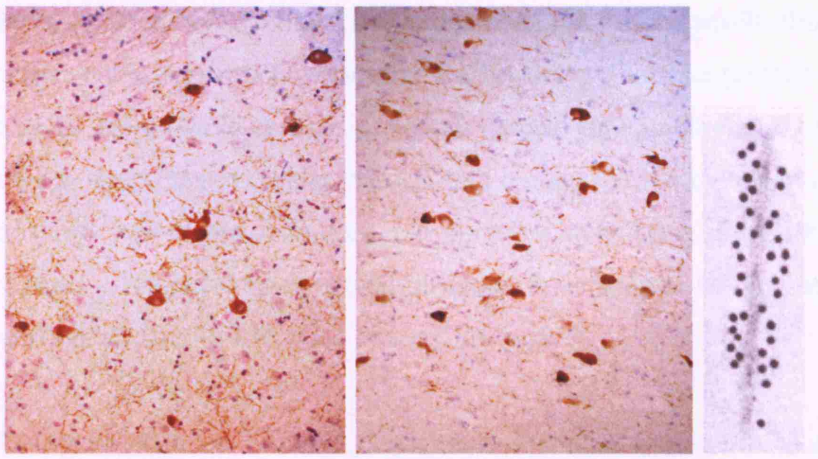

Figure 1.4 Tau pathology

The AT8 antibody is specific for all phosphorylated tau isoforms and the RD4 antibody is specific for 4-repeat tau and they demonstrate tau immunoreactivity in the griseum pontis in the PSP brain (x10). The AT100 antibody is specific for all phosphorylated tau isoforms and the image is an electron micrograph of a paired helical filament from the Alzheimer brain (immunogold labelled). The images are re-produced with permission from de Silva [163].

1.7.2 Genetics

1.7.2.1 *MAPT* mutations and PSP

The identification of missense and splice-site mutations in *MAPT* affirmed a central role of tau dysfunction in neurodegenerative aetiology [91]. Although the other related tauopathies, including AD, CBD and PSP are also defined by fibrillar tau pathology, *MAPT* is not mutated in these diseases. However there are several examples of *MAPT* mutations causing a clinical and pathological phenotype closely resembling that of PSP and in one instance CBD.

Genetic analysis of *MAPT* in a patient clinically diagnosed as having young onset PSP, with onset of symptoms at the age of 40, revealed this patient to have a tau exon 10 +16 (intronic) mutation. Neuropathological examination confirmed the genetic diagnosis of FTDP-17 [164].

A silent mutation in exon 10 *MAPT*, S305S was identified in a subject without significant atrophy or cellular degeneration of the frontal and temporal cortices [165,

166]. The cellular pathology was characteristic of PSP, with neurofibrillary tangles concentrating within subcortical regions of the basal ganglia. The mutation does not give rise to an amino acid change in the tau protein but functional exon trapping experiments revealed that it results in a 4.8-fold increase in the inclusion of exon 10 thus increasing transcripts encoding tau proteins containing four microtubule-binding repeats. The most likely mode of action is thought to be disruption of the stem-loop structure; the -1/+15 (U-A) base pairing [6, 91, 116].

A *MAPT* mutation in exon 1 causing the fifth amino acid arginine to be substituted for histidine, R5H, was identified in an 81-year-old-man whose elder brother had died at 86 years of age with dementia [129]. A microtubule assembly assay was performed with the mutant versus the wild-type histidine. Recombinant 3R tau and 4R tau with the substitution showed a reduced ability to promote microtubule assembly compared with wild-type tau protein. The effect of the R5H substitution on tau filament formation *in vitro* by use of recombinant tau protein was assessed quantitatively by electron microscopy. At 48 hours of incubation in the presence of heparin, the mutant 4-repeat tau yielded abundant straight or twisted tau filaments, whereas wild-type tau gave smaller amounts under the same conditions. Neuropathologically, widespread oligodendroglia-predominat tau deposits were observed, similar to that seen in PSP [129].

Sequence analysis of *MAPT* in a series of 96 subjects with clinically diagnosed progressive supranuclear palsy identified another missense mutation in intron 1 of the gene, R5L, substituting the amino acid arginine for leucine in 1 patient [130]. This mutation was not found in the other PSP subjects or indeed in 198 additional controls. On examination of the diseased brain, it revealed changes fulfilling the pathological criteria for PSP. The functional effect of the R5L mutation was examined using a mictotubule assembly assay and it was also found to alter the ability of tau to promote microtubule assembly [130].

In a Spanish kindred, two brothers born from a third degree consanguineous marriage were both affected with atypical PSP. A homozygous deletion at codon 296 (delN296) was identified in one of the affected siblings [127]. Among the

heterozygous carriers, two members with probable Parkinson's disease were identified, but none of the heterozygous individuals developed the atypical parkinsonism. The mutation itself lies in exon 10, situated in the sequence corresponding to the second tubulin-bindidng repeat of the tau protein and affects one asparagine residue absolutely conserved in other species. The identification of this homozygous deletion in this family indicates that dysfunction of two copies of the tau protein may be responsible for cases of familial atypical PSP, with an autosomal recessive mode of inheritance.

1.7.3.2 Common genetic variation at the *MAPT* locus associated with PSP

In 1997 the first evidence that there may be a genetic predisposition to sporadic PSP involving *MAPT* emerged [167] and in fact predated the discovery of *MAPT* as the underlying genetic lesion for FTDP-17T [91]. The genetic association involved a TG dinucleotide repeat marker located between exon 9 and 10 (within intron 9) of the gene. Specifically, the a_0 allele and a_0/a_0 genotype was significantly over-represented in patients with PSP compared to normal controls (**Table 1.4**). The a_0 allele is defined by the occurrence of 11 TG repeats, and the other alleles, a_1, a_2, a_3 and a_4 are defined the presence of 12, 13, 14 or 15 repeats, respectively. The a_0 is the most common allele in the controls (~70 % in Caucasian controls) and in PSP (~90%), with the remainder being made up by the other less frequent alleles. This finding was subsequently confirmed by four independent studies [168-171]. In the PSP series from the United Kingdom [170] the odds ratio for the a_0a_0 genotype was 4.3 (95% confidence interval 3.6 to 10.1) and in the Spanish PSP series [171] the association was particularly strong (odds ratio for a_0a_0 genotype, 10.62). The implications of these initial findings were either that this polymorphism itself contributes to an increased risk of PSP, or more likely, due to the nature and physical location of the polymorphism that it is in linkage disequilibrium (LD) with the actual causative polymorphism or variant.

The allelic association of *MAPT* with PSP was extended to a series of single nucleotide polymorphisms (SNPs) spanning the entire coding sequence of the gene. These eight SNPs, present within or close to the gene exons (1, 2, 3, 9, 11 and 13)

were identified during routine sequencing of *MAPT* exons in FTDP-17 [172]. Investigation revealed these SNPs to be in perfect linkage disequilibrium (LD) with one another (that is, the correlation of each SNP allele is 100%) and the intron 9 microsatellite. This region of LD was first shown to physically cover ~100 kb of DNA sequence from exon 1 to 13, thus forming two extended haplotypes that were designated H1 and H2 (**Figure 1.5**). Not a single recombination event was identified between any of these loci implying that these two haplotypes were established early in the history of the Caucasian population. The a_0 polymorphism that previously had been shown to be associated with PSP together with the alleles a1 and a2 segregated with the extended H1 haplotype whereas the a_3 and a_4 alleles are inherited along with the extended H2 haplotype.

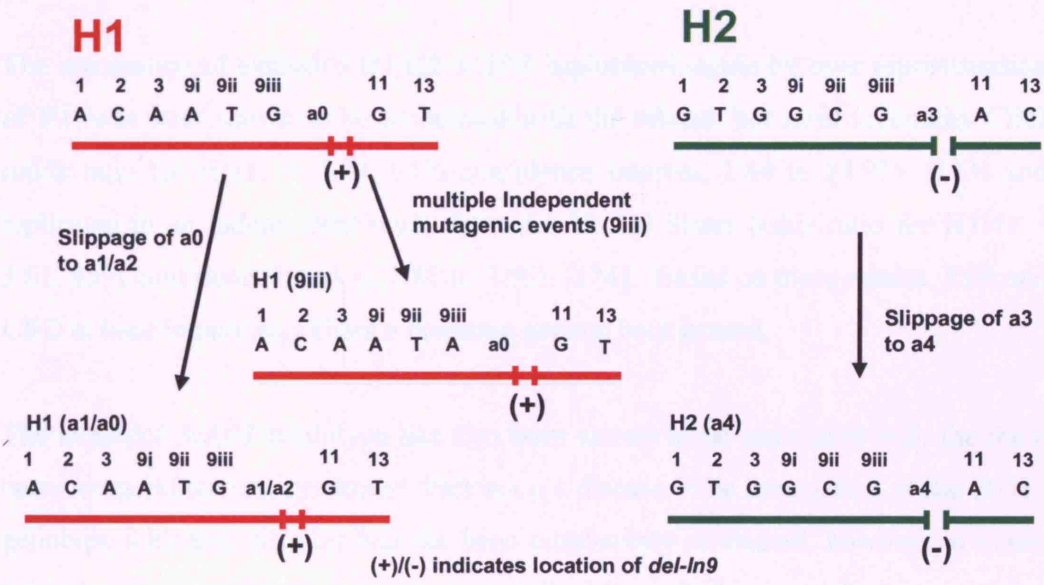

Figure 1.5 H1 and H2 *MAPT* haplotypes

The relationship between the extended H1 and H2 haplotypes and the a_0 polymorphism. The ancient haplotypes are modified by the indicated mutational events but are not altered by recombination. The SNPs that were typed in the original study are numbered 1-13 (referring to which exon number the SNP resides). The 238 bp deletion (*del-In9*) between exons 9 and 10 is shown by a break in the gene in the H2 haplotype and the presence or absence of this region is denoted by a (+) or (-), respectively. Figure modified from Baker [172].

Re-sequencing of selected individuals through the exon-intron boundaries revealed the presence of a 238 bp deletion in Intron 9 (commonly named the intron 9 del/ins

polymorphism, or *del-In9*) that forms part of the extended H2 background [172]. The presence or absence of the 238bp sequence corresponds perfectly to the extended H1 and H2 haplotypes and has often been used singularly as a marker for subsequent *MAPT* genetic association studies.

The frequency of the extended H1 haplotype in the Caucasian control populations is 77 - 80%; in contrast this allele frequency increases to more than 93% in PSP cases. Thus the extended H1 haplotype in *MAPT* is a genetic risk factor for PSP (**Table 1.4**). The odds ratio for the H1H1 susceptibility genotype was 4.08 (95 % confidence interval 1.89 to 8.79, [172]). Of equal importance is the under-representation of the extended H2 in PSP and is suggestive of protective role against PSP for this haplotype.

The association of extended H1/H2 *MAPT* haplotypes, again by over representation of H1, was later shown to be associated with the related, but rarer tauopathy CBD (odds ratio for H1H1 = 5.42, 95% confidence interval, 1.44 to 23.97), [173] and replicated in an independent study from the United States (odds ratio for H1H1 = 3.61, 95% confidence interval, 1.85 to 7.05), [174]. Based on these results, PSP and CBD at least in part may share a common genetic background.

The extended *MAPT* haplotype has also been shown to be associated with the most common parkinsonian syndrome, Parkinson's disease. The association of the H1H1 genotype with this disorder has not been consistently replicated, however, a meta-analysis of all studies suggests an overall risk to Parkinson's disease in those homozygous for the extended H1 haplotype; under a fixed effect model, the summary odds ratio for homozygous individuals of the H1 allele were 1.57 times more likely to develop Parkinson's disease than those carrying a H2 allele (95% confidence interval 1.33 to 1.85), [175]. These findings are surprising since Parkinson's disease is not traditionally associated with tau pathology, but rather aggregated α-synuclein in Lewy bodies [176].

Through examination of expressed sequence tags at the *MAPT* locus, a novel gene named saitohin *(STH)* was identified. It is nested within intron 9 of *MAPT* and

encodes a 128 amino acid protein with no clear homologues [177], though interestingly the tissue expression of *STH* is similar to that of tau. Genetic investigation of the gene identified a single nucleotide polymorphism resulting in an amino acid change from glutamine to arginine (Q7R). This coding polymorphism was used for a case-control study with late onset Alzheimer's disease. The RR genotype was found at a significantly higher frequency in the AD group though this finding has never been replicated in independent studies [178-184]. In fact, the *STH* Q7R polymorphism is in perfect linkage disequilibrium with the *MAPT* extended H1 and H2 haplotypes, respectively and unsurprisingly was found to be highly associated with PSP [185]. The QQ genotype was over represented in PSP (96%) against healthy controls (69%). This finding was replicated in a Spanish PSP series (91% QQ versus 47%), [186]. The function of Saitohin protein is currently unknown, though recently has been shown to interact with peroxiredoxin 6 (*Prdx6*) by a yeast two-hybrid screen and co-immunoprecipitation and glutathione S-tranferase pull-down assays [187]. Interestingly alterations in Prdx6 levels are associated with PiD, dementia with Lewy bodies and sporadic Creutzfeldt-Jacob disease [188-190].

The *MAPT* H1 and H2 haplotypes were extended a further 68 Kb to several SNPs in the promoter region of the gene, all of which were found to be highly associated with PSP [191, 192]. Further to this study, the linkage disequilibrium and association with PSP was mapped to genetic markers lying outside the limits of the *MAPT* gene, implicating the entire *MAPT* gene, flanking untranslated region (UTR) and adjacent genes [193]. Genomic sequencing of the 3' UTR in PSP revealed no mutations [186].

A possible effect of the extended *MAPT* H1 haplotype on tau biochemistry and neuropathology was examined in 25 PSP cases matched for age, sex and post-mortem delay [194]. The H1 haplotype was not found to influence pathological or biochemical phenotype of PSP with respect to tau burden or tau isoform composition.

		Risk Genotype and allele frequency (%)					
Study		**PSP**			**Controls**		**OR**
Tau Intron 9 dinucleotide repeat	**n**	**a0/a0**	**a0**	**n**	**a0/a0**	**a0**	
Conrad et al. [167]	22	95.5	97.7	61	57.4	74.6	15.6
Conrad et al. [195]	31	100	100	67	97	98	n/a
Oliva et al. [171]	30	86.7	93.3	50	52	73	6.5
Morris et al. [170]	53	84.3	90.6	75	53	72.7	n/a
Baker et al. [172]	64	79.7	89.8	139	51.1	70.5	n/a
Bennet et al. [168]	30	86.7	93.3	36	61.1	76.4	4.1
Higgins et al. [196]	24	83.3	91.6	86	53.4	75.6	4.4
H1 haplotype		**H1/H1**	**H1**		**H1/H1**	**H1**	
Baker et al. [172]	64	87.5	93.7	145	62.8	78.4	4.1
Morris et al. [197]	50	98	99	71	63	79	n/a
Tau promoter haplotypes		**C/C**	**C**		**C/C**	**C**	
Ezquerra et al. [198]	35	91.4	94.2	195	49.7	72.5	11.8
		H1p/H1p	**H1p**		**H1p/H1p**	**H1p**	
de Silva et al. [191]	42	97.6	98.8	70	61.4	78.6	n/a
5' tau haplotype		**Hap A/A**	**Hap A**		**Hap A/A**	**Hap A**	
Higgins et al. [199]	52	98	99	54	48	77.7	n/a

Table 1.4 *MAPT* association studies in PSP

Association studies of the *MAPT* locus and identified risk alleles in progressive supranuclear palsy. All studies found a significant association of the H1 haplotype or other marker in populations from North America, Western Europe and Spain. However not all studies are independent, some studies utilised the same PSP population and typed different markers for association. There are five independent study populations in the table.

1.7.3.3 Genetic linkage of autosomal dominant PSP to 1q31.1

Recently, a large Spanish family has been described whose clinical phenotype was broadly similar to that found in patients with sporadic PSP [200, 201]. One member was pathologically confirmed, where neuropathological examination revealed moderate numbers of neurofibrillary tangles without prominent senile plaques in the cortex, and neuronal loss, gliosis and moderate to severe accumulation of tangles in the basal ganglia and brainstem [200].

Clinical information was obtained from 45 individuals and DNA taken for linkage analysis in 31 of the family members. Phenotype was defined not only by the presence of clinical symptoms but also based on the results of ^{18}F-fluoro-dopa (^{18}F-dopa) and ^{18}F-deoxyglucose positron emission tomography (PET) scans measuring presynaptic dopaminergic functions and cerebral glucose metabolism, performed in 10 members of the kindred. Family members with abnormal neuroimaging results were included as affected individuals in the subsequent genetic analysis of the pedigree.

A genome-wide search was performed using 340 mircosatellite markers. Parametric two-point analysis excluded 20% of the genome, including the region at 17q21 in which *MAPT* maps. LOD scores suggestive of linkage were identified at markers D1S428 and D1S1678 and genotyping of additional markers in this area and multipoint calculations yielded a maximum LOD score of 3.53 for the interval between D1S428 and D1S461 with 90% penetrance. Thus, the genetic locus responsible for the disease in this pedigree maps to 1q31.1, between the markers D1S238 and D1S2823, a region of 3.4cM however there are no clear candidate genes thought to be involved in the pathogenesis of neurodegenerative disorders[202].

The locus associated with this PSP family is different from any other region of the human genome involved in PSP or in other related neurodegenerative disorders. At 1q31.1, in the 3.4cM interval, there are three putative genes: *DBCCR1*-like, *ENSG00000185167* and *ENSG00000162670*. *DBCCR1* is a gene that is expressed in some human cancers. In human tissues, it is largely expressed in the brain and is a homologue of BMP/retinoic acid-inducible neural specific protein 3 (*Brinp3*). Retinoic acid signalling does play a role in the differentiation of several neuronal phenotypes and it modulates the survival of dopamine neurones *in vitro* [203]. Little is known about the other putative genes, except that *ENSG00000162670* is also expressed in the brain.

1.8 Corticobasal degeneration

Corticobasal degeneration (CBD) is a progressive neurological disorder characterised by atrophy of multiple brain areas including the cerebral cortex and the basal ganglia [204]. Initial symptoms, which typically appear at the age of around 60, are similar to those found in Parkinson's disease, such as poor coordination, akinesia, rigidity, impaired balance and limb dystonia. Other symptoms such as cognitive and visual-spatial impairments, apraxia, hesitant and halting speech, myoclonus (muscular jerks), and dysphagia (difficulty in swallowing) may also occur.

Neuropathologically, CBD is distinguished from PSP and other dementia by several important features. Most pathology in CBD is in the cerebrum, whereas the basal ganglia, diencephalon, and brainstem are mainly targets of PSP. Histologically, there are ballooned neurons, astrocytosis, and 4R tau-positive neuronal and glial inclusions. The most characteristic neuronal tau pathology in CBD is numerous and widespread wispy, fine filamentous inclusions within neuronal cell bodies, whereas affected neurons in PSP have compact, dense filamentous aggregates [205, 206].

The genetics of CBD had not been widely studied until now because the disease is rare and usually sporadic in occurrence. However, the extended H1 haplotype has also shown to be a genetic risk factor for CBD [173] that was subsequently independently replicated [174].

1.9 Thesis aims and objectives

In this chapter, the rationale and strategies behind genetic association studies for complex disorders were reviewed and their potential power to identify genetic susceptibility variants in genome. The association of the *MAPT* locus with PSP has been firmly established, reflected by the fact that this finding has been replicated in every independent study to date. The extended H1 haplotype is a genetic risk factor for PSP however, the pathogenic basis of this association is not known. Tau dysfunction has already been implicated in FTDP-17T, whereby highly penetrant dominant mutations in *MAPT* cause the disease. Although *MAPT* mutations are not found in the majority of PSP cases, some FTDP-17T mutations can cause clinical and sometimes pathological pictures resembling that of PSP. The association of the H1 *MAPT* haplotype implies tau dysfunction is related to the pathogenesis of PSP and this is also supported by the characteristic tau pathology found in the diseased brain post-mortem.

The hypothesis is that common genetic variation at 17q21.31 affects *MAPT* gene expression and it was the main aim of the work in this thesis is to investigate the genetic and functional basis of the *MAPT* H1 haplotype association with PSP.

At the onset of this work, the extent of the *MAPT* H1 non-recombining haplotype block had not yet been defined. The first aim of this work was to investigate the extent of the H1 LD and its association with PSP so as to define the entire candidate genomic interval at 17q21.31.

On strong evidence from neuropathology and biochemistry that tau dysfunction could play a role in PSP, subsequent work focused on high-resolution analysis of *MAPT* LD structure and haplotype diversity in order to identify potential elements of the gene that could have subtle effects on *MAPT* function, including its expression and/or splicing. At the onset of this work, the genetic diversity of the *MAPT* gene was defined in terms of the H1 and H2 haplotypes and there was evidence of much greater haplotypic diversity. Using population genetic methods, the underlying LD structure and haplotype diversity of the *MAPT* gene was examined. Expanding on

this, a Japanese control group was also examined in order to obtain insight into the *MAPT* LD and haplotype diversity in a population carrying only the H1 haplotype.

The establishment of the detailed architecture of the *MAPT* gene gave the basis for selection of htSNPs for more streamlined and efficient genotyping of the large numbers of PSP, CBD and control samples from the UK and USA. The allele, genotype and haplotype frequencies of the htSNPs will be statistically assessed for differences between PSP cases and controls. Recently, the pathologically confirmed UK PSP series has been show to be distinctly composed of two clinical sub-types of PSP; Richardson's syndrome and PSP-parkinsonism. The *MAPT* association with PSP will be examined in respect of these two groups to see if it modulates these clinical differences. The *MAPT* H1 haplotype had also previously been shown to be associated with the related tauopathy, CBD. The same set of htSNPs will also be genotyped in a cohort of CBD cases to determine if the *MAPT* genetic association is the same for both PSP and CBD.

Identification of the associated *MAPT* haplotypes that confer risk, protection or are neutral provided the basis for a targeted re-sequencing strategy. Homozygote carriers representative of the risk, neutral and protective *MAPT* haplotypes will be re-sequenced to identify candidate polymorphisms on the haplotype backgrounds for functional studies. This work will include investigating the allele-specific effect polymorphisms have on *MAPT* promoter-driven expression by reporter gene experiments.

Recently, a *de novo* microdeletion of 500-650 kb in size encompassing the *MAPT* gene at 17q21.31 has been identified in three individuals with developmental delay and learning disabilities. This is a new microdeletion syndrome that has not been characterised before at the molecular level. The genomic structure of the locus will be examined in terms of the complement of LCRs. In the families, detailed SNP haplotyping with respect to H1 and H2 will be performed. This is in order to determine if there is any relationship between *MAPT* haplotype background and susceptibility to the pathogenic structural alteration.

2 Materials and Methods

2.1 Methods

2.1.1 DNA sample extraction from tissue

DNA was routinely extracted by hand from flash frozen brain material as required. In the laboratory two methods were used, either by use of the DNeasy Tissue Kit (Qiagen) or by Proteinase K/phenol-chloroform extraction. In the latter protocol the frozen brain was first proteolysed with 100 µl of Proteinase K (10mg/ml) and 240 µl of 10% SDS and 2.06 ml of DNA (TE) buffer incubated overnight at 45° C. The following morning 2.4 ml of phenol was added to the lysate, vigorously shaken by hand for 5 minutes and then centrifuged at 3000 rpm for 5 minutes at 10 ° C. The supernatant was then removed, placed in a new tube and 1.2 ml phenol and 1.2 ml of chloroform/isoamyl alcohol (24:1) added and the mixture was shaken again for 5 minutes. The step was repeated for a third time, though this time 2.4 ml of chloroform/isoamyl alcohol was added to the supernatant. The DNA contained within the supernatant fraction was precipitated by addition of 25 µl of 3 M sodium acetate (pH 5.2) and 5 ml of 100 % Ethanol. Upon precipitation, the DNA thread is removed from the solution using a glass hook, washed in 70% ethanol, dried, and re-suspended in 0.5 ml sterile water overnight at 4 ° C.

2.1.2 DNA quantification.

DNA extraction quantity and quality was monitored by UV spectrophotometer absorption (CECIL, CE2021 platform). The absorption at 260 nm indicated the concentration of DNA in the sample; the absorption measurement is multiplied by any dilution factor then by 50 (the constant for double stranded DNA) for the final concentration value in ng/ul. The ratio of absorption values at 260 nm and 280 nm provides an indication of DNA purity of the sample. Ratios of >1.5 indicate a pure DNA sample, ratios <1.5 indicate protein contamination.

2.1.3 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is an *in vitro* method for amplifying defined target DNA sequences present within a source or template of DNA. Usually, the method is designed to permit selective amplification of a specific target DNA sequence within a heterogeneous collection of DNA sequences (e.g. total genomic DNA or a cDNA population). To permit such selective amplification, some prior DNA sequence information from the target sequence is required. This information is used to design two oligonucleotide primers which are specific for the target sequence and which are about 15-25 nucleotides long (usually designed in the computer program Primer3 [http://fokker.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)]. The primers bind specifically to complementary DNA sequences at the target site to denatured template DNA. In the presence of suitably heat-stable DNA polymerase and DNA precursors (the four deoxynucleoside triphosphates, dATP, dCTP, dGTP and dTTP), they initiate the synthesis of new DNA strands which are complementary to the individual DNA strands of the target DNA segment. The PCR is a chain reaction because newly synthesized DNA strands will act as templates for further DNA synthesis in subsequent cycles. After about 25 cycles of DNA synthesis, the products of the PCR will include about 10^5 copies of the specific target sequence. Although PCR was only developed relatively recently, the versatility of the technique have ensured that it is among the most ubiquitous of molecular genetic methodologies, with a wide variety of applications ranging from; DNA sequencing, genotyping, mutation screening, cDNA cloning and gene expression studies.

PCR was performed using the *Taq* DNA polymerase core kit (Qiagen). Typical 50 µl reactions contained 10x reaction buffer, 0.1-0.5 uM of each of forward and reverse primers, 1 unit of *Taq* DNA polymerase, 2 µl of each 10mM dNTPs and distilled water and 25 ng of genomic DNA template. Some PCR reactions required the addition of 5x Q solution (Qiagen) for optimal performance and specificity. For purposes of DNA sequencing and genotyping PCR reactions were routinely carried out in volumes of 10 -20 µl.

PCR temperature cycling was achieved by using an automatic Eppendorf Mastercycler or a Techne Touchgene. The PCR reaction involves an initial denaturing step at 95 °C for 5 minutes, followed by 25-35, 30 second long cycles of denaturation (95 ° C), primer annealing (variable, depending on the annealing temperature of the primers) and extension (72 °C) followed by a final extension of 7 minutes at 72 °C and a refrigeration hold at 4 °C until sample collection.

2.1.4 Genotyping

2.1.4.1 Restriction fragment length polymorphism

The great majority of the nucleotide variation of DNA is not associated with a disease; they often occur within non-coding DNA sequences. As a large number of recognition sequences are known for type II restriction endonucleases, many point mutation polymorphisms will be characterised by alleles which possess or lack a recognition site for a specific restriction endonuclease and therefore display a restriction site polymorphism. Accordingly, individual polymorphisms normally have two detectable alleles (one lacking and one possessing the specific restriction site).

For SNP analysis by restriction fragment length polymorphism (RFLP), firstly PCR primers were designed to amplify the region of genomic DNA surrounding the SNP. Genotyping assays were designed by identifying restriction endonuclease enzymes whereby the cleavage of PCR product was unique to a single nucleotide change of the polymorphism in question, by the program BioEdit Sequence Alignment Editor 6.0.5. Volumes of 15 µl of raw PCR products were incubated with 1 unit of the corresponding restriction endonuclease (New England Biolabs, Hertfordshire, and Promega, Southampton, UK) in a reaction volume of 20 µl at the recommended temperature for at least four hours. Digests were run out on a 4% agarose gel, visualized with ethidium bromide staining and photographed with a Polaroid camera for genotype scoring.

2.1.4.2 Pyrosequencing

Pyrosequencing is a DNA sequencing technique that is based on the detection of released pyrophosphate (PPi) during DNA synthesis. In a cascade of enzymatic reactions, visible light is generated that is proportional to the number of incorporated nucleotides. The cascade starts with a nucleic acid polymerization reaction in which inorganic PPi is released as a result of nucleotide incorporation by polymerase. The released PPi is subsequently converted to ATP by ATP sulphurylase, which provides the energy to luciferase to oxidize luciferin and generate light. Addition of dNTPs is performed one at a time and because the added nucleotide is known, as the process continues the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the pyrogram.

For SNP analysis by Pyrosequencing, firstly PCR primers were designed (by primer3 program) to amplify the region of genomic DNA surrounding the SNP. Then the third primer for the Pyrosequencing assay was designed by the manufacturer's internet Pyrosequencing Primer design program (http://techsupport.pyrosequencing.com/).

For performing the assay itself, 12 µl of the PCR product was first immobilised on Streptavidin-Sepharose^{TM} HP (Amersham Pharmacia Biotech, Piscataway, NJ): The gel slurry (4 µl) is then re-suspended in binding buffer (10 mM Tris-HCl, 2 M NaCl, 1mM EDTA, 0.1 % Tween-20). Template and beads were mixed continuously for >5 min at room temperature. The immobilised DNA template was then transferred to a 96-well filter plate attached to a vacuum manifold (Millipore Inc., Bedford, MA), then immersed for 10 seconds 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 pmoles) was then annealed to the single-stranded template in 12 µl of annealing buffer (20 mM Tris-acetate, 2mM magnesium acetate, pH 7.6) at 80°C for 2 min before cooling to room temperature. Samples were analysed using a PSQ 96 System together with SNP Software and SNP reagent kits (Pyrosequencing Inc., Westbourough, MA) following the manufacturer's instructions. Genotype scoring

was carried out by the SNP software though each individual read (pyrogram) is also visually inspected for quality control purposes.

2.1.4.3 PCR genotyping

For polymorphisms involving large insertions or deletions (>50 bp) of nucleotide sequence, genotyping was carried out by running PCR products on an agarose gel, visualized with ethidium bromide staining and photographed with a Polaroid camera for genotype scoring

2.1.4.4 Micosatellite genotyping

Microsatellites are DNA sequence length polymorphisms. They are genetic markers that are particularly useful for mapping disease genes, usually $(CA)_n$ repeats though tri- and tetranucleotide repeats are also routinely used.

For genotyping repeat polymorphisms, PCR primers (designed in Primer3 program) of which one is fluorescently labeled, routinely with FAM of HEX, were designed to amplify the repeat sequence of interest using genomic DNA as temple. PCR products were then subjected to capillary electrophoresis on either a 310 Genetic analyser or a 3730 (Applied Biosystems) and fragment sizes/allele sizes calculated with Genescan software.

2.1.5 DNA sequencing

The DNA sequencing method routinely employed is a variation of the Sanger sequencing method and typically PCR products amplified from human genomic DNA were used as templates for sequencing. In the sequencing reaction, the PCR product is subjected to linear amplification with a single primer and dNTPs; containing a proportion of dideoxynucleotides that are dye terminators not permitting further extension and are labelled with fluorescent dyes; a different colour for each base. The result of the linear amplification is to produce different product lengths each coloured differently according to the terminating nucleotide.

Automated capillary electrophoresis is then used to resolve the sequence products at a resolution of one base pair, generating a readable trace (chromatogram) of the DNA sequence.

Briefly, PCR primers were first designed to amplify the DNA sequence of interest of the target template DNA, routinely whole genomic DNA. PCR products were 'cleaned' by use of ExoSAP-IT reagent (Amersham). ExoSAP-IT reagent contains Exonuclease 1 for degradation of excess primer and shrimp Alkaline Phosphatase for inactivating unincorporated dNTPs. Typically, 2 μl of reagent was added to 5 ul of PCR product. For the sequencing reaction, 6ul of treated PCR product was made up to a volume of 12 ul of dH_2O including 12.8 pmol of primer and sent for sequencing (Advanced Biotechnology Centre, Imperial College London) using BigDye V3.1 and subjected to capillary electrophoresis on a ABI 3100 capillary sequencer (Applied Biosystems). Sequence traces (ABI chromatograms) were viewed and analysed using BioEdit software.

2.1.6 Statistical analysis in population genetic studies

2.1.6.1 Hardy-Weinberg equilibrium

In a large population, with random mating, Hardy-Weinberg equilibrium (HWE) is the stable frequency distribution of genotypes. For a bi-allelic polymorphism; allele A and a, at frequencies p and q, the stable frequency genotypes will be p^2, 2pq and q^2 that is a consequence of random mating in the absence of mutation, migration, natural selection or random drift. The expected numbers for HWE can be calculated and be compared to observed genotypes of the population in question and deviations can be identified through a chi-squared test. Determination of HWE deviations in the case and control populations was made in the genetics software program TagIT and statistical significance was set at $p<0.05$ for significant deviations.

2.1.6.2 Genetic association studies

2.1.6.2.1 Single-locus analysis

Statistical comparison of the allele and genotype distributions of single loci between cases and control in the cohorts under study was achieved by chi-square test, a non-parametric test of statistical significance for bivariate tabular analysis. For bi-allelic markers, a 2x2 table is used with 2 degrees of freedom for comparison of allele counts between two groups/populations and a 2x3 table with 3 degrees of freedom (df) for comparison of genotype counts between the two groups. Statistical significance was set at $p<0.05$, for these tests and all tests unless otherwise stated.

CLUMP software was routinely used for chi-squared analysis, however using this approach significance is assessed using a Monte Carlo approach; repeated simulations are performed to generate tables having the same marginal totals as the one under consideration; and counting the number of times that a chi-squared value associated (either greater than or equal to) with the real table is achieved, by the randomly simulated data. Typically 1000-10000 simulations were performed or increased further until a satisfactory accurate estimate of the true significance was achieved.

2.1.6.2.3 The odds ratio

The odds ratio (OR) is a way of comparing whether the probability of a certain event is the same for two groups. In terms of case-control genetic studies, this is the ratio of odds of having a particular allele or genotype in the case group divided by the odds of having the allele or genotype in the control group. An OR of 1 implies that the allele or genotype is equal in both groups; an OR greater than 1 implies risk in the case group; an OR less than 1 implies protection of the allele or genotype to the case group. As the calculation is based purely on a sample of the population in question it is essentially only an estimate, thus the accuracy is determined by the size of the sample. For this reason, it is conventional to calculate the 95% confidence interval (CI) for the OR. For interpretation, a proposed allele or genotype acts as a significant risk to disease if the OR is grater than 1 and the lower bound of the CI

dies not go below 1 and *visa versa* for a protective allele or genotype in the case group. An OR and 95% CI calculator (http://www.hutchon.net/ConfidOR.htm) was used for all case-control studies unless otherwise stated.

2.1.6.2.3 Haplotype analysis

The inheritance of closely linked alleles (for example two DNA polymorphisms) as a unit, on a single chromosome is a haplotype. In the case of muti-locus genetic association studies, to potentially increase power, haplotypes were also analysed for association.

Haplotypes and their respective frequencies in the unrelated populations were calculated by use of the expectation-maximization (EM) algorithm. It predicts haplotype phase from genotype data from multiple genetic markers, usually SNPs. A routine algorithm is implemented in SNPHAP software (http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt).

There are also several other forms of EM algorithms available. In TagIt program, this particular EM algorithm is tailored to work with trio population data, as structured in the Utah CEPH trio population (Chapter 4). There have also been EM algorithms developed to handle sets of population data with large numbers of SNP loci that are largely uncorrelated with on another (that have little LD between the individual loci). One such algorithm is the partition ligation-expextation maximization (PL-EM) algorithm. This algorithm first breaks up the SNP loci into 'windows' of smaller subsets of loci, calculates the haplotypes and their respective frequencies within each 'window' by EM and then by 'ligation' assembles the sub-sets of haplotype together for final output .

Distributions of multi-locus haplotypes defined by single loci were compared between case and control groups using WHAP software. This SNP haplotype analysis suite performs a regression based haplotype association test through a likelihood ratio test (LRT), which is a χ^2 test with n-1 df to derive the associated p value, where n is number of haplotypes observed for the data set. This test was used

for omnibus testing of haplotype frequencies and also used for individual haplotype specific tests (df = 1) of association.

2.1.6.3 Tagging single nucleotide polymorphisms

The efficiency of genetic association studies can be increased by typing informative SNP – haplotype tagging SNPs (htSNPs) that are in linkage disequilibrium with several other SNPs thus a small fraction or subset of SNPs at the locus or gene of interest are sufficient to 'capture' the vast majority of the genetic variation. The programs TagIt (version 1.19) and Haploview were routinely used to select htSNPs for genetic association studies. Both programs use the correlation of r^2 (typically haplotype r^2) between loci to determine which SNPs (or indeed combinations of SNPs) can predict the allele state of the other SNPs.

2.1.7 Bioinformatics/Web resources

2.1.7.1 NCBI

The National Centre for Biotechnology Information (NCBI) is a resource for molecular biology and genetics and consists of publicly available databases (http://www.ncbi.nlm.nih.gov/) invaluable for retrieving information such as nucleotide sequence data and polymorphism frequency data (for example the db SNP database). The resource also contains web based bioinformatics programs such as basic local alignment search tool (BLAST), that is used to search and to retrieve sequences homologous to the one of interest and this program was used routinely during the work in this thesis.

2.1.7.2 UCSC

The University of California Santa Cruz (UCSC, http://genome.ucsc.edu/) genome browser is a particularly useful web resource that allows for visualization of an assembled reference human genome (and indeed other organism such as chimpanzee) annotated with such information as the position of genes, polymorphic variation, repeats, cross-species conservation and structural variation. The web

resource also contains some useful programs that allow for identifying the location of nucleotide sequences on the genome (BLAT) and *in silico* PCR, for identifying the PCR products generated of primer-pairs when using genomic DNA as a template.

2.1.7.3 HapMap

The International HapMap project (HapMap) is a web based resource that allows the retrieval of high-density SNP genotype data in a total of 270 individuals in four populations: 30 CEPH (Centre d'Etude du Polymorphisme Humain) -trios (families from Utah, US of Western European origin), 45 unrelated Chinese individuals from Beijing, thirty trios from the Yoruba people of Ibadan, Nigeria and 45 unrelated individuals from Tokyo, Japan.

Downloaded population genotype data can be used to analyse the haplotype diversity of the population in question and one application of such data is to identify htSNPs for candidate gene genetic association studies.

2.1.7.4 mVISTA

mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) is a set of programs for comparing DNA sequences from two or more species and for visualizing these alignments with annotation information [207]. The program is implemented as an on-line server that provides access to global pairwise, multiple and (global with rearrangements) alignment tools. This program is particularly useful for identifying homology between human genes and the mouse homologue.

2.1.7.5 RepeatMasker

RepeatMasker (http://www.repeatmasker.org/) is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences. The output of the program is a detailed annotation of the repeats that are present in the query sequence as well as a modified version of the query sequence in which all the

annotated repeats have been masked (replaced by Ns). On average, almost 50% of a human genomic DNA sequence currently will be masked by the program. This program is useful for removing repeating elements from DNA sequences that are being used for BLAST searches or PCR primer design, since primers that bind to repeat sequences in a genomic DNA template do not amplify well.

2.1.7.6 YASS

YASS (http://bioinfo.lifl.fr/yass/) is a program to perform DNA local alignment of genomic sequences. YASS uses 'seeds' to detect potential similarity regions, and then tries to extend them to actual alignments. YASS is a particularly useful program for comparing long stretches of DNA sequence such as BAC clones and for identifying large regions of sequence homology such as low-copy repeats.

2.1.8 Molecular Biology

2.1.8.1 PCR

The polymerase chain reaction (see section 2.2.3) was routinely used to generate DNA sequence inserts for cloning. Primers were designed to amplify the DNA and secondly to introduce new sequence (typically 6 nucleotides long) at the 5' and 3' end of the PCR product to be recognised by specific restriction endonuclease enzymes for subsequent cloning steps of the DNA insert into the vector of choice.

2.1.8.2 Agarose gel electrophoresis

Agarose gels, used for analysing DNA fragment sizes and quality were made by melting agarose powder (Sigma-Aldrich) in either TAE or TBE buffer in a microwave oven. Gels were made routinely between 0.8 and 4% w/vol. The gels were cast with the addition of 50 ng/mL ethidium bromide and using plastic combs for wells into which samples could be loaded. Once set, gels were submerged in either TAE or TBE buffer in the electrophoresis tank. Samples were pre-mixed with 5x or 6x loading dye/buffer and loaded into the wells of the gel. Samples were subjected to electrophoresis for approximately 30 minutes to 1 hour at 80 to 200 mV

depending on the size and percentage of agarose of the gel. The DNA was visualised in the gel with a UV transilluminator and photographed with a Polaroid camera.

2.1.8.3 DNA fragment purification by agarose gel electrophoresis

In plasmid based cloning, it was necessary to purify DNA from other contaminating nucleic acids and enzymes for subsequent sub-cloning steps. For this purpose, DNA bands of interest visualised in ethidium bromide stained agarose gels were excised with a clean scalpel and the DNA purified with in an ion-exchange chromatography column kit (Qiagen).

2.1.8.4 Restriction digestion

Restriction digestion of plasmid DNA was carried out in a total volume of 50 µl, containing the appropriate 10x reaction buffer, 100x bovine serum albumin (BSA), 1 µg of the plasmid DNA and 1-2 units of each restriction endonuclease (New England Biolabs) in dH_20. Digestion reactions were incubated at the manufacturers recommended temperature, typically at 37°C.

2.1.8.5 DNA ligation into plasmid vectors

Insert DNA generated by PCR was routinely cloned into the plasmid vector pGEM – T (Promega). *Taq* DNA polymerase generates PCR products with 5' adenosine overhangs that are complementary to the 3' thymidine overhangs of the linearised vector. Vector and insert DNA are ligated together with T4 DNA ligase (Promega), typically in 10 ul reaction volumes.

In subsequent sub-cloning steps of the insert DNA, pGEM-T–insert clones were first propagated in *E. coli* then purified from the harvested cells by ion-exchange chromatography (Qiagen). The purified plasmid DNA was then digested with appropriate restriction enzyme(s) to release the insert DNA fragments form the vector and purified (see section 2.2.7.3). Insert DNA fragments were then inserted

into target vectors that had been pre-linearised with the appropriate restriction endonuclease(s).

2.1.8.6 *E.coli* transformations and purification of plasmid DNA

The *E. coli* strain JM109 was transformed with cloning plasmids. Competent cells (Promega) were thawed on ice in 100 µl aliquots and mixed with 5 µl of ligated insert-vector DNA. Following 30 minutes of incubation on ice the cell and DNA mixture was 'heat-shocked' at 42°C for 2 minutes. After a further incubation on ice for 2 minutes, 900 µl of LB media was added and the cells were allowed to recover at 37 °C for 1 hour with gentle agitation. The cells were then gently centrifuged at 3000g and re-suspended in 100 ul of LB media. The cell suspensions were spread onto LB agar plates, containing the appropriate antibiotic selection for the plasmid (e.g. 50 µg/ml ampicillin), using a sterile glass spreader and grown at 37 °C for approximately 16 hours.

Single colonies were aseptically picked from the plate and used to inoculate LB broth (routinely 1-5 mL) supplemented with the appropriated selection antibiotic as contained in the agar plate. After an overnight incubation at 37 °C with vigorous shaking, the cells were harvested by centrifugation. The pelleted cells were then subjected to alkaline lysis and the DNA recovered by ion exchange chromatography by use of plasmid mini-prep spin columns (Qiagen).

For preparation of high-quality plasmid DNA for mammalian cell transfection it was necessary to purify DNA from the propagating *E. coli* host without contaminating endotoxins. In these instances, plasmid preparations were made using the Endofree preparation Kit (Qiagen) according to the manufacturer's instructions. Briefly, the process involves an overnight growth of a 100 mL culture, alkaline lysis of pelleted cells and removal of the lipid and protein fraction using a syringe filter. Endotoxin removal is achieved by incubation with a kit component. The DNA is precipitated with isopropanol and recovered by centrifugation.

2.1.9 Cell culture

2.1.9.1 SH-SY5Y and BE(2)-M17 neuroblastoma cell culture

The SH-SY5Y and BE(2)-M17 human neuroblastoma cells were obtained from the European collection of cell cultures (ECACC). They have a substrate-adherent phenotype and naturally grow processes in culture. SH-SY5Y cells and BE(2)-M17 cells were cultured in SH-SY5Y culture medium, incubated at 37°C in 100% relative humidity and 5% carbon dioxide. The growth media was changed approximately every two to three days.

Cells were propagated until reaching approximately 80% confluency and subcultured by diluting (splitting) at a ratio of 1:5. This process involved discarding the culture medium and rinsing the adherent cells with 1x phosphate buffered saline (PBS) and incubating the adherent cells with a thin layer of trypsin-EDTA solution at 37°C for 5 minutes; until the cells detached from the flask/well. The detached cells were collected in SH-SY5Y culture medium, diluted accordingly for seeding and growth therein of additional culture flaks.

For long term storage of cells, cultures of SH-SY5Y cells and BE(2)-M17 cells were harvested by trypsinisation and spun down at 800g for 5 minutes, re-suspending the cells in 1mL of freezing medium. The suspended cells were then transferred to cryotubes, slowly frozen in a bath of isopropanol in a -80°C freezer and transferred to liquid nitrogen storage.

2.1.9.2 Transfection of cells with plasmid DNA

The transfection of the SH-SY5Y and BE(2)-M17 cell lines was carried out using TransFast Transfection reagent (Promega). The reagent is comprised of synthetic cationic lipid and the neutral lipid, L-dioleoyl phosphatidylethanolamine (DOPE). Briefly, cells are seeded into the growth flask/well in SH-SY5Y media and grown for approximately 24 hours. On the day of transfection, the required amount of Endotoxin-free plasmid DNA preparation was added to Transfast Transfection reagent at a charge ratio of 1:1 in MEM media (typically 1.65 µl of DNA (at

50ng/µl), 5 µl of Transfast transfection reagent and 213.4 µl of MEM media for 5 wells of a 96-well plate). The mixture is incubated for 15 minutes at room temperature. The growth medium is first removed from the target cells and replaced with the transfection medium. The cells and the transfection media are then incubated together at 37°C for one hour. Finally the transfected cells are overlaid with SH-SY5Y culture medium and allowed to recover for approximately 24 hours.

2.1.9.3 Luciferase assay

Transcriptional regulation coupled to reporter gene expression is routinely used to study a wide range of physiological responses, for example, quantifying the action of specific DNA *cis*-elements on promoter transcriptional ectivity. A change in the transcription and subsequent translation of the reporter molecule, for example firefly luciferase, can be tracked in a biological system. A gene isolated from the firefly *Photinus pyralis*, when translated, gives rise to the luciferase enzyme, which acts as the catalyst in the oxidation of luciferin to oxyluciferin; an ATP dependent reaction that produces light in a linear relationship with enzyme activity.

The promoter and regulatory sequences to be studied are inserted into the multiple-cloning site upstream of the luciferase gene sequence. The reporter gene plasmid (*luc2* in pGL4.10, (Promega) was expressed in cell cultures to measure promoter activity. Readings were normalised by co-transfecting with the internal control plasmid pRL-CMV which expresses *renilla* luciferase. The presence of the internal control essentially negates the experimental error associated with well-to-well differences in transfection efficiency. All luciferase assays were carried out using the Dual-Glo Luciferase Assay System (Promega). The system is designed to allow high-throughput analysis of mammalian cells containing both genes for firefly and *Renilla* luciferases in the growth media without preconditioning and is composed of two reagents. The Dual-Glo Luciferase reagent induces cell lysis and acts as a substrate for firefly luciferase and produces a stable luminescent signal than can be read over a period of two hours. Addition of the second reagent, Dual-Glo Stop and Glo reagent quenches the luminescence from the firefly reaction and provides the

substrate for the *renilla* luciferase in a reaction that also can be read within two hours.

Luciferase assays were carried out on mammalian cultures grown in opaque 96-well plates (Greiner), in 200 µl of media. Firstly, 125 µl of media was removed from each well and 75 µl of each reagent was sequentially added to each well. The luminescence was measured in each well sequentially after the addition of each reagent using a 96-well reader luminometer (Tecan GENios).

2.2 Materials

2.2.1 PCR reagents

Taq DNA polymerase kit (Qiagen)

2.2.2 DNA/Genotyping/Sequencing reagents

2.2.2.1 DNA extraction reagents

DNA (TE) buffer (Tris-EDTA):

10 mM Tris-Cl, pH 8.0

1mM EDTA

Chloroform
Ethanol
Isoamyl alcohol
Phenol
Phosphate buffered saline
Proteinase K
RNase A
Sodium dodecyl sulphate (SDS) solution, 10%

2.2.2.2 Genotyping Reagents

Restriction fragment length polymorphisms:

All Restriction endonuclease enzymes were either obtained from New England Biolabs (Heartforshire) and Promega (.

Pyrosequencing:

Pyro Gold reagent (Biotage, Foxboro, Maryland)

2.2.2.3 Sequencing reagents

ExoSAP-IT reagent (Amersham)

2.2.3 Molecular biology reagents

TAE Buffer: 40 mM Tris-HCl pH 8.0 (Sigma-Aldrich)

20 mM acetic acid (Merck)

1mM ethylenediamine tetraacetic acid (EDTA) (Sigma-Aldrich)

TBE Buffer: 45 mM Tris-borate pH 8.0 (Sigma-Aldrich)
1 mM EDTA (Sigma-Aldrich)

TE Buffer: 10 mM Tris-HCL pH 8.0 (Sigma-Aldrich)
1mM EDTA

Luria-Bertani (LB) broth:
20 g/L LB broth powder (Invitrogen) dissolved in deionised water, autoclaved at 121°C before use.

LB agar: 32 g/L LB agar (Invitrogen) dissolved in deionised water, autoclaved at 121°C before use.

2.2.4 Cell culture reagents

All cell culture reagents were supplied by Sigma-Aldrich

SH-SY5Y cell culture growth medium:
42% vol/vol Ham's F12 nutrient mixture (F12)
42% vol/vol Eagle's minimum essential medium (MEM)
15% vol/vol foetal calf serum (FCS)
2 mM L-glutamine
1% vol/vol non-essential amino acids (NEAA)
20 units/mL penicillin
20 mg/mL streptomycin
250 ng/mL amphotericin B

SH-SY5Y cell culture growth media was also used for M17 cell culture.
Trypsin-EDTA solution (1x) (Sigma-Aldrich)
Freezing solution 10% vol/vol dimethy sulphoxide (DMSO) (Sigma-Aldrich)
90% vol/vol SH-SY5Y cell culture medium

2.2.5 Luciferase assay reagents

2.2.5.1 Luciferase vectors

pGL4.10 *luc2* (For promoter constructs; Promega)

pRL-CMV (For internal *renilla* control; Promega)

2.2.5.2 Assay reagents

Dual-Glo Luciferase Assay System (Promega)

2.3 Suppliers

Suppliers of materials and services used throughout this study:

Abgene, Abgene House, Blenheim Road, Epsom, Surrey, KT19 9AP

Advanced Biotechnology Centre, Imperial College London, Faculty of Natural Sciences, 2nd Floor, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ

Amersham Pharmacia Biotech Ltd, Amersham Place, Little Chalfont, Buckinghamshire, HP7 9NA

GE Healthcare UK Ltd, Pollards Wood, Nightingales Lane, Chalfont St. Giles, Bucks, HP8, 4SP

Helena Biosciences Europe, Queensway South, Team Valley Trading Estate, Gateshead, Tyne and Wear, NE11 0SD

Invitrogen Ltd, 3 Fountain Drive, Ichinnan Business Park, Paisly, UK, PA4 9RF

New England Biolabs UK Ltd, 73 Knowl Piece, Wilbury Way, Hitchin, Hertfordshire, SG4 0TY

Promega UK Ltd, Delta House, Chilworth Research Centre, Southampton, SO16 7NS

Qiagen UK Ltd, Flemming Way, Crawly, West Sussex, RH10 9NQ

Sigma-Aldrich Company Ltd, Fancy Road, Poole, Dorset, BH12 4QH

3 Delineation of the PSP associated H1 haplotype at 17q21.31

3.1 Overview

In order to assess the extent of the linkage disequilibrium (LD) around the *MAPT* locus, single nucleotide polymorphisms (SNPs) were genotyped at regular intervals across the *MAPT* locus and flanking regions in subjects with progressive supranuclear palsy (PSP) and normal control subjects. Pairwise LD was found to extend for a total of ~2 Mb in both directions beyond *MAPT*. Association analysis of the genotypes revealed the entire LD block to be associated with PSP. This study delineated the outer limits of the PSP associated *MAPT* H1 haplotype and highlighted the possibility of further candidate pathogenic loci for study.

3.2 Background

Robust genetic association of *MAPT* with PSP and rare reports of families with more than one affected member indicate that genetic factors could play a role in PSP [200]. Conrad and colleagues first proposed that variation in *MAPT* itself could be an important genetic influence in sporadic PSP by demonstrating strong allelic association with PSP of a dinucleotide marker in *MAPT* intron 9 [167]. This association was subsequently confirmed in several other studies [168-171], indicating that the more common allele (a_0) was over-represented in this disorder. This suggested that either this polymorphism itself could contribute to increased risk or that it is in LD with the actual causative variant. Although some *MAPT* mutations in FTDP-17T cause a clinical and pathological phenotype similar to PSP [123, 129, 130, 164, 208], no obvious causative variations of *MAPT* have been identified in clinically and pathologically diagnosed sporadic PSP, which consists of the majority of cases.

The allelic association of *MAPT* with PSP was subsequently extended to a series of polymorphisms spanning nearly 62kb over the *MAPT* coding region. In 200

unrelated Caucasians, these polymorphisms were shown to be in complete LD, thereby forming two extended haplotypes, designated H1 and H2 [172]. The authors suggested that either there was a suppression of recombination, or that recombinants were selected against. They also demonstrated that the more common haplotype, H1, with which the a_0 allele segregated, was significantly over-represented in PSP [172]. Subsequent studies extended the *MAPT* haplotype a further 68kb to the promoter region of *MAPT* where three SNPs, highly associated with PSP, were in complete LD with the rest of the *MAPT* haplotype [198].

The aim of the work described in this chapter was to determine how far the region of LD extends beyond *MAPT* by delineating the outer edges of this extended region of LD. Knowing the extent of the *MAPT* haplotype block would enable further analysis of underlying variation in candidate genes within the H1 haplotype and establish H1-specific subtypes that could flag up potential causal variants.

3.3 PSP and Control Subjects

The control population (n=63; average age of death: 72 years) were of Caucasian origin from a European Brain Bank series and had neither clinical evidence of neurodegenerative disease nor abnormal histopathology. They have been used in previous studies [209]. The PSP cases (n=60; average age at death 77 years) also of Caucasian, western European origin were all pathologically confirmed following standardised criteria [210]. Most of the PSP cases have been used in previous studies and additional pathologically confirmed cases have been added to this study [209]. All cases and controls were collected under approved protocols followed by informed consent. This work was approved by the joint Medical Ethics Committee of the National Hospital of Neurology.

3.4 SNP selection

For this LD analysis, a series of SNPs from the SNP Consortium Database (http://snp.cshl.org) were selected covering regions of 1Mb at intervals of ~50kb,

extending from each edge of *MAPT* (5'-end of *MAPT* exon –1 and 3'-end of *MAPT* exon 14, respectively). Selection preference was given to validated SNPs.

3.5 Single nucleotide polymorphism genotyping

3.5.1 PCR

All genotyping was carried out by PCR-based methods using primers flanking each SNP. The 238bp *MAPT* H2 deletion in intron 9 (*del-In9*) was used to unambiguously assign the tau H1/H2 haplotype status in all individuals [172]. Oligonucleotide primer pairs were designed to specifically amplify by PCR the *MAPT* haplotype SNPs of interest (rs5911, rs895685, rs732589, rs2668643, rs1880748, rs110402, rs1396862, rs916793, rs1468241, rs1528072, rs2240756, rs142167, rs199528, rs70602, rs1662577, rs758391). The reactions contained 25ng of template DNA, 1 unit of Taq DNA polymerase, 10x PCR buffer, 10mM dNTP mix, 10 pmoles of each oligonucleotide primer pair and 5x Q solution. Thermocycling conditions were carried out using the "touchdown" method. Primer details are given in the appendix (Section 10.1) Genotyping of the SNPs was carried out either by Pyrosequencing or by restriction digests (RFLP).

3.5.2 Genotyping

3.5.2.1 Pyrosequencing

The SNPs rs110402, rs1396862, rs916793, rs1468241, rs1528072, rs2240756, rs142167 and rs70602 were analysed by Pyrosequencing. Samples were analysed using a PSQ 96 System together with SNP Software and SNP reagent kits (Pyrosequencing Inc., Westbourough, MA) following the manufacturer's instructions (Section 2.1.4.2).

3.5.2.2 Restriction fragment length polymorphism

For SNP analysis by RFLP, 15 µl of PCR product (SNPs : rs1880748 [*Bsp*H I (C)]; rs199528 [*Pvu* II (C)]; rs758391[*Hph* I (A)]; rs1662577 [*BsrG* I (C)]; rs2668643 [Apo I (A)] and rs894685 [*Acc* I (T)]) was digested by 1 unit of the corresponding

restriction endonuclease. The PCR products are cleaved by the corresponding enzyme once at the indicated (N) allele. Digests were run out on a 4% agarose gel for analysis.

3.5.3 Linkage disequilibrium

The square of the correlation coefficient (r^2) and D' was calculated pair-wise between each marker. The LD was calculated separately in the cases and in controls from the expectation-maximisation (EM) derived haplotype frequencies. Working outwards from the *MAPT* gene to locate the edges of this extended LD block in both the control and PSP cohorts, there was tight LD extending beyond *MAPT* in both directions and this was reflected by the match of genotypes at these loci. Both measures of LD are based upon D, the basic pairwise-disequilibrium coefficient, the difference between the probabilities of observing the markers independently in the population: $D = f(A_1B_1)-f(A_1)f(B_1)$ (ref), 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 other locus can always be predicted by the allele at the other locus. By contrast, D' can reach a value of 1.0 when the allele frequencies vary. An r^2 value of > 0.5 represents a high level of LD, r^2 values <0.1 represent essentially no LD.

3.6 Results

3.6.1 Linkage disequilibrium

The maximum extent of the *MAPT* haploype block was defined as a region of ~2Mb that is in strong LD. The centromeric end of the haplotype block could only be resolved to a minimum region of ~0.4Mb within which there is complete loss of LD.

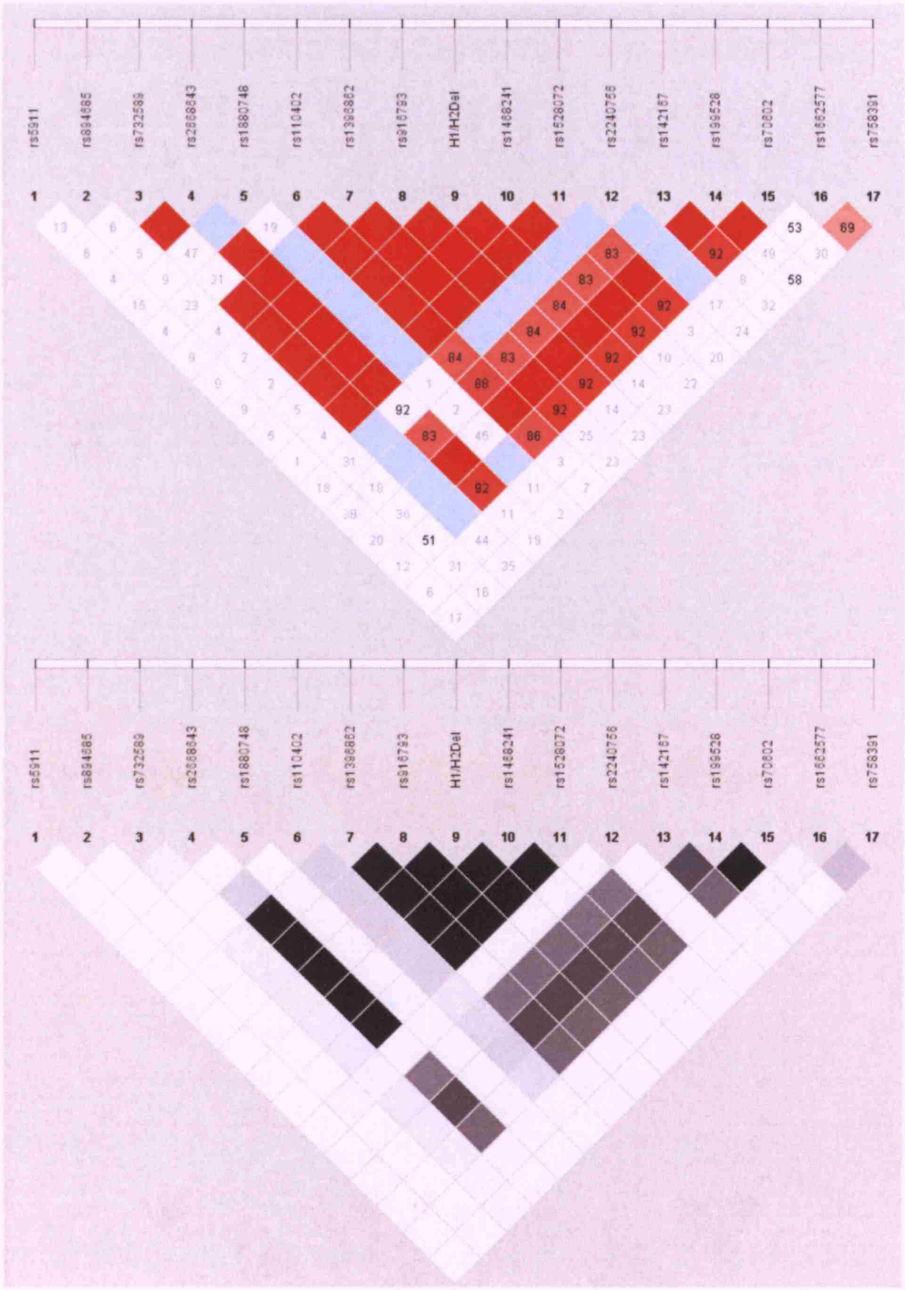

Figure 3.1 LD in the control population

Pair-wise D' (upper) and r^2 (lower) LD analysis of all the SNPs in the control population. The blocks are shaded corresponding to the values which were obtained from the LD analysis program Haploview. The darker the shading indicates a higher extent of LD between markers. Red and Black segments represent perfect pairwise LD.

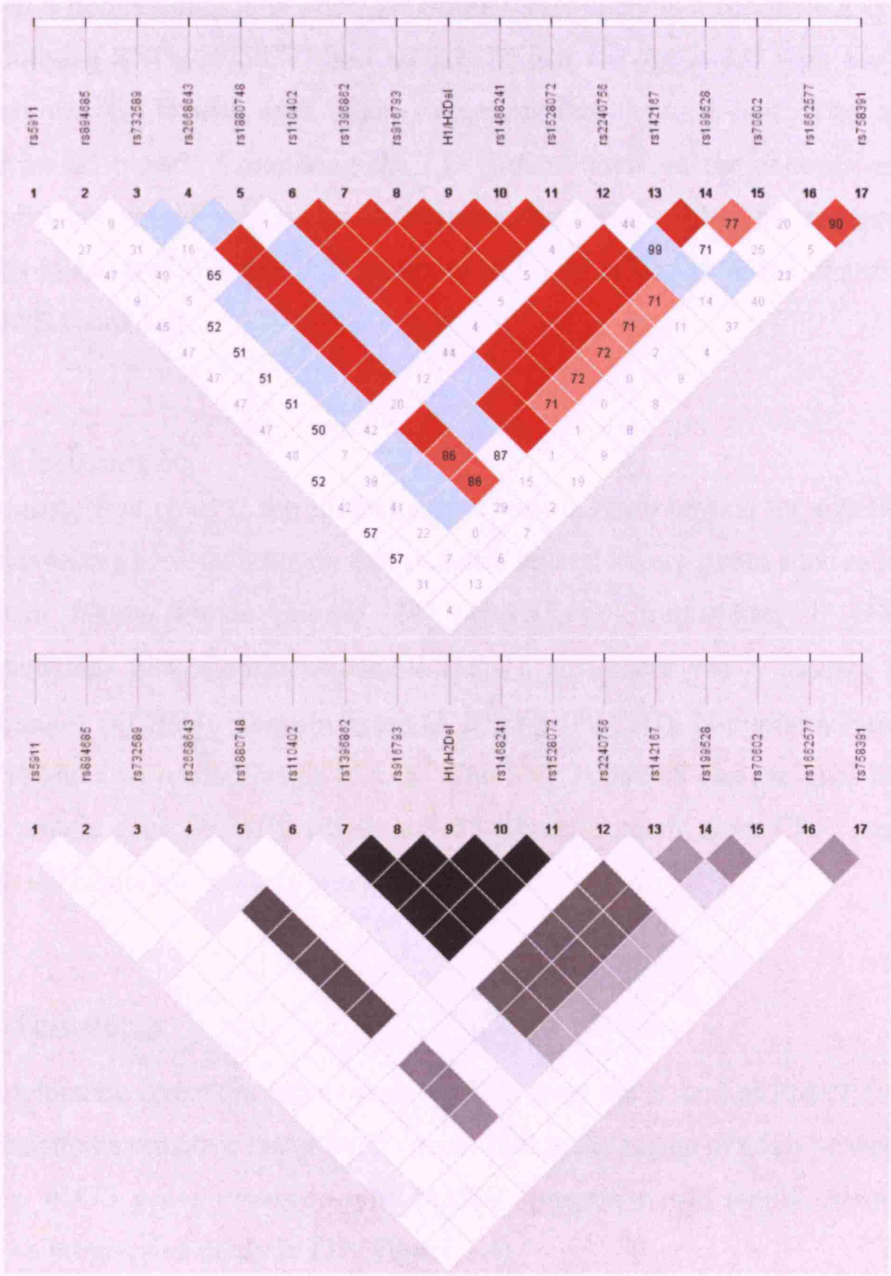

Figure 3.2 LD in the PSP population

Pair-wise D' (upper) and r^2 (lower) LD analysis of all the SNPs in the PSP population. The blocks are shaded corresponding to the values which were obtained from the LD analysis program Haploview. The darker the shading indicates a higher extent of LD between markers. Red and Black segments represent perfect pairwise LD.

By D', the haplotype block extends from rs732589 to rs70602 (**Figure 3.1 and 3.2**). At the telomeric end, which is marked by the *WNT3* gene, the haplotype block was

resolved a minimal region of ~150 kb within which there is a complete loss in LD. Two outlying SNPs, rs758391 and rs1662577 that are not in LD with the H1/H2 deletion, are in LD with each other, suggesting that they are part of an adjacent haplotype/LD block. 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 (**Figure 3.1 and 3.2**).

3.6.2 Centromeric

Centromeric from *MAPT*, the LD extends at least 0.39Mb beyond the *CRHR1* gene and ends within a ~0.4Mb region that includes several known genes such as mitogen activated kinase kinase kinase 14 (*MAP3K14),* formin-like 1 (*FMNL1*), hexamethylene bisacetamide-inducible *(HIS1),* acyl-coenzyme A binding domain containing 4 (*ACBD4*), phospholipase C, delta 3 (*PLCD3*), N-myristoyltransferase (*NMT1*) and C1q-related factor (*CRF*). The SNP rs894685 and the glial fibrillary acidic protein gene (*GFAP*), which is ~ 53 Kb centromeric from CRF, clearly lie outside the haplotype block (**Figure 3.4**).

3.6.3 Telomeric

In the telomeric direction, LD extends ~0.8Mb from the 3'-end of *MAPT* to the *N*-ethlymaleimide sensitive factor (*NSF*) gene. Within the region of 65kb between *NSF* and the *WNT3* genes (wingless-type MMTV integration site family, member 3), there is a progressive decay in LD (**Figure 3.4**).

3.6.4 SNPs in the extended haplotype block

In this study three types of SNPs were encountered:

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, *del-In9*, rs1468241 and rs1528072 had an r^2 value of >0.8 and SNPs rs142167, rs199528 and rs70602 showed a decline in r^2, and, to a lesser extent 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 the H1/H2 deletion polymorphism 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 reveal high LD according to the D' values. Using EM algorithm prediction of haplotypes that could be formed by these SNPs and the *del-In9*, it was noted that multiple haplotypes segregate with the H1 variant of *del-In9* whereas only one major form did 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 of which have relatively high minor allele frequencies) have extremely low LD when considered against one-another by r^2 and have lowered LD when measured with D'. This data suggests there is diversity and DNA recombination within the extended H1 haplotype.

In the third group of SNPs (rs5911, 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 variation and are not uniform relative to the H1 and H2 haplotypes confirming that they are not in LD with the *del-In9*.

			MAF		HWE		Case-Control Association	
SNP	Position	Assay	Controls	Cases	Controls (*p*)	Cases (*p*)	Genotypic (*p*)	Allelic (*p*)
rs5911	42928200	RFLP	0.344	0.31	0.54	0.92	0.473	0.527
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	44263995	RFLP	0.193	0.200	0.917	0.197	0.662	0.893
rs110402	44355458	PYRO	0.492	0.377	0.379	0.749	0.103	0.045
rs1396862	44378418	PYRO	0.194	0.070	0.282	0.139	0.004	0.005
rs916793	44430117	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	44671572	PYRO	0.189	0.070	0.328	0.139	0.007	0.006
rs1528072	44712118	PYRO	0.200	0.070	0.282	0.144	0.004	0.004
rs2240756	44735087	PYRO	0.143	0.313	0.769	0.463	0.004	0.001
rs142167	45269915	PYRO	0.186	0.036	0.078	0.780	0.002	0.002
rs199528	45317762	RFLP	0.196	0.083	0.121	0.324	0.007	0.017
rs70602	45334486	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	0.50	0.466	0.370	0.670	0.567	0.599

Table 3.1 SNPs in the extended *MAPT* haplotype

SNPs used in the analysis of the *MAPT* haplotype block. Approximate positions of the SNPs are given according to chromosome 17 co-ordinates from the July 2003 assembly. Also shown are comparisons of minor allele frequencies between PSP cases and controls together with P-values to fit HWE and case-control association of each SNP.

3.6.5 Association of the extended haplotype block with PSP

There was no significant deviation from Hardy-Weinberg equilibrium (HWE; significance level was set at $p \leq 0.05$) in the controls and PSP cases at any of the loci (**Table 3.1**). A case-control, locus-by-locus association study of all the SNPs used in this study was performed and all of the SNPs that were in LD with *del-In9* forming the extended haplotype block were associated with PSP (**Figure 3.3**). In addition, some of the H1-specific SNPs (rs2240756, rs110402) but not the others

(rs732589, rs1880748) showed significant allelic and genotypic association with PSP. The associated H1-specific SNPs flank *MAPT* and could be part of a sub-class of the H1 haplotype that includes the pathogenic variant(s). These SNPs would prove important for future study of the underlying architecture of the H1 haplotype in order to be able to progressively refine the candidate regions responsible for PSP.

Recent work suggests that r^2 is viewed more favourably as a measure of association for the causal variant than that of D'. 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 [211].

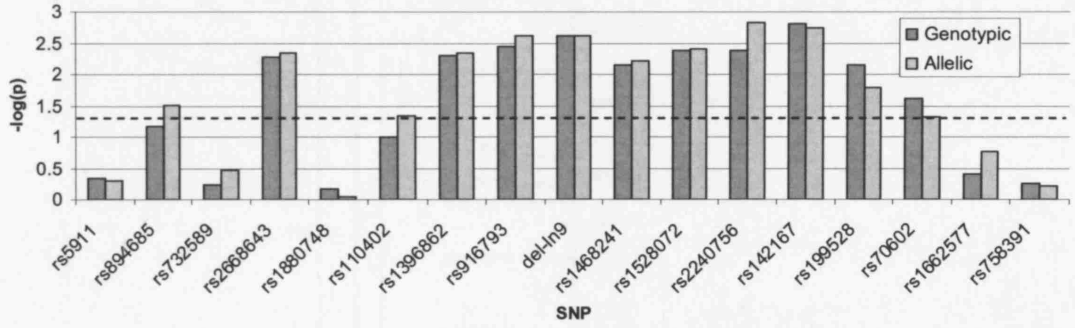

Figure 3.3 Associations of SNPs with PSP

Plot of –log(P) for case-control test of both allele and genotype association of the SNPs with PSP. The threshold of significance (α=0.05) is marked with a dotted line.

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. At the centromeric end of the haplotype, the outlying SNP rs894685 also shows evidence of allelic and genotypic association with PSP. This is surprising since 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. However, genotyping a further SNP (rs5911) that lies more centromeric from rs894685 revealed that it was completely out of LD and not associated with PSP.

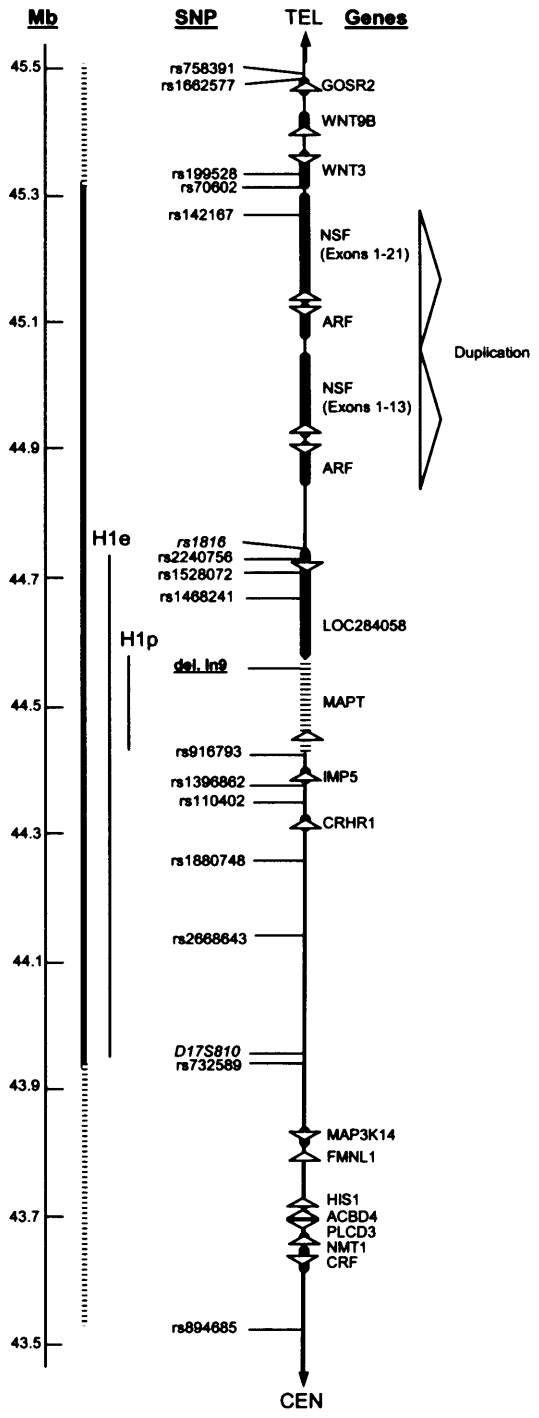

Figure 3.4 The extended haplotype block at 17q21.31

The 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 upon the July 2003 draft of the human genome sequence. Relative positions of the SNPs and confirmed genes are indicated. Arrowheads on genes indicate the direction of transcription. The extent of the previously reported *MAPT* haplotypes H1*p* [191] and H1e [193] are indicated by bars on the left. CEN, centromeric; TEL, telomeric.

3.7 Discussion

The *MAPT* haplotype block was defined as a region of ~2Mb on chromosome 17q21.31 that is in complete LD both pair-wise and relative to the H1/H2 defining *del-In9*. This block provides the candidate region for more detailed analysis of the underlying architecture of the H1 haplotype in order to pinpoint any variant involved in PSP pathogenesis. The extension and delineation of the outer limits of the *MAPT* haplotype block significantly increases the candidate region in both PSP and CBD, and all genes within this block are potential pathogenic candidates since they are all associated with PSP (**Figure 3.4**). 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 directly involved with tau function. These include: corticotrophin releasing hormone receptor 1 (*CRHR1*), presenlilin homologue 2 (*PS2*), Saitohin (*STH*), *NSF* and *WNT*3.

At this stage of the work it would however have been premature to speculate on the importance of the association with PSP of any single gene or polymorphism as it was clear that the entire *MAPT* haplotype block and the genes it contains are associated with PSP. However, based on the considerable evidence from tau related pathology in both PSP and FTDP-17T, the *MAPT* locus remained the main focus of attention in the remainder of this work in order to identify pathogenic polymorphism(s) on the H1 haplotype background.

There is some evidence to suggest that *CRHR1* is not involved in PSP pathogenesis [212]. The expression of the gene was measured in the globus pallidus by mRNA real-time PCR in PSP subjects compared to control groups and the expression pattern of *CRHR1* was found to be similar in all groups. In addition, PSP cases were sequenced and no associated coding polymorphisms were identified [212].

In a large panel of publicly available DNA samples representative of most racial groups, population genetic studies of the H1 and H2 haplotype frequencies (as defined by the *del-In9*) showed that the H2 haplotype is found almost exclusively in the and Caucasian ancestry, with European populations having H2 allele frequencies

of ~25%, central Asian (including Finnish) populations having H2 allele frequencies of ~5% and other populations (African, East Asian and native American) having H2 allele frequencies of essentially zero [213]. The pattern of linkage disequilibrium over the locus is shared by the different ethnic groups indicative this haplotype block is indeed ancient and at an estimated 3 million years old [214], predating the separation of the Caucasians [215, 216].

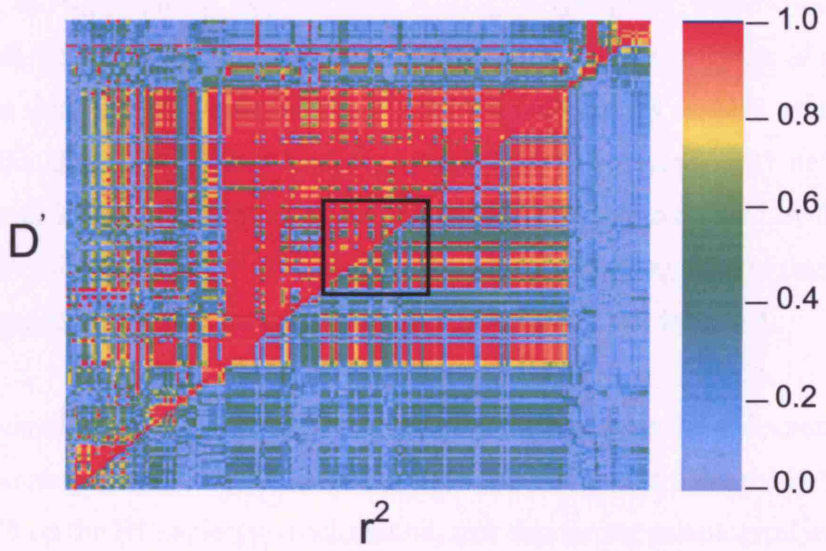

Figure 3.5 HapMap LD data at 17q21.31

Pair-wise LD analysis of HapMap project data at 17q21.31. The *MAPT* gene in the centre of the LD block is boxed.

After completion of the work in this chapter, the genotype data for SNPs in this area of chromosome 17q21.31 was made available on HapMap. The analysis of data for 178 SNPs covering a region of 2 Mb confirmed the validity of the work in this chapter and refines the original sizing of the LD block to 1.8 Mb.

The results of the work described in this chapter were published in *Human Molecular Genetics*:

Pittman,A., Myers,A.J., Duckworth,J., Bryden,L., Hanson,M., Abou-Sleiman,P., Wood,N.W., Hardy,J., Lees,A., de Silva,R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics 13(12), 1267-1274.

4 Dissection of the genetic association of *MAPT* gene haplotypes with progressive supranuclear palsy and corticobasal degeneration

4.1 Overview

In order to refine the association of *MAPT* with progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD), a systematic framework of genetic analysis was devised to investigate the common haplotype structure of the gene for genetic study in PSP case-control cohorts from both the UK and US, and CBD cases from the US. The LD and haplotype structure was analysed in 27 defined CEPH trios with 24 SNPs in relation to the H1 and H2 haplotypes defined by the bi-allelic *del-In9* polymorphism. A set of six haplotype-tagging SNPs (htSNPs) were subsequently identified for use in the association study that followed.

Two common haplotypes were shown to be associated with these disorders, defining a candidate region of ~56 kb spanning sequences from upstream of *MAPT* exon 1 to intron 9 on the H1 haplotype background, thus supporting pathological evidence that underlying variations in *MAPT* could contribute to disease pathogenesis possibly by subtle effects on gene expression, mRNA stability and/or splicing.

4.2 Background

The extended H1 haplotype, at the *MAPT* locus, is highly associated with PSP and CBD. Identifying the functional basis of the H1 haplotype association will be important in providing an insight into the aetiopathogenesis of these disorders.

As previously described in Chapter 3, the outer limits of the H1 haplotype were identified to physically cover a region of ~2 Mb [217]. Although all the genes within this multi-gene haplotype block by definition are associated with PSP, the hallmark tau pathology of these disorders strongly implicates *MAPT* itself. Controversially, recent work also shows a weak association of the H1 haplotype with sporadic Parkinson's disease (PD) and association with Norwegian PD cases of a sub-H1 haplotype spanning the 5' half of *MAPT* [218]. It is possible that

variations in *MAPT* increase neuronal vulnerability in PD, but without the formation of tau pathology.

The aim of this work was to exhaustively analyse the *MAPT* haplotype association with PSP in order to identify non-coding variants in H1 that could affect *MAPT* expression, splicing or processing. Secondary aims included assessing the association of the locus with CBD, examining the *MAPT* LD and haplotype structure in the Japanese population, that completely lacks the H2 haplotype and re-examining the association with PSP in respect of two clinical subtypes; Richardson's syndrome (RS) and PSP-parkinsonism (PSP-P) [219].

4.3 *MAPT* haplotype diversity in the CEPH-trios

For the primary genetic analysis of *MAPT*, genotype data from SNPs at *MAPT* in 27 CEPH-trios was downloaded from the International HapMap project web site (http://www.hapmap.org/). The CEPH-trios are families from Utah, USA, of northern and western European origin. The raw SNP genotype data was analysed using TagIt, a software suite for identifying and evaluating htSNPs, which also contains algorithms for inferring haplotypes from trio data, as structured in the CEPH families. SNPs that had a minor allele frequency of <5%, broke with Hardy-Weinberg equilibrium (HWE) or that contained inconsistencies through the parental-offspring relationship of the trios were removed from further analysis. A resulting set of 24 SNPs and the *del-In9* covering the entire *MAPT* gene from upstream of the core promoter to beyond exon 13 were used to infer haplotypes (**Table 4.1**). For ease of interpretation, the bi-allelic *del-In9* was designated a SNP. The average density of the markers was one SNP every 6.7 Kb.

MAPT haplotypes and their respective frequencies were obtained by using the EM algorithm tailored to analyse trio data as structured in the CEPH-trios. The algorithm combines information from resolved and unresolved chromosomes and, in those that are unresolved, restricts the set of possible haplotypes to those consistent with known data from both parents and child [220]. The EM-predictions depicted a total of 14 different *MAPT* haplotypes exceeding a frequency of 1%, three of which

are common, having a frequency of greater than 10% and the remainder less than 5% (**Table 4.2 and Figure 4.1**). Only one of the common predicted haplotypes is representative of the extended H2 as defined by the *del-In9*. The other two common variants are based upon H1 and differ from one another at multiple SNP loci. A further 10 rare variants of H1 are predicted and it is noteworthy that there is a single rare variant of H2 with an occurrence at less than 1%.

SNP	Position	dbSNP ID	Alleles	Ancestral	MAF	HWE
1	44411059	rs962885	C/T	T	0.361	0.572
2	44421540	rs1078830	C/T	C	0.189	0.426
3	44427146	rs2055794	A/G	A	0.189	0.442
4	44443848	rs7210728	A/G	A	0.259	0.248
5	44453262	rs1864325	C/T	C	0.189	0.426
6	44453969	rs1560310	A/G	G	0.189	0.442
7	44455965	rs3885796	G/T	C	0.189	0.426
8	44461645	rs1467967	A/G	A	0.352	0.851
9	44468843	rs3785880	G/T	T	0.462	0.709
10	44474041	rs1467970	G/T	T	0.189	0.442
11	44474234	rs767058	A/G	C	0.189	0.442
12	44481288	rs1001945	C/G	G	0.454	0.301
13	44494232	rs2435205	A/G	A	0.407	0.251
14	44495212	rs242557	A/G	G	0.396	0.854
15	44502238	rs242562	A/G	G	0.189	0.684
16	44528923	rs2217394	A/G	G	0.189	0.442
17	44529908	rs3785883	A/G	G	0.204	0.524
18	44531122	rs754512	A/T	T	0.189	0.442
19	44538720	rs2435211	C/T	C	0.368	0.061
20	44549365	rs1052553	A/G	G	0.185	0.442
21	44551539	rs2471738	C/T	C	0.287	0.335
22	44562589	*del-In9*	+/-	+	0.177	0.617
23	44565039	rs733966	C/T	C	0.189	0.442
24	44577047	rs9468	C/T	C	0.189	0.442
25	44580881	rs7521	A/G	G	0.434	0.569

Table 4.1 *MAPT* HapMap SNPs

The 24 SNPs and *del-In9* used for the LD and haplotype structure analysis of *MAPT* in the CEPH-trios. The analysis was performed on the genotype data for these SNPs from HapMap (http://www.hapmap.org/). In addition the *del-In9* was genotyped 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).

By resolving parental chromosomes, 34 haplotypes were phased, representing 42% of the total number of haplotypes (**Table 4.2**). Although 'phase-known', the frequency of these haplotypes is likely to be a distortion since they are only

representative of a fraction of the whole population, thus EM-trio haplotype estimation was used for subsequent htSNP selection because of a full utilization of the CEPH-trio population data.

ID	Haplotype	EM (%)	Resolved
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	Y
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	Y
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	Y
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	N
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	Y
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	Y
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	...	Y
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	...	Y
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	...	Y
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	...	Y
K	0 1 1 1 0 1 1 1 1 1 1 0 1 1 0 0 0 0 1 1 0 **1** 0 0 1 1	...	Y
L	0 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 0 0 **0** 0 1 1	...	Y
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	...	Y
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	...	Y
O	1 1 1 1 0 1 1 1 1 1 1 0 1 0 0 0 0 0 1 1 0 **1** 0 0 1 1	...	Y
P	1 1 1 1 0 1 1 1 1 1 1 0 1 0 1 1 0 1 1 0 0 0 **0** 0 1 0	...	Y
Q	1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 1 1 0 **1** 0 0 1 1	1.9	N
R	1 1 1 1 0 1 1 1 1 1 1 0 1 0 0 0 0 1 1 0 0 0 **0** 0 1 0	1.9	N
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	N
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	N
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	N
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	N
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	N
X	1 1 1 1 0 1 1 1 1 1 1 0 1 0 0 0 0 1 1 0 0 0 **0** 0 1 1	1.9	N
...	1 0 0 0 0 1 - 0 1 1 - 1 0 1 1 1 1 1 1 0 1 0 **0** 0 0 1	**Ancestral**	

Table 4.2 *MAPT* haplotype structure in the CEPH-trios

The haplotype structure of the *MAPT* gene in CEPH-trios based upon the 25 markers in **Table 4.1**. 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 Expectation-Maximisation (EM-trio) inferred haplotype frequency exceeded 1%. Bold allele is the *del-In9* (0=H1, 1=H2). Additionally presented is the build of the ancestral haplotype (Chimpanzee).

The ancestral (chimpanzee; *Pan troglodytes*) haplotype was constructed based on that of the 24 SNPs and the *del-In9* (**Table 4.2**). This appears not to resemble any haplotype present in the CEPH-trios, though is closest to H2, but still differing by ten loci. The remaining loci are either consistent with the extended H1 (SNPs 1, 5, 6, 10, 12, 18, 23, for rs numbers please see **Table 4.1**), including the presence of the 238 bp *del-In9*, or the ancestral allele is not observed in *Homo sapiens* (SNPs 7 and

11). The ancestral allele at each locus was determined by direct sequence comparison of the SNP loci in human and chimpanzee *MAPT* and by searching for the ancestral allele in genome database (NCBI).

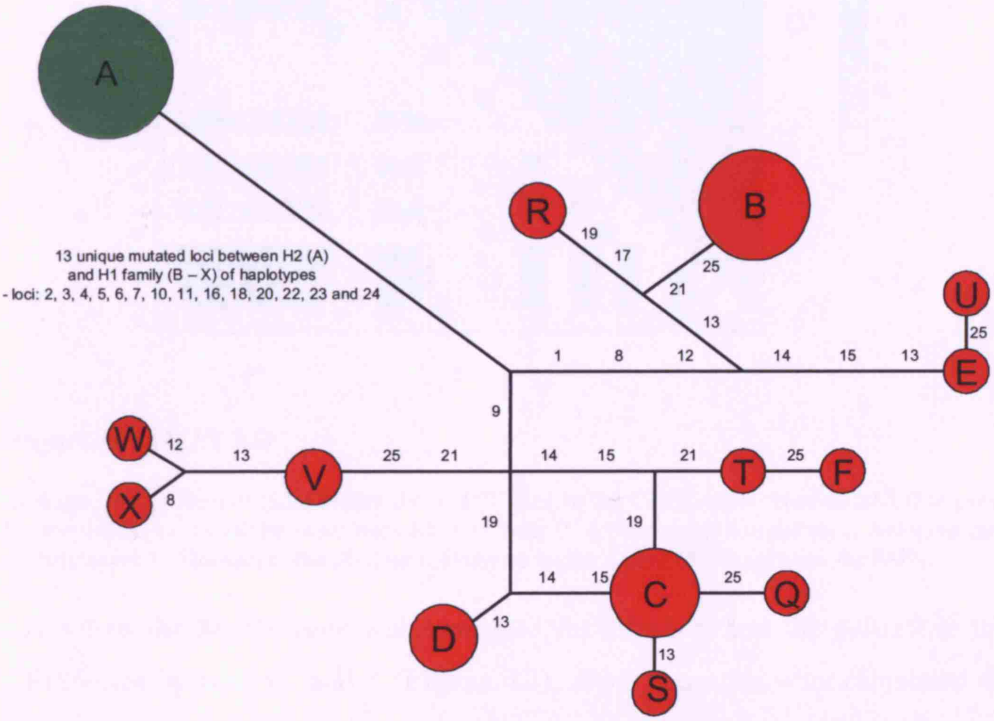

Figure 4.1 *MAPT* haplotype network

Reduced median network of *MAPT* haplotypes (frequencies that exceed 1% by EM - from **Table 4.2**) in the CEPH-trios. Node size is proportional to haplotype frequency. Any two given haplotypes differ by the SNP(s) (as numbered from **Table 4.1**) along the lines that connect them. The sole H2 haplotype (A) is green and the H1 family is represented in red.

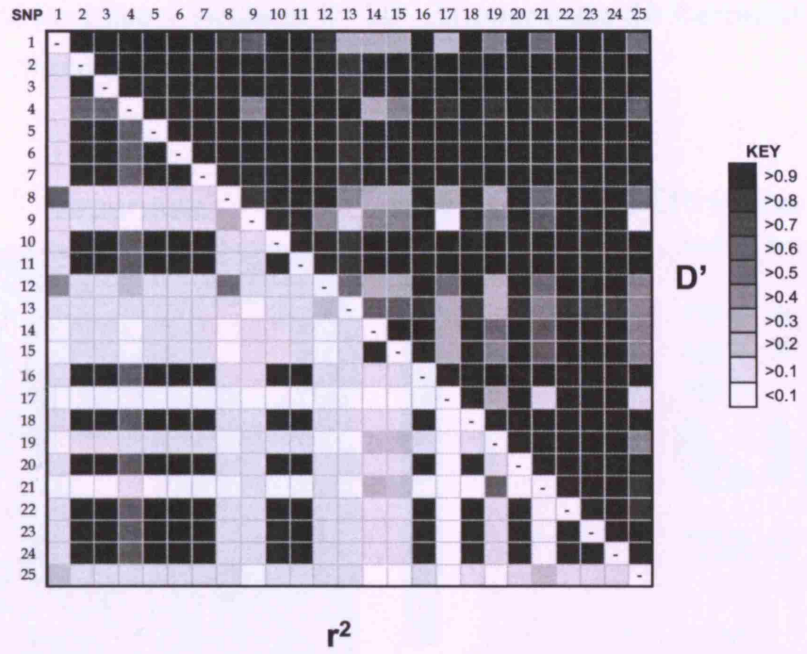

Figure 4.2 *MAPT* LD

Linkage Disequilibrium (LD) across the *MAPT* gene in the CEPH-trios. Numerical LD is presented by grey-scale, pair-wise between each SNP by both D' (upper right) and the more stringent measure r^2 (bottom left). The darker the shading indicates a higher extent of LD between the SNPs.

LD across the *MAPT* gene was evaluated for all SNPs and the *del-In9* in the 27 CEPH-trios by both D' and r^2 (**Figure 4.2**). Both measures were calculated firstly by estimating pair-wise haplotype frequencies by EM-trio, then assessing the statistical strength of association via a likelihood ratio test (LRT) by comparing the EM frequencies with haplotype frequencies, assuming no LD. The entire *MAPT* gene is featured by significant LD evident by the measure of D', though by the more stringent measure of r^2 it appears more fragmented with evidence of variability on the background of the extended H1. LD correlation by D' between many SNPs is low, suggesting a degree of linkage equilibrium between them and an unremarkable pattern of LD across extended H1 in *MAPT*.

4.4 The *MAPT* locus and the Japanese population

The *MAPT* LD pattern in the Japanese population is very different from that observed in Caucasians (**Figure 4.4**). This is because there is no H2 present in this population and it changes the LD structure remarkably; the LD is fragmented across

the *MAPT* gene into 4 blocks of D' LD. In comparison the Caucasian population *MAPT* is characterised by a single large block of D' LD.

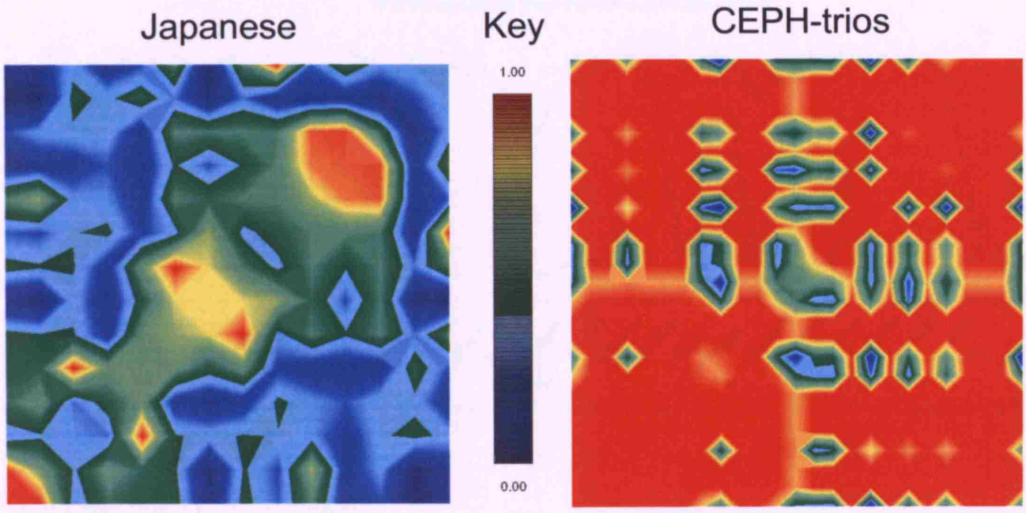

Figure 4.3 Comparison of the Japanese LD pattern versus the CEPH-trios

Linkage disequilibrium (LD) across *MAPT* in the Japanese population compared to that of the CEPH-trios (European ancestry). The data shown is a graphical overview (Image generated from pairwise D' LD statistics in GOLD program) of the LD and demonstrates a substantial difference between the two populations.

4.5 Selection of haplotype-tagging SNPs

From the EM-trio predicted haplotypes in the CEPH-trios, htSNPs were selected to capture the 'known' genetic diversity of the original 24 SNPs and the *del-In9* in the CEPH-trios. The approach focuses on the coefficient of determination in a linear regression (haplotype r^2, criterion 5 in TagIt, version 1.19), which uses the haplotypes defined by the htSNPs to predict the state of the tagged-SNPs. Criterion 5 in the programme is specifically designed to select htSNPs for use in genetic association studies. Six htSNPs (SNPs 8, 14, 17, 21, 22 (*del-In9*) and 25, see **Table 4.1**) are sufficient to represent all the HapMap SNPs in the CEPH trio-population with a high coefficient of determination.

Using the TagIt program, a minimal set of 9 htSNPs was selected from the original 15 SNPs in the Japanese population. (SNPs: 2, 4, 5, 6, 8, 10, 12, 14 and 15, **Table**

4.3, average haplotype r^2 =0.99). These can be utilised for future *MAPT* association studies in the Japanese population.

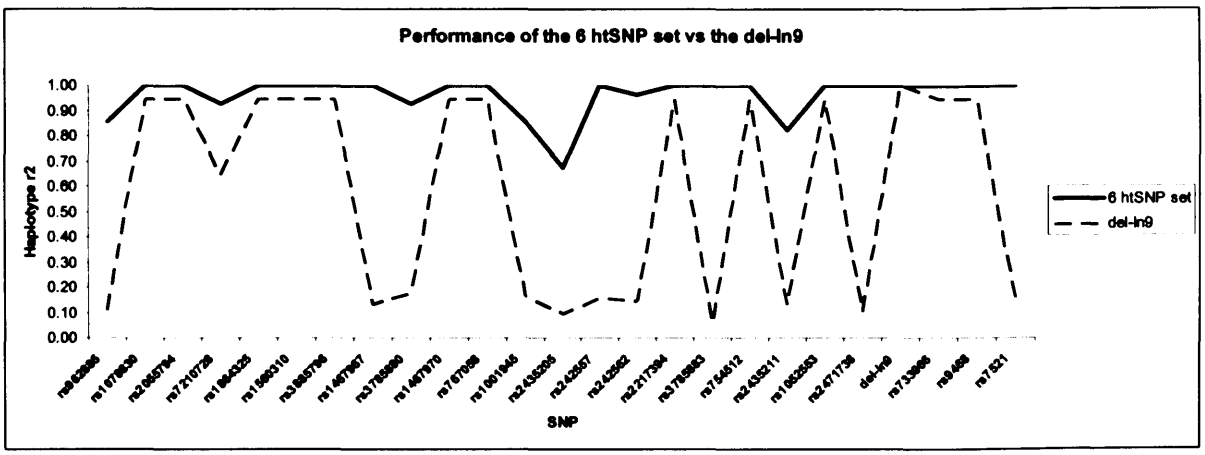

Figure 4.4 htSNP performance

Performance plot of the 6 htSNP set and the *del-In9* marker alone against the original set of 25 SNPs analyzed in the CEPH-trios.

Five of these SNPs are H1-specific that is, they vary only on the H1 background. The bi-allelic *del-In9* marker distinguishes haplotypes of the H1 and H2 type. The performance value of the set of six htSNPs is interpreted at an average haplotype r^2 value of 0.95 (95%) and as the minimum single locus value of 0.68 (68%), rs2435205; SNP number 13 (**Figure 4.3**). Interestingly, excluding the *del-In9* from the htSNP set results in a loss of performance of just 3%, with the overall performance down to 92% with the remaining 5 H1-specific SNPs. This is because a particular allelic combination of the 5 H1-specific SNPs is representative of the extended H2 haplotype and thus this illustrates the rationale of haplotype r^2. The performance value of just the *del-In9* against all the known SNPs in the CEPH-trios is an inadequate 50% (the average haplotype r^2).

4.6 Case-control samples

4.6.1 US series

The unrelated US PSP cases (n=238; 50% males) from the Mayo Clinic Brain Bank, Jacksonville, consisted of cases pathologically confirmed by standard criteria, with

an average age at death of 75.3 years. The unrelated US CBD cases (n=44; 50% males) also consisted of cases pathologically confirmed, by standard criteria with an average age at death of 71.3 years. The unrelated US control population consisted of individuals (n=131; 50% males) without a history of neurological disease, free of abnormal histopathology and an average age at death of 79.9 years.

4.6.2 UK series

The unrelated PSP cases, (n=83) from the Queen Square Brain Bank for Neurological Disorders, were all Caucasian, of western European origin and all pathologically confirmed. The majority of these cases have been used in previous studies [209, 217]. The unrelated British control population (n=169), all Caucasian, were taken from brain bank tissue with no clinical evidence of neurodegenerative disease and no abnormal histopathology, from the MRC, MRC Building, Newcastle, UK. The samples were age matched, where the average age at death was 73.5 years for the PSP cases (63% males) and 76 years for the controls (51% males). 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.

4.6.3 Japanese control series

The unrelated Japanese controls obtained from the RIKEN Brain Research Institute, Saitama, Japan consisted of healthy individuals (n=95) with a male to female ratio of 50% and an average age of 47 years.

4.7 SNP amplification and genotyping

4.7.1 The htSNPs for genetic association study

The htSNPs (dbSNP ID numbers: rs1467967, rs242557, rs3785883, rs2471738, rs7521 and the *del-In9*) were genotyped in the PSP and CBD case-control cohorts as follows. PCR primer pairs were designed by the Primer3 program and used to

amplify regions containing each polymorphism. The *del-In9* polymorphism was genotyped by PCR followed by 2% agarose gel electrophoresis and visualization with ethidium bromide staining, for determining the presence or absence of the 238 bp deletion. Genotyping of the SNPs, rs1467967, rs242557, rs3785883, rs2471738 and rs7521 was conducted by by RFLP; the following restriction endonuclease cutting the PCR product once at the (N) allele, respectively of each SNP: *Dra I* (A), *Apal I* (A), *BsaH I* (G), *BstE II* (T) and *Pst I* (A). PCR products were incubated with 1-2 units of the corresponding restriction enzyme at the manufacturers recommended temperature and analysed on 4% agarose gels. Accuracy of interpretation of genotypes was achieved by re-genotyping sets of htSNPs and samples for comparisons.

4.7.2 SNPs to determine the *MAPT* LD structure in the Japanese population.

To genotype the Japanese control population for haplotype and LD analysis, *MAPT* SNPs were selected from the JP SNP database (http://snp.ims.u-tokyo.ac.jp/) and/or selected because they were found to be polymorphic on the H1 background in the Caucasian population. SNPs that defined the H1 and H2 haplotype distinctions, as expected, were not polymorphic in the Japanese population and were excluded from further analysis. **Table 4.3** summarises the final set of SNPs used in the analysis with their minor allele frequencies (MAF), Hardy-Weinberg equilibrium (HWE and the restriction enzyme's used to genotype the SNPs. Genotyping the *del-In9* confirmed the complete absence of H2 chromosomes.

The haplotype frequencies formed by these SNPs were estimated by using an EM algorithm, specifically the PL-EM (partition ligation-expectation maximization) algorithm [221]. Predicted haplotypes were used to calculate pairwise D' LD between individual SNPs via a LRT test in TagIt and the results are summarised in **Figure 4.4**.

SNP	Position	dbSNP ID	Alleles	MAF	HWE	Enzyme
1	41210672	rs4074461	(A/G)	0.194	0.149	*Sma I*
2	41291377	rs2301689	(A/G)	0.404	0.683	*Fsp I*
3	41298590	rs242928	(C/T)	0.348	1.000	*Mbo II*
4	41335518	rs930119	(A/G)	0.451	0.421	*Bsm I*
5	41340709	rs2280004	(C/T)	0.180	0.867	*BstFS I*
6	41342006	rs1467967	(A/G)	0.418	0.284	*Dra I*
7	41375548	rs242557	(A/G)	0.400	0.268	*ApaL I*
8	41381498	rs242559	(A/G)	0.042	1.000	*Apo I*
9	41382599	rs242562	(A/G)	0.403	0.279	*Xho I*
10	41410268	rs3785883	(A/G)	0.269	0.604	*BsaH I*
11	41414515	rs2435207	(A/G)	0.253	0.966	*Bcl I*
12	41422969	rs2258689	(C/T)	0.500	0.958	*BmGB I*
13	41431900	rs2471738	(C/T)	0.280	0.423	*BstE II*
14	41448663	rs11568306	(C/A)	0.399	1.000	*Pvu II*
15	41461242	rs7521	(A/G)	0.097	0.805	*Pst I*

Table 4.3 *MAPT* SNPs in the Japanese population

SNPs genotyped a Japanese control group. Included is the db SNP ID number, the SNP alleles, the minor allele frequency of the SNP in the population, *p*-values for test to fit HWE and the restriction enzymes used for genotyping assays.

4.7.3 ApoE genotyping

ApoE genotyping was performed as previously described [222]. Briefly, the SNP-containing region is amplified by PCR (Forward: 5'-TAAGCTTGGCACGGCTGTCCAAGG-3' and Reverse: 5'-ACAGAATTCGCCCCGGCCTGGTACACTGCC-3') followed by restriction enzyme incubation (*Hha I*) of the PCR product to generate the specific combination from 5 allele discriminating fragments. The ApoE region is GC rich and requires a modified PCR protocol: this includes a longer initial denaturising step of 10 minutes at 94° C and an additional 10 seconds in the denaturising steps in each cycle. A primer annealing temperature of 56° C is used. Digested PCR products are run on a 4% low melting agarose gel and stained with ethidium bromide to visualize the allele discriminating bands.

4.8 Results

4.8.1 Single locus association

Single locus association results are summarized in **Table 4.4**. In all groups there were no significant deviations from HWE at any of the htSNP loci. The association of the H1/H2 *del-In9* with PSP was recapitulated in both the UK and US cohorts (p=1.14x10^{-5} and 4.02x10^{-8}, respectively, **Table 4.4**). This trend was observed in CBD but the difference was not significant, probably due to a small sample size. No evidence of association was found for htSNPs 8, 17 and 25 in the studies except for CBD, where htSNP 17 is associated (p=0.019). The odds-ratios (ORs) and their 95% confidence intervals were calculated for each htSNP by comparison of each minor allele verses each major allele. The H2 haplotype as defined by the *del-In9* is a significant protective factor. The H1-specific SNPs rs242557 and rs2471738 both have significant ORs and are highly associated with both PSP and CBD and thus the single locus results implicate genetic association of variability on the H1 background in *MAPT*.

4.8.2 Study power

For each associated htSNP in each study, power was assessed and is summarised in **Table 4.5**. In the US PSP study, power was well above 80% for each associated htSNP and approached 100% for both the *del-In9* and rs242555 thus giving confidence to the findings. In the UK PSP series power exceeded 80% for the *del-In9*; however, both rs242557 and rs2471738 fell short of this. Ideally, a larger UK case-control cohort would give greater confidence for the association study in this population though the results do represent a direct replication of the findings in the more powerful US PSP series.

In the US CBD series power falls short of 80% for all loci and so these findings should be regarded with some caution. These findings should be regarded as preliminary and hypothesis generating, rather than hypothesis testing, and should be replicated in a larger CBD study cohort.

US PSP

		Frequency (*F*1)		Association (*p*)		Odds Ratio (MA)	
htSNP	**dbSNP ID**	**Cases**	**Controls**	**Allelic**	**Genotypic**	**OR**	**95% CI**
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 3.255
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

		Frequency (*F*1)		Association (*p*)		Odds Ratio (MA)	
htSNP	**dbSNP ID**	**Cases**	**Controls**	**Allelic**	**Genotypic**	**OR**	**95% CI**
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**	**1.8153**	**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**	**2.1421**	**1.368 to 3.355**
del-In9	…	93.2	76.6	**$1.14ex^{-5}$**	**$5.31ex^{-5}$**	**0.2148**	**0.099 to 0.466**
25	rs7521	51.2	45.7	0.546	0.814	0.7729	0.505 to 1.183

US CBD

		Frequency (*F*1)		Association (*p*)		Odds Ratio (MA)	
htSNP	**dbSNP ID**	**Cases**	**Controls**	**Allelic**	**Genotypic**	**OR**	**95% CI**
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

Table 4.4 Single locus association

Allele frequencies (*F*1) and *p*-values of single-locus association in the three studies. 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; http://www.smd.qmul.ac.uk/statgen/dcurtis/software.html).

**A genotypic test was not performed for the *del-In9* in intron 9 in the CBD series, since there were no rare homozygotes in the CBD cases, thus preventing a valid test. Significant single-locus associations of htSNPs are indicated in bold. Odds ratios and their 95% confidence interval are presented for the minor allele (MA) verses the major allele for all htSNPs.

Study	*del-In9*	rs242557	rs2471738
US PSP	97%	97%	88%
UK PSP	96%	60%	72%
US CBD	23%	62%	51%

Table 4.5 Study power

Calculations for power in all three studies. Power calculation were made in PS (power and sample size calculator; http://www.mc.vanderbilt.edu/prevmed/ps/index.htm) based upon the obtained ORs, the study sample sizes, the frequency of the risk variants and at a two sided significance threshold of α=0.05.

4.8.3 Haplotype association

The six identified htSNPs capture 95% of the common haplotypic diversity of *MAPT* and EM estimated haplotype frequency differences between cases and controls was statistically assessed in an omnibus test of all haplotypes >1% in frequency. The haplotype distribution was significantly different in the UK PSP cohort (p=9.75x10^{-5}, d.f =19) and in the US PSP cohort (p=7.40x10^{-12}, d.f = 20) but not in the US CBD cohort (p=0.120, d.f. =17).

In addition to the omnibus test, haplotype-specific tests (i.e. d.f = 1) were performed to identify individual haplotypes responsible for the global association. Two common haplotypes, IDs A and C, were found to be strongly associated with PSP in both the UK and US series (**Table 4.6**). Haplotype A, which derives from the *del-In9* H2 is the most common haplotype in the controls and is significantly under-represented in both PSP groups (p=2.01x10^{-7}, and p=2.77x10^{-4}, d.f. =1 for the US and UK populations respectively). Haplotype C, a variant of H1 is significantly overrepresented on PSP and is the most common haplotype in PSP but not in the controls (p=6.42x10^{-7}, and p=0.022, d.f. =1 for the US and UK populations respectively). The most common H1-derived haplotype in the control population is haplotype B, and is not associated with either PSP or CBD. These trends were observed in CBD, though on correction for multiple comparisons, no *MAPT* haplotype is significantly associated with this disorder. In both PSP cohorts, after strict correction of p-values according to the number of tests performed, both

haplotypes A and C remain significant. Haplotypes A and C, derived from H2 and H1 respectively, differ in alleles by only two H1 specific SNPs, 14 and 21 (rs242557 and rs2471738, respectively, **Table 4.1**) that, in addition to the *del-In9*, also show significant single locus associations. Haplotypes A and C do not have allelic differences by htSNPs 8 and 25 (rs1467967 and rs7521, respectively, **Table 4.1**), 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 (H1c).

Table 4.6 Haplotype association

Association of common *MAPT* haplotypes with PSP and CBD (next page). The 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. The statistical comparisons were made in WHAP software and the frequencies of the haplotypes were calculated in SNPHAP.

ID	htSNP Haplotypes								UK PSP				US PSP				US CBD			
	rs1467967	rs2422557	rs3785883	rs2471738	del-Ins9	rs7521			Frequency (%)		Association (*LRT*)		Frequency (%)		Association (*LRT*)		Frequency (%)		Association (*LRT*)	
									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)	

To assess whether the significant association with PSP of any of the H1-specific htSNPs are independent to that of the *del-In9*; H1/H2, each htSNP was analysed after entering the *del-In9* as an explanatory factor to the logistic regression model. Significant associations of single-locus htSNPs were found for 14, 17 and 21 ($p=9.00x10^{-6}$, $2.87x10^{-3}$ and $2.73x10^{-3}$, respectively) for the US PSP cases, htSNP 21 ($p=0.0421$) for the UK PSP cases and 14 and 21 ($p=0.0183$ and 0.0436) for the CBD cases. Haplotype analysis on the sub-set of htSNPs 14, 17 and 21 (db SNP ID rs242557, rs3785883 and rs2471738, respectively) with the *del-In9* entered as an explanatory factor revealed significant association in both PSP (UK and US) and CBD.

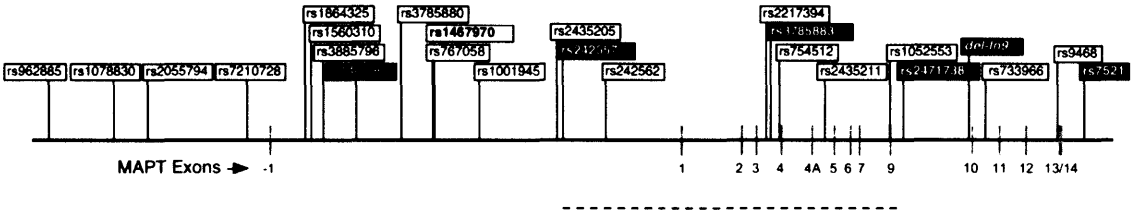

Figure 4.5 Distribution of the SNPs in *MAPT*

Genomic organisation of *MAPT* and the distribution of the 25 markers used in the LD and haplotype analysis in the CEPH-trios. Relative positions of the promoter, coding exons and SNP markers in the HapMap CEPH-trios are to scale. Chromosome co-ordinates (bp) are according to the March 2004 build of the Human Genome Sequence (http://genome.ucsc.edu). The htSNPs selected for the study are indicated by the black boxes. The minimal candidate region as identified in this study as to where a potential causal variant(s) may reside on the H1-background is indicated by the blue dotted line.

This three-locus haplotype remained significant when the *del-In9* was not included as an explanatory factor. Haplotype b (A-G-T) was greatly over-represented in each group and the haplotype a (G-G-C) under-represented (**Table 4.7**). The constituent SNPs 14, 17 and 21 (db SNP ID rs242557, rs3785883 and rs2471738, respectively) are H1-specific SNPs and although they vary only on the H1 background, haplotype a is representative of H2 in addition to H1-derived variants. These SNP haplotypes span a region, from rs242557 to rs2471738 on the H1 background. This minimal candidate region incorporates ~56.3kb of sequence (dotted line, **Figure 4.5**), from upstream of exon 1 downstream to intron 9 that could harbour potential causal variant(s) that are in the LD with these SNPs or are indeed themselves causal.

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)
a	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)**
b	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)
c	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)
d	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)
e	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)
f	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)
g	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)
h	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)

Table 4.7 Sub-haplotype association

Association of the sub-set of htSNP haplotypes with PSP and CBD. The haplotype analysis was based on a sub-set of H1-specific htSNP-defined haplotypes that showed evidence of association after consideration of the *del-In9*. After correction of *p*-values for multiple testing (bracketed *p* values), haplotypes **a** and **b** in both PSP studies and haplotype **a** in the CBD study are significant. The statistical comparisons were made in WHAP software and the frequencies of the haplotypes were valculated using SNPHAP.

4.8.4 PSP clinical subtypes and the H1c haplotype

A notable proportion of pathologically diagnosed PSP cases who prove difficult to diagnose during life do not develop the characteristic signs and clinical symptoms are labelled as atypical PSP [223]. In 103 consecutive cases of pathologically confirmed PSP two distinct clinical phenotypes were identified which were named Richardson's syndrome (RS) and PSP-parkinsonism (PSP-P) [219]. Cases of RS made up 54% of all cases and were characterised by the early onset of postural instability and falls, supranuclear gaze palsy and cognitive dysfunction. A second group (32%) were characterised by asymmetric onset, tremor, a moderate initial therapeutic response to levodopa and were frequently confused with Parkinson's disease (PSP-P). A third group (14%) could not be separated according to either group.

Of the cases from the Queen Square Brain Bank with a pathological diagnosis of PSP, 52 patients were classified as RS, 19 as PSP-P [219]. In the other cases, there was insufficient information to assign a clinical phenotype or they could not be classified into either group. Genotypes and allele frequencies were significantly different from controls in both RS and PSP-P for the three previously associated SNPs tested (**Table 4.8**).

	del-In9			rs242557 (A/G)			rs2471738 (C/T)		
	MAF	***p***	**OR**	**MAF**	***p***	**OR**	**MAF**	***p***	**OR**
Controls	27 (H2)			35 (A)			20 (T)		
All PSP	4 (H2)	<0.001	0.1 (0.04-0.24)	49 (G)	0.002	1.9 (1.3-2.8)	34 (T)	0.001	2.2 (1.5-3.3)
RS	3 (H2)	<0.001	0.08 (0.02-0.25)	48 (G)	0.006	2.0 (1.3-3.2)	37 (T)	0.001	2.4 (1.5-3.8)
PSP-P	5 (H2)	<0.001	0.15 (0.05-0.49)	48 (A)	0.060	1.7 (0.9-3.0)	37 (T)	0.006	2.3 (1.3-4.2)

Table 4.8 *MAPT* association with RS and PSP-P

Significant allelic association of *MAPT* SNPs (*del-In9,* rs242557 [A/G] and rs2471738 [C/T]) with RS and PSP-P. Shown are the minor allele frequencies (MAF, %), p-values and odds ratios (OR) and their 95% confidence intervals.

	Allele Counts		
	2 (%)	**3 (%)**	**4 (%)**
All PSPs	15 (11)	107 (79)	14 (10)
RS	6 (9)	52 (82)	5 (8)
PSP-P	3 (8)	30 (83)	3 (8)
Controls	25 (9)	206 (77)	37 (14)

	Genotype Counts					
	2.2 (%)	**2.3 (%)**	**2.4 (%)**	**3.3 (%)**	**3.4 (%)**	**4.4 (%)**
All PSPs	1 (1)	11 (16)	2 (3)	43 (63)	10 (15)	1 (1)
RS	1 (3)	4 (13)	1 (3)	22 (69)	4 (13)	0 (0)
PSP-P	0 (0)	3 (16)	0 (0)	13 (72)	1 (6)	1 (6)
Controls	0 (0)	20 (15)	5 (4)	78 (58)	30 (22)	1 (1)

Table 4.9 ApoE genotype and RS and PSP-P

Summary of the allele (upper) and genotype (lower) counts and frequency in controls, all PSPs, RS and PSP-P. No significant association was found by performing 10,000 permutations in CLUMP software.

A constrained test of association was performed of the H1c haplotype alone to negate the multiple testing confounds of testing each individual haplotype for association in 1 d.f. tests. The H1c haplotype was significantly more frequent in both RS ($p<0.001$) and PSP-P ($p=0.022$) than in controls. Thus the genetic risk

factor associated with H1c *MAPT* haplotype previously identified in pathological PSP is present in both clinical subtypes of PSP; RS and PSP-P.

The distribution of ApoE alleles and genotypes was investigated in the two clinical subtypes to see if it could account for the clinical differences. The results are summarized in **Table 4.9**; no significant differences (p>0.05) were found between the PSP and the control group or between either RS and PSP-P and the controls.

4.9 Discussion

The underlying LD and haplotype diversity of *MAPT* was successfully investigated using high-density genotype data from HapMap in 27 CEPH-trios. In addition a panel of SNPs were genotyped in a Japanese population sample. The LD structure is clearly different between the Caucasians and Japanese populations in that the absence of H2 in the latter results in a fragmented *MAPT* LD and haplotype structure compared to the monolithic nature of the *MAPT* LD pattern in Caucasians.

Using association-based criteria, a set of six htSNPs were assigned for genotyping in Caucasian PSP and CBD target populations from the US and UK. This provided a substantially streamlined and economical protocol for association study by using a minimal set of htSNPs that effectively account for the majority of diversity across the whole locus.

In both PSP populations two common haplotypes were associated, firstly under-representation of the 'classical' H2 haplotype and secondly over-representation of an H1-derived haplotype, the H1c haplotype. In an attempt to further minimise the candidate domain of *MAPT* that potentially harboured any pathogenic variants(s), subsets of htSNP haplotypes were analysed, with consideration of the H1/H2 divide (*del-In9*). This approach implicated a minimal candidate region starting upstream of exon 1 and extending downstream to intron 9.

Although similar trends were observed in the small number of CBD cases, they were not significant due to the smaller size of this cohort. Assuming that these findings

can be confirmed in a larger CBD cohort, they suggest that the causative variant(s) in *MAPT* may be the same for both disorders.

Due to the unusual nature of the H1/H2 dichotomy in the Caucasian population it would have been particularly useful to determine if the *MAPT* H1c association with PSP would have been present in a population with no H2. Through collaborations with a Japanese group (Professor Yoshikuni Mizuno, Juntendo University, Tokyo, Japan), together we have been collecting Japanese PSP samples but have not yet got sufficient for this work. However, a collection of 95 control DNA samples were obtained and the LD analysis carried out these samples and the set of identified htSNPs would form the basis of future studies of *MAPT* in the Japanese population.

The H1c *MAPT* haplotype is associated with both PSP clinical subtypes; RS and PSP-P. This suggests that genetic variation at the *MAPT* locus is not responsible for these differences in the clinical presentation of pathologically proven PSP. The distribution of ApoE genotypes was also investigated in respect of RS and PSP-P and there was no significant difference between the two groups and thus additional unidentified modifying factors, either genetic or environmental, must be responsible for these clinical differences.

The *MAPT* haplotypes that confer protection (H2) risk (H1c) or are neutral (H1b) in PSP and CBD provided the basis for targeted re-sequencing strategies for *MAPT* described in the next chapter. It is clear that there are no obvious pathogenic missense or splice site mutations in *MAPT* in the large majority of sporadic PSP cases. It is plausible, however, that the associated SNPs identified in this study that confer the greatest risk or protection are indeed causative or are in LD with variants that could cause subtle changes either in the alternative splicing or overall expression levels of *MAPT*. It is possible that each neuronal sub-group is dependent on a particular tau isoform profile and expression level and aberrations in this homeostasis could affect one neuronal sub-group more than one another and lead to the selective and disease specific neuronal death and tau pathology.

The results of the work described in this chapter were published in the *Journal of Medical Genetics* and *Brain*:

Pittman,A.M., Myers,A.J., Abou-Sleiman,P., Fung,H.C., Kaleem,M., Marlowe,L., Duckworth,J., Leung,D., Williams,D., Kilford,L., Thomas,N., Morris,C.M., Dickson,D., Wood,N.W., Hardy,J., Lees,A.J., de Silva,R. (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the *tau* gene in progressive supranuclear palsy and corticobasal degeneration. Journal of Medical Genetics 42(11), 837-846.

Williams,D.R., de Silva,R., Paviour,D.C., Pittman,A., Watt,H.C., Kilford,L., Holton,J.L., Revesz,T., Lees,A.J. (2005). Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. Brain 128(6), 1247-1258.

Williams,D.R., Pittman,A., Lees,A.J., de Silva, R (2006). Genetic variation at the tau locus and clinical syndromes associated with progressive supranuclear palsy. Movement Disorders (in press).

5 Identifying candidate functional SNPs responsible for the *MAPT* H1c association

5.1 Overview

In order to investigate the functional basis of the *MAPT* H1c haplotype association, a re-sequencing strategy of individuals homozygous for the risk, neutral and protective haplotypes was devised to identify candidate polymorphic loci on the H1c background that could potentially alter *MAPT* gene behaviour. A total of 216 polymorphic sites were identified and 12 SNPs were short listed because they constitute the H1c haplotype background.

Cross-species genomic sequence comparison revealed that two SNPs (rs7201728 and rs242557) on the H1c background are situated in evolutionarily conserved islands in the vicinity of the promoter region of the *MAPT* gene. Allele-specific luciferase reporter gene studies revealed differences in *MAPT* promoter driven expression of these variants, in particular rs242557 (A/G). SNPs and haplotypes of the variants associated with greater expression are over-represented in progressive supranuclear palsy (PSP) thus implicating increased *MAPT* expression as the possible functional basis of the genetic association of the H1c haplotype with this disorder. These findings are now supported by *in vitro* data from brains, in a collaborative study.

5.2 Background

Common variation at the *MAPT* locus is a genetic risk factor for PSP. As described in Chapter 2, the region making up the extended H1 haplotype of (~2 Mb at 17q21.3) is significantly over-represented in PSP. Although the interval is relatively gene rich, containing corticotrophin releasing hormone receptor 1 (*CRHR1*), presenilin homologue 2 (*PSH2*), saitohin (*STH*), N-ethymaleimide sensitive factor (*NSF*), wingless-type (*WNT3*), overwhelming pathological evidence suggests *MAPT* is responsible for the association. This is supported by the identification of missense

and splice mutations in *MAPT* causing frontotemporal dementia with parkinsonism with tau pathology linked to chromosome 17 (FTDP-17T), affirming a central role for tau dysfunction in some neurodegenerative diseases. Although the other related tauopathies including Alzheimer's disease (AD), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) are defined by fibrillar tau pathology, *MAPT* is not mutated in these disorders; however, some mutations in FTDP-17 cause a clinical and pathological picture closely resembling that of PSP.

As described in Chapter 3, the study of association of the H1 haplotype with PSP and CBD was focused on *MAPT* and refined to high resolution. A systematic framework of genetic analysis was employed that firstly examined the LD and common haplotype diversity of the gene followed by selection of haplotype-tagging SNPs (htSNPs) that, capturing the genetic diversity of the gene, were used for a case-control study in PSP and CBD. At this level of resolution multiple *MAPT* haplotypes were identified that could not be entirely accounted for by the H1/H2 definition alone. Many common variants of the H1 haplotype were identified and of these, the H1c haplotype was shown to be preferentially associated with PSP compared to all other H1 haplotypes that were not associated. In addition to this, the single *MAPT* H2 haplotype was negatively associated with this disorder. In fact, the decrease in frequency of *MAPT* H2a was almost entirely accounted for by the increase in the risk haplotype H1c.

The haplotypes identified from that study that conferred protection, risk or are neutral provided the basis for targeted re-sequencing strategies for *MAPT*. The most strongly associated SNPs which defined the H1c haplotype (rs242557 and rs2471738) are likely themselves to cause subtle changes either in the overall expression levels or alternative splicing of the gene, or both, or they are in LD with such variants.

Recently, comparative DNA sequence analysis of a human PAC (201 kb) and a mouse BAC (161 kb) containing the entire *MAPT* and *Mapt* genes, respectively, revealed over 100 conserved non-repeat potential *cis*-acting regulatory sequences

within or close to *MAPT* [224]. The conserved islands with greater than 67% nucleotide identity ranged in size from 20 to greater than 1700 nucleotides. These included the coding exons and the core promoter (exon -1). As the susceptibility site or causal variant(s) associated with H1c is currently unknown, it was necessary to determine if any of the polymorphisms examined in the previous study, and any novel variants identified in this study reside within these potential regulatory sequences identified in *MAPT*.

Previous functional work has been restricted to the *MAPT* core promoter. The *MAPT* promoter was originally defined by deletion analysis and reporter gene expression assays and the majority of the activity was found to reside within a ~0.5 kb fragment (41,327,240-41,327,840; UCSC May 2006; NCBI build 35) next to the 5' end of exon -1 [225]. Interestingly, no TATA or CAAT boxes are found within this region although there is a cluster of three predicted AP2 transcription factor binding-sites around the transcription start site and the location of an additional 7 AP2 sites correlate with promoter activity. The region is GC rich (74%) and the *MAPT* promoter region is not neuronal specific [226].

Polymorphisms residing within the promoter of the gene have been identified; some of which affect the predicted transcription factor binding sites in the core promoter, including a C/G SNP at position -226, a A/G SNP at position -45 (numbering relative to the first nucleotide of exon -1). Linkage disequilibrium analysis has revealed that these polymorphisms correspond to the extended H1 and H2 haplotypes [192]. The former (C nucleotide) creates a c-myb site on the H1 background (H1*p*; promoter) and the latter (G nucleotide) creates and additional AP2 site on the H2 background (H2*p*; promoter) [191]. The relative transcriptional activity of these two promoter variants has previously been examined using a luciferase reporter assay and the data suggested the H1*p* variant was more efficient (~1.2 fold greater) at driving gene expression than H2*p*. However, this cannot fully explain the association of the H1c haplotype [227].

The aim of this work was to identify novel candidate polymorphic variants on the H1c background responsible for the risk associated with this haplotype that could alter *MAPT* gene function.

5.3 *MAPT* re-sequencing

5.3.1 Design

In order to identify additional polymorphisms on the H1c background, three individuals were selected for re-sequencing and also compared to the published reference sequence (UCSC May 2004 assembly). The three individuals were selected on the basis of homozygosity for the three most common *MAPT* haplotypes; H1c, H1b and H2, that is, the risk, neutral and protective haplotypes (Chapter 4), respectively. Thus the *MAPT* gene in six chromosomes in total were sequenced, two of each haplotype. The previous haplotype association study of *MAPT*, based on the set of 6 htSNPs revealed the strongest association to be a minimal candidate region of ~56 kb of sequences spanning upstream of *MAPT* exon 1 (intron 0) downstream to intron 9 (**Figure 4.5**). This region was re-sequenced in the three selected individuals and compared to the published sequence. Conserved islands were identified by comparison of *MAPT* to the mouse homologue, *Mapt*. All conserved regions were sequenced in intron 0 and completely from intron 2 to 10 regardless of conservation. Genomic sequences of low complexity that could not generate a PCR product for direct sequencing were not sequenced. This strategy, far from exhaustive for identifying all variation in the *MAPT* locus, was thought to be the most economical strategy of identifying candidate polymorphic loci on the risk H1c background within the scope of this work.

5.3.2 Sequencing results

A total of 216 polymorphic sites were identified including 200 SNPs and 16 other polymorphisms that included deletions ranging from 1 nt to 40 nt in length. These findings are summarized in the Appendix 10.5 and includes db SNP identification numbers where available, the single nucleotide variant of each of the three

individual samples and of the published sequence. Single nucleotides unique to the individual/haplotype background are shaded.

The majority of the polymorphisms identified (n=195) correspond to distinctions between that of H2 (H2a) and H1 lineage (H1c and H1b). However, 21 distinctions were identified between the two H1 haplotype individuals sequenced. In these cases the polymorphisms were either specific to the H1c (12 polymorphisms) or H1b background (9 polymorphisms). All polymorphisms identified in the sequence traces were homozygous This was expected since the genomic DNA templates used were individuals selected on the basis of *MAPT* haplotype homozygosity. One heterozygous SNP was identified in intron 10 in the individual homozygous for H2a thus representing rare intra-H2 haplotype diversity.

5.3.3 Candidate H1c SNP shortlist

Out of the SNPs identified from the re-sequencing effort, 12 polymorphisms were present specifically on the H1c background (**Table 5.1**). Of these SNPs, two (rs242557 and rs2471738) were used in the original htSNP study detailed in Chapter 4, and allelic variants of these SNPs in part define the H1c haplotype which adds confidence to the sequencing strategy. In addition a further 7 novel SNPs, 1 microdeletion and two dinucleotide polymorphisms were identified that represent equally good candidates for further study.

	MAPT Region	dbSNP ID	Context Sequence	I1	I2	I3	P
1	Intron 0	rs242557	GCTTCGCCCAGGGTa/gCACCAGGACACGGT	A	G	G	A
2	Intron 0	rs242562	ACCAGCCCGACTCGa/gGCTTCCCCTGAGCC	A	G	G	A
3	Intron 2	rs9896485	AAACAAGTCGCAATc/gCCCTCTGGACCCTG	G	C	C	C
4	Intron 4	rs2435207	CACTGTCTGCCGTGa/gTCAGTGACCTGGAT	A	G	G	A
5	Intron 4	…	ATGTGTGACATGCTD/ITGTGCCCAACCAGA	D	I	I	I
6	Intron 4	…	GCAGGGCACACACAtt/gcAGGTGACACACACA	TT	GC	GC	TT
7	Intron 4	rs2435207	GTATAGTTGGAAAAc/tCTCTGGGCATACTC	T	C	C	T
8	Intron 7	rs2435214	TTAGGGTTGTTTGCa/gAGAGAATGAAAGAA	A	G	G	A
9	Intron 9	rs2471738	AACTCACAGGTGACc/tCAGGGCCTCTGTCCT	T	C	C	T
10	Intron 9	rs12946693	AAAATTAGCTGGGCAc/tGGTGGCTCACGCCT	C	T	T	C
11	Intron 9	…	CTGGGATGTGACTcg/ca/tgACCTCCCGTCACT	CG	CA	TG	CG
12	Intron 10	rs12150530	GAGCACATGCCAGCc/tGACACCAAAATAAT	C	T	T	C

Table 5.1 Candidate H1c polymorphisms

Candidate polymorphisms identified on the H1c background. In these cases this is where a SNP is unique to the H1c background; unique to individual 1 (I1). Individual 2 (I2) is H1b, individual 3 (I3) is H2 and shown is the published nucleotide (P).

5.3.4 SNP evolutionary conservation

The regions containing the SNPs from the previous study and the candidate H1c SNPs were investigated for evolutionarily conservation. The published *MAPT* sequence [May 2006 UCSC human reference assembly (NCBI build 36.1)] was compared to the mouse homologue (*Mapt*) [February 2006 UCSC mouse (*Mus musculus*) reference assembly (NCBI build 36] with the sequence alignment program mVISTA. All SNPs were assessed in relation to the conserved islands identified between human and mouse.

The SNP rs242557 (A/G) resides within one such region of 262bp in length and with an average homology of 76% between mouse and human (**Figure 5.1a and Figure 5.2**). This conserved island itself resides between exon -1 and the first coding exon 1 (intron 0) ~50kb downstream from the core promoter and due to the high degree of conservation is strongly predicted to be a putative gene control element. Similarly the SNP rs7210728 (A/G) that was used in the previous htSNP study (in the primary analysis in the CEPH-trios) resides within an evolutionarily conserved island, 303bp in length with an average homology of 77% ~3kb 5' upstream of the core promoter

and is also predicted to be another putative gene control element (**Figure 5.1a and Figure 5.2**).

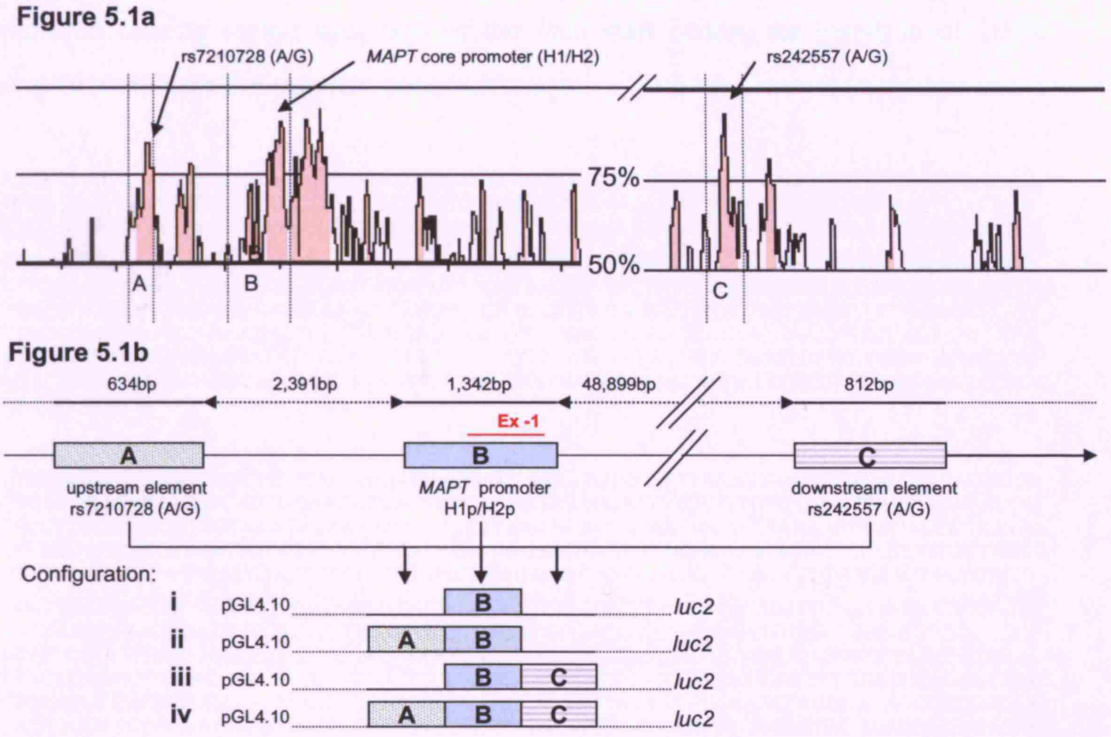

Figure 5.1 *MAPT* promoter constructs

5.1a Sequence conservation of the single nucleotide polymorphisms rs7210728, rs242557, the *MAPT* core promoter and surrounding regions of interest.

5.1b Genomic organisation of the *MAPT* core promoter (B) in relation to the putative upstream (A) and downstream (C) elements for construction of the *MAPT* promoter and surrounding elements for genetic reporter gene study. The modules (A, B and C) are differentially inserted into the pGL4.10 *luc2* vectors to form the different configurations (i, ii, iii and iv). Different genetic variants of each module of A, B and C were constructed in each configuration, as detailed in **Table 5.3**.

5.4 Functional assessment of *MAPT* promoter SNP haplotypes

5.4.1 *MAPT* promoter luciferase reporter gene constructs

In order to assess the potential function of the two conserved domains and if their allelic variants affected *MAPT* promoter driven transcription, the respective variants [rs7210728 (A/G) and rs242557 (A/G)] of the domains in context with the H1*p* and

H2*p* variants of the core promoter and exon (-1) were cloned into pGL4.10 *luc2* reporter vector (Promega). The pGL4.10 luciferase reporter vector contains the synthetic firefly (*Photinus pyralis*), gene luc2 without a promoter but instead with a multiple cloning region upstream of the *luc2* start codon, for insertion of DNA sequences to measure reporter gene expression.

Figure 5.2 Sequence conservation of modules A and C

Shown is the sequence of the two modules A (upper block) and C (lower block) cloned together with the *MAPT* core promoter into pGL4.10 *luc2* vector. The two SNPs rs7201728 (A/G) and rs242557 (A/G) are shaded blue and the nucleotides conserved between human and the mouse are shaded yellow.

The respective domains were generated by PCR from the genomic DNA of genetically defined homozygous individuals (**Table 5.2**). Primer pair design included additional bases at the end of each oligonucleotide to introduce unique restriction enzyme sites for subsequent sub-cloning steps; rs7210728 (A/G) in module A (634bp fragment; 41,323,936-41,324,569), H1/H2 promoter (H1p/H2p) in module B (1342bp fragment; 41,326,960-41,328,301) and rs242557 (A/G) in module C (812bp fragment; 41,375,200-41,376,011), where the base positioning corresponds to the May 2006 UCSC human reference assembly (NCBI build 36.1; see **Figure 5.1b**).

Module	Forward Primer (5'-3')	Reverse Primer (5'-3')
A	**ggtacc**TCAGGACCCAACTCTCCA (*Kpn I*)	**gagctc**CAGCGTCCTGTATCCTGA (*Sac I*)
B	**gagctc**CAAATGCTCTGCGATGTGTT (*Sac I*)	**gctagc**GGACAGCGGATTTCAGATTC (*Nhe I*)
C	**gctagc**TGGGACAGATCCTCAGTG (*Nhe I*)	**gatatc**GGCTGTCGATGAACCCTA (*EcoR V*)

Table 5.2 Primer sequences for module construction

Primers sequences and the introduced restriction sites (**bold sequence**) used to clone the fragments.

The PCR fragments were first cloned into the pGEM-T PCR cloning vector (Promega) in the *E.coli* host JM109. Correct clones were identified by analytical restriction digestion and agarose gel electrophoresis. Fragments for each module were released from the pGEM-T vector using the appropriate pair of restriction endonucleases (New England Biolabs), separated on an agarose gel and purified using the Qiagen gel extraction kit.

The allele-specific fragments of each module were sequentially sub-cloned into pGL4.10 *luc2* vector to assemble the configurations and the haplotypes formed of the SNPs as in **Table 5.3**. The haplotypes of the constructs reflect those found in the original samples.

Correct clones were verified by PCR, analytical restriction digestion and agarose gel electrophoresis and each completed construct was genotyped with respect to H1p/H2p, rs7201728 and rs242557. Correct pGL4.10 constructs were prepared for transfection by amplification in *E.coli* followed by purification of the plasmid DNA by the Qiagen Endotoxin-free Maxi-prep kit.

	Construct/element allelotypes			
Configuration	**rs7210728**	**MAPT promoter**	**rs242557**	**Name**
I	...	H2*p*	...	H2*p*
I	...	H1*p*	...	H1*p*
Ii	A	H2*p*	...	A-H2*p*
Ii	A	H1*p*	...	A-H1*p*
Ii	G	H1*p*	...	G-H1*p*
Iii	...	H2*p*	G	H2*p*-G
Iii	...	H1*p*	G	H1*p*-G
Iii	...	H1*p*	A	H1*p*-A
Iv	A	H2*p*	G	A-H2*p*-G
Iv	G	H1*p*	A	G-H1*p*-A

Table 5.3 *MAPT* promoter constructs

The *MAPT* core (module B) corresponds either to H1*p* or H2*p*. The allelotype of elements A and B are in respect of the two SNPs rs7210728 (A/G) and rs242557 (A/G) and also that of H1*p* and H2*p*. Construct genotypes account for correct polymorphic distinctions between the H1 and H2 clades and are designed to be representative of common haplotypes naturally occurring in *Homo sapiens*.

5.4.2 Cell culture and transfection

Human neuroblastoma SH-SY5Y and BE(2)-M17 cells were grown in SH-SY5Y culture medium, maintained at 37°C in 100% relative humidity and 5% carbon dioxide. Cultured cells were seeded in opaque 96-well tissue culture (Greiner) plates at 2.7 x 10^4 cells per well, in 200µl of media and grown for approximately 24 hours until 80% confluent. Transient transfection was carried out with Transfast transfection reagent (Promega) at a charge ratio of 1:1 of transfection reagent to DNA in serum free medium supplemented with 2mM L-glutamine. In each well cells were transfected with 150ng of pGL4.10-construct DNA and 3ng of the internal control plasmid DNA (pRL-CMV, Promega) that encodes the *Renilla* luciferase gene under the control of the CMV promoter in the presence of Transfast transfection reagent and serum free medium. The co-transfection of this additional plasmid works as an internal control for standardizing the transfection of the

pGL4.10-constructs, correcting for differences in well-to-well transfection efficiencies.

Five wells of cells, both SH-SY5Y and BE(2)-M17 were transfected by each reporter construct. Each transfection experiment was independently replicated, a further two occasions for SH-SY5Y and one further occasion for BE(2)-M17.

5.4.3 Luciferase assay

The transiently transfected cells were grown for 24 hours, following which the luciferase assay was carried out using the Dual-Glo luciferase assay system (Promega). Firefly luciferase (from the pGL4.10-*MAPT* constructs, **Table 5.3**) and *renilla* luciferase (from the pRL-CMV internal control) were measured sequentially on a 96-well GENios (Tecan) plate reader set to measure luminescence at the maximal dynamic range. Firstly 125µl of growth medium was removed from each well and 75µl of Dual-Glo Luciferase Reagent added and incubated at room temperature for 15 minutes and the luciferase-generated luminescence from the pGL4.10 *luc2* constructs measured. Quenching of the first luminescent signal was achieved by addition of the second reagent; Dual-Glo Stop and Glo, which itself also provides substrate for the second pRL-CMV *renilla* luciferase measurement. After a second incubation at room temperature the *renilla* luciferase was measured.

The ratio of luminescence from the experimental reporter to the luminescence from the control reporter was calculated for each sample, defined as the relative luciferase activity. Statistical significance was calculated on the relative activity of each transfection experiment by student's t-test.

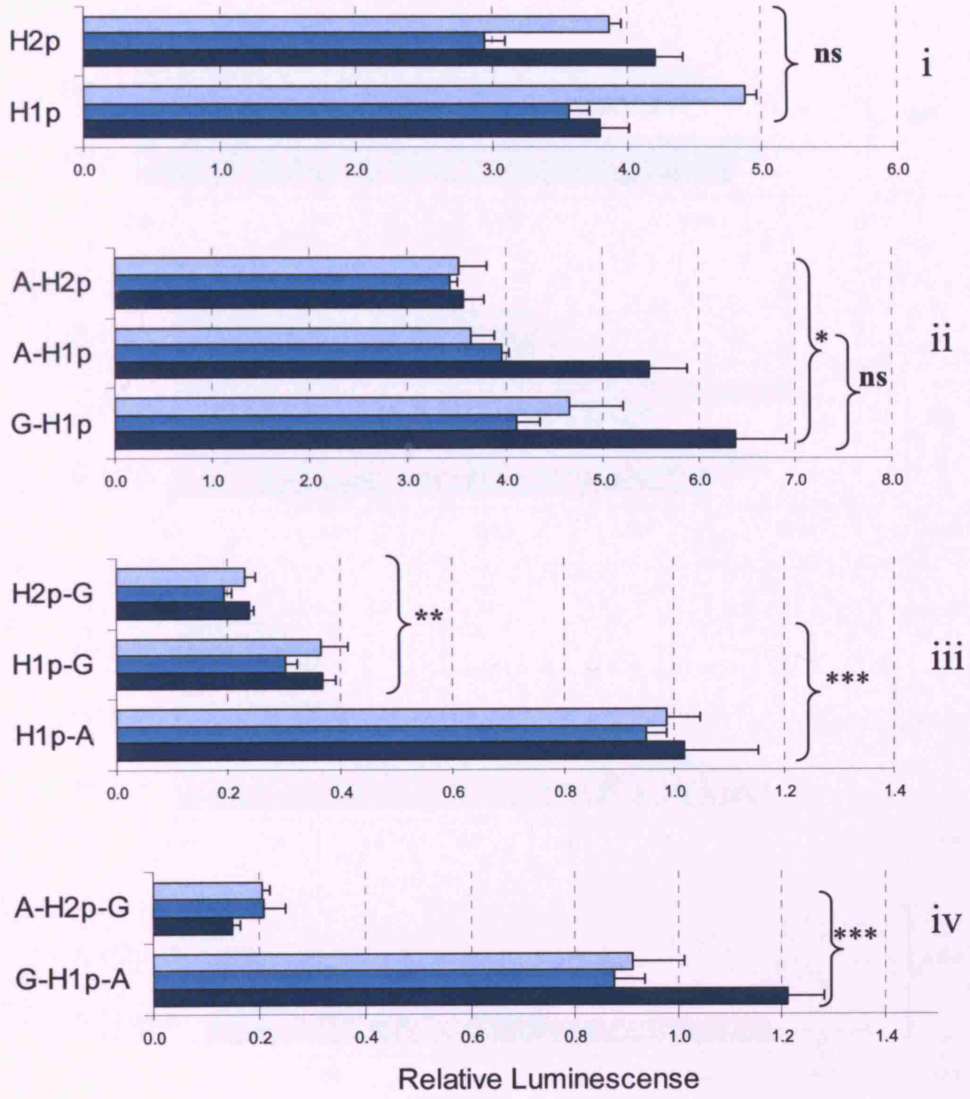

Figure 5.3 Relative luminescence in SH-SY5Y cells

Relative luminescence of the pGL4.10 *MAPT* constructs in SH-SY5Y neuroblastoma cells. Relative luminescence is defined as the pGL4.10-*MAPT* construct luminescence signal divided by the internal control pRL-CMV *renilla* signal. Statistical comparison was made within configuration (i – iv) for allele–specific differences (**Table 5.3**); **p* <0.05, ***p* <0.01, ****p* <0.0001, ns = not significant. Error bars represent the standard error of the mean. These results were replicated in a further two independent experiments as shown with darker shadings.

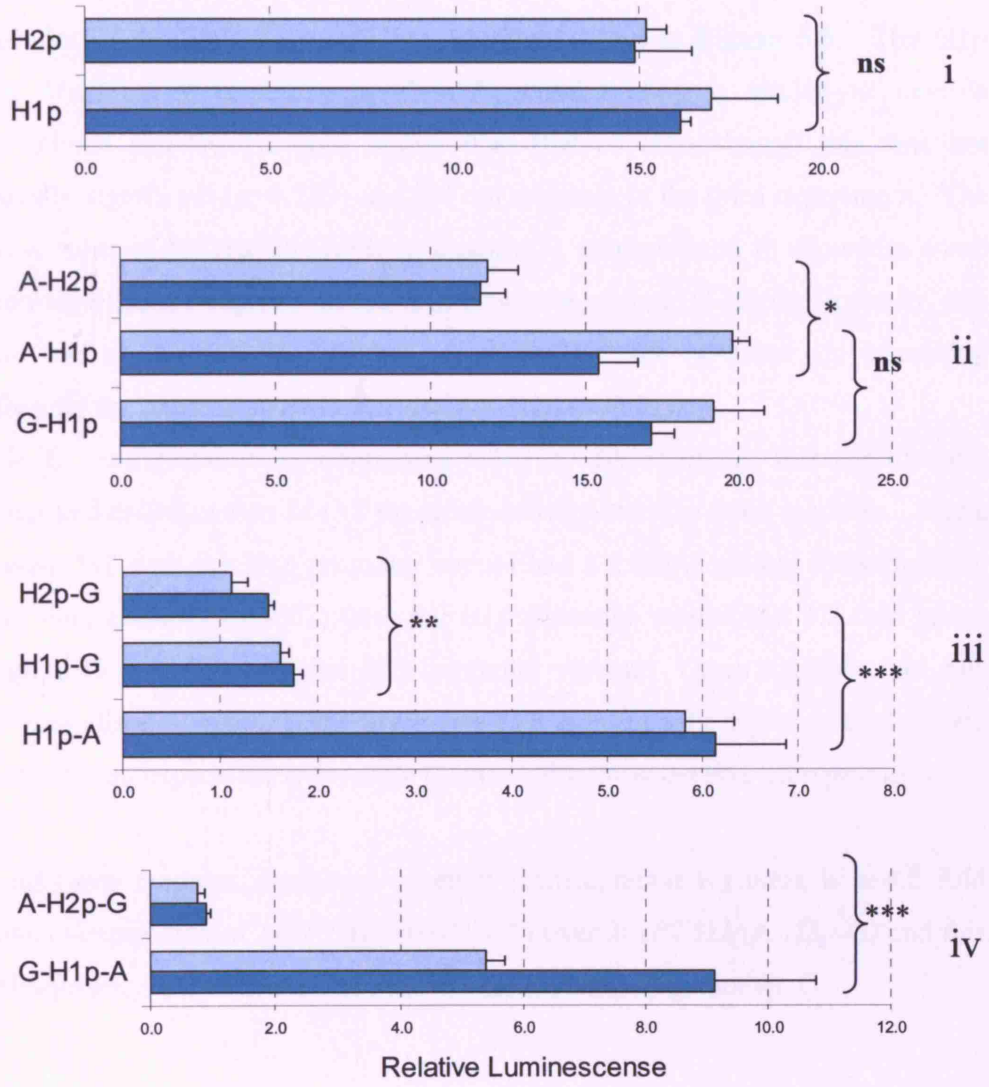

Figure 5.4 Relative luminescence in BE(2)-M17 cells

Relative luminescence of the pGL4.10 *MAPT* constructs in BE(2)-M17 neuroblastoma cells. Statistical comparison was made within configuration (i – iv) for allele–specific differences (**Table 5.3**); **p* <0.05, ***p*<0.01, ****p* <0.0001, ns = not significant. Error bars represent the standard error of the mean. These results were replicated in one further independent experiment as shown with darker shading.

5.5 Results

5.5.1 Relative luminescence in SH-SY5Y neuroblastoma cell lines

The results of the SH-SY5Y cell line are summarised in **Figure 5.3**. The H1*p* variant *MAPT* core promoter (module B, configuration i) tended to have a transcriptional activity 1.2 fold higher than that of H2p, though this was not statistically significant (p=0.253) and did not replicate in the third experiment. The allelic variants of rs7210728 (A/G) in module A, configuration ii, showed a small and non-significant influence on the H1*p* promoter variant (G-H1p). However, this domain had a significantly different effect on the core promoter by increasing significantly the expression of G-H1*p* over A-H2*p* (p<0.01).

Module C, configuration iii, containing rs242557 (A/G) alleles had significantly more marked difference on *MAPT* promoter activity than the other modules. Allele A of rs242557 with the H1*p* promoter variant had a 2.7-fold greater transcriptional activity than allele G (p<0.01) with the H1*p* promoter variant and 4.2 fold grater than allele G (p<0.01) with the H2*p* promoter variant. Taken together this data shows that allele A of rs242557 gives rise to a significantly higher rate of *MAPT* promoter driven expression over allele G, even with the same H1*p* core promoter.

With all three modules combined together (configuration iv) there is a 4.2 fold increase in expression of *MAPT* H1 (G-H1*p*-A) over *MAPT* H2 (A-H2*p*-G) and this is predominantly due to the action of SNP rs242557 (A/G) in module C.

5.5.2 Relative luminescence in BE(2)-M17 neuroblastoma cell lines

A very similar distribution of results and level of statistical significance were obtained with BE(2)-M17 neuroblastoma cells (**Figure 5.4**) as observed with the SH-SY5Y cells but with a few differences. Consistently higher luminescence values were obtained, possibly representing greater transfection efficiency for this particular cell line and the H1*p* core promoter was consistently ~1.13 fold higher than H2*p*, however, and the effect size and direction of the allelic variants of rs242557 (A/G) was the same.

US PSP

	Frequency (%)				
dbSNP ID	**Cases**	**Controls**	**Allelic association (*p*)**	**OR**	**95% CI**
rs7210728	18.2	31.0	1.95ex-5	2.023	1.420 to 2.882
H1/H2	91.6	77.1	4.02ex-8	3.252	2.102 to 5.032
rs242557	54.4	31.0	2.91ex-9	2.356	1.706 to 3.255

US PSP

dbSNP ID	**Genotypic association (*p*)**	**OR**	**95% CI**
rs7210728	8.40ex-5	2.478	1.593 to 3.854
H1/H2	1.00ex-5	3.834	2.340 to 6.281
rs242557	2.29ex-8	3.800	1.957 to 7.389

Table 5.4 Promoter SNP association in US PSP

Genetic association of *MAPT* promoter SNPs with PSP from the US. *p* values were determined empirically by 100,000 Simulations (CLUMP software). Odds ratios are presented for the risk allele (upper) and risk genotype (lower) with their confidence intervals of 95%.

UK PSP

	Frequency (%)				
dbSNP ID	**Cases**	**Controls**	**Allelic association (*p*)**	**OR**	**95% CI**
rs7210728	14.2	35.8	1.00ex-6	3.372	2.043 to 5.565
H1/H2	93.2	76.6	1.14ex-5	4.655	2.148 to 10.089
rs242557	47.9	35.7	0.012	1.815	1.209 to 2.726

UK PSP

dbSNP ID	**Genotypic association (*p*)**	**OR**	**95% CI**
rs7210728	1.00ex-6	5.467	2.981 to 10.028
H1/H2	5.31ex-5	5.599	2.363 to 13.262
rs242557	0.016	2.916	1.303 to 6.527

Table 5.5 Promoter SNP association in UK PSP

Genetic association of *MAPT* promoter SNPs with PSP from the US. *p* values were determined empirically by 100,000 Simulations (CLUMP software). Here odds ratios are presented for the risk allele (upper) and risk genotype (lower) with their confidence intervals of 95%.

5.5.3 Genetic association of functional *MAPT* promoter haplotypes

The SNPs constituting the promoter haplotypes are highly associated with PSP in both the US and UK populations (**Tables 5.4 and 5.5**). Single locus associations were calculated empirically by use of CLUMP software. All alleles and genotypes have significant risk odds ratios.

The promoter SNP haplotypes are highly associated with PSP; globally, the significance is p=7.05ex^{-12}, p=4.12ex^{-7} for the US and UK populations respectively. The haplotype specific tests highlight two haplotypes out of the five whose frequencies are statistically different in the control group verses PSP (**Table 5.6**); Hap2 and Hap3, increased and decreased, respectively in PSP. Hap2 corresponds to the configuration iv construct G-H1p-A and Hap3 A-H2p-G. Thus the higher expressing *MAPT* promoter haplotype Hap2 is over-represented in PSP and the lower expressing haplotype Hap3 is under represented.

US PSP

ID	Haplotype	Control (f)	Cases (f)	Association (p)
Hap1	G-H1p-G	42.1	34.3	0.20
Hap2	G-H1p-A	26.3	45.2	**2.05ex^{-7}**
Hap3	A-H2p-G	20.4	5.4	**1.30ex^{-9}**
Hap4	A-H1p-G	5.9	4.3	1.00
Hap5	A-H1p-A	4.0	8.2	0.12

UK PSP

ID	Haplotype	Control (f)	Cases (f)	Association (p)
Hap1	G-H1p-G	36.0	41.8	0.71
Hap2	G-H1p-A	27.0	42.3	**5.00ex^{-3}**
Hap3	A-H2p-G	22.0	2.5	**2.2ex^{-7}**
Hap4	A-H1p-G	5.8	4.5	1.00
Hap5	A-H1p-A	9.1	7.6	1.00

Table 5.6 *MAPT* promoter haplotype association

Genetic association *MAPT* promoter haplotypes with PSP from the US (upper) and the UK (lower). Globally, the distribution of haplotypes between cases and controls is highly significant; p = 7.05 ex-12 for US PSP and p = 4.12ex-7 for UK PSP. Performing tests of association of each specific haplotype revealed the global difference was due to a significant frequency increase in Hap2 and a significant decrease in Hap3 in PSP, consistent in both populations. All reported p-values are adjusted according to the number of tests performed. The analysis was performed on all haplotypes exceeding 2% in frequency; accounting for >97% of Expectation-Maximization predicted haplotypes.

5.6 Discussion

The functional basis of the PSP associated H1c haplotype was investigated. A re-sequencing strategy was devised and implemented to identify non-coding polymorphisms constituting the H1c background that could explain the basis of the

association. From the re-sequencing data, a list of candidate polymorphisms was drawn up for further functional study. Two candidate SNP polymorphisms (rs242557 and rs7210728) that both resided within evolutionary conserved islands flanking the promoter region, were tested for their relative effect on *MAPT* H1*p* and H2*p* promoter driven expression using a luciferase assay. In the first instance the relative expression of the core *MAPT* H1*p* and H2*p* promoter variants was investigated. Although H1*p* tended to give a higher expression (1.13 fold in BE(2)-M17 cells) than that of H2*p* confirming previous findings [227], the difference did not reach significance in this experimental design.

By far, the most striking finding from the luciferase assays was the change in the relative expression associated with the allelic variants of rs242557 (A/G). Allele A of this SNP, residing in the putative control element gave consistently increased expression (3.8-4.2 fold) over allele G under the same H1*p* *MAPT* core promoter. This data suggests that this genetic variation within the H1 haplotype, specifically the H1c allele (A), gives rise to higher level of *MAPT* gene expression.

Although the luciferase serves as a good model for investigating the relative change in expression of the SNP alleles in question the limitations of these experiments should be highlighted so not to over-interpret the results. The human neuroblastoma cell line has the advantage of doubling approximately every day. The continuous division and the immortality of these cells is the result of the oncogenic mutations that has given rise to the original neuroblastoma line and thus tau protein function is likely to be quite different from wild-type non-dividing neuronal cells in a real biological system.

The luciferase assay constructs themselves are indeed artificial and are inadequate for studying the full dynamics of *MAPT* gene expression. Only a small proportion of the putative functional control elements of the *MAPT* gene together with the core promoter were included into the luciferase reporter gene constructs. It is highly likely that *MAPT* expression is regulated at multiple sequences within and around the promoter and the gene and so the results obtained in these experiments are likely

to be an exaggeration of the functional effect of the SNPs identified. Nonetheless, these results do quite clearly demonstrate allele-specific differences in *MAPT* expression that is associated with PSP.

These findings have been validated *in vivo* by our collaborators at the NIA/NIH whereby in brains of individuals with the H1c, overall tau expression is significantly higher when compared to the other haplotypes (see section 8.3) and have been submitted in a paper to *Neurobiology of Disease.*

6 Microdeletion encompassing the *MAPT* gene is associated with developmental delay and learning disabilities

6.1 Overview

Three individuals that phenotypically presented with developmental delay and learning disabilities were shown to have *de novo* deletions of ~650 kb in size at 17q21.31 encompassing the *MAPT* gene. An in-depth SNP haplotype analysis was undertaken in the three affected individuals and their parents with respect of the extended H1 and H2 haplotypes. In two of the three family triads (father, mother and child) it was established that the parents-of-origin were heterozygous for the H1/H2 inversion polymorphism. The region shows complex genomic architecture with low-copy repeats (LCRs) and consequently a mechanism of non-allelic homologous recombination (NAHR) in inversion heterozygote carriers and H2 homozygote carriers was proposed.

6.2 Background

Whilst much of the previous work detailed in previous chapters was underway several new findings came to light regarding the *MAPT* locus. The extended region of linkage disequilibrium (LD) at 17q21.31 of the two non-recombining H1 and H2 haplotypes defined in Chapter 1, was shown to be due to an ancient inversion of a region of ~900 kb that includes *MAPT* [214]. The inversion was identified by generating chromosome-specific assemblies from RP11 BAC clones originating in a DNA sample from one individual heterozygous for H1 and H2. By genotyping the RP11 clones at 17q21.31 for 60 microsatellite markers and assembling the clones into chromosome specific contigs, it was found that the H2 haplotype was structurally different, relative to the Build 34 (H1 orientation) assembly. One implication of this finding is that the orientation of the *MAPT* gene is different on the H2 background relative to the H1 background.

The global population distribution of *MAPT* H1/H2 haplotype was examined in publicly available reference samples representative of most racial groups. The analysis showed that the H2 *MAPT* haplotype has a geographically restricted distribution; it is almost exclusively Caucasian in origin, with European populations having H2 allele frequencies of ~25%, central Asian (including Finnish) populations of ~5% and other populations (African, East Asian and native American) having H2 allele frequencies of essentially zero [213]. In fact, the incidence of H2 in non-European groups was in populations with historical admixture with Europeans.

Detailed sequence analysis of 17q21.31 revealed the presence of low-copy repeats (LCRs), which are the basis of the inversion polymorphism and genomic complexity of this region by forming the substrates for non-allelic homologous recombination [228]. The genomic complexity around the *MAPT* locus, mediated by the multiple LCRs, is underlined not only by the presence of the inversion but also by complex arrangements of duplications and triplications close to the *NSF* gene [214], see **Figure 6.3**. Another consequence of this complexity is the deletion of a ~650 kb region containing *MAPT* in three children with learning disability and mental retardation that was identified by array-based comparative genomic hybridization (array-CGH).

The array-CGH technique has been used to investigate DNA copy number changes because it has many potential advantages over other methods currently available. Routine karyotype analysis is not sensitive enough to detect small chromosome rearrangements less that 5 Mb in size. The sensitivity of array-CGH has been refined to permit the detection of chromosome deletions as small as 3 Mb and the information it provides is directly linked to the physical and genetic maps of the human genome (see section 1.3.1).

In a recent study, 50 patients with learning disability/mental retardation and dysmorphism were investigated for DNA copy number changes using array-CGH. The microarray was constructed from large insert BAC clones spaced approximately at 1Mb intervals across the genome [229]. Twelve copy-number abnormalities were

identified in 12 patients (24% of the total); seven deletions (six apparently *de novo* and one inherited from a phenotypically normal parent) and five duplications (one *de novo* and four inherited from phenotypically normal parents). One of the individuals in this report had a *de novo* deletion at 17q21.31 involving a single clone, RP5-843B9 that encompasses the *MAPT* gene [229, 230]. Subsequently, a further two similarly affected individuals were identified with the same deleted clone.

The aim of this work was to analyse the DNA from the three affected individuals and their parents to determine the parental origin of the deleted allele and to see whether or not there was any relationship between the *de novo* deletion and H1/H2 haplotype background. The haplotype analysis described in this chapter was carried out in collaboration with Dr. Charles Shaw-Smith at Addenbrookes Hospital, Cambridge and Dr. Nigel Carter at the Sanger Centre, Cambridge who were responsible for the findings described in the next section (6.3)

Figure 6.1 Clinical photographs of the three affected individuals

6.3 Background data

6.3.1 Array-CGH

6.3.1.1 Whole genome array-CGH

All three patients were ascertained through research projects designed to identify submicroscopic chromosomal imbalances in patients with learning disabilities. The single clone, RP5-843B9 was found to be deleted in all three cases by whole-genome array-CGH. There are some clinical similarities between the three cases.

All were born close to term but birth weights were at or below the 2nd centile. Hypotonia and poor feeding were noted in the neonatal period. Motor and speech development were delayed. Two of the cases had pectus excavatum (a depression of the sternum into a concave shape) and facial features are mildly dysmorphic in each case **(Figure 6.1)**. Patient 3 was the subject of a recent clinical report suggesting a phenotypic resemblance to Angelman syndrome [230].

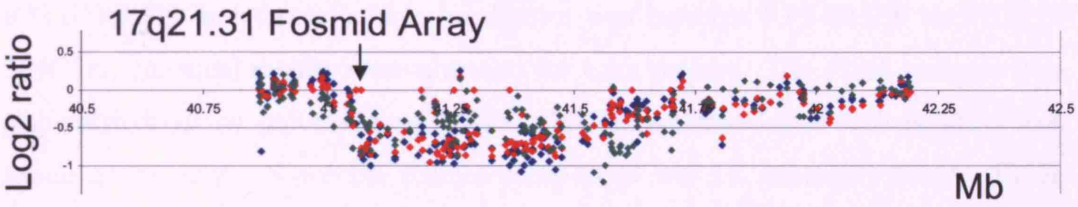

Figure 6.2 Fosmid array-CGH

High-resolution fosmid array-CGH analysis at 17q21.31 of the patients 1-3 (individual 1 is green; individual 2 is blue and individual 3 red). The panel depicts an overlay of results for all three cases and illustrates in particular the common centromeric breakpoint for all three cases (arrowed). This data is from Charles Shaw-Smith.

6.3.1.2 Fosmid array-CGH at 17q21.31

To asses the extent of the deletion in the three patients with higher resolution, a redundant tiling array comprising 227 fosmid clones that spanned the region from 40.84 Mb to 42.17 Mb (UCSC, May 2004 assembly) was constructed. The results of the three cases are shown in **Figure 6.2**. The three cases appeared by this analysis to have a common, identical centromeric breakpoint at approximately 41.05 Mb (estimated to reside in a 15 Kb region; 41,055,000 to 41,070,000). The telomeric breakpoint could not readily be determined from this analysis as there is no clear transition from deletion to absence of deletion. An unambiguous interpretation of the fosmid data is hindered by the following factors: first, the fosmid sequences are not single copy throughout the region which contains numerous LCRs. Second, the display of the data does not take into account the fact that fosmids are displayed according to the H1 assembly. Nonetheless, taking these limitations into account,

the fosmid data does confirm the close correlation in size of the deletion in the three patients and reveals a common breakpoint at the centomeric side.

6.3.2 Fluorescence *in situ* hybridization

Detailed mapping by fluorescence *in situ* hybridization (FISH) with BAC and P1-derived artificial chromosomes (PAC) clones indicated that the deletion was between 500 and 650 kb in size. The centromeric breakpoint was within clone RP11-707O23 and the telomeric breakpoint was between RP5-843B9 and RP11-259G18. Identical results were obtained for each patient. The FISH analysis was also carried out on parental samples and this demonstrated that each deletion had arisen as *de novo*. Since the fosmid array-CGH was H1 assembly biased, FISH studies were undertaken with clones from the RP11 library that had previously been shown to be specific for H1 and H2 haplotypes. The results of each individual clone is shown on **Figure 6.3** and are distributed the same for each case. The signal from the FISH probes, relative to the other homologue/chromosome; show either a normal signal (green; non-deleted), no-signal (red; deleted) or a diminished signal (blue; partially deleted). With the detailed LD and haplotype map of the 17q21.31 region involved in this deletion, obtained by work described in previous chapters, the aim at this point was to investigate the mechanisms leading to the deletion by attempting to map the break points of the deletion on which parental haplotype the deletion took place.

6.4 Haplotype analysis

As described previously, the 238 bp *MAPT* H2 deletion in intron 9 (*del-In9*) was used to determine unambiguously the tau H1/H2 haplotype in the triads [172]. In addition, a further panel of informative SNPs (as characterised Chapter 1) that are in perfect linkage disequilibrium (D'=1, r^2=1) with the *del-In9*; rs2696425, rs2668643, rs2532418, rs1528072, rs1468241, rs916793, rs1396862, rs1528072 and rs199528 were genotyped. Genotypes were obtained by direct sequencing using BigDye V3.1

and run out on an Applied Biosystems 3100 capillary analyser, and by restriction digestion as previously described in Chapter 1 [217].

A mixed *MAPT* tetra-nucleotide repeat (*MAPT*-repeat-t) was genotyped in the three triads to distinguish intra-H1 and H2 parental chromosomes. The alleles were sized by analysing fluorescently labelled (FAM) PCR products on an Applied Biosystems 3730 96-well capillary analyser and the genotypes were scored with Genemapper software version. 3.0.

Genotypes obtained from the *del-In9*, the SNPs and *MAPT*-repeat-t were used together to assemble haplotypes in the triads. The SNP alleles were converted to '1' or '2' corresponding to H1 and H2 alleles respectively for ease of interpretation. H1/H2 SNP parental and affected offspring 'wild-type' haplotypes were constructed in accordance with the H1/H2 doctrine. Deleted affected offspring haplotypes are largely inferred as the most likely configuration on account the array-CGH and the FISH results combined with the observed SNP marker genotypes.

Assay	Forward Primer (5'-3')	Reverse Primer (5'-3')	Bp
rs2696425	GCAAGAGGAAAATGGCTCTA	GCTAGCTTCTGCCAGTGCTC	346
rs2668643	GAGCTTAGGGCCTGATTCTG	GGTTGCGGTGAACTGAGATT	431
rs1396862	TGTGGACCAAGGTTCACTGCTC	TGGAATGTATGCCCTACGCCAG	134
rs916793	TGCCCTGAATTTGGTGT	AATGCTAAGAGGCCATCA	179
rs1468241	AGCCGACGGATAAACAA	ATGGATAAGGCCCATGA	151
rs1528072	AACAGCCTTTCCCATCA	TTGATTTGGCAAGTACCAG	241
rs2532418	TTATGTGTGGCCCAAGACAG	AATTACATGCGAGGAAATGG	467
rs199528	AACAGCTAACGACTCTGTCC	GTTCTGTGTAGGGTTGTCCT	250
del-In9	GGAAGACGTTCTCACTGATCTG	AGGAGTCTGGCTTCAGTCTCTC	484 (H1)
MAPT-repeat-t	CATTTTCCAAGCAGAACTTCATTTCC -[FAM]	TGGACAACATTTATGTCCTGAAACAG	283-325
(1)	TGTGCTTGTTTCCATGCGTA	CATCCAGCAGCCCTAAGAAC	439
(2)	AGGTTCTCTTAGTTCAGATGTAGGAC	CACTCAACAGCTGCCTGAAC	471
(3)	TCACATGAACTACTGCCAAAGG	GTCTCTGCAGCATTCCATACT	276
(4)	TAATATGGTGCAGGGCTGTG	AGAGGGAGGCAGACATCTCA	424
(5)	TGCCTCTGTTGTTTTCAATG	CACAGGCAGAGAAAGTTGGA	151
(6)	ACTATGAACGCGGCAGAGAT	CAGGATTGCAACAGCTGAAC	455

Table 6.1 Genotyping and sequencing assay primer sequences

Primer sequences for SNP genotyping, *MAPT*-repeat-t and sequencing (1-6) assays in this study.

6.5 Heterozygosity mapping

A re-sequencing strategy designed to map the presence or absence of H1/H2 heterozygosity at the breakpoint region was undertaken to try to refine and narrow down the region containing the breakpoints. Regions targeted for sequencing are contained within the 15 kb region predicted to contain the centromeric breakpoints from the fosmid array-CGH results (41,055,000 to 41,070,000, UCSC May 2004). Sequencing primers used are numbered 1-6 in **Table 6.1**.

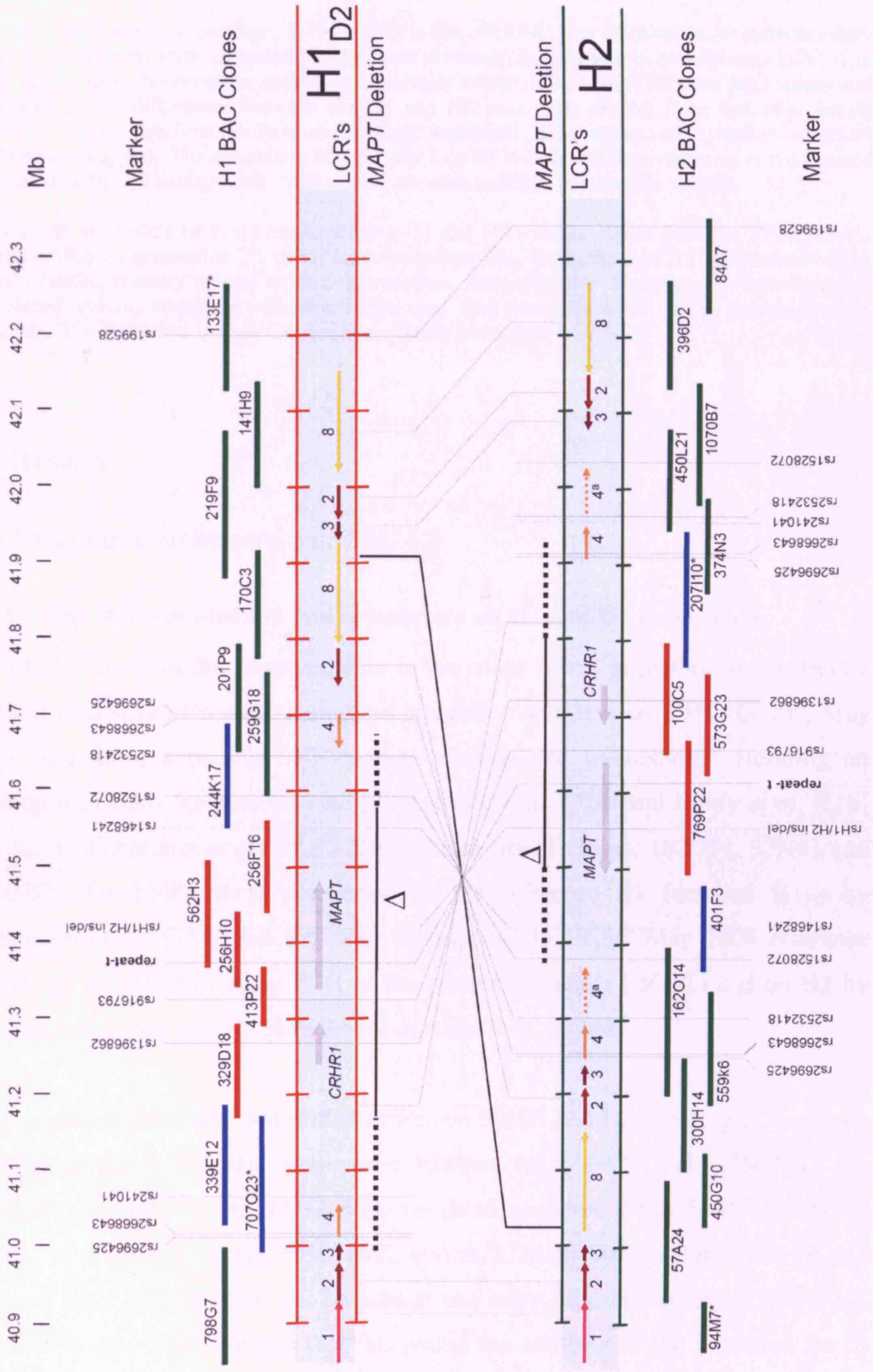
Mb
40.9
41.0
41.1
41.2
41.3
41.4
41.5
41.6
41.7
41.8
41.9
42.0
42.1
42.2
42.3
Marker
H1 BAC Clones
LCR's
H1D2
MAPT Deletion
MAPT Deletion
LCR's
H2
H2 BAC Clones
Marker
798G7
339E12
707O23*
329D18
413P22
562H3
256H10
256F16
244K17
259G18
201P9
170C3
219F9
141H9
133E17*
MAPT
CRHR1
rs2696425
rs2668643
rs2410141
rs2532418
rs1528072
rs1468241
rsH1/H2 ins/del
repeat-t
rs916793
rs1396862
rs1995528
94M7*
57A24
450G10
300H14
559k6
162O14
401F3
769P22
573G23
100C5
207I10*
374N3
450L21
1070B7
396D2
84A7

Figure 6.3 Inverted *MAPT* haplotypes at 17q21.31

The figure is on the previous page. H1 assembly is that of UCSC May 2004 and is for reference only since H1 haplotypes show considerable structural diversity/copy number polymorphisms (CNPs) in the 3' LCR cluster, however the orientation is unlikely to change since the CNPs are duplications and triplications. The differences between the H1 and H2 assemblies are far from that of a simple inversion; other structural differences include segmental duplications and smaller inversion duplications (e.g. 4a). The orientation of LCR pair 4 on H1 is switched from opposing to tandem and duplicated on the H2 background. SNP marker orientation differs between H1 and H2.

Shown also are results of FISH analysis using H1 and H2 specific clones from the RP11 library. Clones have been assigned to H1 or H2 haplotypes according to Steffanson [214]. Green rectangle: signal of equal intensity present on each homologue. Blue rectangle: Signal on one homologue of diminished intensity compared with other homologue. Red rectangle: Signal on one homologue only (deleted). The FISH data is from Charles Shaw-Smith and Lionel Willatt.

6.6 Results

6.6.1 Genomic architecture at 17q21.3

6.6.1.1 SNP marker position and orientation on H1 and H2 haplotypes

In order to interpret the genotype data in the triads it was necessary to construct a detailed assembly of the H2 haplotype in relation to a reference H1 (UCSC, May 2004 assembly), depicting SNP marker position and orientation. Building on existing data from Stefansson *et al.* [214], Cruts *et al.* [228] and Hardy *et al.* [216] and the available sequences of H2 BAC clones (RP11 clones: 162O14, 37N43 and 107OB7), the SNPs were positioned on the reference H1 (denoted $H1_{D2}$ by Stefansson *et al.* [214] and this also corresponds to UCSC May 2004 reference assembly) by use of *in silico* PCR of the primer sequences (UCSC) and on H2 by aligning the primer sequences to the H2-specific BAC clones.

SNP marker position and orientation differs on the H1 and H2 haplotypes, as shown on **Figure 6.3**. The H1 centromeric markers rs22696425 and rs2668643 are duplicated on both H1 and H2, flanking the deletion region. Some SNP markers are present at one copy on H1 (rs1528072 and rs2532418) but are duplicated on H2 whereas other SNP markers are present at one copy on each haplotype (rs241041 and rs199528). SNP markers that fall within the confines of the inversion are in reverse order on H2 relative to H1 (rs1396862, rs916793, H1/H2 238bp *del-In9*

polymorphism and rs1468241). Correct positioning of the SNP markers is required to construct H1 and H2-specific haplotypes in the three triads.

6.6.1.2 LCR position and orientation on H1 and H2 haplotypes

In addition to the SNP marker orientation on the H1 and H2-specific assemblies, other architectural features were added that include the relative positions of two genes: *MAPT* and *CRHR1*, and the structure of the flanking segmental duplications, or low-copy repeats (LCRs).

The reference H1 LCR structure is in keeping with Cruts *et al.* [228] and Stefansson *et al.* [214] and the UCSC May 2004 assembly, NCBI build 35. It is a reference only since H1 haplotypes themselves show DNA copy number variation in the *MAPT* telomeric LCR cluster [214]. To determine the position of the LCRs on H2 in relation to the H1 reference assembly, fully sequenced H2-spcific BAC clones from the RP11 library: 300H14, 162O14, 37N3 and 107OB7, were aligned with H1-specific BAC clones. For example aligning the H1 clone RP11-707O23 with the H2 clone RP11-374N3 identifies the LCR subunit number 4 on H2, **Figure 6.4**. H2-specific LCR sequences (from the BAC clones) were aligned using the BLAT server (UCSC) the reference assembly to determine the orientation of the LCRs. **Figure 6.3** summarises the position and orientation of H1 and H2-specific LCRs.

The structural differences between the H1 and H2 assemblies are far from that of a simple inversion; other structural differences include further segmental duplications (constituting the LCRs) and smaller duplications in inverted orientation. The LCR subunit 4 that flanks *MAPT*, lies in inverted orientation on the H1 haplotype, but the telomeric LCR subunit has undergone an inversion event on the H2 haplotype such that the LCR pair now flank *MAPT* in direct orientation (See **Figure 6.3**). The LCR subunit 4a is present as a single copy on H1 but is duplicated on H2 and is in direct orientation flanking *MAPT*.

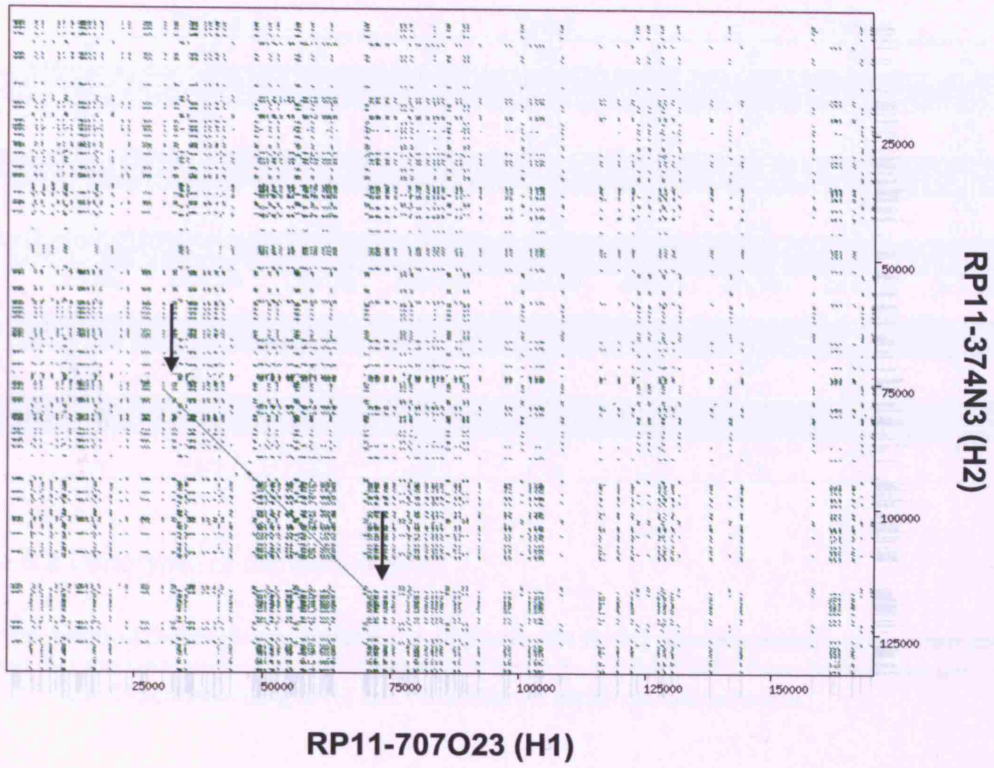

Figure 6.4 Dot plot depicting sequence homology

Dot plot depicting sequence homology between the H1 clone RP11-707O23 and the H2 clone RP11-374N3. These results were generated in YASS [231]. The two arrows mark a region of ~40 kb that corresponds to LCR subunit number 4 in direct orientation.

6.6.2 Genotype analysis

The SNP marker genotypes together with the *del-In9* for the three triads are summarised in **Table 6.2**. In cases 1 and 3 there were inconsistencies in the H1/H2 dichotomy. For example, in case number one there was H1/H2 heterozygosity at markers rs2696425 and rs2668643 but not at any other. In case 3 there was H1/H2 heterozygosity in the markers that flank the deletion (rs2696425, rs2668643, rs2532418, rs1528072 and rs199528) but marker homozygosity within the confines of the deleted region, as per the other wild-type chromosome (H1). In case 2, all marker genotypes were consistent with H2. In all the parents of the three cases, all marker genotypes were compatible with H1/H2 dichotomy; there were no inconsistencies.

	Trio 1			Trio2			Trio 3		
SNP	**P**	**M**	**A**	**P**	**M**	**A**	**P**	**M**	**A**
rs2696425	G/T (1/2)	T/T (2/2)	G/T (1/2)	T/T (2/2)	G/T (1/2)	T/T (2/2)	G/T (1/2)	G/G (1/1)	G/T (1/2)
rs2668643	C/G (1/2)	G/G (2/2)	C/G (1/2)	G/G (2/2)	C/G (1/2)	G/G (2/2)	C/G (1/2)	C/C (1/1)	C/G (1/2)
rs241041	C/T (1/2)	C/C (2/2)	C/C (2/2)	C/C (2/2)	C/T (1/2)	C/C (2/2)	C/T (1/2)	T/T (1/1)	T/T (1/1)
rs1396862	C/T (1/2)	T/T (2/2)	T/T (2/2)	T/T (2/2)	C/T (1/2)	T/T (2/2)	C/T (1/2)	C/C (1/1)	C/C (1/1)
rs916793	C/T (1/2)	T/T (2/2)	T/T (2/2)	T/T (2/2)	C/T (1/2)	T/T (2/2)	C/T (1/2)	C/C (1/1)	C/C (1/1)
MAPT-repeat -t	(325/283)	(299/255)	(255/255)	(283/283)	(283/308)	(283/283)	(283/308)	(287/312)	(312/312)
H1/H2 ins/del	(H1/H2)	(H2/H2)	(H2/H2)	(H2/H2)	(H1/H2)	(H2/H2)	(H1/H2)	(H1/H1)	(H1/H1)
rs1468241	C/T (1/2)	C/C (2/2)	C/C (2/2)	C/C (2/2)	C/T (1/2)	C/C (2/2)	C/T (1/2)	T/T (1/1)	T/T (1/1)
rs1528072	G/T (1/2)	T/T (2/2)	C/C (2/2)	C/C (2/2)	G/T (1/2)	C/C (2/2)	G/T (1/2)	G/G (1/1)	G/G (1/1)
rs2532418	A/G (1/2)	G/G (2/2)	G/G (2/2)	G/G (2/2)	A/G (1/2)	G/G (2/2)	A/G (1/2)	A/A (1/1)	A/G (1/2)

Table 6.2 Genotypes in the three triads

Genotype results of the single nucleotide polymorphisms, the H1/H2 insertion/deletion polymorphism (*del-In9*) and the tetra-nucleotide *MAPT*-repeat-t in the three triads. SNP's have been converted to either a '1' or a '2' corresponding to H1 and H2 alleles for haplotype interpretation.

Since the structural and positional differences of the SNP markers between H1 and H2 have been established, it was possible to reliably construct the parental and the affected offspring's wild-type non-deleted chromosomes. In the case of the affected individual's deleted chromosomes, these were inferred as the most likely configuration on account of the genotype data results combined with positional information on the deletion. This is under the assumption that no additional re-arrangements have taken place apart from the deletion itself

6.6.3 Haplotype analysis

6.6.3.1 Trio 1

The parental origin of the *MAPT* deletion is the heterozygote H1/H2 father and a normal H2 haplotype was inherited from the H2/H2 homozygote mother. This was established by including the *MAPT*-repeat-t in the haplotype analysis. Of particular interest is that chromosome carrying the deletion appears to be composed of both H1

and H2 alleles and therefore for the first time suggesting a mitotic recombination between H1 and H2 haplotypes (**Figure 6.5**).

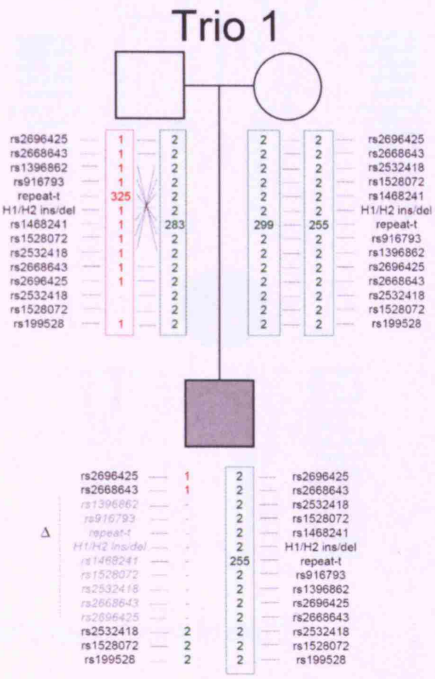

Figure 6.5 Extended *MAPT* haplotypes in trio 1

The phased parental haplotypes (boxed) have been constructed in accordance with strict H1/H2 ruling from the SNP marker genotypes and the deletion haplotype of the affected offspring have been inferred (unboxed) from the genotype data of **Table 6.2**. Markers in faint italics lie within the confines of the deleted region (Δ).

6.6.3.2 Trio 2

It was not possible to determine the parental origin of the deletion in this triad. The parental H2 haplotypes all rigidly carried the same size repeat alleles. The father is homozygous for H2 whereas the mother is a H1/H2 heterozygote. Genotypes of all the markers in the affected individual are consistent with H2 and so in this case origin from either parent cannot be ruled out (**Figure 6.6**).

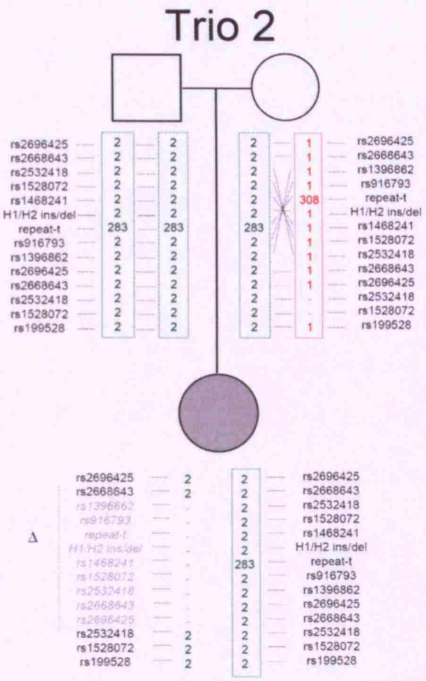

Figure 6.6 Extended *MAPT* haplotypes in trio 2

The phased parental haplotypes (boxed) have been constructed in accordance with strict H1/H2 ruling from the SNP marker genotypes and the deletion haplotype of the affected offspring have been inferred (unboxed) from the genotype data of **Table 6.2**. Markers in faint italics lie within the confines of the deleted region (Δ).

6.6.3.3 Trio 3

It was possible to establish the parental origin of deletion as the H1/H2 heterozygous father in this triad, the same as observed for trio 1; however, the chromosome carrying the deletion in the affected individual was solely composed of H2 alleles and not an apparent H1/H2 recombinant (**Figure 6.7**).

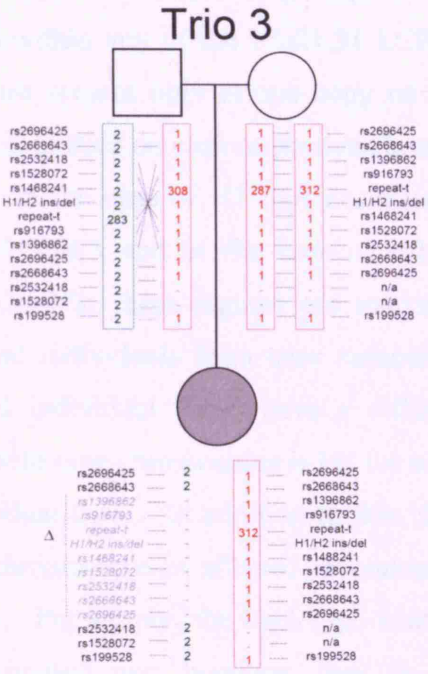

Figure 6.7 Extended *MAPT* haplotypes in trio 3

The phased parental haplotypes (boxed) have been constructed in accordance with strict H1/H2 ruling from the SNP marker genotypes and the deletion haplotype of the affected offspring have been inferred (unboxed) from the genotype data of **Table 6.2**. Markers in faint italics lie within the confines of the deleted region (Δ).

6.6.4 Heterozygosity mapping

A re-sequencing strategy designed to map the presence or absence of H1/H2 heterozygosity at the predicted breakpoint regions was undertaken to try to refine and narrow down the breakpoints. This could only be applied to affected individuals from trios 1 and 3, since in the affected individual in trio 2, the wild-type chromosome is H2 with respect to the deleted chromosome that is also H2 and thus there is no such H1/H2 heterozygosity to exploit. The sequences of the affected individuals were each compared to that of their unaffected H1/H2 paternal heterozygote.

The region targeted for sequencing was within the 15 kb region predicted to contain the centromeric breakpoints (41,055,000 to 41,070,000). A portion of LCR subunit number 4 (LCR4) is contained within this region and is represented in sequencing

amplicons 1, 2 and 3. The other sequencing amplicons 4, 5 and 6 are unique sequences not contained within any of the 17q21.31 LCRs. Although sequencing amplicons 4, 5, and 6 are present only at one copy on each of H1 and H2, the physical regions that are amplified on each respective chromosome are inverted with respect to one another. In the case of H1 chromosomes, the amplified region is contained within RP11-707O23 and in the case of H2 the amplified region is contained within RP11-374N3; these regions are marked with black arrows on **Figure 6.8**. The affected individuals from trios number one (affected individual one) and three (affected individual three) have a different haplotype make up, **Figure 6.5** and **6.7**; the wild type chromosome is H2 for affected individual one and is H1 for affected individual three. In addition to this, the proximal chromosome segment of the deleted chromosome of affected individual one is H1 and is H2 for affected individual three. Put simply, the haplotype make up of individual one is opposite to that of individual two, however, they do share distal H2 deleted chromosome segments.

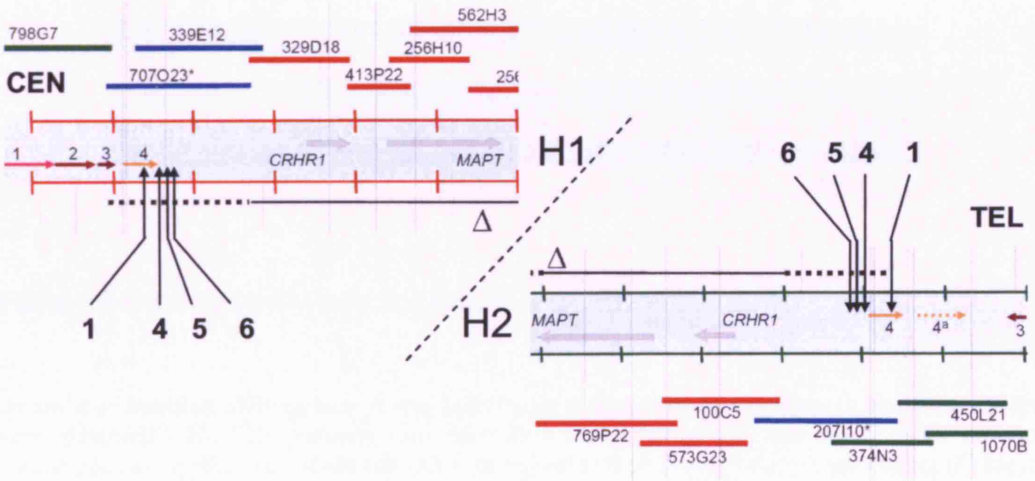

Figure 6.8 Heterozygosity mapping

Shown are the relative positions of the heterozygosity mapping sequencing amplicons (1, 4-6) on H1 and H2 haplotypes, as **Figure 6.3**. CEN; centromeric, TEL; telomeric.

Considering all of this information together; the haplotype make up of the affected individuals and the physical regions in which are sequenced on H1 and H2 for the heterozygosity mapping, this approach should be useful in mapping the centromeric

breakpoint of affected individual number one and the telomeric breakpoint of affected individual number three.

The results of the heterozygosity mapping are shown in **Table 6.3**. There is homozygosity observed in amplicons 4, 5 and 6 for the affected individuals compared to heterozygosity in the paternal samples. The two individuals are homozygous for opposite alleles, as expected since the individuals other WT chromosomes are different. These sequence results show that the deleted region extends at least to within ~ 4 kb of LCR4. Taken together with the other SNP/haplotype data of **Table 6.2**, this implies that the loss of heterozygosity, or indeed the gain of homozygosity occurs within a position to which only one LCR subunit maps: LCR4. Assuming the centre of LCR 4s as the breakpoints, this would size the deletion to 650 kb.

	Trio 1		**Trio 2**	
	P	**A**	**P**	**A**
1	(A/G) + (C/T)	(A/G)+**(C/C)**	(A/G) + (C/T)	**(G/G)**+ (C/T)
2	-	-	-	-
3	-	-	-	-
4	(A/G) + (A/C)	**(A/A)** + **(C/C)**	(A/G) + (A/C)	**(G/G)** +**(A/A)**
5	(A/G)	**(A/A)**	(A/G)	**(G/G)**
6	(A/G)	**(G/G)**	(A/G)	**(A/A)**

Table 6.3 Heterozygosity mapping results

Shown are identified SNP variants in the individuals sequenced. In amplicons 1 and 4 two SNPs were identified. No SNP variants were identified in amplicons two and three. SNP variants homozygous in the affected individuals (A) with respect to their H1/H2 heterozygous father (P) are in bold.

The sequence results of amplicon number 1 are also of interest. The PCR primers amplify a region of LCR4; two copies on both H1 and H2. On inspection of the RP11 clones encompassing all LCR 4 subunits, it would appear that the two SNPs identified (**Table 6.3)** correspond to allele differences in LCR 4 subunits: RP11-707O23 (G-T), RP11-259G18 (G-C), RP11-162O14 (G-C) and RP11-374N3 (A-C). Thus a normal individual heterozygous for H1/H2 should always be heterozygous at

both of these SNP loci: (A/G) and (C/T). The two affected individuals are not heterozygous at both of these loci, and it appears that individual one has lost an H1 allele (T from RP11-707O23) and that individual two has lost an H2 allele (A from RP11-374N3). This suggests that there is a copy number change (from 4 to 3 subunits) of LCR 4 in these two individuals.

6.7 Proposed mechanisms of non-allelic homologous recombination

The genomic architecture associated with the H1 and H2 haplotypes at 17q21.3 present possibilities for the generation of microdeletions by the mechanism of non-allelic homologous recombination (NAHR) [46, 47, 49, 232-235].

LCR4 maps closest to the deletion breakpoints on both the H1 and H2–specific FISH analyses. Furthermore, the clear-cut centromeric breakpoint region from the fosmid array-CGH data and the heterozygosity mapping data would appear to place the deletion breakpoints at the LCR4 subunits.

There are two copies of LCR 4 on each of both the H1 and H2 backgrounds. On H1, the orientation of the two is opposing whereas on H2 (LCR subunits 4/4a) they are situated in tandem. In H1/H2 heterozygotes, the change in orientation of LCR 4 on the H2 background means that NAHR between these subunits is now possible, (as shown in **Figure 6.8**), generating a deletion predicted to be approximately 650 kb in size. A mitotic cross at this position is consistent with Trio 1 since the observed SNP genotypes do not conform to H1/H2 dichotomy between the two SNP markers rs2668643 and rs2532418, wherein lies LCR4. The LCR4 subunits share homology of 98% and are largely composed of interspersed repeating element (58% in total) consisting of 40% SINES (predominantly ALU's), 10% LINES, 7% LTR and a small amount (1%) of other DNA elements.

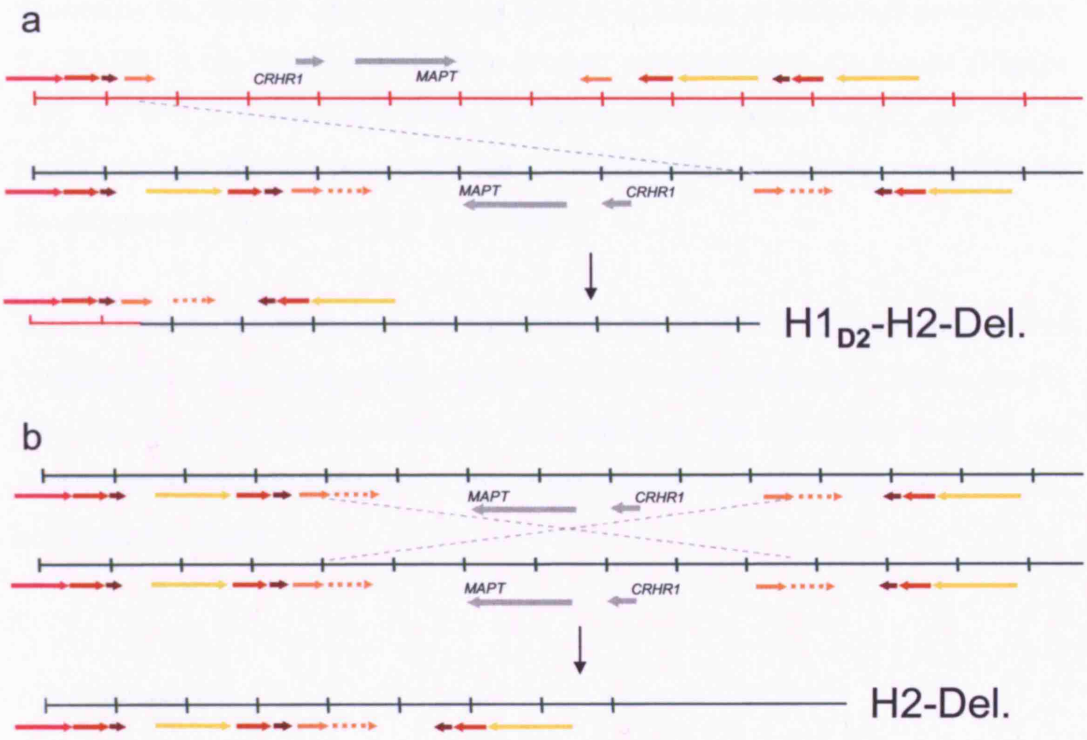

Figure 6.9 Proposed mechanism of NAHR

(**a**) Possible NAHR event in individual heterozygous for the H1 and H2 haplotypes. The product of this event is a deletion approximately 650 kb in size, and a 'hybrid' chromosome of H1 type centromerically and H2 type telomerically. Representation of LCRs is as for **Figure 6.3**.

(**b**) Possible NAHR event in individual homozygous for H2 haplotypes. The product of this event is a deletion approximately 650 kb in size. LCR substrates for NAHR are joined by the dotted lines.

One prediction to arise from this is that NAHR between H1 and H2 chromosomes should always give rise to a 'hybrid' H1/H2 chromosome, consisting of proximal H1-type alleles and distal H2-type alleles. This prediction is borne out by the observed genotype data for Trio 1, but not for Trios 2 and 3. In Trio 3, centromeric and telomeric markers are of type H2, despite an apparently heterozygous parent of origin. An explanation of this could be a possible gene conversion event involving the H1 alleles at this position, thus although there may be a mitotic H1/H2 cross, the genotyped markers are not informative enough to detect this. Alternatively, the NAHR event could have occurred intra-chromosomal, that is, in sister chromatids in the H2 chromosome; the apparent size of the deletion on the H2 background is

flanked by the directly oriented repeats (LCR4/4a) and as an additional possible site for NAHR, it can generate a deletion product consistent with the results (**Figure 6.8**). In Trio 2, it was not possible to distinguish between an H1/H2 and H2/H2 parent of origin, thus the possibility that the NAHR event could have occurred in the homozygous H2 father cannot be excluded.

NAHR between LCRs on the H1 background (as in H1/H1 homozygotes) is not consistent with the genotype/haplotype data and the orientation of LCRs on the H1 background are inversely orientated, not satisfying the conditions required for NAHR; no deleted chromosome in any of the affected individuals has both proximal and distal H1 alleles.

6.8 Discussion

Three patients had been identified with deletions at 17q21.31. The deletion encompasses the gene encoding *MAPT* and these cases represent haploinsuffciency of *MAPT* in humans. The deletions are of similar size (approximately 650 kb), and all appear to have a common centromeric breakpoint. The deletion is encompassed by a 900 kb inversion polymorphism that represent the H1 and H2 lineages, which diverged as long as 3 Million years ago and have shown no evidence of having recombined over that period [214].

Consistent with the genotype/haplotype data presented, individuals who are H1/H2 heterozygous and possibly that are H2 homozygous, but not H1 homozygotes are predicted to be susceptible to deletions in this region through the mechanism of NAHR. A prediction from this present study is that if the presence of the H2 allele predisposes to deletions at 17q21.31, the incidence of these deletions should be restricted to those populations in which the H2 haplotype occurs, namely those of Europe and the Middle East [213].

Missense and splice-site mutations in *MAPT* cause frontotemporal dementia with parkinsonism linked to chromosome 17 with tau pathology (FTDP-17T) and

common genetic variation of *MAPT* is associated with progressive supranuclear palsy, corticobasal degeneration (See introduction and Chapters 4 and 5). The possible role of *MAPT* in development has been studied in mice lacking the *Mapt* gene (*tau* $^{-/-}$). The mice are phenotypically and developmentally normal, however, histological examination revealed a reduced microtubule density and stability in some small calibre axons and more recent phenotypic analysis of these mice has revealed muscle weakness in the wire-hanging test, hyperactivity in a novel environment and an impairment of contextual fear.

The deleted region also contains other genes: corticotrophin releasing hormone receptor 1 (*CRHR1*), intramemebrane protease 5 (*IMP5*) and the predicted genes NP_689679.1, NP_787078.1 and KIAA1267. Of these five genes, there is some evidence *CRHR1* is important in central nervous system development [236] but at this stage it is not possible to say if this or any of these genes might contribute through haploinsufficiency to the clinical phenotype.

The results of the work in this Chapter were published in *Nature Genetics*:

Pittman,A.M.,Shaw-Smith,C., Willatt,L., Martin,H., Rickman,L., Gribble,S., Curley,R., Cumming,S., Dunn,C., Kalaitzopoulos,D., Porter,K., Prigmore,E., Krepischi-Santos,A.C., Varela,M.C., Koiffmann,C.P., Lees,A.J., Rosenberg,C., Firth,H.V., de Silva,R., Carter,N.P. (2006). Microdeletion encompassing *MAPT* at chromosome 17q21.3 is associated with developmental delay and learning disability. Nature Genetics 38(9), 1032-1037.

7 Identification of a novel *MAPT* mutation in FTDP-17T

7.1 Overview

A novel *MAPT* mutation, S305I, was identified in a Hungarian individual with a pathological diagnosis of frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17T).

7.2 Background

After the first identification of *MAPT* mutations in FTDP-17T, the description of over 35 mutations in more than 100 families worldwide has established the importance of tau dysfunction as central in the aetiopathogenesis of this disorder. The mutations affecting *MAPT* in FTDP-17 can be divided into two classes, the missense and the exon 10 splicing mutations. The chief effect of the former is in the biochemistry of the tau protein, where in most cases, the MT-binding capacity is reduced, leading to a greater abundance of unbound tau species that may lead to the aggregation and the formation of the insoluble tau inclusions that characterise these disorders. Other missense mutations enhance tau-tau interactions thereby increasing its tendency to form fibrils and filaments. The exon 10 splice site mutations effect tau at the RNA level and mainly cause increased incorporation of exon 10, thereby increasing the relative levels of 4R-tau isoforms.

Although classified as FTDP-17T, these neurodegenerative disorders have diverse clinical phenotypes including movement disorders such as progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD), memory dysfunction similar to that found in Alzheimer's disease (AD) as well as typical clinical pictures observed in Pick's disease (PiD) and Frontotemporal lobar degeneration (FTLD) [80]. In fact there are reports of the same mutation, P301L causing either CBD or FTDP-17T in the same family [237]. It is therefore clear that the manifestations of each mutation are subject to potential epigenetic and environmental factors.

The aim of this work was to describe the clinical and pathological phenotype associated with the novel S305I mutation.

7.3 Samples

The DNA sample with the identified mutation was obtained from a Hungarian individual who had on autopsy had pathological changes consistent with a tauopathy. The control individuals consisted of 95 (n=190 chromosomes) unrelated Hungarian individuals free of neurological disease.

7.4 *MAPT* sequencing

For the individual *MAPT* exon numbers 1, 9, 10, 11, 12 and 13 were sequenced. The other exons were not sequenced since no mutations have ever been identified in any other exon of *MAPT*. The primer sequences are shown in **Table 7.1**. PCR products were amplified from the individual sample DNA, subjected to DNA sequencing (both forward and reverse strands) and the sequence traces were analysed in BioEdit software (section 2.1.5). The presence of mutations in the sequences was checked by visual inspection of the chromatograms.

Exon	Forward Primer (5'-3')	Reverse Primer (5'-3')	bp
1	CAACACTCCTCAGAACTTATC	CAGTGATCTGGGCCTGCTGTG	229
9	AGTGGTGAGCCTGGGAATG	ATGCACAGTCCCACGACTC	492
10	GGTGGCGTGTCACTCATCC	GTACGACTCACACCACTTCC	200
11	AGCAGTTCCAGCCTCACCAG	TTGGCAGAATTTCGACAACAC	371
12	GTCCTGTCATTGTCTTCTTC	ACCCACTGGATGCTGCTGAG	437
13	CTTTCTCTGGCACTTCATCTC	CCTCTCCACAATTATTGACCG	299

Table 7.1 *MAPT* sequencing primers

7.5 Results

7.5.1 Identification of the *MAPT* S305I exon 10 mutation

A novel mutation in exon 10 2 bp before the exon10/intron10 splice-junction was identified in the pathological Hungarian sample (**Figure 7.1**). The mutation substitutes a G nucleotide to a T and changes the amino acid serine to isoleucine

(S305I). This variant was not identified in the 95 control samples from Hungary indicating that this variant is associated with the FTDP-17 phenotype and does not occur in healthy individuals.

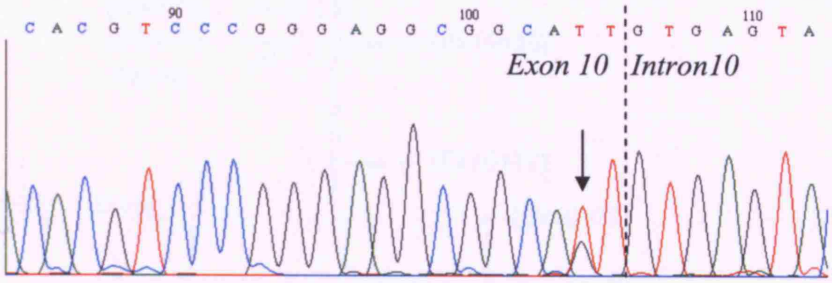

Figure 7.1 Chromatogram S305I *MAPT* mutation

The chromatogram depicts G to T nucleotide transition resulting in the S305I exon 10 *MAPT* mutation.

7.5.2 Possible effect of the S305I mutation on tau biochemistry

The novel S305I mutation appears to have an effect at both the RNA and the protein level (**Figure 7.2**). Firstly, the G to T (U in the transcribed RNA) nucleotide transition is directly opposite the nucleotide at the +16 position and disrupts the stem-loop structure. *MAPT* mutations at the same codon (S305S [166] and S305N [115]) also disrupt the stem-loop structure and cause FTDP-17T [91]. These mutations increase the proportion of exon 10 inclusion and thus 4-repeat tau protein and it is likely that the S305I mutation has this same effect.

Secondy, the amino acid serine is substituted to isoleucine. Serine is an amino acid with an uncharged polar side chain and isoleucine has a hydrophobic side chain. This change could have a detrimental effect on tau protein function, for example, decreased microtubule binding and/or a greater tendency for tau to assemble into filaments. Thus, the S305I mutation is interesting because it not only disrupts the balance of 3-repeat to 4-repeat tau but may also alter tau-microtubule and/or tau-tau interactions.

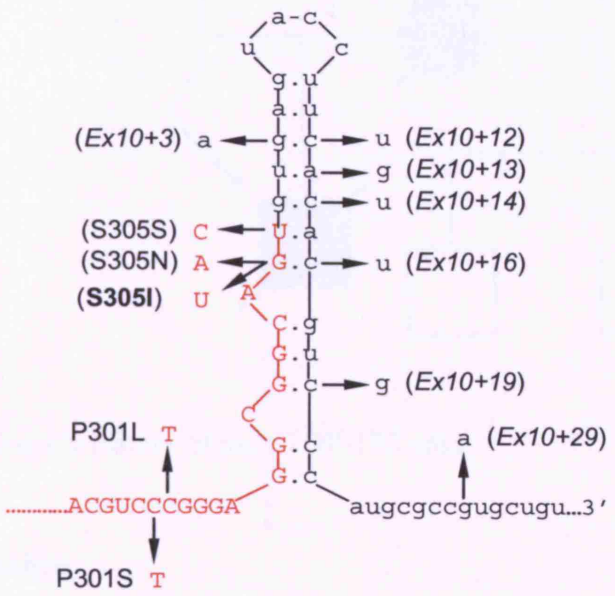

Figure 7.2 The S305I mutation in relation to the exon 10 stem-loop structure

Shown is the S305I (**bold**) mutation in relation to other exon/intron 10 mutations.

7.5.3 The clinical and pathological phenotype associated with S305I

The male patient first developed personality changes at the age of 39 and later developed parkinsonism, dementia, speech disorder, anxiety and later depression. SPECT (single photon emission computed tomography) demonstrated frontotemporal hyperfusion. The patient died at the age of 40 and the clinical duration was approximately 1.5 years. The patient had a family history of neurological disease (**Figure 7.3**) in that the mother died of a neurological disease at the age of 45. No data on the patient's father was available. The patient's mother has two more sons from another father; they are around 40 years of age and both are chronic alcohol abusers. No data on the grandparents was available.

Pathological examination of the diseased patient revealed a tauopathy with abundant oligodendroglial and astroglial tau pathology in subcortical structures, frontal and temporal cortex. Furthermore, pre-tangles were also observed in the limbic areas.

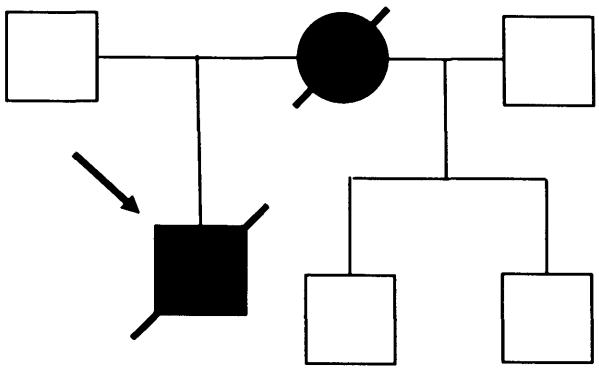

Figure 7.3 Family history of the FTDP-17T case

7.6 Discussion

The novel *MAPT* mutation, S305I, was identified in a deceased Hungarian individual, confirming the clinical and pathological diagnosis of FTDP-17T. FTDP-17T is inherited as an autosomal dominant condition characterised by behavioural, cognitive, and motor disturbance with an onset typically between 45 to 65 years of age and average disease duration of 4 years. This new case had typical clinical symptoms associated with FTDP-17T, though with a much earlier disease onset at the age of 39 and a rapid disease course of 1.5 years. At autopsy, patients with FTDP-17T display frontotemporal atrophy and intra-neuronal tau inclusions are present on histological examination consistent with the changes observed in this case.

The mutation itself is similar to two previously described; S305S and S305N [115]. The former is a silent mutation that does not alter an amino acid, however, it does disrupt base-pairing in the stem-loop structure opposite the +15 position and consequently increases the incorporation of exon10-containg mRNA [165]. The S305N mutation changes the amino acid at this position from serine to asparagine and in addition disrupts the stem-loop in a similar way, at the +16 position [115].

Three cases with the FTDP-17T S305S mutation mimicked progressive supranuclear palsy (PSP) [165]. The subject had no significant atrophy or cellular degeneration of the frontal and temporal cortices and the cellular pathology was characteristic of

PSP, with neurofibrillary tangles concentrating within the subcortical regions of the basal ganglia. Two additional affected family members presented with symptoms of dementia and later developed neurological deficits including abnormality of vertical gaze and extrapyramidal signs. A case with the S305N mutation closely resembled Pick's disease [238] though additional individuals studied noted features typical with FTDP-17T [124, 239].

This work identified a novel *MAPT* mutation that causes FTDP-17T and provides new data on the associated clinical and pathological phenotype, in particular early disease onset and rapid disease duration. This could reflect the nature of the new mutation in which it appears to affect the balance of 3R and 4R tau isoforms and could alter tau interactions at the microtubule binding domain.

8 Discussion

8.1 Summary of results

In the first place the outer extents of the H1 haplotype block associated with progressive supranuclear palsy (PSP) were defined by genotyping a series of single nucleotide polymorphisms (SNPs) at regular intervals (~50 kb) away from *MAPT*. The entire haplotype block was in tight linkage disequilibrium (LD) and was found to cover a region of ~2 Mb at 17q21.31 which was subsequently refined to 1.8 Mb. Locus-by-locus association analysis revealed that the entire interval was associated with PSP. With the full extent of the PSP-associated H1 haplotype defined, this significantly increased the candidate region to include several other genes in addition to *MAPT*, including corticotrophin releasing hormone receptor 1 (*CRHR1*), presenilin homologue 2 (*IMP5*), saitohin (*STH*), nitrate-ethylamalemide sensitive factor (*NSF*) and wingless-type MMTV integration site family member 3 (*WNT3*). Although the region contained numerous genes, within the scope of this work, *MAPT* gene was selected as the prime candidate potentially harbouring the pathogenic basis of the H1 association and was targeted for high-resolution association mapping utilizing dense publicly available HapMap project data. Genotype data in 27 CEPH-trios for 24 SNPs across the gene was downloaded to explore the *MAPT* haplotype diversity. Several common variants of H1 and the sole H2 haplotype were identified. With the common haplotypes of the *MAPT* defined, a set of 6 haplotype-tagging SNPs (htSNPs) were selected that together represented the genetic diversity of the gene. These markers were then genotyped in the target populations; pathologically confirmed PSP cases from both the UK and the US and one corticobasal degeneration (CBD) cohort from the US. Association analysis revealed that two common haplotypes were associated with both PSP with the same trend in CBD; over-representation of a variant of H1 (H1c) and under-representation of H2. A series of H1-specific SNPs that were tightly associated with PSP defined a minimal candidate region that could harbour potential causal variant(s) an the background of *MAPT*, a region starting upstream of exon 1 downstream to intron 9.

The genetic risk of pathologically proven PSP conferred by H1c was investigated in the two distinct clinical subtypes of PSP; Richardson's syndrome (RS) and PSP-parkinsonism (PSP-P). The H1c haplotype was significantly associated with both groups suggesting that variations of the *MAPT* locus do not account for these clinical differences. This suggested that additional, unidentified genetic or environmental factors were responsible, for example a study of the ApoE locus showed no significant difference between RS and PSP-P eliminating a potential role of ApoE aetiology of either PSP sub-type.

In order to investigate the functional basis of the *MAPT* H1c association with PSP, a re-sequencing strategy was devised to identify candidate polymorphic loci on the H1c background that could potentially affect *MAPT* gene behaviour. Several SNPs identified from the sequencing were short-listed because they defined the H1c background, and of these, two SNPs, rs7210728 and rs242557, were located in evolutionarily conserved islands in the vicinity of the promoter region of the gene. Allele-specific luciferase reporter gene studies revealed allele-specific differences in *MAPT* promoter driven expression of different variants, in particular those of SNP rs242557 (A/G). SNPs and haplotypes of the variants associated with greater gene expression are over-represented in PSP and thus implicating that, at least in part, that the functional basis of the H1c association is due to increased *MAPT* expression.

Based on the high resolution LD and haplotype map of the *MAPT* region obtained during this work, a *de novo* microdeletion encompassing *MAPT*, in three children presenting with developmental delay and learning disabilities, a detailed SNP haplotype analysis was undertaken in respect of their unaffected parents. This was in order to determine if any relationship existed between the microdeletion and H1/H2 haplotypes at 17q21.31 and to pinpoint the breakpoints of the deletion to unravel a possible structural mechanism.

In two of the three family triads it was established that parents-of-origin were both heterozygous for H1/H2. However, in one case it was not possible to determine the parent-of-origin. All three cases shared distal H2 chromosome segments at the telomeric side of the deletion and by sequencing two affected individuals close to

the suspected breakpoint regions it was possible to demonstrate that the breakpoints were likely contained within highly homologous (>98%) flanking low-copy repeat (LCR) subunits. Analyzing the architecture at 17q21.31 revealed that H1/H2 heterozygotes and also possibly H2/H2 homozygotes were susceptible to the microdeletion and consequently a mechanism of non-allelic homologous recombination (NAHR) was proposed.

A novel mutation; S305I in exon 10, was identified in a case of FTDP-17T. This data provided valuable new information regarding the clinical phenotype associated with the respective mutations and, in particular, information on the pathological changes associated with the novel S305I mutation.

8.2 General discussion

8.2.1 Extended haplotypes at 17q21.31

Mapping the extent of the region of complete LD revealed that it extended for nearly 2 Mb and included several other genes besides *MAPT* [217] and has now been shown to be the longest region of LD in the human genome [240]. Analysis of the global distribution of the H2 haplotype revealed that it is restricted to Caucasians [213, 214] and the structure of the extended haplotypes is shared by individuals [215]. Sequence analysis of the human assembly across the locus revealed the presence of LCRs [145] and these sites are associated with structural rearrangements: these include a 3 million year old 900 kb inversion polymorphism perfectly corresponding to the H1/H2 lineages and a complex arrangement of polymorphic duplications close to the *NSF* gene on the H1 background [214] and the *de novo* deletion described in this work [241]. These recent findings explain that the perfect H1/H2 LD at 17q21.31 [217] is due to a physical constraint suppressing recombination between the two lineages and identifying other such high regions of LD in the human genome could aid in the identification candidate regions harbouring novel inversion polymorphisms [33].

8.2.2 The *MAPT* association with PSP

Although by default all genes contained within the extended H1 haplotype are associated with PSP, the hallmark tau pathology of this disorder and the identification of dominant mutations in the related tauopathy, FTDP-17T, strongly implicate *MAPT* itself. In the present study a common variant of the H1 haplotype, H1c, was found to be preferentially associated with PSP and interestingly another common variant of the H1 haplotype, H1b, was neutral. The interval with the strongest association defined a minimal candidate region on the H1 background that remained significant even after consideration of the H1/H2 (*del-In9*) association and suggest that common variation within the extended H1 clade in *MAPT* is a risk factor for PSP. The strongest associated H1-specific SNP in the US population was rs242557 (A/G) and in the UK population was rs2471738 (C/T) though both associations were highly significant in both populations. The associated H1-specific SNPs that defined the minimal candidate region identified in this work implies that the functional basis of the association could be due either to general expression levels of the gene, an effect on alternative splicing (or perhaps both) or to differences in RNA stability.

The association of rs242557 with PSP was later identified in an independent study [242] with a sample series that overlapped the US series detailed in Chapter 4. However, in this study the association was also replicated in an independent cohort from the US. Interestingly in the study by Rademakers *et al.* [242], the population was stratified by including only H1 haplotype carriers thus removing the confounding effect of the associated H2 haplotype. Unfortunately, the authors did not ascertain the effects removing H2 on the allele frequencies of the other SNP loci tested for association. In the present study (detailed in Chapter 4), removing H2 carriers from the control and PSP populations resulted in significant heterogeneity ($p<0.05$) in both the US and UK control populations at rs242557 and rs2471738, respectively, [243] and thus removal of the H2 chromosomes from the present study would have changed the allele frequencies of these SNP loci in the case and control populations. This could have led to erroneous association study results and requires unbiased inclusion of all samples regardless of H1/H2 status. A valid H1-only

approach would be to examine the association of the *MAPT* locus in an H1-only population such as the Taiwanese or Japanese [213].

These findings also highlight the importance of high resolution association mapping in candidate genes, taking into account the diversity of the entire locus rather than just genotyping just a few scattered polymorphisms throughout the gene in question. This is to maximize the chances of identifying potential causal variants and indeed to exclude non-associated genes with confidence [211].

The identification of the H1c haplotype as a risk factor for PSP refines and elaborates on the association of the *MAPT* locus and provided the basis for a targeted re-sequencing strategy in order to identify functional variant(s) responsible for the association.

8.2.3 *Is there a general MAPT association with all tauopathies?*

The *MAPT* locus has been tested for association with a variety of related diseases including Alzheimer's disease (AD), Parkinson's disease (PD), frontotemporal lobar degeneration (FTLD), amyotrophic lateral sclerosis and parkinsonism dementia complex of Guam (ALS-G and PDC-G, respectively) and CBD. The tests of association were facilitated for the most part by H1/H2 haplotyping and on this basis both ALS-G and PDC-G showed associations [244] and PD, AD and FTLD have not shown a consistent association of the H1 haplotype.

In PD a total of nine case-control association studies have evaluated the association of *MAPT* [170, 175, 245-250]. However, the results of individual studies for the large part are largely underpowered and thus, altogether failed to give a definite answer. In the largest study to date [175] a significant association was identified (OR: 1.57, 95% CI: 1.33-1.85) and when combined with all previous studies in a meta-analysis, there was an overall association for homozygosity of the H1 haplotype [175]. In one study of PD cases and controls from Trondheim, Norway a very strong effect of *MAPT* H1 on disease risk was identified (OR: 5.52, 95% CI: 2.64-11.10, [246]). The magnitude of this association was grater than any other

previously observed in PD studies, and the effect size is similar to that seen in PSP. This could be because the Trondheim population, close to the Artic Circle, has a low immigration rate and represents a geographically remote and genetically homogeneous population and as such is ideal for case-control studies.

In AD, the association studies of the H1 haplotype, restricted only to H1/H2 haplotypes, have largely been negative [251-254]. In a recent study utilizing the set of six *MAPT* haplotype-tagging SNPs (htSNPs) detailed in Chapter 4 of this thesis, a significant increase of the H1c haplotype was identified in a late-onset pathologically confirmed AD series [255]. This finding has since been replicated in an independent cohort (Dr. Amanda Myers, personal communication). The H1c haplotype association in AD is the same direction as in PSP although to a much smaller extent. In both AD series, the H2 haplotype was not significantly under-represented and explains why the original studies were inconclusive when considering H1/H2 haplotypes alone. It also underlines the importance of analysing underlying population-wide diversity of the whole gene in order to identify associations.

In sporadic FTLD (cases that are negative for *MAPT* mutations) the association studies have again been restricted to the H1/H2 haplotypes [209, 256-258]. These studies represent both positive and negative findings for H1. In one study of a sample series that contained familial FTD samples that were negative for *MAPT* mutations, an over-representation of the H2 haplotype was found [259]. However this was a very underpowered study and the increase in H2 could be explained by the fact that a prevalent mutation in progranulin at 17q21.31 is associated with an ancestral H2 haplotype [260] and thus, the haplotype itself may not be the risk factor for familial FTD in this cohort, rather a high frequency of progranulin mutations. Far more studies with larger case numbers providing greater power are needed to provide a definite answer.

In CBD, for the most part there has been a consistent association of the *MAPT* locus [173, 174, 209, 243], with most studies finding a significant elevated frequency of either H1 or a H1 derived haplotype. The association with CBD (OR: 3.56, CI 95%

1.53 - 8.43, [261]) would appear to be almost as strong as for that seen in PSP. The present study in this thesis (Chapter 4) would suggest that this is due to an over-representation of the H1c haplotype and an under-representation of the H2 haplotype; however, due to the small sample size employed, this would require replication in a larger more powerful cohort.

Taken together the *MAPT* locus appears to be a strong genetic risk factor for PSP and CBD, and to a lesser extent with PD and AD. The current findings hold that the locus could also be a risk factor for FTLD, ALS-G and PDC-G but additional studies are required. The association with PD is somewhat surprising since this disease is not traditionally associated with tau pathology.

8.2.4 The functional basis of the *MAPT* association

8.2.4.1 Expression

The identification of the H1c haplotype as a risk factor for neurodegeneration provided the basis of a more targeted approach in dissecting the functional basis of the *MAPT* H1 association. On the H1c background the region that gave the greatest association signal was from the large intron upstream of *MAPT* exon 1 to the 3' region of intron 9 [243] suggesting that the functional basis of the association may relate to general expression or to alternative splicing of the gene, or even a combination of the two. A basis for increased expression is the functional H1-specific SNP rs242557 (A/G), which is located within intron 0 of the gene in a 262 bp evolutionarily conserved island (76% conservation between mouse and human), approximately 47 kb downstream from the *MAPT* core promoter at exon -1. Luciferase reporter gene experiments showed that the allelic variants of rs242557 (A/G) in the context of it function as a gene control element alter *MAPT* promoter driven expression. The allele A (the variant that constitutes the H1c haplotype) has a significantly (~4 fold) higher expression level than the G allele (the variant constituting the H1b and the H2 haplotypes).

In a previous luciferase reporter gene study of the *MAPT* core promoter, the H1 variant (H1*p*) was found to express 1.2 fold higher than that of H2 (H2*p*) [227] however this difference was not significant in the present study. In another study the allelic variants of rs242557 (A/G) in the context of the H1*p* core promoter were investigated for effect on expression though in this case, the allele G was found to be a significantly higher expresser than the A allele [242]. These results do appear to conflict with the present study but in this case the experimental design was quite different, most notably the putative element containing rs242557 was cloned upstream of the *MAPT* core promoter as opposed to the correct downstream context as in the present study. This demonstrates that the relative position and spacing of a gene promoter in relation to nearby control elements in such experiments is essential in mimicking the *in vivo* situation, though taken together both studies highlight a significantly different effect in the expression level associated with each rs242557 (A/G) allele.

In PSP, the H2 haplotype is significantly under-represented and this suggests it protects against the disease. From the results obtained in the present study, the strongest effects on *MAPT* promoter driven expression were not SNP loci distinguishing the H1 and H2 haplotypes and it is highly plausible that the inversion itself alters the high order structure of the region and leads to differential expression levels of H1 and H2 haplotypes due to positional effects. Such an effect could not be investigated with luciferase reporter gene experiments.

8.2.4.2 Splicing

Another important functional consideration is alternative *MAPT* splicing. The present study does not address this potentially important effect of the H1c haplotype. The interval in the gene showing the association included the region of the alternatively spliced exons 1 and 2, and intron 9 in sequences that could potentially affect the regulation of exon 10 alternative splicing. PSP and CBD are predominantly 4-repeat (4R) tauopathies and a disruption of the balance of 3R and 4R tau expression could potentially lead on the pathway to neurodegeneration. There is evidence to suggest this pathogenic mechanism, in FTDP-17T,

approximately half of the mutations affect the alternative splicing of exon 10 at the RNA level, predominantly increasing levels of 4R tau protein and neurofibrillary tangles. In some instances, these mutations can even cause a clinical picture resembling that of PSP, for example a +16 intron 10 mutation in an individual presenting clinically as sporadic young onset PSP [164].

8.2.4 Tau dysfunction and neurodegeneration

Two contradictory tau hypotheses have been suggested [262]:

- Aggregated tau represents the toxic species.
- Aggregated tau represents a biologically inert pool to sequester and inactivate a toxic soluble tau species.

Transgenic mice over-expressing human tau have consistently demonstrated neurological defects and neuron loss appearing with neurofibrillary tangles (NFTs) [263-266]. In addition to this approach, numerous transgenic mice have been produced that express human tau with FTDP-17T mutations [267-270].

Transgenic mice, which overexpress 3R tau exhibit intraneuronal filaments of hyperphosphorylated tau by six months of age [271] and a reduced axonal transport and degeneration is observed. The aged transgenic mice also develop hyperphosphorylated NFTs in the hippocampus, entorhinal cortex and the amygdala [272]. This demonstrates that NFTs can develop with wild-type tau in the absence of additional exogenous factors. Moderate over-expression of 4R human tau in mice leads to hyperphophorylation of tau, but not the development of filament or tangles [273]. These mice show dilated axons in the spinal cord and the cerebral cortex and go on to develop behavioral deficits. Interestingly the neurodegeneration in this system is likely caused by an excess of tau binding to the microtubules since it can be rescued by co-expression of glycogen synthase kinase 3 beta (GSK3β), a tau kinase that phosphorylates tau thus reducing its affinity for microtubules.

The majority of studies of tau dysfunction in animal models are those expressing tau associated with FTDP-17T mutations. Several of these models recapitulate many of

the features of tauopathies, including the age dependent accumulation of filamentous tau pathology associated with behavioural abnormalities, neuronal loss and defects in axonal transport. In transgenic mice expressing mutant tau (P301L), filaments and tangles are formed and apoptotic events are observed [265, 274, 275]. In one recent study of mice with repressible human P301L tau expression, the mice developed age-related NFTs, neuronal loss and behavioural impairments [276]. After the suppression of transgenic tau, memory function recovered and neuron numbers stabilized but NFTs continued to accumulate. Thus, in this model of tauopathy, NFTs are not sufficient to cause cognitive decline or neuronal death and thus challenges the idea that pathogenic tau is due to the formation of NFTs alone.

This new data suggests that if the aggregation of abnormal tau filaments into NFTs is associated with a toxic gain-of-function, this alone may not be sufficient to cause the degeneration of neurons in the tauopathies. Thus, rather than being a consequence of a toxic gain-of-function by NFTs, pathological tau may induce neurodegeneration pre-tangle formation. NFTs could just be an incidental marker for the actual neurotoxic cascade in the transgenic mice representing a protective neuronal response, allowing sequestering of neurotoxic species into a less harmful and stable form.

The effect of over-expressing tau on the trafficking of vesicles and organelles in primary cortical neurons and neuroblastoma cells has been studied. Tau over-expression in these cells inhibits kinesin-dependent transport of peroxisomes, neurofilaments and Golgi-derived vesicles into neurites and inhibits the forward and reverse transport of APP along the axons [277]. The tau inhibition of these organelles, vesicles and neurofilaments is consistent with the view that these components are carried down the axon by kinesin-dependent transport along microtubules [278] and tau-transfected cells have shorter neurites [277]. This transport inhibition has serious consequences for the growth and survival of cell processes and tau-transfected cells are sensitive to oxidative stress, namely H_2O_2. This is presumably due to the absence of peroxisomes, since exogenous catalase provides protection [277].

It is possible that different toxic states of tau exist, which induce neurodegeneration by different mechanisms and is supported by the spectrum of *MAPT* mutations in FTDP-17T that affect tau dynamics in a variety of ways.

In an asymptomatic carrier from a three generation FTDP-17T family bearing the P301L mutation, the clinical and biochemical features were examined [279]. The subject was examined at 50 years of age and had a normal neurological examination, with no extrapyramidal deficit or frontal signs. The subject was leading an appropriate social life, had had no complaints about behavioural disturbances or mood disorders and was holding down a demanding job. On cognitive evaluation, the scores were normal on all tests, except for a notable reduction on the Verbal Fluency test for letters. The patient's brain computed tomography (CT) scans were normal though there was an elevated level of both total tau protein and p-SER181-tau (tau protein phosphorylated at serine residue 181) in the CSF (cerebrospinal fluid) compared to control subjects. Interestingly, in overt FTD patients, tau in CSF is almost undetectable [280] as abnormally phosphorylated soluble tau might be sequestered into tangle formation in the disease process, although preclinically there is a greater abundance on total and phosphorylated tau.

8.2.5 Increased *MAPT* expression and the risk of PSP

The data from the work in this thesis suggests that specific SNP variants increase *MAPT* expression and that this is the risk factor for PSP. It would be particularly useful to extend the reporter gene experiments described in this work to quantitative measurements of *MAPT* expression in post-mortem tissue from the PSP brain (see section 8.3.2). In fact, preliminary data from collaboration with the Laboratory of Neurogenetics (Professor John Hardy and Dr. Amanda Myers), NIA, NIH, strongly suggests that the H1c haplotype gives rise to higher *MAPT* expression and a grater proportion of exon 10-containing transcripts compared to the H2 and H1b haplotypes (Dr. Amanda Myers, personal communication). This new *in vitro* data supports the *in vivo* data described in this thesis and strongly supports the notion that increased *MAPT* expression is associated with PSP and also suggests a role for an imbalance of the 3R and 4R ratio.

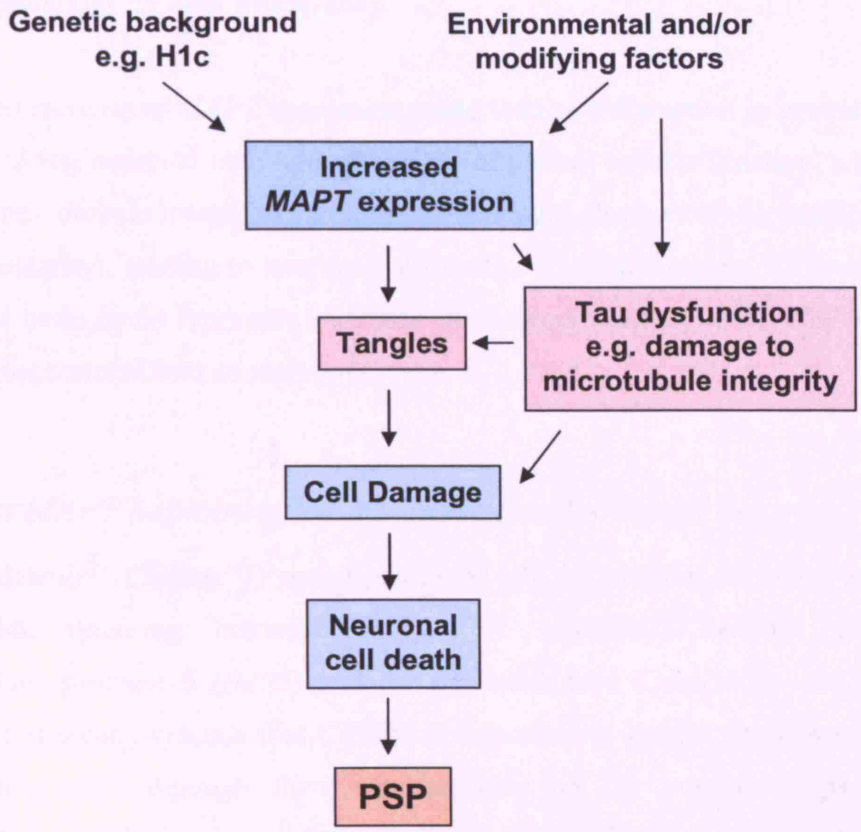

Figure 8.1 Increased *MAPT* expression and PSP

Schematic depicting two possible pathways in which increased tau expression could lead to PSP.

In **Figure 8.1**, two speculative models are proposed as to how altered *MAPT* expression may contribute to a cascade of events and the development of PSP, based upon the current tau hypothesis (see section 8.2.4). In the first place, increase in *MAPT* expression could directly lead to a greater abundance of unbound tau protein that could in turn lead to its deposition into filaments and NFTs. The formation of NFT inclusions in the neuronal cells could then lead to cellular damage, cell death and the neurodegenerative process in PSP. This theme, whereby the excess production and deposition of the pathological protein is deposited is common in many neurodegenerative diseases [281]. In PD, multiplications in α-synuclein and genetic variability in expression affects the risk of developing the disease [282, 283]. This is also consistent with the observation that individuals with trisomy 21

(resulting in three copies of the amyloid precursor protein gene) inevitably develop Alzheimer pathology by their fifties [284].

Secondly, an increase in *MAPT* expression could lead to a disruption in normal tau function in living neuronal cells and disruption of normal cellular function, e.g. by damage to microtubule integrity (that may, for example, disrupt cellular trafficking or axonal integrity), leading to neuronal cell death. The pathological NFTs in the PSP disease brain could represent a defence mechanism whereby toxic soluble tau species are sequestered into an inert pool.

8.2.6 *Does MAPT haploinsufficiency cause developmental delay?*

The microdeletion (Chapter 7) encompasses *MAPT* in addition to other genes: corticotrophin releasing hormone receptor 1 (*CRHR1*), Saitohin (*STH*), intramembrane protease 5 (*IMP5*) and the predicted gene KIAA1267. Of these genes, there is some evidence that *CRHR1* is important in central nervous system development [236], although there are available on the possible impact of haploinsufficiency of this gene on that process. KIAA1267 is a large gene adjacent to tau encoding a hypothetical protein of unknown function. *STH* has an mRNA expression pattern similar to that of tau [177] but the protein product itself and its distribution has never been studied in humans. *IMP5*, which is 47 kb from *MAPT*, is a presenilin homologue and has structural homology to presenilin 1 and 2, that are mutated in early onset familial AD [285]. The function of *IMP5* is unknown and also no pathological mutations have been identified in FTLD [285].

The possible role of tau in development has been studied in mice lacking the *Mapt* gene (*tau* $^{-/-}$). The mice are phenotypically and developmentally normal, however, histological examination revealed a reduced microtubule density and stability in some small calibre axons [286] and more recent phenotypic analysis of these mice has revealed muscle weakness in the wire-hanging test, hyperactivity in a novel environment and an impairment of contextual fear [287]. Thus, there could be sufficient redundancy among the remaining microtubule associated proteins. However several key *in vitro* experiments support an essential role for tau in axonal

elongation and maintenance. A series of antisense oligonucleotides added to culture neuron cells suppressed tau accumulation and neurite outgrowth, suggesting a key role in neuronal growth, presumably by affecting microtubules properties. However, this result was not supported by the studies in mice. It could be that tau protein has a more important role in human brain development than in mouse and haploinsufficiency of *MAPT* is sufficient to disrupt normal human brain development.

8.2.7 Structural variation and neurological disease

Dominant point mutations in the α-synuclein gene (*SNCA*) cause PD [288] and subsequently fibrillar α-synuclein was identified as the major component of Lewy bodies, the diseases pathological hallmark [289]. In a large family with autosomal dominant PD (with an average age of onset at 34 years) with a clinical phenotype ranging from dementia with Lewy bodies to typical PD, a triplication of the *SNCA* locus was identified that segregated with the disease phenotype [282], thus each affected individual caries 4 fully functional copies of the *SNCA* gene. This finding is also consistent with haplotype data that genetic variability in the *SNCA* promoter contributes to the risk of developing PD [283, 290-294] and thus an increase in expression of the protein is sufficient, at least in some cases to cause disease. Likewise, mutations in the amyloid precursor protein (*APP*) gene on chromosome 21 cause autosomal dominant AD [295] and *APP* duplications cause autosomal dominant early onset AD with cerebral amyloid angiopathy [296-298].

8.2.8 *Could multiplications of the MAPT locus cause neurodegeneration?*

Microdeletions of the *MAPT* gene are associated with low-copy repeat (LCR) rearrangements likely by a mechanism of non-allelic homologous recombination (Chapter 7) and it stands to reason that duplications could occur by the same mechanism, since one recombination product would carry a deleted segment and the other would carry the duplicated segment. The proposed mechanism for the generation of the microdeletion (Chapter 7) predicts that duplications can also occur,

though such a variant has not yet been identified and it may not even exist; it could even be a lethal allele. Identifying the duplicated variant is hampered by uncertainty of any associated phenotype, if any at all. However, since multiplications of the *SNCA* locus causes PD and duplications of *APP* case AD, a duplication of the *MAPT* locus may cause a tauopathy, by overproduction of the fibrillogenic protein. Some transgenic mouse models over expressing human tau protein develop neurofibrillary tangles.

8.3 Future research

8.3.1 Genetic association studies of *MAPT*

The present study identified the association of the H1c haplotype with PSP and CBD and it would be particularly important to extend this study by genotyping the same set of htSNPs in related tauopathies, for example FTLD, PiD and parkinsonian syndromes such as PD. In PD it appears that there is an association of the *MAPT* locus by over-representation of H1/H1 genotype but it is currently unclear if this is H1c or another H1 derived haplotype. It is possible that it could be a different haplotype with a different functional basis, for example, the alternative splicing of exons 2 and 3. The association of the H1c haplotype with CBD should be independently replicated in a larger, more powerful series.

It would be very important to extend the association studies of the *MAPT* locus in additional populations since the majority of studies have been limited to Caucasian populations in Western Europe and North America. The inverted H2 haplotype is absent in populations in Asia such as the Chinese, Taiwanese and Japanese and it would be particularly interesting to map the association of the *MAPT* gene in these H1-only populations and ascertain if a minimal region or a single locus can be implicated. The set of htSNPs identified in Chapter 4 specific for the Japanese population could be used for this very purpose.

8.3.2 *MAPT* expression studies

The results of the present study suggest that *MAPT* expression levels are different depending on genetic background, and that this could at least in part be the functional basis of the association. However, these results were generated in an artificial *in vitro* system that does not represent the full dynamics of *MAPT* gene expression. To expand upon these findings and to provide more evidence, it would prove very interesting if *MAPT* gene expression could be investigated at the RNA level in post-mortem PSP and control brains. This experiment would involve designing quantitative PCR (qPCR) assays that could directly measure the relative amount of *MAPT* mRNA. A large pathological cohort whose relative *MAPT* expression has been quantified could be tested for correlation with the individual *MAPT* genotypes/haplotypes based upon the htSNP data to see if the H1c *MAPT* expression level is greater than H1b and H2 etc. This could be carried out by comparing *MAPT* expression from individuals that are homozygous for the respective *MAPT* haplotypes. Another method would be to capitalise on the fact that *MAPT* contains SNP polymorphisms in the coding sequence, for example there are such SNPs present in Exons 1, 2, 3, 9, 11 and 13 that can be used to distinguish mRNA from the H1 or H2 alleles. Assays based on qPCR could be designed to quantify the allele-specific expression levels, for example the TaqMan assay. Individuals heterozygous for H1 and H2 could be selected and, using the H2 expression level as the internal control (the standard), measure the relative expression of H1 compared to that of H2. Heterozygous individuals could be selected who differ by H1 haplotypes, for example H2/H1b, H2/H1c, H2/H1d etc. and the expression of H1 relative to that of the H2 could be compared between individual samples that carry different H1 haplotypes. It would also be useful to examine the regional expression using these methods in both the PSP and the control brain, to see if genetic effects on expression are related to disease status or any particular brain region.

Another functional aspect of *MAPT* that requires further investigation is alternative splicing. The present study has not identified any obvious candidate variants that could affect the alternative splicing of the gene but it would be interesting to

examine the relative expression level of 4R and 3R encoding *MAPT* mRNA in respect of genotype/haplotype background. A qPCR assay could be designed to measure exon 10-containg mRNA rather than total *MAPT* mRNA. This could be achieved by designing the qPCR primers to generate a product across the exon 9 and exon 10 boundaries and this assay could also incorporate the allele-specific method described above, that is, to compare the relative amount of exon 10-containg *MAPT* mRNA of each distinct H1 haplotype verses H2. If an effect on splicing is identified, this could be further investigated by exon-trapping experiments using mini-gene constructs of *MAPT*. Such a method would be useful to identify specific polymorphisms that could influence splicing.

8.3.3 Functional studies

It would be relevant also to identify the proteins and transcription factors that differentially bind the rs242557 (A/G) alleles in the module C. It is likely that there is a differential binding of proteins for each SNP variant. This could be investigated with electophoretic mobility shift assays (EMSA), using double-stranded DNA probes of the region containing the different SNP variants. To identify the actual binding proteins themselves, the same probes could be used as bait for affinity chromatography to purify and subsequently identify the proteins by proteomics and mass spectrometry. Identifying proteins and transcription factors that influence *MAPT* gene expression would be important to understand regulation of *MAPT* expression their role in the pathogenesis of the tauopathies.

8.3.4 Refine the proposed mechanism of the LCR mediated NAHR

The proposed mechanism described in Chapter 7 needs to be confirmed by cloning and sequencing the deletion breakpoints in the affected individuals. The identification of a further 7 individuals with the same *de novo* deletion of the region [299, 300] indicates that this syndrome is likely to be a relatively common cause of mental retardation (estimated to account for 1% of mental retardation cases, [299]) and these and any other individuals that have been identified should be haplotyped with H1/H2 SNPs in respect of their unaffected parents. This would provide

additional data to help understand the specific genetic makeup required to transmit the deleted chromosome. It would also prove useful to screen for the deleted chromosome in patients with a similar phenotype in populations that are H1-only, such as the Japanese, since the current hypothesis is that the deletion cannot occur in individuals homozygous for H1.

8.3.5 Screen individuals for duplications of the *MAPT* locus

It would be particularly interesting to identify the reciprocal product of the *MAPT* locus deletions, namely duplications. Identifying such duplications would be the final piece of the puzzle and prove the proposed mechanism of NAHR. It would also be interesting to observe the associated phenotype since such individuals would have 3 copies of the *MAPT* gene as opposed to 2 and as such, are likely to have higher *MAPT* gene expression, and this could be a genetic risk factor for a tauopathy. A relatively straightforward assay could be designed based on qPCR to easily and cheaply screen a large number of samples.

8.4 Genetic association studies post-human genome project

Since the completion of the human genome sequence [301], the identification of millions of SNPs and deposition into public databases, the International HapMap project (HapMap) and rapid advances in SNP genotyping technology, genome-wide association studies are now feasible. These studies are now possible because of the rapidly falling cost of genotyping and the db SNP database contains approximately 9 million SNPs, representing the majority (>80%) of the variation with allele frequencies of more that 1% that is estimated to exist. The genotyping of haplotype-tagging SNPs based on the LD structure of the genome from the HapMap project provides yet more economical advantage. Thus the genome-wide approach therefore represents a 'giant' genetic association study that surveys most of the genome for causal genetic variants either by typing causal variants themselves or by typing them indirectly through LD.

The first genome-wide association studies are beginning to emerge and one recent study in PD and highlights potential weaknesses of the procedure [302]. In the study a two tier approach was used: For tier 1, 443 sibling pairs discordant for PD were genotyped for 198,345 SNPs. For tier 2a, 1,793 SNPs (p <0.01 in tier 1) were genotyped in 332 matched case-unrelated control pairs and for the eleven SNPs that were associated with PD in both tier 1 and tier 2a, this data was combined to derive p values, odds ratios and confidence intervals (tier 2b). A SNP (rs7702187) within the semaphorin 5A gene (*SEMA5A*) had the lowest combined p value (p= $7.62ex^{-6}$, OR: 1.74, 95% CI 1.36 to 2.24). Hot on the heels of the identification of this novel gene from the whole-genome association study, four independent replication studies [303-306] failed to replicate the associations with *SEMA5A* or even the eleven other SNPs associated in both tier1 and tier2a. As independent replication in genetic association studies is considered the gold-standard in giving confidence to true associations, the original study findings are sadly likely to be spurious. One criticism of the original study is that the PD cases were all selected on the basis that they were sporadic, and thus a substantial proportion of the cases in tier 1 and 2 may have little or no genetic basis for disease. Case identification in tier 1 should focus on the selection of those most likely to carry the inherited form of the disease, whereas controls should be likely non-gene carriers drawn from the same population [307].

There has been one notable success with a whole-genome study. In age-related macular degeneration (AMD), a leading cause of blindness in the developed world, a genome-wide screen was performed in 96 cases and 50 controls for 116,204 SNPs and an intronic and common variant in the complement factor H gene (*CFH*) was found to be strongly associated (OR:7.4, 95%CI 2.9-19 [308]). Re-sequencing revealed a polymorphism in linkage disequilibrium with the risk allele representing a tyrosine-histidine change at amino acid 402. The *CFH* gene is located on chromosome 1 in a region that has been repeatedly linked to AMD in family based studies. The association of the *CFH* locus with AMD was rapidly and convincingly replicated in three independent studies [309-311], however the actual causative variant has not been identified.

Whole-genome studies are now feasible for PSP due to the large collections of pathologically confirmed samples that now exist worldwide. In such a study it would be interesting to see if the approach can identify the association of the *MAPT* locus to validate the method and indeed to identify novel associated loci.

8.6 Conclusions

The *MAPT* locus is an established genetic risk factor for PSP though the causal variant(s) and the functional basis of the association had not yet been established. The aim of the work in this thesis was to identify the functional basis of the *MAPT* association with PSP in order to provide the most promising prospect of understanding the pathogenesis of this disorder and the tauopathies in general.

This included delineating the outer limits of the extended H1 haplotype, high-resolution SNP association mapping of the *MAPT* gene, re-sequencing the *MAPT* gene to identify candidate variants and functional assays to investigate the effect specific variants had on *MAPT* expression. Secondary aims of the thesis included extending the association study of *MAPT* to the related disorder, CBD, and to investigate the molecular genetic basis of haploinsufficiency a new microdeletion deletion syndrome at 17q21.31 that encompasses the *MAPT* gene.

In the first instance, the PSP associated H1 haplotype was defined as a haplotype block of 1.8 Mb at 17q21.31 and defined the candidate region, or associated interval that was associated with PSP. The defined block provided the candidate region for a more detailed analysis of the underlying architecture of the H1 haplotype and to identify all of the genes within the block. Although the region contained other genes, *MAPT* was selected for high-resolution association mapping with PSP due the overwhelming pathological evidence that itself was responsible for the association.

For the high-resolution *MAPT* association study, HapMap project data was used to investigate the haplotype diversity and the LD across the gene. In addition, the haplotype diversity was examined in the Japanese population; to compare the genetic structure of the locus in two distinct racial groups. The Japanese population

had a very different genetic structure to that of the Caucasian group. This re-inforces the idea that different ethnic groups can have large differences in linkage disequilibrium and haplotype diversity and this should always be an important consideration when performing association studies in different populations; the genetic diversity of a gene or locus may indeed be quite different from one ethnic group to another.

Appreciating the genetic diversity of the *MAPT* region allowed the selection of htSNPs that could be genotyped in larger populations consisting of PSP cases and controls. This methodology also represents a modern, streamlined and economical protocol for candidate gene association studies and capitalising on the publicly available HapMap project data for future candidate gene association studies and will greatly speed up the identification of novel loci predisposing to common diseases.

Identifying the H1c haplotype as the major risk haplotype with PSP provided the basis of a targeted re-sequencing strategy to identify novel candidate loci in *MAPT* that could have the potential to be causal. Through this approach, candidate variants that resided within sequences that were conserved between human and mouse were identified and these were tested for effects on *MAPT* promoter driven expression. Variants that were associated with an increased expression level were found to be significantly over-represented in PSP, suggesting that an increased level of expression of *MAPT* formed the functional basis of the association with PSP. This data suggests that a mis-regulation of *MAPT* gene expression and downstream tau protein dynamics could lead to the development of PSP. This is in accordance with mutation data in FTDP-17T that demonstrates that tau dysfunction is associated with the development of the tauopathies. This however, is the first time it has been suggested that the actual level of *MAPT* gene expression could potentially be pathogenic.

The H1c haplotype was also identified as a risk factor for the related tauopathy CBD, though as this study was hypothesis generating, this finding should be independently replicated in a larger, more powerful series. However, these preliminary findings do suggest that the same *MAPT* haplotype could be associated

with both disorders and that expression changes in *MAPT* could also be predisposing factor for this tauopathy. The findings from these studies could have potential therapeutic implications for the tauopathies; intervening compounds that could reduce *MAPT* expression may be able to slow the progression of disease.

Analysing the genetic architecture of 17q21.31 revealed the presence of LCRs flanking the *MAPT* gene that are responsible, by a mechanism of NAHR, for a microdeletion that is associated with developmental delay and learning disabilities.

These studies of the *MAPT* locus also suggest that there is more to be found; identification of microdeletions that encompass the *MAPT* gene would suggest that duplications of the gene are also possible and it would be of great interest to explore any associated phenotype in light of the findings that increased *MAPT* expression is a risk factor for at least one tauopathy.

9 References

1. Weeks, D.E. and G.M. Lathrop, *Polygenic disease: methods for mapping complex disease traits.* Trends Genet, 1995. **11**(12): p. 513-9.
2. Elston, R.C., *The genetic dissection of multifactorial traits.* Clin Exp Allergy, 1995. **25 Suppl 2**: p. 103-6.
3. Burton, P.R., M.D. Tobin, and J.L. Hopper, *Key concepts in genetic epidemiology.* Lancet, 2005. **366**(9489): p. 941-51.
4. Morton, N.E., *Sequential tests for the detection of linkage.* Am J Hum Genet, 1955. **7**(3): p. 277-318.
5. Chotai, J., *On the lod score method in linkage analysis.* Ann Hum Genet, 1984. **48**(Pt 4): p. 359-78.
6. Hutton, M., *Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms.* Neurology, 2001. **56**(11 Suppl 4): p. S21-5.
7. Risch, N.J., *Searching for genetic determinants in the new millennium.* Nature, 2000. **405**(6788): p. 847-56.
8. Reich, D.E. and E.S. Lander, *On the allelic spectrum of human disease.* Trends Genet, 2001. **17**(9): p. 502-10.
9. Roses, A.D., et al., *Apolipoprotein E E4 allele and risk of dementia.* Jama, 1995. **273**(5): p. 374-5; author reply 375-6.
10. Strittmatter, W.J. and A.D. Roses, *Apolipoprotein E and Alzheimer disease.* Proc Natl Acad Sci U S A, 1995. **92**(11): p. 4725-7.
11. Altshuler, D., et al., *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.* Nat Genet, 2000. **26**(1): p. 76-80.
12. Davies, J.L., et al., *A genome-wide search for human type 1 diabetes susceptibility genes.* Nature, 1994. **371**(6493): p. 130-6.
13. Arya, R., et al., *Linkage of high-density lipoprotein-cholesterol concentrations to a locus on chromosome 9p in Mexican Americans.* Nat Genet, 2002. **30**(1): p. 102-5.

14. Myers, A., et al., *Susceptibility locus for Alzheimer's disease on chromosome 10.* Science, 2000. **290**(5500): p. 2304-5.

15. Altmuller, J., et al., *Genomewide scans of complex human diseases: true linkage is hard to find.* Am J Hum Genet, 2001. **69**(5): p. 936-50.

16. Devlin, B. and N. Risch, *A comparison of linkage disequilibrium measures for fine-scale mapping.* Genomics, 1995. **29**(2): p. 311-22.

17. Lewontin, R.C., *The Interaction of Selection and Linkage. Ii. Optimum Models.* Genetics, 1964. **50**: p. 757-82.

18. Cordell, H.J. and D.G. Clayton, *Genetic association studies.* Lancet, 2005. **366**(9491): p. 1121-31.

19. Hirose, S., et al., *Genetics of idiopathic epilepsies.* Epilepsia, 2005. **46 Suppl 1**: p. 38-43.

20. Hattersley, A.T. and M.I. McCarthy, *What makes a good genetic association study?* Lancet, 2005. **366**(9493): p. 1315-23.

21. Ioannidis, J.P., T.A. Trikalinos, and M.J. Khoury, *Implications of Small Effect Sizes of Individual Genetic Variants on the Design and Interpretation of Genetic Association Studies of Complex Diseases.* Am J Epidemiol, 2006.

22. Cardon, L.R. and L.J. Palmer, *Population stratification and spurious allelic association.* Lancet, 2003. **361**(9357): p. 598-604.

23. Pritchard, J.K. and P. Donnelly, *Case-control studies of association in structured or admixed populations.* Theor Popul Biol, 2001. **60**(3): p. 227-37.

24. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.

25. Corder, E.H., et al., *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.* Science, 1993. **261**(5123): p. 921-3.

26. Roses, A.D., *Apolipoprotein E genotyping in the differential diagnosis, not prediction, of Alzheimer's disease.* Ann Neurol, 1995. **38**(1): p. 6-14.

27. Roses, A.D., *Apolipoprotein E and Alzheimer's disease. A rapidly expanding field with medical and epidemiological consequences.* Ann N Y Acad Sci, 1996. **802**: p. 50-7.

28. Morrow, J.A., et al., *Differences in stability among the human apolipoprotein E isoforms determined by the amino-terminal domain.* Biochemistry, 2000. **39**(38): p. 11657-66.

29. Weisgraber, K.H., S.C. Rall, Jr., and R.W. Mahley, *Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms.* J Biol Chem, 1981. **256**(17): p. 9077-83.

30. Ashford, J.W., *APOE genotype effects on Alzheimer's disease onset and epidemiology.* J Mol Neurosci, 2004. **23**(3): p. 157-65.

31. Levi, O., et al., *ApoE4 impairs hippocampal plasticity isoform-specifically and blocks the environmental stimulation of synaptogenesis and memory.* Neurobiol Dis, 2003. **13**(3): p. 273-82.

32. Harris, F.M., et al., *Carboxyl-terminal-truncated apolipoprotein E4 causes Alzheimer's disease-like neurodegeneration and behavioral deficits in transgenic mice.* Proc Natl Acad Sci U S A, 2003. **100**(19): p. 10966-71.

33. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome.* Nat Rev Genet, 2006. **7**(2): p. 85-97.

34. Jacobs, P.A., et al., *Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding.* J Med Genet, 1992. **29**(2): p. 103-8.

35. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.

36. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome.* Science, 2004. **305**(5683): p. 525-8.

37. Feuk, L., et al., *Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies.* PLoS Genet, 2005. **1**(4): p. e56.

38. Tuzun, E., et al., *Fine-scale structural variation of the human genome.* Nat Genet, 2005. **37**(7): p. 727-32.

39. Feuk, L., et al., *Structural variants: changing the landscape of chromosomes and design of disease studies.* Hum Mol Genet, 2006. **15 Spec No 1**: p. R57-66.

40. Pinkel, D., et al., *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.* Nat Genet, 1998. **20**(2): p. 207-11.

41. Solinas-Toldo, S., et al., *Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances.* Genes Chromosomes Cancer, 1997. **20**(4): p. 399-407.

42. Lucito, R., et al., *Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation.* Genome Res, 2003. **13**(10): p. 2291-305.

43. Kennedy, G.C., et al., *Large-scale genotyping of complex DNA.* Nat Biotechnol, 2003. **21**(10): p. 1233-7.

44. Fredman, D., et al., *Complex SNP-related sequence variation in segmental genome duplications.* Nat Genet, 2004. **36**(8): p. 861-6.

45. Sharp, A.J., et al., *Segmental duplications and copy-number variation in the human genome.* Am J Hum Genet, 2005. **77**(1): p. 78-88.

46. Shaw, C.J. and J.R. Lupski, *Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R57-64.

47. Stankiewicz, P. and J.R. Lupski, *Genome architecture, rearrangements and genomic disorders.* Trends Genet, 2002. **18**(2): p. 74-82.

48. Stankiewicz, P., et al., *Serial segmental duplications during primate evolution result in complex human genome architecture.* Genome Res, 2004. **14**(11): p. 2209-20.

49. Inoue, K. and J.R. Lupski, *Molecular mechanisms for genomic disorders.* Annu Rev Genomics Hum Genet, 2002. **3**: p. 199-242.

50. Ewart, A.K., et al., *Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome.* Nat Genet, 1993. **5**(1): p. 11-6.

51. Osborne, L.R., et al., *A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome.* Nat Genet, 2001. **29**(3): p. 321-5.

52. Giglio, S., et al., *Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation.* Am J Hum Genet, 2002. **71**(2): p. 276-85.

53. Bergemann, A.D., F. Cole, and K. Hirschhorn, *The etiology of Wolf-Hirschhorn syndrome.* Trends Genet, 2005. **21**(3): p. 188-95.

54. Weingarten, M.D., et al., *A protein factor essential for microtubule assembly.* Proc Natl Acad Sci U S A, 1975. **72**(5): p. 1858-62.

55. Binder, L.I., A. Frankfurter, and L.I. Rebhun, *The distribution of tau in the mammalian central nervous system.* J Cell Biol, 1985. **101**(4): p. 1371-8.

56. Fujii, T., *[Structure and function of mammalian brain microtubule-associated proteins].* Yakugaku Zasshi, 1994. **114**(7): p. 435-47.

57. Gu, Y., F. Oyama, and Y. Ihara, *Tau is widely expressed in rat tissues.* J Neurochem, 1996. **67**(3): p. 1235-44.

58. LoPresti, P., et al., *Functional implications for the microtubule-associated protein tau: localization in oligodendrocytes.* Proc Natl Acad Sci U S A, 1995. **92**(22): p. 10369-73.

59. Andreadis, A., W.M. Brown, and K.S. Kosik, *Structure and novel exons of the human tau gene.* Biochemistry, 1992. **31**(43): p. 10626-33.

60. Goedert, M., et al., *Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease.* Neuron, 1989. **3**(4): p. 519-26.

61. Brandt, R., J. Leger, and G. Lee, *Interaction of tau with the neural plasma membrane mediated by tau's amino-terminal projection domain.* J Cell Biol, 1995. **131**(5): p. 1327-40.

62. Goedert, M., et al., *Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau.* Proc Natl Acad Sci U S A, 1988. **85**(11): p. 4051-5.

63. Gustke, N., et al., *Domains of tau protein and interactions with microtubules.* Biochemistry, 1994. **33**(32): p. 9511-22.

64. Trinczek, B., et al., *Domains of tau protein, differential phosphorylation, and dynamic instability of microtubules.* Mol Biol Cell, 1995. **6**(12): p. 1887-902.

65. Couchie, D., C. Charriere-Bertrand, and J. Nunez, *Expression of the mRNA for tau proteins during brain development and in cultured neurons and astroglial cells.* J Neurochem, 1988. **50**(6): p. 1894-9.

66. Couchie, D., et al., *Primary structure of high molecular weight tau present in the peripheral nervous system.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4378-81.

67. Chen, W.T., W.K. Liu, and S.H. Yen, *Expression of tau exon 8 in different species.* Neurosci Lett, 1994. **172**(1-2): p. 167-70.

68. Nelson, P.T., et al., *Molecular evolution of tau protein: implications for Alzheimer's disease.* J Neurochem, 1996. **67**(4): p. 1622-32.

69. Kanemaru, K., et al., *Fetal-type phosphorylation of the tau in paired helical filaments.* J Neurochem, 1992. **58**(5): p. 1667-75.

70. Kopke, E., et al., *Microtubule-associated protein tau. Abnormal phosphorylation of a non-paired helical filament pool in Alzheimer disease.* J Biol Chem, 1993. **268**(32): p. 24374-84.

71. Alonso, A.C., et al., *Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease.* Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5562-6.

72. Lindwall, G. and R.D. Cole, *Phosphorylation affects the ability of tau protein to promote microtubule assembly.* J Biol Chem, 1984. **259**(8): p. 5301-5.

73. Spillantini, M.G. and M. Goedert, *Tau protein pathology in neurodegenerative diseases.* Trends Neurosci, 1998. **21**(10): p. 428-33.

74. Goedert, M., S.S. Sisodia, and D.L. Price, *Neurofibrillary tangles and beta-amyloid deposits in Alzheimer's disease.* Curr Opin Neurobiol, 1991. **1**(3): p. 441-7.

75. Ikeda, S., et al., *Variability of beta-amyloid protein deposited lesions in Down's syndrome brains.* Tohoku J Exp Med, 1994. **174**(3): p. 189-98.

76. Yoshimura, N., et al., *Down's syndrome in middle age. Topographical distribution and immunoreactivity of brain lesions in an autopsied patient.* Acta Pathol Jpn, 1990. **40**(10): p. 735-43.

77. Togo, T., et al., *Argyrophilic grain disease is a sporadic 4-repeat tauopathy.* J Neuropathol Exp Neurol, 2002. **61**(6): p. 547-56.

78. Pearce, J.M., *Pick's disease.* J Neurol Neurosurg Psychiatry, 2003. **74**(2): p. 169.

79. Burn, D.J. and A.J. Lees, *Progressive supranuclear palsy: where are we now?* Lancet Neurol, 2002. **1**(6): p. 359-69.

80. Rademakers, R., M. Cruts, and C. van Broeckhoven, *The role of tau (MAPT) in frontotemporal dementia and related tauopathies.* Hum Mutat, 2004. **24**(4): p. 277-95.

81. Nirmalananthan, N. and L. Greensmith, *Amyotrophic lateral sclerosis: recent advances and future therapies.* Curr Opin Neurol, 2005. **18**(6): p. 712-9.

82. Vanier, M.T. and G. Millat, *Niemann-Pick disease type C.* Clin Genet, 2003. **64**(4): p. 269-81.

83. Greenberg, S.G. and P. Davies, *A preparation of Alzheimer paired helical filaments that displays distinct tau proteins by polyacrylamide gel electrophoresis.* Proc Natl Acad Sci U S A, 1990. **87**(15): p. 5827-31.

84. Lee, V.M., M. Goedert, and J.Q. Trojanowski, *Neurodegenerative tauopathies.* Annu Rev Neurosci, 2001. **24**: p. 1121-59.

85. Neary, D., J. Snowden, and D. Mann, *Frontotemporal dementia.* Lancet Neurol, 2005. **4**(11): p. 771-80.

86. Knopman, D.S., et al., *Dementia lacking distinctive histologic features: a common non-Alzheimer degenerative dementia.* Neurology, 1990. **40**(2): p. 251-6.

87. Neary, D., J.S. Snowden, and D.M. Mann, *Classification and description of frontotemporal dementias.* Ann N Y Acad Sci, 2000. **920**: p. 46-51.

88. Rascovsky, K., et al., *Rate of progression differs in frontotemporal dementia and Alzheimer disease.* Neurology, 2005. **65**(3): p. 397-403.

89. Piguet, O., et al., *Similar early clinical presentations in familial and non-familial frontotemporal dementia.* J Neurol Neurosurg Psychiatry, 2004. **75**(12): p. 1743-5.

90. Rosso, S.M., et al., *Frontotemporal dementia in The Netherlands: patient characteristics and prevalence estimates from a population-based study.* Brain, 2003. **126**(Pt 9): p. 2016-22.

91. Hutton, M., et al., *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature, 1998. **393**(6686): p. 702-5.

92. Rademakers, R., et al., *Tau negative frontal lobe dementia at 17q21: significant finemapping of the candidate region to a 4.8 cM interval.* Mol Psychiatry, 2002. **7**(10): p. 1064-74.

93. Wilhelmsen, K.C., et al., *Localization of disinhibition-dementia-parkinsonism-amyotrophy complex to 17q21-22.* Am J Hum Genet, 1994. **55**(6): p. 1159-65.

94. Hosler, B.A., et al., *Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22.* Jama, 2000. **284**(13): p. 1664-9.

95. Brown, J., et al., *Frontotemporal dementia linked to chromosome 3.* Dement Geriatr Cogn Disord, 2004. **17**(4): p. 274-6.

96. Skibinski, G., et al., *Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia.* Nat Genet, 2005. **37**(8): p. 806-8.

97. Foster, N.L., et al., *Frontotemporal dementia and parkinsonism linked to chromosome 17: a consensus conference. Conference Participants.* Ann Neurol, 1997. **41**(6): p. 706-15.

98. D'Souza, I., et al., *Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5598-603.

99. Poorkaj, P., et al., *Tau is a candidate gene for chromosome 17 frontotemporal dementia.* Ann Neurol, 1998. **43**(6): p. 815-25.

100. Goedert, M., *Tau gene mutations and their effects.* Mov Disord, 2005. **20 Suppl 12**: p. S45-52.

101. Hasegawa, M., M.J. Smith, and M. Goedert, *Tau proteins with FTDP-17 mutations have a reduced ability to promote microtubule assembly.* FEBS Lett, 1998. **437**(3): p. 207-10.

102. Hong, M., et al., *Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17.* Science, 1998. **282**(5395): p. 1914-7.

103. Dayanandan, R., et al., *Mutations in tau reduce its microtubule binding properties in intact cells and affect its phosphorylation.* FEBS Lett, 1999. **446**(2-3): p. 228-32.

104. Lu, M. and K.S. Kosik, *Competition for microtubule-binding with dual expression of tau missense and splice isoforms.* Mol Biol Cell, 2001. **12**(1): p. 171-84.

105. Matsumura, N., T. Yamazaki, and Y. Ihara, *Stable expression in Chinese hamster ovary cells of mutated tau genes causing frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17).* Am J Pathol, 1999. **154**(6): p. 1649-56.

106. Nagiec, E.W., K.E. Sampson, and I. Abraham, *Mutated tau binds less avidly to microtubules than wildtype tau in living cells.* J Neurosci Res, 2001. **63**(3): p. 268-75.

107. Barghorn, S., et al., *Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias.* Biochemistry, 2000. **39**(38): p. 11714-21.

108. Gamblin, T.C., et al., *In vitro polymerization of tau protein monitored by laser light scattering: method and application to the study of FTDP-17 mutants.* Biochemistry, 2000. **39**(20): p. 6136-44.

109. Goedert, M., R. Jakes, and R.A. Crowther, *Effects of frontotemporal dementia FTDP-17 mutations on heparin-induced assembly of tau filaments.* FEBS Lett, 1999. **450**(3): p. 306-11.

110. Nacharaju, P., et al., *Accelerated filament formation from tau protein with specific FTDP-17 missense mutations.* FEBS Lett, 1999. **447**(2-3): p. 195-9.

111. von Bergen, M., et al., *Tau aggregation is driven by a transition from random coil to beta sheet structure.* Biochim Biophys Acta, 2005. **1739**(2-3): p. 158-66. Epub 2004 Nov 12.

112. DeTure, M., et al., *Tau assembly in inducible transfectants expressing wild-type or FTDP-17 tau.* Am J Pathol, 2002. **161**(5): p. 1711-22.

113. Goedert, M., et al., *Reduced binding of protein phosphatase 2A to tau protein with frontotemporal dementia and parkinsonism linked to chromosome 17 mutations.* J Neurochem, 2000. **75**(5): p. 2155-62.

114. Grover, A., et al., *Effects on splicing and protein function of three mutations in codon N296 of tau in vitro.* Neurosci Lett, 2002. **323**(1): p. 33-6.

115. Hasegawa, M., et al., *FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10.* FEBS Lett, 1999. **443**(2): p. 93-6.

116. Spillantini, M.G., et al., *Mutation in the tau gene in familial multiple system tauopathy with presenile dementia.* Proc Natl Acad Sci U S A, 1998. **95**(13): p. 7737-41.

117. Spillantini, M.G., J.C. Van Swieten, and M. Goedert, *Tau gene mutations in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17).* Neurogenetics, 2000. **2**(4): p. 193-205.

118. Yoshida, H., R.A. Crowther, and M. Goedert, *Functional effects of tau gene mutations deltaN296 and N296H.* J Neurochem, 2002. **80**(3): p. 548-51.

119. Grover, A., et al., *5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10.* J Biol Chem, 1999. **274**(21): p. 15134-43.

120. Malkani, R., et al., *A MAPT mutation in a regulatory element upstream of exon 10 causes frontotemporal dementia.* Neurobiol Dis, 2006. **22**(2): p. 401-3.

121. Yasuda, M., et al., *A novel mutation at position +12 in the intron following exon 10 of the tau gene in familial frontotemporal dementia (FTD-Kumamoto).* Ann Neurol, 2000. **47**(4): p. 422-9.

122. Stanford, P.M., et al., *Mutations in the tau gene that cause an increase in three repeat tau and frontotemporal dementia.* Brain, 2003. **126**(Pt 4): p. 814-26.

123. Stanford, P.M., et al., *Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene: expansion of the disease phenotype caused by tau gene mutations.* Brain, 2000. **123**(Pt 5): p. 880-93.

124. Iijima, M., et al., *A distinct familial presenile dementia with a novel missense mutation in the tau gene.* Neuroreport, 1999. **10**(3): p. 497-501.

125. Rizzu, P., et al., *High prevalence of mutations in the microtubule-associated protein tau in a population study of frontotemporal dementia in the Netherlands.* Am J Hum Genet, 1999. **64**(2): p. 414-21.

126. Clark, L.N., et al., *Pathogenic implications of mutations in the tau gene in pallido-ponto-nigral degeneration and related neurodegenerative disorders*

linked to chromosome 17. Proc Natl Acad Sci U S A, 1998. **95**(22): p. 13103-7.

127. Pastor, P., et al., *Familial atypical progressive supranuclear palsy associated with homozigosity for the delN296 mutation in the tau gene.* Ann Neurol, 2001. **49**(2): p. 263-7.

128. Spillantini, M.G., et al., *A novel tau mutation (N296N) in familial dementia with swollen achromatic neurons and corticobasal inclusion bodies.* Ann Neurol, 2000. **48**(6): p. 939-43.

129. Hayashi, S., et al., *Late-onset frontotemporal dementia with a novel exon 1 (Arg5His) tau gene mutation.* Ann Neurol, 2002. **51**(4): p. 525-30.

130. Poorkaj, P., et al., *An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype.* Ann Neurol, 2002. **52**(4): p. 511-6.

131. Rizzini, C., et al., *Tau gene mutation K257T causes a tauopathy similar to Pick's disease.* J Neuropathol Exp Neurol, 2000. **59**(11): p. 990-1001.

132. Grover, A., et al., *A novel tau mutation in exon 9 (1260V) causes a four-repeat tauopathy.* Exp Neurol, 2003. **184**(1): p. 131-40.

133. Hogg, M., et al., *The L266V tau mutation is associated with frontotemporal dementia and Pick-like 3R and 4R tauopathy.* Acta Neuropathol (Berl), 2003. **106**(4): p. 323-36.

134. Yasuda, M., et al., *Phenotypic heterogeneity within a new family with the MAPT p301s mutation.* Ann Neurol, 2005. **58**(6): p. 920-8.

135. van Herpen, E., et al., *Variable phenotypic expression and extensive tau pathology in two families with the novel tau mutation L315R.* Ann Neurol, 2003. **54**(5): p. 573-81.

136. Pickering-Brown, S.M., et al., *Frontotemporal dementia with Pick-type histology associated with Q336R mutation in the tau gene.* Brain, 2004. **127**(Pt 6): p. 1415-26.

137. Spillantini, M.G., et al., *Tau pathology in two Dutch families with mutations in the microtubule-binding region of tau.* Am J Pathol, 1998. **153**(5): p. 1359-63.

138. Lippa, C.F., et al., *Frontotemporal dementia with novel tau pathology and a Glu342Val tau mutation.* Ann Neurol, 2000. **48**(6): p. 850-8.

139. Nicholl, D.J., et al., *An English kindred with a novel recessive tauopathy and respiratory failure.* Ann Neurol, 2003. **54**(5): p. 682-6.

140. Neumann, M., et al., *Pick's disease associated with the novel Tau gene mutation K369I.* Ann Neurol, 2001. **50**(4): p. 503-13.

141. Ghetti, B., et al., *Progress in hereditary tauopathies: a mutation in the Tau gene (G389R) causes a Pick disease-like syndrome.* Ann N Y Acad Sci, 2000. **920**: p. 52-62.

142. Goedert, M., R.A. Crowther, and M.G. Spillantini, *Tau mutations cause frontotemporal dementias.* Neuron, 1998. **21**(5): p. 955-8.

143. Mackenzie, I.R., et al., *A family with tau-negative frontotemporal dementia and neuronal intranuclear inclusions linked to chromosome 17.* Brain, 2006. **129**(Pt 4): p. 853-67.

144. Kertesz, A., et al., *Familial frontotemporal dementia with ubiquitin-positive, tau-negative inclusions.* Neurology, 2000. **54**(4): p. 818-27.

145. Cruts, M., et al., *Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region.* Hum. Mol. Genet., 2005. **14**(13): p. 1753-1762.

146. Baker, M., et al., *Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17.* Nature, 2006.

147. Cruts, M., et al., *Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21.* Nature, 2006.

148. He, Z. and A. Bateman, *Progranulin (granulin-epithelin precursor, PC-cell-derived growth factor, acrogranin) mediates tissue repair and tumorigenesis.* J Mol Med, 2003. **81**(10): p. 600-12.

149. Vance, C., et al., *Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3.* Brain, 2006. **129**(Pt 4): p. 868-76.

150. Gydesen, S., et al., *Chromosome 3 linked frontotemporal dementia (FTD-3).* Neurology, 2002. **59**(10): p. 1585-94.

151. Yancopoulou, D., et al., *Tau protein in frontotemporal dementia linked to chromosome 3 (FTD-3).* J Neuropathol Exp Neurol, 2003. **62**(8): p. 878-82.

152. Cannon, A., et al., *CHMP2B mutations are not a common cause of frontotemporal lobar degeneration.* Neurosci Lett, 2006. **398**(1-2): p. 83-4.

153. Rizzu, P., et al., *CHMP2B mutations are not a cause of dementia in Dutch patients with familial and sporadic frontotemporal dementia.* Am J Med Genet B Neuropsychiatr Genet, 2006.

154. Richardson, J.C., J. Steele, and J. Olszewski, *Supranuclear Ophthalmoplegia, Pseudobulbar Palsy, Nuchal Dystonia and Dementia. a Clinical Report on Eight Cases of "Heterogenous System Degeneration".* Trans Am Neurol Assoc, 1963. **88**: p. 25-9.

155. Steele, J.C., J.C. Richardson, and J. Olszewski, *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**: p. 333-59.

156. Litvan, I., et al., *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): p. 1-9.

157. Golbe, L.I., et al., *Prevalence and natural history of progressive supranuclear palsy.* Neurology, 1988. **38**(7): p. 1031-4.

158. Kawashima, M., et al., *Prevalence of progressive supranuclear palsy in Yonago, Japan.* Mov Disord, 2004. **19**(10): p. 1239-40.

159. Nath, U., et al., *The prevalence of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome) in the UK.* Brain, 2001. **124**(Pt 7): p. 1438-49.

160. Hauw, J.J., et al., *Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy).* Neurology, 1994. **44**(11): p. 2015-9.

161. Flament, S., et al., *Abnormal Tau proteins in progressive supranuclear palsy. Similarities and differences with the neurofibrillary degeneration of the Alzheimer type.* Acta Neuropathol, 1991. **81**(6): p. 591-6.

162. Golbe, L.I., *Progressive supranuclear palsy in the molecular age.* Lancet, 2000. **356**(9233): p. 870-1.

163. de Silva, R., et al., *Pathological inclusion bodies in tauopathies contain distinct complements of tau with three or four microtubule-binding repeat domains as demonstrated by new specific monoclonal antibodies.* Neuropathol Appl Neurobiol, 2003. **29**(3): p. 288-302.

164. Morris, H.R., et al., *Tau exon 10 +16 mutation FTDP-17 presenting clinically as sporadic young onset PSP.* Neurology, 2003. **61**(1): p. 102-4.

165. Stanford, P.M., et al., *Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene: Expansion of the disease phenotype caused by tau gene mutations.* Brain, 2000. **123**(5): p. 880-893.

166. Wszolek, Z.K., et al., *Progressive supranuclear palsy as a disease phenotype caused by the S305S tau gene mutation.* Brain, 2001. **124**(Pt 8): p. 1666-70.

167. Conrad, C., et al., *Genetic evidence for the involvement of tau in progressive supranuclear palsy.* Ann Neurol, 1997. **41**(2): p. 277-81.

168. Bennett, P., et al., *Direct genetic evidence for involvement of tau in progressive supranuclear palsy. European Study Group on Atypical Parkinsonism Consortium.* Neurology, 1998. **51**(4): p. 982-5.

169. Higgins, J.J., et al., *An extended 5'-tau susceptibility haplotype in progressive supranuclear palsy.* Neurology, 2000. **55**(9): p. 1364-7.

170. Morris, H.R., et al., *The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases.* J Neurol Neurosurg Psychiatry, 1999. **66**(5): p. 665-667.

171. Oliva, R., et al., *Significant changes in the tau A0 and A3 alleles in progressive supranuclear palsy and improved genotyping by silver detection.* Arch Neurol, 1998. **55**(8): p. 1122-4.

172. Baker, M., et al., *Association of an extended haplotype in the tau gene with progressive supranuclear palsy.* Hum. Mol. Genet., 1999. **8**(4): p. 711-715.

173. Di Maria, E., et al., *Corticobasal degeneration shares a common genetic background with progressive supranuclear palsy.* Ann Neurol, 2000. **47**(3): p. 374-7.

174. Houlden, H., et al., *Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype.* Neurology, 2001. **56**(12): p. 1702-1706.

175. Healy, D.G., et al., *Tau gene and Parkinson's disease: a case-control study and meta-analysis.* J Neurol Neurosurg Psychiatry, 2004. **75**(7): p. 962-5.

176. Bennett, M.C., *The role of alpha-synuclein in neurodegenerative diseases.* Pharmacol Ther, 2005. **105**(3): p. 311-31.

177. Conrad, C., et al., *A polymorphic gene nested within an intron of the tau gene: Implications for Alzheimer's disease.* PNAS, 2002. **99**(11): p. 7751-7756.

178. Streffer, J.R., et al., *Saitohin gene is not associated with Alzheimer's disease.* J Neurol Neurosurg Psychiatry, 2003. **74**(3): p. 362-363.

179. Lynnette Cook, C.E.B., Douglas Easton, John Grimley Evans, John Xuereb, Nigel J. Cairns, David C. Rubinsztein,, *No evidence for an association between <I>Saitohin</I> Q7R polymorphism and Alzheimer's disease.* Annals of Neurology, 2002. **52**(5): p. 690-691.

180. Clark, L.N., et al., *The Saitohin 'Q7R' polymorphism and tau haplotype in multi-ethnic Alzheimer disease and Parkinson's disease cohorts.* Neurosci Lett, 2003. **347**(1): p. 17-20.

181. Verpillat, P., et al., *Is the saitohin gene involved in neurodegenerative diseases?* Ann Neurol, 2002. **52**(6): p. 829-32.

182. Johansson, A., et al., *TAU haplotype and the Saitohin Q7R gene polymorphism do not influence CSF Tau in Alzheimer's disease and are not associated with frontotemporal dementia or Parkinson's disease.* Neurodegener Dis, 2005. **2**(1): p. 28-35.

183. Oliveira, S.A., et al., *The Q7R Saitohin gene polymorphism is not associated with Alzheimer disease.* Neurosci Lett, 2003. **347**(3): p. 143-6.

184. Cook, L., et al., *No evidence for an association between Saitohin Q7R polymorphism and Alzheimer's disease.* Ann Neurol, 2002. **52**(5): p. 690-1.

185. de Silva, R., et al., *Strong association of the Saitohin gene Q7 variant with progressive supranuclear palsy.* Neurology, 2003. **61**(3): p. 407-409.

186. Ezquerra, M., et al., *Sequence analysis of tau 3'untranslated region and saitohin gene in sporadic progressive supranuclear palsy.* J Neurol Neurosurg Psychiatry, 2004. **75**(1): p. 155-157.

187. Gao, L., et al., *Saitohin, which is nested in the tau locus and confers allele-specific susceptibility to several neurodegenerative diseases, interacts with peroxiredoxin 6.* J Biol Chem, 2005. **280**(47): p. 39268-72.

188. Krapfenbauer, K., et al., *Aberrant expression of peroxiredoxin subtypes in neurodegenerative disorders.* Brain Res, 2003. **967**(1-2): p. 152-60.

189. Krapfenbauer, K., et al., *Expression patterns of antioxidant proteins in brains of patients with sporadic Creutzfeldt-Jacob disease.* Electrophoresis, 2002. **23**(15): p. 2541-7.

190. Power, J.H., et al., *Nonselenium glutathione peroxidase in human brain : elevated levels in Parkinson's disease and dementia with lewy bodies.* Am J Pathol, 2002. **161**(3): p. 885-94.

191. de Silva, R., et al., *Strong association of a novel Tau promoter haplotype in progressive supranuclear palsy.* Neurosci Lett, 2001. **311**(3): p. 145-8.

192. Ezquerra, M., et al., *Identification of a novel polymorphism in the promoter region of the tau gene highly associated to progressive supranuclear palsy in humans.* Neuroscience Letters, 1999. **275**(3): p. 183-186.

193. Pastor, P., et al., *Further extension of the H1 haplotype associated with progressive supranuclear palsy.* Mov Disord, 2002. **17**(3): p. 550-6.

194. Liu, W.K., et al., *Relationship of the extended tau haplotype to tau biochemistry and neuropathology in progressive supranuclear palsy.* Ann Neurol, 2001. **50**(4): p. 494-502.

195. Conrad, C., et al., *Differences in a dinucleotide repeat polymorphism in the tau gene between Caucasian and Japanese populations: implication for progressive supranuclear palsy.* Neurosci Lett, 1998. **250**(2): p. 135-7.

196. Higgins, J.J., R.L. Adler, and J.M. Loveless, *Mutational analysis of the tau gene in progressive supranuclear palsy.* Neurology, 1999. **53**(7): p. 1421-.

197. Morris, H.R., et al., *Multiple system atrophy/progressive supranuclear palsy: alpha-Synuclein, synphilin, tau, and APOE.* Neurology, 2000. **55**(12): p. 1918-20.

198. Ezquerra, M., et al., *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**(3): p. 183-6.

199. Higgins, J.J., et al., *An extended 5'-tau susceptibility haplotype in progressive supranuclear palsy.* Neurology, 2000. **55**(9): p. 1364-1367.

200. de Yebenes, J.G., et al., *Familial progressive supranuclear palsy. Description of a pedigree and review of the literature.* Brain, 1995. **118**(Pt 5): p. 1095-103.

201. Rojo, A., et al., *Clinical genetics of familial progressive supranuclear palsy.* Brain, 1999. **122**(Pt 7): p. 1233-45.

202. Raquel Ros, P.G.G., Michio Hirano, Yen F. Tai, Israel Ampuero, Lídice Vidal, Ana Rojo, Aurora Fontan, Ana Vazquez, Samira Fanjul, Jaime Hernandez, Susana Cantarero, Janet Hoenicka, Alison Jones, R. Laila Ahsan, Nicola Pavese, Paola Piccini, David J. Brooks, Jordi Perez-Tur, Torbjorn Nyggard, Justo G. de Yébenes,, *Genetic linkage of autosomal dominant progressive supranuclear palsy to 1q31.1.* Annals of Neurology, 2005. **57**(5): p. 634-641.

203. Mena, M.A., et al., *Effects of dibutyryl cyclic AMP and retinoic acid on the differentiation of dopamine neurons: prevention of cell death by dibutyryl cyclic AMP.* J Neurochem, 1995. **65**(6): p. 2612-20.

204. Gibb, W.R., P.J. Luthert, and C.D. Marsden, *Clinical and pathological features of corticobasal degeneration.* Adv Neurol, 1990. **53**: p. 51-4.

205. Dickson, D.W., *Neuropathologic differentiation of progressive supranuclear palsy and corticobasal degeneration.* J Neurol, 1999. **246 Suppl 2**: p. II6-15.

206. Komori, T., *Tau-positive glial inclusions in progressive supranuclear palsy, corticobasal degeneration and Pick's disease.* Brain Pathol, 1999. **9**(4): p. 663-79.

207. Mayor, C., et al., *VISTA : visualizing global DNA sequence alignments of arbitrary length.* Bioinformatics, 2000. **16**(11): p. 1046-7.

208. Pau Pastor, E.P., Cristóbal Carnero, Rosario Vela, Teresa García, Guillem Amer, Eduardo Tolosa, Rafael Oliva,, *Familial atypical progressive supranuclear palsy associated with homozigosity for the delN296 mutation in the tau gene.* Annals of Neurology, 2001. **49**(2): p. 263-267.

209. Morris, H.R., et al., *The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases.* J Neurol Neurosurg Psychiatry, 1999. **66**(5): p. 665-7.

210. Litvan, I., et al., *Validity and reliability of the preliminary NINDS neuropathologic criteria for progressive supranuclear palsy and related disorders.* J Neuropathol Exp Neurol, 1996. **55**(1): p. 97-105.

211. Goldstein, D.B., et al., *Genome scans and candidate gene approaches in the study of common diseases and variable drug responses.* Trends Genet, 2003. **19**(11): p. 615-22.

212. Campdelacreu, J., et al., *No evidence of CRHR1 gene involvement in progressive supranuclear palsy.* Neurosci Lett, 2006.

213. Evans, W., et al., *The tau H2 haplotype is almost exclusively Caucasian in origin.* Neuroscience Letters, 2004. **369**(3): p. 183-185.

214. Stefansson, H., et al., *A common inversion under selection in Europeans.* Nat Genet, 2005. **37**(2): p. 129-37. Epub 2005 Jan 16.

215. Fung, H.C., et al., *The architecture of the tau haplotype block in different ethnicities.* Neuroscience Letters, 2005. **377**(2): p. 81-84.

216. Hardy, J., et al., *Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens.* Biochem Soc Trans, 2005. **33**(Pt 4): p. 582-5.

217. Pittman, A.M., et al., *The structure of the tau haplotype in controls and in progressive supranuclear palsy.* Hum. Mol. Genet., 2004. **13**(12): p. 1267-1274.

218. Skipper, L., et al., *Linkage disequilibrium and association of MAPT H1 in Parkinson disease.* Am J Hum Genet, 2004. **75**(4): p. 669-77.

219. Williams, D.R., et al., *Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism.* Brain, 2005. **128**(6): p. 1247-1258.

220. Weale, M.E., et al., *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**(3): p. 551-65.

221. Qin, Z.S., T. Niu, and J.S. Liu, *Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms.* Am J Hum Genet, 2002. **71**(5): p. 1242-7.

222. Hixson, J.E. and D.T. Vernier, *Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI.* J Lipid Res, 1990. **31**(3): p. 545-8.

223. Morris, H.R., et al., *Pathological, clinical and genetic heterogeneity in progressive supranuclear palsy.* Brain, 2002. **125**(Pt 5): p. 969-75.

224. Poorkaj, P., et al., *A genomic sequence analysis of the mouse and human microtubule-associated protein tau.* Mamm Genome, 2001. **12**(9): p. 700-12.

225. Andreadis, A., et al., *A tau promoter region without neuronal specificity.* J Neurochem, 1996. **66**(6): p. 2257-63.

226. Heicklen-Klein, A. and I. Ginzburg, *Tau promoter confers neuronal specificity and binds Sp1 and AP-2.* J Neurochem, 2000. **75**(4): p. 1408-18.

227. Kwok, J.B., et al., *Tau haplotypes regulate transcription and are associated with Parkinson's disease.* Ann Neurol, 2004. **55**(3): p. 329-34.

228. Cruts, M., et al., *Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region.* Hum Mol Genet, 2005. **14**(13): p. 1753-62.

229. Shaw-Smith, C., et al., *Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features.* J Med Genet, 2004. **41**(4): p. 241-8.

230. Varela, M.C., et al., *A 17q21.31 microdeletion encompassing the MAPT gene in a mentally impaired patient.* Cytogenet Genome Res, 2006. **114**(1): p. 89-92.

231. Noe, L. and G. Kucherov, *YASS: enhancing the sensitivity of DNA similarity search.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W540-3.

232. Lupski, J.R. and P. Stankiewicz, *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.* PLoS Genet, 2005. **1**(6): p. e49.

233. Lupski, J.R., *Genome structural variation and sporadic disease traits.* Nat Genet, 2006. **38**(9): p. 974-6.

234. Stankiewicz, P., et al., *Genomic disorders: genome architecture results in susceptibility to DNA rearrangements causing common human traits.* Cold Spring Harb Symp Quant Biol, 2003. **68**: p. 445-54.

235. Stankiewicz, P. and J.R. Lupski, *Molecular-evolutionary mechanisms for genomic disorders.* Curr Opin Genet Dev, 2002. **12**(3): p. 312-9.

236. Bishop, G.A. and J.S. King, *Corticotropin releasing factor in the embryonic mouse cerebellum.* Exp Neurol, 1999. **160**(2): p. 489-99.

237. Mirra, S.S., et al., *Tau pathology in a family with dementia and a P301L mutation in tau.* J Neuropathol Exp Neurol, 1999. **58**(4): p. 335-45.

238. Kobayashi, K., et al., *Another phenotype of frontotemporal dementia and parkinsonism linked to chromosome-17 (FTDP-17) with a missense mutation of S305N closely resembling Pick's disease.* J Neurol, 2003. **250**(8): p. 990-2.

239. Boeve, B.F., et al., *Longitudinal characterization of two siblings with frontotemporal dementia and parkinsonism linked to chromosome 17 associated with the S305N tau mutation.* Brain, 2005. **128**(Pt 4): p. 752-72. Epub 2004 Dec 22.

240. Hinds, D.A., et al., *Whole-genome patterns of common DNA variation in three human populations.* Science, 2005. **307**(5712): p. 1072-9.

241. Shaw-Smith, C., et al., *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability.* Nat Genet, 2006.

242. Rademakers, R., et al., *High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy.* Hum Mol Genet, 2005. **14**(21): p. 3281-92.

243. Pittman, A.M., et al., *Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration.* J Med Genet, 2005. **42**(11): p. 837-46.

244. Poorkaj, P., et al., *TAU as a susceptibility gene for amyotropic lateral sclerosis-parkinsonism dementia complex of Guam.* Arch Neurol, 2001. **58**(11): p. 1871-8.

245. de Silva, R., et al., *The tau locus is not significantly associated with pathologically confirmed sporadic Parkinson's disease.* Neuroscience Letters, 2002. **330**(2): p. 201-203.

246. Farrer, M., et al., *The Tau H1 Haplotype is associated with Parkinson's disease in the Norwegian population.* Neuroscience Letters, 2002. **322**(2): p. 83-86.

247. Golbe, L.I., et al., *The tau A0 allele in Parkinson's disease.* Mov Disord, 2001. **16**(3): p. 442-7.

248. Hoenicka, J., et al., *The tau gene A0 allele and progressive supranuclear palsy.* Neurology, 1999. **53**(6): p. 1219-25.

249. Maraganore, D.M., et al., *Case-Control study of the extended tau gene haplotype in Parkinson's disease.* Ann Neurol, 2001. **50**(5): p. 658-61.

250. Pastor, P., et al., *Significant association between the tau gene A0/A0 genotype and Parkinson's disease.* Ann Neurol, 2000. **47**(2): p. 242-5.

251. Baker, M., et al., *No association between TAU haplotype and Alzheimer's disease in population or clinic based series or in familial disease.* Neurosci Lett, 2000. **285**(2): p. 147-9.

252. Crawford, F., et al., *No genetic association between polymorphisms in the Tau gene and Alzheimer's disease in clinic or population based samples.* Neurosci Lett, 1999. **266**(3): p. 193-6.

253. Green, E.K., et al., *A polymorphism within intron 11 of the tau gene is not increased in frequency in patients with sporadic Alzheimer's disease, nor does it influence the extent of tau pathology in the brain.* Neurosci Lett, 2002. **324**(2): p. 113-6.

254. Lilius, L., et al., *Tau gene polymorphisms and apolipoprotein E epsilon4 may interact to increase risk for Alzheimer's disease.* Neurosci Lett, 1999. **277**(1): p. 29-32.

255. Myers, A.J., et al., *The H1c haplotype at the MAPT locus is associated with Alzheimer's disease.* Hum Mol Genet, 2005. **14**(16): p. 2399-404.

256. Hughes, A., D. Mann, and S. Pickering-Brown, *Tau haplotype frequency in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Exp Neurol, 2003. **181**(1): p. 12-6.

257. Ingelson, M., et al., *Increased risk for frontotemporal dementia through interaction between tau polymorphisms and apolipoprotein E epsilon4.* Neuroreport, 2001. **12**(5): p. 905-9.

258. Kowalska, A., et al., *Genetic analysis in patients with familial and sporadic frontotemporal dementia: two tau mutations in only familial cases and no association with apolipoprotein epsilon4.* Dement Geriatr Cogn Disord, 2001. **12**(6): p. 387-92.

259. Ghidoni, R., et al., *The H2 MAPT haplotype is associated with familial frontotemporal dementia.* Neurobiol Dis, 2006. **22**(2): p. 357-62.

260. van der Zee, J., et al., *A Belgian ancestral haplotype harbours a highly prevalent mutation for 17q21-linked tau-negative FTLD.* Brain, 2006. **129**(Pt 4): p. 841-52.

261. Houlden, H., et al., *Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype.* Neurology, 2001. **56**(12): p. 1702-6.

262. Brandt, R., M. Hundelt, and N. Shahani, *Tau alteration and neuronal degeneration in tauopathies: mechanisms and models.* Biochim Biophys Acta, 2005. **1739**(2-3): p. 331-54.

263. Gotz, J., *Tau and transgenic animal models.* Brain Res Brain Res Rev, 2001. **35**(3): p. 266-86.

264. Ishihara, T., et al., *Age-dependent emergence and progression of a tauopathy in transgenic mice overexpressing the shortest human tau isoform.* Neuron, 1999. **24**(3): p. 751-62.

265. Lewis, J., et al., *Neurofibrillary tangles, amyotrophy and progressive motor disturbance in mice expressing mutant (P301L) tau protein.* Nat Genet, 2000. **25**(4): p. 402-5.

266. Tatebayashi, Y., et al., *Tau filament formation and associative memory deficit in aged mice expressing mutant (R406W) human tau.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13896-901. Epub 2002 Oct 4.

267. Allen, B., et al., *Abundant tau filaments and nonapoptotic neurodegeneration in transgenic mice expressing human P301S tau protein.* J Neurosci, 2002. **22**(21): p. 9340-51.

268. Perez, M., et al., *The FTDP-17-linked mutation R406W abolishes the interaction of phosphorylated tau with microtubules.* J Neurochem, 2000. **74**(6): p. 2583-9.

269. Tanemura, K., et al., *Formation of filamentous tau aggregations in transgenic mice expressing V337M human tau.* Neurobiol Dis, 2001. **8**(6): p. 1036-45.

270. Tanemura, K., et al., *Neurodegeneration with tau accumulation in a transgenic mouse expressing V337M human tau.* J Neurosci, 2002. **22**(1): p. 133-41.

271. Ishihara, T., et al., *Age-dependent emergence and progression of a tauopathy in transgenic mice overexpressing the shortest human tau isoform.* Neuron, 1999. **24**(3): p. 751-62.

272. Ishihara, T., et al., *Age-dependent induction of congophilic neurofibrillary tau inclusions in tau transgenic mice.* Am J Pathol, 2001. **158**(2): p. 555-62.

273. Spittaels, K., et al., *Prominent axonopathy in the brain and spinal cord of transgenic mice overexpressing four-repeat human tau protein.* Am J Pathol, 1999. **155**(6): p. 2153-65.

274. Katsuse, O., et al., *Neurofibrillary tangle-related synaptic alterations of spinal motor neurons of P301L tau transgenic mice.* Neurosci Lett, 2006.

275. Murakami, T., et al., *Cortical Neuronal and Glial Pathology in TgTauP301L Transgenic Mice: Neuronal Degeneration, Memory Disturbance, and Phenotypic Variation.* Am J Pathol, 2006. **169**(4): p. 1365-1375.

276. Santacruz, K., et al., *Tau suppression in a neurodegenerative mouse model improves memory function.* Science, 2005. **309**(5733): p. 476-81.

277. Stamer, K., et al., *Tau blocks traffic of organelles, neurofilaments, and APP vesicles in neurons and enhances oxidative stress.* J Cell Biol, 2002. **156**(6): p. 1051-63.

278. Shea, T.B. and L.A. Flanagan, *Kinesin, dynein and neurofilament transport.* Trends Neurosci, 2001. **24**(11): p. 644-8.

279. Alberici, A., et al., *Frontotemporal dementia: impact of P301L tau mutation on a healthy carrier.* J Neurol Neurosurg Psychiatry, 2004. **75**(11): p. 1607-10.

280. Vanmechelen, E., et al., *Quantification of tau phosphorylated at threonine 181 in human cerebrospinal fluid: a sandwich ELISA with a synthetic phosphopeptide for standardization.* Neurosci Lett, 2000. **285**(1): p. 49-52.

281. Singleton, A., A. Myers, and J. Hardy, *The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R123-6. Epub 2004 Feb 19.

282. Singleton, A.B., et al., *alpha-Synuclein locus triplication causes Parkinson's disease.* Science, 2003. **302**(5646): p. 841.

283. Maraganore, D.M., et al., *Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease.* Jama, 2006. **296**(6): p. 661-70.

284. Olson, M.I. and C.M. Shaw, *Presenile dementia and Alzheimer's disease in mongolism.* Brain, 1969. **92**(1): p. 147-56.

285. Ponting, C.P., et al., *Identification of a novel family of presenilin homologues.* Hum Mol Genet, 2002. **11**(9): p. 1037-44.

286. Harada, A., et al., *Altered microtubule organization in small-calibre axons of mice lacking tau protein.* Nature, 1994. **369**(6480): p. 488-91.

287. Ikegami, S., A. Harada, and N. Hirokawa, *Muscle weakness, hyperactivity, and impairment in fear conditioning in tau-deficient mice.* Neurosci Lett, 2000. **279**(3): p. 129-32.

288. Polymeropoulos, M.H., et al., *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.* Science, 1997. **276**(5321): p. 2045-7.

289. Spillantini, M.G., et al., *Alpha-synuclein in Lewy bodies.* Nature, 1997. **388**(6645): p. 839-40.

290. Farrer, M., et al., *{alpha}-synuclein gene haplotypes are associated with Parkinson's disease.* Hum. Mol. Genet., 2001. **10**(17): p. 1847-1851.

291. Mueller, J.C., et al., *Multiple regions of alpha-synuclein are associated with Parkinson's disease.* Ann Neurol, 2005. **57**(4): p. 535-41.

292. Chiba-Falek, O., et al., *Regulation of alpha-synuclein expression by poly (ADP ribose) polymerase-1 (PARP-1) binding to the NACP-Rep1*

polymorphic site upstream of the SNCA gene. Am J Hum Genet, 2005. **76**(3): p. 478-92. Epub 2005 Jan 25.

293. Chiba-Falek, O., J.W. Touchman, and R.L. Nussbaum, *Functional analysis of intra-allelic variation at NACP-Rep1 in the alpha-synuclein gene.* Hum Genet, 2003. **113**(5): p. 426-31. Epub 2003 Aug 16.
294. Chiba-Falek, O. and R.L. Nussbaum, *Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system.* Hum Mol Genet, 2001. **10**(26): p. 3101-9.
295. Goate, A., et al., *Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.* Nature, 1991. **349**(6311): p. 704-6.
296. Cabrejo, L., et al., *Phenotype associated with APP duplication in five families.* Brain, 2006.
297. Sleegers, K., et al., *APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy.* Brain, 2006.
298. Rovelet-Lecrux, A., et al., *APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy.* Nat Genet, 2006. **38**(1): p. 24-6.
299. Koolen, D.A., et al., *A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism.* Nat Genet, 2006. **38**(9): p. 999-1001.
300. Sharp, A.J., et al., *Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome.* Nat Genet, 2006. **38**(9): p. 1038-42.
301. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
302. Maraganore, D.M., et al., *High-resolution whole-genome association study of Parkinson disease.* Am J Hum Genet, 2005. **77**(5): p. 685-93.
303. Clarimon, J., et al., *Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease.* Am J Hum Genet, 2006. **78**(6): p. 1082-4; author reply 1092-4.

304. Farrer, M.J., et al., *Genomewide association, Parkinson disease, and PARK10.* Am J Hum Genet, 2006. **78**(6): p. 1084-8; author reply 1092-4.

305. Goris, A., et al., *No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening.* Am J Hum Genet, 2006. **78**(6): p. 1088-90; author reply 1092-4.

306. Li, Y., et al., *A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan.* Am J Hum Genet, 2006. **78**(6): p. 1090-2; author reply 1092-4.

307. Myers, D.R., *Considerations for genomewide association studies in Parkinson disease.* Am J Hum Genet, 2006. **78**(6): p. 1081-2.

308. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration.* Science, 2005. **308**(5720): p. 385-9.

309. Hughes, A.E., et al., *A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration.* Nat Genet, 2006. **38**(10): p. 1173-7.

310. Li, M., et al., *CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration.* Nat Genet, 2006. **38**(9): p. 1049-1054.

311. Maller, J., et al., *Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration.* Nat Genet, 2006. **38**(9): p. 1055-1059.

10 Appendices

10.1 Chapter 3 genotyping assays:

SNP	Forward	Reverse	bp
rs5911	GGAAGATCTGTCTGCGATCC	TCCTGTCAACCCTCTCAAGG	214
rs894685	CCAAGTCTCCCCAAACAACA	TTGCTGAGCTTCTTGGAGGT	234
rs732589	TGGGCTCCAGTTACAATGG	TAAAGGGCAGGTCCAGGCTTCAAG	420
rs2668643	GAGCTTAGGGCCTGATTCTG	GGTTGCGGTGAACTGAGATT	431
rs1880748	TGGCTTGGGCTGATGGAATG	TGCTGACTTTTCTGTGACGATGC	296
rs110402	AAACCACCTCGGCACAAAGC	TCACAAACCTTCCACAGAGCAAG	378
rs1396862	TGTGGACCAAGGTTCACTGCTC	TGGAATGTATGCCCTACGCCAG	134
rs916793	TGCCCTGAATTTGGTGT	AATGCTAAGAGGCCATCA	179
del-In9	GGAAGACGTTCTCACTGATCTG	AGGAGTCTGGCTTCAGTCTCTC	484(H1)
rs1468241	AGCCGACGGATAAACAA	ATGGATAAGGCCCATGA	151
rs1528072	AACAGCCTTTCCCATCA	TTGATTTGGCAAGTACCAG	241
rs2240756	GGAGGAAACACGAATCAA	CCCCATTGTATTCCACAA	232
rs142167	ATAATGTGTCGTTCGGGGAGTGGG	TTGTTGTCTTGCTGAAGGCTGAG	206
rs199528	AACAGCTAACGACTCTGTCC	GTTCTGTGTAGGGTTGTCCT	250
rs70602	GCGGAAATCTAACCATCTGTGC	GAACGGCTTCTTGACCTAAGTGG	414
rs1662577	AGGAGAGGCTGGAGGGTAAG	ATCCCTGGGAGGTGGATAAG	247
rs758391	CCACCGTGGAGACATCTGTA	GCTGGTAGGCCTGTGGTAAA	232

Table 10.1 PCR primer pairs, used to for genotyping the SNPs in Chapter 1

SNP	bp	Pyrosequencing primer	Enzyme
rs5911		...	*Fok I* (C)
rs894685		...	*Acc I* (T)
rs732589		...	*BsrG I* (C)
rs2668643		...	*BstE II* (T)
rs1880748		...	*Dra III* (C)
rs110402		ACAGAGGACTGGTGTTG	...
rs1396862		AGGAAAAAGCACCT	...
rs916793		CATAAAAAGGCATTTTCTAC	...
del-In9		...	...
rs1468241		GGAAGCAGTGCACTGAC	...
rs1528072		TTGTCCAGAATTCTTTCTAC	...
rs2240756		GCATGAATCAGGTGGT	...
rs142167		GCATGATGGAAAGCTG	...
rs199528		...	*Pvu II* (C)
rs70602		TCTCCTGTGGTCATTTT	...
rs1662577		...	*Apo I* (A)
rs758391		...	*Hph I* (G)

Table 10.2 Genotyping assays for the SNP polymorphisms in Chapter 1

Genotyping assays for the SNP polymorphisms, either by Pyrosequencing or by RFLP. In the case of the restriction digests, the enzyme cuts at the (N) allele.

10.2 Chapter 4 genotyping assays

SNP	Forward	Reverse	bp	Enzyme
rs1467967	GAAGGGAGGAGCTCACACAG	CCACCCTTCAGTTTTGGATG	365	*Dra I* (A)
rs242557	ACAGAGAAAGCCCCTGTTGG	ATGCTGGGAAGCAAAAGAAA	384	*Apa L I* (A)
rs3785883	CATTGCCATCACCTTGTCAG	AGTTTCCTGGAAGCCATGTG	293	*Bsa H I* (G)
rs2471738	GAACACAGGAGGGAGGGAAG	GAACCGAATGAGGACTGGAA	292	*Bste II* (C)
rs7521	ACCTCTGTGCCACCTCTCAC	AGGTGAGGCTCTAGGCCAGT	232	*Pst I* (A)
del-In9	GGAAGACGTTCTCACTGATCTG	AGGAGTCTGGCTTCAGTCTCTC	484	…

Table 10.3 Haplotype tagging SNPs for the *MAPT* association study

SNP	Forward	Reverse	bp	Enzyme
rs4074461	TCAGAAGGGTCGGATATTGG	GCTTGTCACCCTCATCATCC	291	*Sma I* (G)
rs2301689	GGAAGCCAGGTAGATTCTCT	GGGAACTGGGAATTAGAAAG	238	*Fsp I* (A)
rs242928	TTAAGGAAGCACCCATGACAGCC	AAACAGTTCTGTGGAATTTCACCCTG	461	*Mbo II* (T)
rs930119	GGTCCTGGAGATGAACATAA	TCTGAGAAATTCCTGTGTCC	158	*Bsm I* (G)
rs2280004	TAACGAGGCAATGGTTTTAG	GTTCCTTTGCCCTACTTCA	385	*BstFS I* ()
rs1467967	GAAGGGAGGAGCTCACACAG	CCACCCTTCAGTTTTGGATG	365	*Dra I* (A)
rs242557	ACAGAGAAAGCCCCTGTTGG	ATGCTGGGAAGCAAAAGAAA	384	*ApaL I* (A)
rs242559	TATCCCAATGGCCTGAC	CCTTCCCCACAAGACAA	274	*Apo I* (A)
rs242562	ACTGAAGGGATAAGGAGGAC	GGGAGCTGACGTTCATTA	227	*Xho I* (A)
rs3785883	CATTGCCATCACCTTGTCAG	AGTTTCCTGGAAGCCATGTG	293	*BsaH I* (G)
rs2435207	AGCAAGCTGTGTGACCAG	CCCATTCTCTGACAGATTTG	238	*Bcl I* (A)
rs2258689	AGACATCCACACGTTCCTC	CAAACCACAGCAGAGCAG	292	*BmGB I* (C)
rs2471738	GAACACAGGAGGGAGGGAAG	GAACCGAATGAGGACTGGAA	292	*BstE II* (C)
rs11568306	GACTGATAGGTGGGAGGTGGCTGC	CAGCAGCTCGGACGTGAG	454	*Pvu II* (AA)
rs7521	ACCTCTGTGCCACCTCTCAC	AGGTGAGGCTCTAGGCCAGT	232	*Pst I* (A)

Table 10.4 Primer pairs of the Japanese SNPs

SNPs used to determine the linkage disequilibrium structure in the Japanese population

10.3 Chapter 5 genotyping assay: rs7201728 (A/G)

SNP	Forward	Reverse	bp	Enzyme
rs7201728	GAGGGAATGCATTTTCTTGG	TTCACATTTTTCCCCCTCTAAG	248	*HpyCH4 IV* (G)

Table 10.5 rs7201728 (A/G) SNP assay

10.4 *MAPT* re-sequencing primers

Amplicon	Forward	Reverse	bp
1	TTACCCCCTCCCCACTACTTTCGG	CCCCTGTCGTCTCCTCTTCAAG	642
2	CCATTCCATCCTGAACCTGCC	CTCATCACTGGGGAGAGTTTGTG	684
3	CCACTTTAGGTGCTATCAAGGGG	TGTGTCTGAGAGGCTTCTGGGTC	527
4	CGGGAAAACTGAAGGGCATC	CCAACAGAAAGACCAAAGAACAGC	697
5	CTGCCTTCCTTCCTTCCTGC	GCCAAAACCGTGTCCTGGTG	668
6	TTCTTCTTCCTTACAAAGCAGTTGG	CATACCAGGGCAAGCATACACATC	748
7	CTAAGTGAAGGATTCGGAGTGGAC	CAGTGGGGTAGGTGGGAGTTTG	782
8	TTCTGTCCCTCCAAGACGGTG	GTCCCAACTTCTGCCCTGATG	582
9	CCTTCCTTGTCTTCCCTGTGC	GGTGGAGGCTTTGGCTTCAC	620
10	GCCCAAGAAAGACAGTTTTTTTTGC	TAAGACAGAGTCCCCCTCGGTC	507
11	CACGGTGGCTCACACCTGTAATAC	AATCAACGATTGTATGAGGGCATAG	602
12	ACATCAGCAGCAACCATCCG	TTGTGTCCCCTCCTCTCTCCTG	379
13	AGAGAGGAGGGGACACAAAAGG	AGGTGGGTGAATCATCTGAGGTC	367
14	TGTTGGTCAGGCTGGTCTCG	TAGGGAATGGGAGTGGGGAAGG	367
15	AACTGGCAAGGGATGTTGACTG	TGCTGCGTGCTGTTTGGAAC	548
16	CAAACAGCACGCAGCATTCC	CACACCTGAAATCCCAGCACTTC	709
17	ACCCCATCTCCATTGTTGTCAG	CACCAAACCCAGAAGAAAGGC	680
18	TATTTTCCTGTCTGTGTGTCTCGC	AGATGGGGTTTCACTATGTTGCC	742
19	GAAGGTTGAGAGAAGTGGCTTGG	GGACAGGAAATGCTCGTAGGTTC	711
20	CAGGTAGGTCCCAAAGCATTAGC	GTCAGGAGCAGGAGAATCAAAGG	533
21	TCTTGGCTAACACGGTGAAACC	CGATTGGATGGGTAGAGAGCAAC	647
22	AAAACGCAGTCACGGGGAAGAG	AGGCTGGAGAATCGCTTGAACC	760
23	AATCCAACAGAAGGCAGAGGGG	GACAGGCTGGTAGTGAGGAAAGTTC	617

Table 10.6 *MAPT* Intron 0 re-sequencing primers

Amplicon	Forward	Reverse	bp
1	AAGGGGTGGCGGGAACAGTTTG	GGAAAGAGAGAGGAAGAGGCGAAG	766
2	GCCAAATTGACTTCGCCTCTTC	TGTTTGCCATTTTGCCTCCTTAG	691
3	TCTAAGGAGGCAAAATGGCAAAC	GGCTGGAGCAGTCAGAAAGCAC	539
4	TGGGTTGGGGGTAGGTGAAG	TCAAGCAATCCTCCTGCCGC	536
5	CCCCCTGTTTTCTCCTTTGAATC	TTTCCCCCAGCAGCACCAAC	641
6	GATGTGTGAAGTGAGGGGTGGTAG	TAGGACGGAGTCTCGCTCTGTC	548
7	AGGAGAATCACTTGAATCTGGGAAG	CACGGGCTGTTGTGAGCATC	548
8	TGATGCTCACAACAGCCCG	AGATGATTCTCCCACCTCAGCCCCG	641
9	CCTGTAGTCCCAGCTACTTGCG	ACTCTTTGTTCTGGCATTTAACGTG	589
10	GACTTGGCAGAAGGTGTCACTTTAC	TTTCAGGATGGGGCTGGTTC	691
11	CTCCACTTGCCTTTTTTTCATCTTC	CTCCCCTTCCTCCTTTTCCAG	654
12	TGAGACCTTTACACCCTCCGC	AGTCACATACCCCGTCAAGCAC	568
13	TTCCTATCCAATGACTTCCCTTAGC	TGGTTGTGTGAGGCTGAGCG	642
14	AATCACTGCCACTTATTGGGGAC	GGGTGGAGGGAGGAAGATATGG	596
15	CTGGAGAGGCTTTTCTGGGC	GGATGGGTGATGTGTAGATGATTCG	787
16	ATTCCTTCCTCCATCCATCCC	TTCAGTTCTGCCCCACCTGG	618
17	TCATCCATCCATCCGTCCG	TTGTCCCTTCCCTGTGGCAC	711
18	TGCTCAGCCCTGGATACTGC	GAAAATCGCTCCCTGTGGAAAC	748

Table 10.7 *MAPT* Intron 1 re-sequencing primers

Amplicon	Forward	Reverse	bp
1	ATGCCTCCAGCCTGTGCTTAG	GAGCCCCCAAAATGAAGGATAC	387
2	GCTCCATACGATTGTCCTCCCAC	CAACTTGTTCCTGCTTCAAAGGC	560
3	CTCTTGCTTCCCATCTGTGTCAG	CGTTCTTCCCCCTTGGTTTATTC	462
4	GCATCACATCTGCCTTTGAAGC	CAACTTGTAAAACCGTCCCGAC	613
5	GAGCATTTTGCCGTCCAACC	GAGGGGATTGCGACTTGTTTG	552
6	CACTTCCTGCTCATCTTGTCGG	TGATTGGACGGTTTCCTCATCTG	649
7	GGACAGGAGCGGACCTATTTATTG	CCCTCACCTTTGCTTCTTGCG	532
8	GCGTGTTTATCCTCCTGTGGG	CAAAAGGTGTGGTCCAGTCATCTG	362
9	AGACCCGTGCCTTCTTGACAC	CTCTCCACTGCTGGTGAAATCC	506
10	GTTTGGTGATTTGGAGGGTGC	GTGATGGAAAAGTCAAGAGTTGGC	458
11	GCCAACTCTTGACTTTTCCATCAC	TTTGCTTGCCAGCGTGTTTCCCG	531
12	ACTGAGTTGGTCCTGGTGTGAGC	TGGTGTGTGTCCCCTCTGGTTAG	480
13	ACCACAGTCCACCAGGTCCCTAAC	GCTAAATGCCAGAGTCCTTGAGAC	639
14	TTCTGGGGTCAAGGGGTGGCAAAG	GTGAAGAGTGTAGGCAAGGGAGAG	536
15	AACCTTTCTCTCCCTTGCCTACAC	GTTACATTTCCTTCCTCCCAACTTC	534
16	GAAATGCTGGTGATTCTGGTGG	TGGTTGTGGAGGCTGTGTCTAAG	605
17	GAAGTTGGGAGGAAGGAAATGTAAC	GTATGTGTCAGCAAATGAAGGGG	616

Table 10.8 *MAPT* Intron 2 and 3 re-sequencing primers

Amplicon	Forward	Reverse	bp
1	CTGGAATTGCCTGCCATGAC	TGCTCCTGGATCAGCCACAG	630
2	TGAGGCAGGACAATCGCTCG	TGGGGGCAGGAATGTGTAGTTAG	606
3	ATTTTAGGAAATGGGAGACCTATGG	CATGATTCCAGGCTGCTGTCAG	681
4	TAGGAGGCAGGAGAATTGCTTG	GGTGGTTGACATGGGCTCATC	624
5	GATGCCAGGAACAGGAATTGG	CCATAAGACTCCTCGAAAAAGAAGC	611
6	AATTCTCCTGTTCTTAATATGCGGC	CTTTAACTCCAGGGCTCAAACAAC	604
7	CAGAAGTCAAGTATGGTGGCTCATC	TTCCTCTCCCTACCTCTCACTGTTC	638
8	TGTAGCCCTCCCGCTTTCCCAGTAG	CCGCCTTTTTGTCCCCACTC	645
9	CAAACCAGAGGAGTTGAGAGTTCC	TTGGAGGGAGGGGAGTCTTG	654
10	AGGAGGTGGATGAAGACCGC	CCAGAGAGTCAAACAAGGTGGG	751
11	GTGCTCCCCACCTTGTTTGAC	ACACCAACTCTCCTCTCCCTGATAC	647
12	GGGGAGGCTGAAGCAGGATAAG	GGACAACTGGGTCTCTTGGAGC	634
13	GGGGTAAGCCTGGGAGTCTTTTAG	CCTTGGAAACTGCCTGGAACC	668
14	AAGACAACCCTGCCAGCCTCTC	TCATCGCTTCCAGTCCCGTC	750
15	AAGCGATGACAAAAAAGCCAAG	GCTCTCCAAATCCCCACAGG	762
16	TGAGCATTCAGTGGTGTGTTGAG	TGGAGGAGAAGGGGGTGTTTG	691
17	TTTGTTTCCCGAGAGTGTGACC	CCAGGCAAGGTGCTTTTTTACC	712
18	GGGGGTGGGTGTAACCTAAGTC	AGTAAGAACAGGAAAGTGACCGTCC	715
19	AAGAAAAAGAACTGTGATTGGGGAG	CGTATTTAGGAGAGGAAGGTGGC	353
20	TAGCCCCAAACACCCCACTC	CTCTTGCCAAAATCAGGAATGG	748
21	CCCAGAGGCAGAGGTTGTAGTTAG	CCAAGTTCCCTGAGGACATTTCC	721
22	ATCTGGGTAGGACAGCATCACGCC	ATAGAGGGGTTTTTCACCATTGG	750
23	ACAGAGCAGCAGCCTGGATAAG	TCACAGCCCGCAGCAAGTGCCATAG	723
24	TCCATTGCCCCAGTGAAGTATTC	AGCCACCGAAGAGCAGGTAAG	797
25	GGGTCTTACCTGCTCTTCGGTG	GCTCTGTGGCAATAGGACAAACG	742
26	AGCCTGAAGATGGAGCAGTCCG	CCCAGCCAGCACCAGTTTTTAG	623
27	GCCCCATCTGTTGTTTCCTTATG	CAGCTCTTTGAAGCACCCATGTAG	707
28	GACTCCGTCTCAAAAAAAAAACCAC	TGTCCACACGAGCCATACCG	714
29	CTGTTGAAGTGTAACGGGGGC	CCAGGAAAGCAATGGAAAACC	672

Table 10.9 *MAPT* Intron 3 to intron 9 re-sequencing primers

Amplicon	Forward	Reverse	bp
1	TGAGCACCCCCGCATAACAC	TTGGAAGAAACCCACCCGCCCACAC	553
2	TCTGGATTCTGGGGAGATTGG	TCGCTCTGTTGCTCAAACTGG	639
3	AGGAGGCTGAGGCAGGAGAATC	CAGGGTTTGAGGCAAGGTGC	583
4	GGTGAGAAGGACCACAGCTC	CCGAGTCTTCTCCACCTCTG	416
5	TCGGGATGGAAATGGACGG	GGAGAATGAAAGTCACGGTGGTC	552
6	AGGGAGGGAAGGAAGAGAACATC	ATGGGACACAGGGCAGGGTATG	602
7	CCTGCCTTTTCTCCCATTGC	CCCACACCTTCCAAGCCAAC	583
8	CCCCTGTAAACTCTGACCACACTG	TCTGCCACTTTTCCACACTTCG	587
9	GGCAGCACGTCCAAATCTAC	TGACTTGGCTCCTGCTCTTT	477
10	GCTGGCGGAGTCTGAAGAATC	GGTTGAGGTTGAGGGTAGGTGTC	602
11	CTGGGTTCTTACTGCTGACACCTAC	CTTTTTTGATACAGGGTCTCGCC	552
12	AATGAGCCATGTTCACACCA	GCTCAGAAGTGCCCCATAAA	576
13	GACCAGGAGTTCAAGGTTGC	TGAACTCTCCTAGCGCTCAA	547
14	GAGGGTGTGTGTTCAGCACTAAGG	ACAGCATTCTATTGCCATCTTGTG	650
15	ACCTGAGCAGTGGGGCTTTG	ATCGGAGAGATGGGCAGAGG	550
16	GCCTTGCTCTGTCTCTACCC	ACATCTCAGCCCAAGCTCAC	587
17	CGTTTGCCTCTCCTCAATCAGG	TGGTTTTGTTTGGGTTCAGTTCTC	611
18	CAAAGCACAGCACAGCGAGC	CAAGGGTATTTACAAAGCAGACCTG	576
19	AACCAAGTGGACACACACCCCC	GGAGGCTGAGACAGGAGAATAGC	559
20	CCATTCCTCTTTTCTCTCCCTTCTG	CGCCTGTAATCTCAGCACTTTGG	479
21	TGTGGGGTTAGAACAAGAAAAAGC	CATTAGAAGAAGGGAAGGGCATC	617
22	AGTGCTGAGATTACAGGCGTGAG	CCACATAAACTCCTTCTTTGCGG	590
23	TGCCCTTCCCTTCTTCTAATGTTAC	TTCCGCAACTTTCTCAGGTCAG	637
24	GACCTGAGAAAGTTGCGGAACC	ACTGTTGGCTGTCGTCACCCTGCTG	602
25	GGAGCAGGGAGAAAGTAGGG	TGAGATAGAATCTTGCCTTGTCG	481
26	GGCGACAAGGCAAGATTCTA	AGAGACAAGCTCTCACTCTGTCA	584
27	CTTGAGCCCAGGAGTTTGAG	TGCATTGCTGGCTTATTCTG	367
28	CCATCAGCAAAAAGGGGGTAAAG	CTGGGTGACAGAATGAGACTTCG	581
29	AGCAATGTGGTGAGGCTGGC	GCTACTCGGGAGGTTGAGGAAG	534
30	TGCTTCTGATGACCGTAGACCATAG	AACCGTTTTCTTACCACCCTAACAC	502
31	CTTCCTCAACCTCCCGAGTAGC	TCATCCCGTTTTTTTTCCCC	495

Table 10.10 *MAPT* Intron 9 re-sequencing primers

Amplicon	Forward	Reverse	bp
1	TTATTAGGAAGTGGTGTGAGTGCG	GCTGAGCAGGACAGTAAGACCTTG	483
2	GACAGAATAGGGCAGATGACGG	ACTCCTGGGCTCAAGCGAAC	533
3	CCAAACCAAAGTGCGACAGGTC	CGGGGACATTTAGAGGAAGCC	430
4	GAGACTAAGGCGGGAGGTTC	GTGCCCAGTCGACATAGACC	425
5	CAACTGACCCACCCGATAAGC	CACTTTGAGGAAGCCAGAGAACC	600
6	ACGCTTCCTCACCAGTGTCT	TCCCAGGTAGAGAAGGAGCA	686
7	TGATGCCCAAAGACGCTGC	CCAGAAAGGAGAGGAGTGGATGG	481
8	TGCCCTCTGAGCACCTTCAC	GGTCTAAACTGGAACCATCCCG	613
9	GCCATCCACTCCTCTCCTTTCTG	AGCCACCTCCCACCTATCAGTC	503
10	TTCGGGATGGTTCCAGTTTAGAC	ACAGTGGCAGACAGACAGGTCC	321
11	GGGGACTCCAGGAATGTCAATC	GCACAGCATCTGAAAAATCTTGC	552
12	AACTCCGCTGGGTCCTGCTTAC	AGGGTGGAAAAATGTGTTGTCG	423

Table 10.11 *MAPT* Intron 10 re-sequencing primers

10.5 The SNPs identified from re-sequencing *MAPT*

MAPT region	dbSNP ID	Context Sequence	I1	I2	I3	P
Intron 0		ACAGCCATTCCATCa/cTGAACCTGCCCAATC	C	C	A	C
Intron 0		AGCTGTGTGTCCTTg/tGGCAAGTTACTTCAC	G	T	G	G
Intron 0		CCATCCAGCCCAGg/tTAAGGAATAGTTGGA	G	G	T	G
Intron 0		CAAGGAGTTCCTCGa/gGTCAATTGCTTTTTC	A	A	G	A
Intron 0		CTGACCAAACCAGCc/tTCTGCAGGCCAGCTG	T	T	C	T
Intron 0		GAGGCCACAACTCCc/tACCTACGGGAAAACT	T	T	C	T
Intron 0		CAAAAATGTAGACAc/gTAATAAGGGAAAATC	C	C	G	C
Intron 0		TACTGTTTTCACTTa/gATGATGCATGGACCT	G	n/a	A	G
Intron 0	rs242557	GCTTCGCCCAGGGTa/gCACCAGGACACGGT	A	G	G	A
Intron 0		TGCCTCAGCCTCCCa/gAGTAGCTGGGATTAT	A	A	G	A
Intron 0		TCCTGGGTGCATTAc/tCCGAAAGCTGAGACT	C	C	T	C
Intron 0		TTGAAGTTTACCTTc/tGCTCCACCACTTCTG	C	C	T	C
Intron 0		TCAGAGGAAACTCCg/tTCTCATCCACTTGGG	T	T	G	T
Intron 0		GCCAGGCGTGGTGGI/ΔGTGCCTGTAGTCCC	Δ	Δ	I	Δ
Intron 0		CGTGGTGGGTGCCTa/gTAGTCCCAGCTACTT	G	G	A	G
Intron 0		AGTCTTACTGAGCAc/gCTCCCAGTGGCTGCC	C	C	G	C
Intron 0		AGCACCTCCCAGTGg/tCTGCCTGTGTGCTGG	G	G	T	G
Intron 0	rs242562	ACCAGCCCGACTCGa/gCTTCCCCTGAGCC	A	G	G	A
Intron 0		ACCCACCTAGGGTCg/tCCATCCTGCCCACC	G	G	T	G
Intron 0		GTGAACTCTGCTGCa/gAAGACTCTATTCTTT	A	A	G	A
Intron 0		GTTCTTGGAGGCCCI/ΔAGCCTTGGCAAAATG	I	I	Δ	I
Intron 0		AGTTCCCATACTTTg/tAATTTTTAAAATTTT	T	T	G	T
Intron 0		AAAGCTGGCCGGGCa/gTGGTGGCTCACGCCT	A	A	G	A
Intron 0		GTAAAACGCAGTCAc/gGGGGAAGAGAGGGCA	C	C	G	C
Intron 0		CCTATTTCCCAACCa/gTCCAAGCCCACCTCT	G	G	A	G
Intron 0		CAGTCAGCTCCCTAc/tGAACAGTTGCTCTCT	T	T	C	T
Intron 0		CAGCCACCTCACCCa/gCCTACAGGTGAGTGG	A	A	G	A
Intron 0		TTCCTTTTTTTTTTI/ΔTTCTTTTTCCTTTTT	Δ	Δ	I	Δ
Intron 0		TTTGAAAGGAGTCTc/tACTCTATTGCCCAGG	C	C	T	C
Intron 0		GCAGTGGTGCAATCc/tCAGCTCACTGCAACC	T	T	C	T
Intron 0		TCACTGCAACCTCCa/gCCTCCTAGGTTCAAG	G	G	A	G
Intron 2	rs9896485	AAACAAGTCGCAATc/gCCCTCTGGACCCTG	G	C	C	C
Intron 2	rs10514889	TGTGCTGTCTTAAGa/gATCAAATAGTTGTA	G	A	G	A
Intron 2	rs17572248	TCCTGATTTTGACCa/gTCTTCTCTTGCTTC	A	A	G	A
Intron 2		CTGTGTCAGTCAAGa/gCTGCATTTGGCTGT	A	A	G	A
Intron 2		AAGGCTGGGACATTa/gAGCATTTTGCCGTC	G	G	A	G
Intron 2	rs17651134	GAAGAACGTTAATAa/gTGGCTTTTGAGTCA	G	G	A	G
Intron 2		CTTTCCATCCTCCCa/cATGGAGTTTCAAGG	C	C	A	C
Intron 2		TGGATCTTGTGTCTc/tTCAAGTGCCCGCCT	T	T	C	T
Intron 2		TGCTTCCTGGGGAGct/ggGGTAGCAGGGGCTC	CT	CT	GG	CT
Intron 2		GGGCTCCGGGCCCTa/gCTTCAGGGCTGCTT	A	A	G	A
Intron 3	rs17572361	CGATGCCTTGGTGCt/cCAGTTACCTCCCTG	T	T	C	T
Intron 3	rs17651243	GACCACACCTTTTGa/gTGATCTTAATGAGG	G	G	A	G
Intron 3		GGGAAGCCTGCAGCa/gTTCTCTAGGGCTCT	A	A	G	A
Intron 3		CCTCGAGGGCTGGCa/gGGGAGGGGAGGGGT	A	A	G	A
Intron 3		TTTGAATTCTTCTAA/tTCTCTTACCCCAGCC	Δ	Δ	I	Δ
Intron 3		TCTGAGGTCCTGCTa/gCTCTTACCCCAGCC	A	A	G	A
Intron 3		CCTTGCAGGTTCCTa/gGTTGCCACCTCTTA	G	G	A	G
Intron 3	rs17651285	CAGTGACCGTGTCCa/gAGGAAGCCTCCTCA	G	G	A	G
Intron 3	rs17572467	GTTGAGCCAACACTc/tGCTCAGCGATATTG	T	T	C	T
Intron 3	rs17572495	GCTACCACATGGCTg/tCCAGGAAACTCGAG	T	T	G	T
Intron 3		CCAGCGGTGGGGATc/tCGGCATTCCTATCT	C	C	T	C
Intron 3	rs754593	ATCTTAGGTGCAGAa/gAGTGACTGACTCAT	A	A	G	A
Intron 3		GGATGTGACTTTCCa/tGAATGTTTAAGGGA	T	T	A	T
Intron 3		TTTCCTGAATGTTTa/tAGGGAAAATGCCCG	A	A	T	A
Intron 4		GGCTGCCCTGTGGCtg/caATCCAGGAGCAAGG	TG	TG	CA	TG
Intron 4		TTCTTTTCTCCCGAa/gTGATGGGATCTGTC	G	G	A	G
Intron 4		GGAGACTCTGCCTCg/aTTGAGGATATTTTG	G	G	A	G
Intron 4		GCTGGTGAGTTGCAa/tAGGTTGTTGCGTTG	T	n/a	A	T
Intron 4		AAAAAACATAAAGGg/tTATTTATTAGGAGA	G	n/a	T	G
Intron 4		AGGGTATTTATTAGI/ΔAACCAAGGAGTGTA	I	n/a	Δ	I
Intron 4		GAGAACCAAGGAGTg/tTAATTCTCCTGTTC	G	n/a	T	G
Intron 4		TGATCAGCTGGGTTa/cTGTCTGCCGTGCAG	C	n/a	A	C
Intron 4	rs2435207	CACTGTCTGCCGTGa/gTCAGTGACCTGGAT	A	G	G	A
Intron 4		AGACTCCATTCTTGa/tTGATCACACATTTT	A	T	A	A
Intron 4		GGAGACCCCGTCTCc/tACAAAAATGAAAA	C	C	T	C
Intron 4		AAAAAAAAAAAAAa/gAAAGAAAAAAGGA	G	G	A	G
Intron 4		AAAAAAAAAGAAAA/TGAAAAAAGGAAAA	Δ	Δ	I	Δ
Intron 4		AAAAAGAAAGAAAa/gAAAGGAAAAAAA	A	G	A	A
Intron 4		ATGTGTGACATGCTΔ/ITGTGCCCAACCAGA	Δ	I	I	I
Intron 4		AATGCGCCCTTGCTa/gGGCCGGCAGAGGAA	G	G	A	G
Intron 4		CCAGGAGGGGCCGCc/tGCTGAAGGGGGCAG	C	C	T	C
Intron 4		GGATGACGGGTGCCc/tCACTTTGCAGATGA	C	C	T	C
Intron 4		GCAGGGCACACACAtt/gcAGGTGACACACACA	TT	GC	GC	TT
Intron 4		GCACACACATTAGGc/tGACACACACAGCAT	T	T	C	T
Intron 4		ACAGCAGCCCTGCAg/tGGAGGGATTGGTGG	T	G	T	G
Intron 4		AAACAGGCTTTTCTg/tCTGCAGGACTTGTC	G	G	T	G
Intron 4		CTGGGGCTCTTACTa/gAAGTGCAGGTTCCC	A	A	G	A
Intron 4	rs2435211	GTATAGTTGGAAAAc/tCTCTGGGCATACTC	T	C	C	T
Intron 4		TTTCAGCAGTGACTa/gTCAAGAGAAGGAGA	G	G	A	G
Intron 4		CCTTTGAAGACCAGc/tAAGTGGCAAGAAAC	T	T	C	T
Intron 4		TGCCAGCCTCTCCTa/tTTCTGGAGGCGTGT	A	A	C	A
Intron 4		ATGGGTTGGTCAGCa/gTAAGATGGCCAGGA	A	A	G	A
Intron 4		GGCCAGGAAGGTGGa/gAATCAGGACTGCTG	A	A	G	A
Intron 4		AAGATGGCCAGGAAΔt/ttGGTGGAAATCAGGA	Δ	Δ	I	Δ
Intron 4		GGTTGCGCTAATCAa/gTGACCCCAGTGTGC	G	G	A	G
Intron 4		CAGAGAAATCCAGCc/tTCACCACAGCTGGC	T	T	C	T
Ex 5 - Ex 9		TGTTTGAAAACCAAa/cCCCAGGACAGTAAA	A	A	C	A
Ex 5 - Ex 9		GACAGTAAATGAAGI/ΔTGTGTTTGAAAACC	Δ	Δ	I	Δ
Ex 5 - Ex 9		TGCACCCAGATCTTc/tGCCCATCCTTCTCA	T	C	T	T
Ex 5 - Ex 9		TGCACATTTAGTGCc/tGGGGGCAGGGGGCA	T	T	C	T
Ex 5 - Ex 9		ATCTATTACCCTGGa/gCCTCGGCCAGCTGG	G	G	A	G
Ex 5 - Ex 9		CAGTGAAGTCTGGGa/tCAGGACCCTTCTCT	A	A	T	A
Ex 5 - Ex 9		CCCAGCCCAAGAGTa/tCTAGACCAGCCTGG	T	T	A	T
Ex 5 - Ex 9		GGAGGATCACTTGAa/gCCCAAAGGGTGGAG	A	A	G	A
Ex 5 - Ex 9		AACAAACAAACAAAI/ΔAGAGTTAACATTGG	I	I	Δ	I
Ex 5 - Ex 9		AAAATACAAAAAAAa/cGAGCCGGGCGTGGT	C	C	A	C
Ex 5 - Ex 9		AAAATACAAAAAAAAI/ΔGAGCCGGGCGTGGT	Δ	Δ	I	Δ
Ex 5 - Ex 9		CAAAAAAACGAGCCa/gGGCGTGGTGGTGGG	G	G	A	G
Ex 5 - Ex 9		CAGTCCAGCCTGGGc/tGACAGAGCGAGACT	C	C	T	C
Ex 5 - Ex 9		TCTCTTCTGTCACTc/tCCCGAACTGGCAGT	T	T	C	T
Ex 5 - Ex 9		TGTGGGGTTTGTTTa/gTTTGTTTTTTAGCC	A	A	G	A
Ex 5 - Ex 9		TTTCTTCAAATGAGc/tTCTGGCATAGAAGC	T	T	C	T
Ex 5 - Ex 9		GGGACCTTGGTTTCa/gAAATCGCTACTCTA	G	G	A	G
Ex 5 - Ex 9		TAAAAATACAAAAAc/tTAGCCGGGCGTGGT	T	T	C	T
Ex 5 - Ex 9		CCGTGAGCTCACTGc/gCCCCTCAACTAGCT	C	C	G	C
Ex 5 - Ex 9		GACTGTGCCATGCCc/tGACTCTGCCAGCTT	C	C	T	C
Ex 5 - Ex 9		CTCCTCGGGTGCTCa/gCCTGGTCTGAGCAC	G	G	A	G
Ex 5 - Ex 9		CGGGTGCTCGCCTGc/gTCTGAGCACACCCA	G	G	C	G
Ex 5 - Ex 9		CCGTGAGCTCACTGc/gCCCCTCAACTAGCT	C	C	G	C
Ex 5 - Ex 9		CTAGCTTGTGGTCCc/tTTGGTTAATGTCAG	T	T	C	t
Ex 5 - Ex 9		GACTGTGCCATGCCc/tGACTCTGCCAGCTT	C	C	T	C
Ex 5 - Ex 9		CTCCTCGGGTGCTCa/gCCTGGTCTGAGCAC	G	G	A	G
Ex 5 - Ex 9		CGGGTGCTCGCCTGc/gTCTGAGCACACCCA	G	G	C	G
Ex 5 - Ex 9		TTAGGGTTGTTTGCa/gAGAGAATGAAAGAA	A	G	G	A
Ex 5 - Ex 9		GATGTCTGCATATGtg/caTAGTGCACTAATGT	TG	TG	CA	TG
Ex 5 - Ex 9	rs11656151	TGTGTAGTGCACTAa/gTGTCCTGCCGATGA	A	G	A	A
Ex 5 - Ex 9		CTGATCATTGCTGTa/gAGCTCTGCTGGCCT	G	G	A	G
Ex 5 - Ex 9		AAATAAGAGGCTTCa/gTTTGTCCTATTGCC	n/a	n/a	A	G
Ex 5 - Ex 9		ACATAGTCAGACCCa/cGTCTCTATAAAAAA	n/a	n/a	A	C
Ex 5 - Ex 9		AAAAACATTATTTTa/tAAAAAAGACATGGA	n/a	n/a	A	T
Ex 5 - Ex 9		GTCTCAAAAAAAAa/cCCACAAAAAAACAA	n/a	n/a	C	A
Ex 5 - Ex 9		CCACCACCTGCTGCa/gTGTCTTTGCGGTGG	G	G	A	G
Ex 5 - Ex 9		CACACATGCCGTGCa/gGTGGCACCCCCAAG	A	A	G	A
Ex 5 - Ex 9		GCTAATTTTTGTATa/tTTTAGTAAAGACAG	T	T	A	T
Ex 5 - Ex 9		TCTCTTTAAGCGACc/tAAGCAAGTCCAGAG	T	T	C	T
Ex 5 - Ex 9		GACTGTGCCATGCCc/tGACTCTGCCAGCTT	n/a	n/a	T	C
Ex 5 - Ex 9		CTCCTCGGGTGCTCa/gCCTGGTCTGAGCAC	n/a	n/a	A	G
Ex 5 - Ex 9		CTCCTCGGGTGCTCc/gTCTGAGCACACCCA	n/a	n/a	C	G
Ex 5 - Ex 9	rs2435214	TTAGGGTTGTTTGCa/gAGAGAATGAAAGAA	A	G	G	A

MAPT region	dbSNP ID	Context Sequence	I1	I2	I3	P
Ex 5 - Ex 9		GATGTCTGCATATG**ca/tg**TAGTGCACTAATGT	n/a	n/a	CA	TG
Ex 5 - Ex 9		CTGATCATTGCTGT**a/g**AGCTCTGCTGGCCT	n/a	n/a	A	G
Ex 5 - Ex 9		CGACCACTCTGTAA**a/g**TAAGAGGCATGAAA	A	A	G	A
Ex 5 - Ex 9		CACACTTTGGAAAA**c/t**ACAGACCCATGAGA	T	T	C	T
Ex 5 - Ex 9		GATAGAATACCAGA**c/g**TGTTGAAGTGTAAC	C	C	G	C
Ex 5 - Ex 9		GCATGGAGAGGAGA**a/g**GAGACCCCCCAGCA	A	A	G	A
Ex 5 - Ex 9		GGAAGGGTGCGCTA**a/g**TAACTGCTATGCAT	G	G	A	G
Ex 5 - Ex 9		GCATGGAGAGGAGA**a/g**GAGACCCCCCAGCA	A	A	G	A
Ex 5 - Ex 9		GGAAGGGTGCGCTA**a/g**TAACTGCTATGCAT	G	G	A	G
Ex 5 - Ex 9		TGGGGAACTGTACG**g/t**CAGGGCACAGCAGC	T	T	G	T
Ex 5 - Ex 9		ATCCTGTTCTGAAG**g/t**CCAAATCCCACAGA	G	G	C	G
Ex 5 - Ex 9		CCTGAAGAATTGTAC**t/t**AGAGGTTGGCAAA	T	T	C	T
Intron 9	rs12051901	GAAGCACAGACTTC**a/g**GGGGCCTGGCCTGGA	G	A	G	G
Intron 9	rs2471738	AACTCACAGGTGAC**c/t**CAGGGCCTCTGTCCT	T	C	C	T
Intron 9		TGCCACCCATAGGT**c/g**AGCATCTCCACTCGT	C	C	G	C
Intron 9		GTGGAGAAGCACAG**a/g**CTTCGGGGGCCTGGC	A	A	G	A
Intron 9		GGATCCCTTTCCCC**a/g**TTCCTGTCCCTGTGC	A	A	G	A
Intron 9		CTTTCCCCATTCCT**a/g**TCCCTGTGCCCCTCG	G	G	A	G
Intron 9		GGACAGACAGTGGC**c/t**CCCCTGTGGCTGTGG	T	T	C	T
Intron 9		AGGGAAGGAAGAGA**a/g**CATCTCAGAATCTCC	A	A	G	A
Intron 9		AACTCACAGGTGAC**c/t**CAGGGCCTCTGTCCT	T	T	C	T
Intron 9		CTGGGTTCTGCCAT**a/g**CCCTGCCCTGTGTCC	A	A	G	A
Intron 9		TGAGGGTGGAGGCC**a/g**AGTCTCATGCATTTT	A	A	G	A
Intron 9		CTACAGTGAACATA**g/t**ATTGCTTTGGGGCCC	G	G	T	G
Intron 9		GCAGGAAGAGCCC**gt/aa**AGGCCCCGCCGATCT	GT	GT	AA	GT
Intron 9		ACTTAAAAAAAAAAI/**Δ**TGAAGAAAACACAAC	I	I	Δ	I
Intron 9		AAAAAAAAAATGAA**g/a**AAAACACAACCAGTA	G	G	A	G
Intron 9		TCCTGGCTGGGCGC**g/a**GTGGCGCATGCCTGT	G	G	A	G
Intron 9		GCAGTGAGCTATGA**a/t**CATACCACTACACCC	A	A	T	A
Intron 9		CAGAGGGTGGTGGA**c/t**GGTGGCCCTGAAGAG	C	C	T	C
Intron 9		GAAACATAGAGGAGA**a/g**CTGAAGAAGCAGGG	A	A	G	A
Intron 9		GGCTTTGAAAGCCT**g/t**GGGGCGGGGGGTGCA	T	T	G	T
Intron 9		GTTTGCCTCTCCTC**a/g**ATCAGGGCCAGGGA	A	A	G	A
Intron 9		CTTCTGAAAAATCTI/**Δ**GTGTGAATCAAATC	Δ	Δ	I	Δ
Intron 9		TGAAAAATCTGTGT**a/g**AATCAAATCATTTT	G	G	A	G
Intron 9		GTATTGCACTGGGC**c/t**TGCTGAGCCTGGTG	T	T	C	T
Intron 9		GGAGACAGGCCAAG**c/t**CTGGGCCAGCCTGG	T	T	C	T
Intron 9		GAGATTACAGGCGT**a/g**AGCCACCGTGCCCG	G	G	A	G
Intron 9		GGCGTGAGCCACCG**c/g**CCCGGCCAGAAAA	T	T	C	T
Intron 9		GAGCCACCGTGCCC**a/g**GCCAGAAAAAAACA	G	G	A	G
Intron 9		TCCTTTGATTCATT**c/t**GTTTAATTTCTTTA	T	T	C	T
Intron 9		TGTTTAATTTCTTT**a/t**ATGGATTAGTACAG	A	A	T	A
Intron 9		GATAAGGAGAGTTA**a/c**ATAACCCATGACAT	A	A	C	A
Intron 9		AACTATTTTTTTTTI/**Δ**ATTTAAAAAATAAT	Δ	Δ	I	Δ
Intron 9		TTAAAAAATAATCC**a/g**CAAAGAAGGAGTTT	G	G	A	G
Intron 9		CCTTGAGAAGCAGT**a/g**TGAGAAAGCCTGTC	G	G	A	G
Intron 9		TGATGTGGCACTAGA**a/g**ATTTATCCATCTCT	G	G	A	G
Intron 9		AACTAAAAGAGTGTAA**c/t**TGGACTGGCTAACA	T	T	C	T
Intron 9		ATACAAAAATTAGC**c/t**GGGCACGGTGGCTC	T	T	C	T
Intron 9	rs12946693	AAAATTAGCTGGGCA**c/t**GGTGGCTCACGCCT	C	T	T	C
Intron 9		GCCGGGCATGGTGG**c/t**GGGCGCCTGTAGTC	T	T	C	T
Intron 9		TACAGGAGCCACTG**c/t**ACCCAGCCTAGGAT	T	T	C	T
Intron 9		TATATAAATGAGACA**t/t**CAAGGAAATACAA	T	T	A	T
Intron 9		GTTCCAACCCCAGA**c/t**CCGACAGCCCAGGC	T	n/a	C	T
Intron 9		CTGGGATGTGACT**cg/ca/tg**ACCTCCCGTCACT	CG	CA	TG	CG
Intron 10		CAAGTGCAACTGAC**c/g**CACCCGATAAGCTGA	C	G	n/a	C
Intron 10		CATCATCCCCTGGC**c/t**GGTCTATGTCGACTG	C	C	T	C
Intron 10	rs17652449	GGCCGGTCTATGTC**c/g**ACTGGGCACCCGAGG	G	G	C	G
Intron 10		CTGTGGCTCCGGGGC**t/t**GGGCTGTCTCCAGAA	T	T	C	T
Intron 10	rs733966	CTGTGCTGCTGAGG**c/t**GGAGCTGGGGAGAGA	C	C	T	C
Intron 10	rs733967	TGTCCCGCGCTGTG**c/t**TCATCTGCGGGCTAG	T	T	C	T
Intron 10	rs733968	GCTAGCCTGTGACC**c/t**GCGCTGTGCTCGTCT	C	C	T	C
Intron 10		CCCGCGCTGTGCTC**a/g**TCTGCGGGCTAGCCT	G	G	A	G
Intron 10		GCCCTAGCCTCTCC**c/t**CCAGCCTGGGGGCCC	C	C	T	C
Intron 10		TGTCCCGCGCTGTG**c/t**TCATCTGCGGGCTAG	T	T	C	T
Intron 10		GCTAGCCTGTGACC**c/t**GCGCTGTGCTCGTCT	C	C	T	C
Intron 10	rs733969	CCCGCGCTGTGCTC**a/g**TCTGCGGGCTAGCCT	G	G	A	G
Intron 10	rs4283261	GCCCTAGCCTCTCC**c/t**CCAGCCTGGGGGCCC	C	C	T	C
Intron 10		TGTCCCCCTCCCTG**c/g**CCTCCAGGTATCCTG	C	C	G	C
Intron 10		TCACGAGTTACCTC**a/c**CCCAGGCCAGAGGGA	A	A	C	A
Intron 10	rs4306559	CACCCCCTCAGTGC**c/t**GGCCTCAGAGCACAG	T	T	C	T
Intron 10		CTGACACACATGGG**c/t**CTCAGCCATCTCTGC	T	T	C	T
Intron 10	rs12150530	GAGCACATGCCAGC**c/t**GACACCAAAATAATT	C	T	T	C
Intron 10	rs12150254	AGAAACCACCACCC**c/t**GCCCTGCACCTTGTT	T	T	C	T
Intron 10	rs12150506	GGAAGCTCCAGGGC**a/g**GTGCTTCCTCGGGAG	G	G	A	G
Intron 10		AAGCTCCAGGGCGG**g/t**GCTTCCTCGGGAGCT	T	T	G	T
Intron 10	rs12150170	CTCCAGGAATGTCA**a/g**TCAGTGACCTGCCCC	A	A	G	A
Intron 10	rs12150515	ACACAGCCATGGCT**g/t**CATAGAGGCCTGCCT	G	G	T	G
Intron 10		GAGGCCTGCCTCCA**a/g**GGGACCTGTCTGTCT	A	A	G	A
Intron 10		GGAGTCCCTACAGC**a/g**TGCCCCCCACAGGGG	G	G	A/G	G
Intron 10	rs17573858	CTGGCAGCTCAGGG**c/t**CATCATTCCCAGAGG	T	T	C	T
Intron 10		CTCATGTGTGTCTG**c/t**GTCCGCTCGAGCTTA	C	C	T	C
Intron 10		TGTGTGTCTGCGTC**c/t**GCTCGAGCTTACTGA	C	C	T	C
Intron 10		GGAATAAAGCCCAC**c/g**CATCGACCCTCACAG	G	G	C	G
Intron 10		ACTCATCATATTGA**c/t**CACTTTTTTCTTCAT	T	T	C	T
Intron 10		CCCTGGGGCAGGGTI/**Δ**GGCAGAATTTCGACA	I	I	Δ	I

I1 : Homozygous carrier for the H1c *MAPT* haplotype, I2 : Homozygous carrier of the H1b MAPT haplotype

I3 : Homozygous carrier of the H2 *MAPT* haplotype, P : Allele derived from published sequence

Unique homozygous alleles in shaded squares

I/Δ : insertions/deletions of nucleotides (in the range of 1bp to 40bp)

Table 10.12 The SNPs identified from re-sequencing *MAPT*

10.6 Selected reprints of publications arising from the work in this thesis

Pittman, A., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., de Silva, R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics 13(12), 1267-1274.

Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L., Thomas, N., Morris, C.M., Dickson, D., Wood, N.W., Hardy, J., Lees, A.J., de Silva, R. (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the *tau* gene in progressive supranuclear palsy and corticobasal degeneration. Journal of Medical Genetics 42(11), 837-846.

Pittman, A.M.,Shaw-Smith, C., Willatt, L., Martin, H., Rickman, L., Gribble, S., Curley, R., Cumming, S., Dunn, C., Kalaitzopoulos, D., Porter, K., Prigmore, E., Krepischi-Santos, A.C., Varela, M.C., Koiffmann, C.P., Lees, A.J., Rosenberg, C., Firth, H.V., de Silva, R., Carter, N.P. (2006). Microdeletion encompassing *MAPT* at chromosome 17q21.3 is associated with developmental delay and learning disability. Nature Genetics 38(9), 1032-1037.

Pittman, A.M., Fung, H.C., de Silva, R. (2006). Untangling the tau gene association with neurodegenerative disorders. Human Molecular Genetics 15, R188-R195.

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

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

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.

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

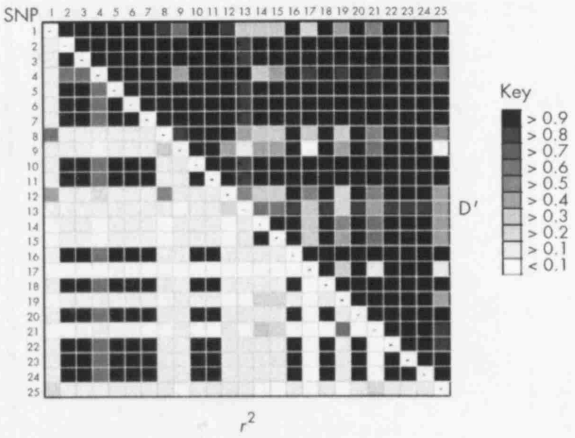

Figure 1 Linkage disequilibrium (LD) across the *MAPT* gene. Numerical LD is presented by grey scale, pairwise between each single nucleotide polymorphism (SNP) by both D' (upper right) and the more stringent measure r^2 (bottom left). The darker the shading indicates a greater extent of LD between the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

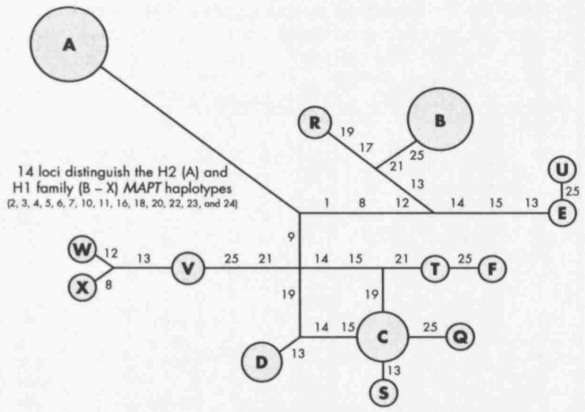

Figure 2 Reduced median network of *MAPT* haplotypes (frequencies that exceed 1% by expectation maximisation, from table 2) in the CEPH trios. Node size is proportional to haplotype frequency. Any two given haplotypes differ by the single nucleotide polymorphism(s) (as numbered from table 1) along the 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): $r^2_{[hap]i} = 1 - R'_i/D_i$, 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 3 Allele frequencies (*F1*) 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 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

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,

Human Molecular Genetics, 2006, Vol. 15, Review Issue No. 2 **R188–R195**
doi:10.1093/hmg/ddl190

Untangling the tau gene association with neurodegenerative disorders

Alan M. Pittman[1], Hon-Chung Fung[1,2,3] and Rohan de Silva[1,*]

[1]Reta Lila Weston Institute of Neurological Studies, University College London, 1 Wakefield Street, London WC1N 1PJ, UK, [2]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Porter Neuroscience Building, 35 Convent Drive, Bethesda, MD 20892, USA and [3]Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taipei, Taiwan

Received July 5, 2006; Revised and Accepted July 25, 2006

Pathological tau protein inclusions have long been recognized to define the diverse range of neurodegenerative disorders called the tauopathies, which include Alzheimer's disease (AD), progressive supranuclear palsy (PSP) and frontotemporal lobar degeneration. Mutations in the tau gene, *MAPT*, cause familial frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17), and common variation in *MAPT* is strongly associated with the risk of PSP, corticobasal degeneration and, to a lesser extent, AD and Parkinson's disease (PD), implicating the involvement of tau in common neurodegenerative pathway(s). This review will discuss recent work towards the unravelling of the functional basis of this *MAPT* gene association. The region of chromosome 17q21 containing *MAPT* locus is characterized by the complex genomic architecture, including a large inversion that leads to a bipartite haplotype architecture, an inversion-mediated deletion and multiplications resulting from non-allelic homologous recombination between the *MAPT* family of low-copy repeats.

INTRODUCTION

The microtubule (MT)-associated protein, tau, was first identified as a 'factor essential for MT assembly', a heat stable protein that induced the assembly of MTs from purified tubulin and belonging to the family of MT-associated proteins (1). Tau is abundantly expressed in both the peripheral and central nervous system (2), where it is enriched in the axons of mature and growing neurones. Low levels of tau are also present in oligodendrocytes and astrocytes (3,4). Tau is a phosphoprotein with developmentally regulated phosphorylation profiles at up to 38 phosphorylation sites (reviewed in 5). The level of protein phosphorylation is highly elevated in fetal tau and pathological tau found within the insoluble, fibrillar inclusions that define tauopathies, compared with normal adult brain tau (6).

The human tau gene, *MAPT* (MIM 157140), spanning ~150 kb of nucleotide sequence on chromosome 17q21.3, consists of one non-coding- and 14 coding exons (7–9) (Fig. 1). In the healthy adult human brain, tau protein exists as six major isoforms produced by the alternative splicing of exons 2, 3 and 10 (10) (Fig. 1). The alternative splicing of exon 10 produces tau isoforms with either three MT-binding repeats (3R-tau) due to exclusion of exon 10 or four repeats (4R-tau) due to exon 10 inclusion. It is now widely recognized that several tauopathies are associated with aberrant splicing of exon 10, causing imbalances in the 3R-tau:4R-tau ratios. For example, the insoluble tau deposits in the different tauopathies have different tau-isoform compositions; in Pick's disease (PiD), the classical Pick bodies consist mainly of 3R-tau isoforms (11,12), whereas in progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and argyrophilic grain disease, both neuronal and glial inclusions contain mostly 4R-tau isoforms (13–16), and roughly equal amounts of 3R- and 4R-tau make up the paired helical filaments and straight filaments observed in Alzheimer' disease (AD) (16,17).

CHROMOSOME 17q21: A FRONTOTEMPORAL DEMENTIA HOTSPOT

After the first identification of mutations in the tau gene (*MAPT*) in FTDP-17 (18,19), over 35 mutations in more than 100 families to date have established the importance of

*To whom correspondence should be addressed. Tel: +44 2076794264; Fax: +44 2076794236; Email: rsilva@ion.ucl.ac.uk

failed to reveal any mutations (30), which presented the possibility of genomic rearrangements affecting *MAPT*. However, two groups have now shown that FTDU-17 is caused by at least 13 different null-mutations in the progranulin (*GRN*) gene, just 1.5 Mb away from *MAPT* (31,32). Identification of mutations not only in the FTDU-17 cases but also in a Belgian FTD patient series also showed that *GRN* mutations are 3.5 times more frequent cause of FTD when compared with *MAPT* (32).

Progranulin is a multifunctional growth factor, expressed in many tissues, including the brain, in both neurons and glia, and has been shown to be involved in several physiological and pathological processes, including wound repair, inflammation and activation of other growth factors such as vascular endothelial growth factor. Over-expression of *GRN* is associated with tumorigenesis (31,32).

Interestingly, most of the mutations identified cause haploinsufficiency and reduced levels of progranulin by creating null alleles (31,32). Nonsense, frameshift and splice-site mutations cause premature termination and degradation of the mutant mRNA by nonsense-mediated mRNA decay (31)—the mRNA from mutation carriers consisted almost entirely of the wild-type mRNA and very little mutant mRNA (31). Two mutations in the first methionine codon destroy the Kozak sequence (31,32).

The comparison of cellular pathways involved in tau and progranulin-related neurodegeneration will provide us with useful insights into the pathogenesis of the clinically similar FTDs and whether both would affect the same pathogenic pathways but different pathological outcomes.

ASSOCIATION OF *MAPT* WITH SPORADIC TAUOPATHIES

Conrad *et al.* (33) first reported a genetic predisposition to sporadic PSP involving *MAPT*. A single allele (a_0) and its genotype (a_0/a_0) of a TG dinucleotide repeat marker located within intron 9 of the gene was significantly over-represented in PSP cases, compared with normal controls (33). A similar association was shown in the rarer tauopathy, CBD (34,35). The finding has now been confirmed in several independent studies (36–39), and Baker *et al.* (40) extended the association with PSP to the H1 haplotype, defined by a region of complete linkage disequilibrium (LD) spanning the entire coding sequence of *MAPT*. The H1 haplotype and its allelic counterpart, H2, were defined by a series of single nucleotide polymorphisms (SNPs), and a 238 bp deletion in intron 9 (*del-In9*) found only on the H2 background (40). The latter is now routinely used to unambiguously assign H1 and H2 haplotypes in *MAPT* genetic association studies (40). In Caucasian populations, the frequency of H1 varies between 70 and 80%; in PSP cases, this frequency is usually over 90% (Table 1). More controversially, the *MAPT* H1 haplotype and H1/H1 genotype have been shown to be associated with sporadic Parkinson's disease (PD) (41). However, this association has not been consistently replicated, although meta-analysis of all studies suggests that homozygosity of H1 contributes to increased risk of PD (41). These findings are surprising, as PD is not traditionally associated with the tau pathology.

The association of the *MAPT* H1 haplotype, in the absence of coding mutations, suggests that underlying variation within the H1 haplotype clade plays a role in what possibly are common pathogenic pathways contributing to the complex aetiologies of these disorders. The most obvious effects could be on *MAPT* transcription, splicing or transcript stability—the 4R-tau dominant pathology seen in PSP and CBD brains would suggest aberrations in exon 10 alternative splicing.

The full extent of the LD and association of H1/H2 with PSP were mapped to cover a region of ~1.8 million base-pairs (42) and can now be confirmed by the analysis of genotype data for 17q21 from HapMap (http://www.hapmap.org). Indeed, it is the longest region of LD identified to date (43) and includes *MAPT* in the centre of this region, with other genes, including corticotrophin-releasing hormone receptor 1 (*CRHR1*), *N*-ethylmaleimide sensitive factor (*NSF*), *IMP5* (44) and predicted genes of unknown function. This raises the possibility that a gene other than *MAPT* could be the actual culprit. For example, a novel gene, *STH*, coding for saitohin, is nested within intron 9 of *MAPT*. A Q7R polymorphism, which is in perfect LD with the H1 and H2 haplotypes, was shown to be associated with AD (45). However, like the H1 and H2 haplotypes, independent studies have failed to confirm this, and unsurprisingly, the Q variant and QQ genotype are over-represented in PSP (46,47). Further study of this gene and its protein would be required to ascertain its function and any disease-related effects.

THE GENOMIC ARCHITECTURE OF THE *MAPT* LOCUS AT 17q21.3

The extended region of complete LD due to the two non-recombining haplotype clades is due to an ancient inversion of a region of ~900 kb that includes *MAPT* (48–50). Detailed analysis of this region revealed multiple low-copy repeats (LCRs), which are the basis of the inversion and genomic complexity of this region by forming the substrates for non-allelic homologous recombination (Fig. 2) (30,49,50). The genomic complexity around the *MAPT* locus, mediated by the multiple LCRs, is underlined not only by complex arrangements of duplications close to the *NSF* gene (50) but also in a recently identified *de novo* microdeletion of 500–600 kb of the locus in individuals with developmental delay and learning disabilities (51,52). It is as yet unclear whether the developmental deficits in these cases were due to haploinsufficiency of the *MAPT* locus or the other affected genes, which include *CRHR1* and *IMP5* (44).

Of particular interest is that the H2 haplotype is limited to the Europeans and other population groups with historical admixture with the Europeans. H2 is completely absent or very rare, in other African, Asian or native American populations (53). Genetic analysis of the H1 clade shows it to be variable and to have a normal pattern of LD (54,55). In contrast, the H2 haplotype is almost invariant, suggesting that it derives from a single founder. Analysis of the sequences on the H1 and H2 backgrounds and comparison of these sequences with those of the chimpanzee (*Pan troglodytes*) sequence show that although both H1 and H2 sequences are

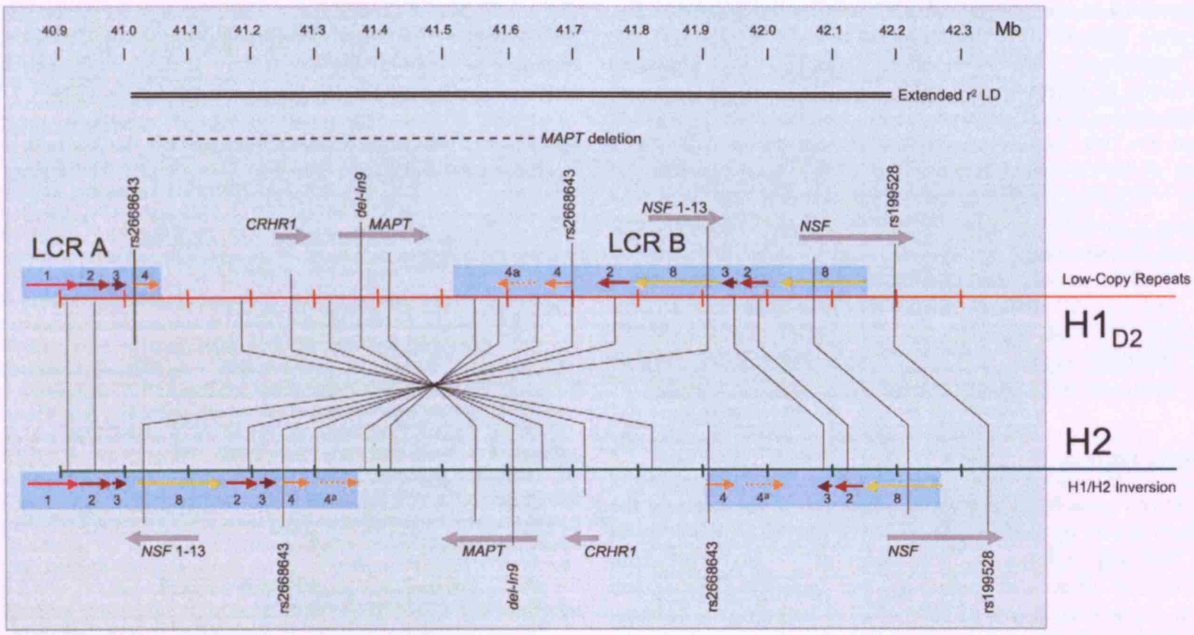

Figure 2. Inverted *MAPT* haplotypes at 17q21.31. H2 assembly is that of Stefansson *et al.* (50). $H1_{D2}$ assembly is based on the human genome assembly of May 2004 (http://genome.ucsc.edu) and is for reference only, as H1 haplotypes show considerable structural diversity/copy number polymorphisms in the region of the 3′ LCR B. Also shown are selected markers (30), the extended region of r^2 LD (42), the *MAPT* microdeletion and genes (51,81); MT associated protein tau (*MAPT*), *CRHR1* and *NSF*. The differences between the H1 and H2 assemblies are far from that of a simple inversion; other structural differences include segmental duplications and segmental duplications in inverted orientation.

associated with PSP both in UK and US case–control cohorts. The H2 haplotype has a strong negative association with PSP, suggesting a protective function (40,54). Using a different approach, Rademakers *et al.* (55) also identified the association of a single common H1 sub-haplotype with PSP.

Although previous studies assessing the H1/H2 division in AD have been at best inconclusive (57–60), we have now shown that the H1c haplotype is also associated with sporadic AD (61), a result that we have now replicated in a separate pathologically confirmed case–control population study (Dr Amanda Myers, personal communication). This supports the notion that underlying variation within the H1c haplotype influences *MAPT* gene at the level of transcription, transcript stability and/or alternative splicing of exon 10. This influence could either be due to allele-specific differences caused by one or more SNPs or larger scale genomic polymorphisms involving *MAPT*. Likewise, some studies have shown an intriguing association of the H1 haplotype with PD, but many have failed to replicate this (38,62–66), although a meta-analysis of all these studies suggests a marginally increased risk (41). In a recent fine-mapping study of the H1 haplotpye in PD, Rademakers *et al.* (67) reported that in the absence of association of the extended H1 clade, sub-haplotype mapping revealed significant associations of a region around exon 10 in early-onset PD (age of onset <55 years) and a region upstream of exon 1 in late-onset PD. A 90 kb interval of the 5′ end of *MAPT* was implicated in a study of the Norwegian PD cases (68).

These findings in AD and PD have also underlined the importance of more detailed analysis of the LD and haplotype pattern of an entire candidate locus using stringently matched case–control cohorts in order to confidently assess any potential association with disease. Consideration of the *MAPT* H1/H2 clade definition alone failed to establish a conclusive association, as they did not take into account the considerable inter-H1 variability.

THE FUNCTIONAL BASIS OF THE *MAPT* ASSOCIATION

The identification of the H1c haplotype of *MAPT* as a modulator of risk of PSP, CBD, AD and PD provided the basis of a more targeted approach in dissecting the functional basis of the *MAPT* H1 association. Sequence polymorphisms within the H1c sub-haplotype could affect gene function, for example, by affecting one or more of the several (>100) evolutionarily conserved, potential *cis*-acting regulatory regions in the 5′- and 3′-untranslated regions of *MAPT* (69). On the basis of the most highly associated H1-specific tagging SNPs on the H1c background (rs242557, rs3785883 and rs2471738), we earmarked a minimal candidate region from the large intron upstream of *MAPT* exon 1 to the 3′ region of intron 9 (54), encompassing potential regulation of both *MAPT* transcription and alternative splicing. Indeed, we have now shown that both transcription and splicing are affected; real-time allele-specific PCR quantitation in postmortem brain reveals that both total

17. Sergeant, N., David, J.P., Goedert, M., Jakes, R., Vermersch, P., Buee, L., Lefranc, D., Wattez, A. and Delacourte, A. (1997) Two-dimensional characterization of paired helical filament-tau from Alzheimer's disease: demonstration of an additional 74-kDa component and age-related biochemical modifications. *J. Neurochem.*, **69**, 834–844.
18. 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.
19. 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.
20. D'Souza, I. and Schellenberg, G.D. (2005) Regulation of tau isoform expression and dementia. *Biochim. Biophys. Acta*, **1739**, 104–115.
21. Goedert, M. (2005) Tau gene mutations and their effects. *Mov. Disord.*, **20**(suppl.), S45–S52.
22. Rademakers, R., Cruts, M. and van Broeckhoven, C. (2004) The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum. Mutat.*, **24**, 277–295.
23. Reed, L.A., Wszolek, Z.K. and Hutton, M. (2001) Phenotypic correlations in FTDP-17. *Neurobiol. Aging*, **22**, 89–107.
24. Casseron, W., Azulay, J.P., Guedj, E., Gastaut, J.L. and Pouget, J. (2005) Familial autosomal dominant cortico-basal degeneration with the P301S mutation in the tau gene: an example of phenotype variability. *J. Neurol.*, **252**, 1546–1548.
25. Baba, Y., Tsuboi, Y., Baker, M.C., Uitti, R.J., Hutton, M.L., Dickson, D.W., Farrer, M., Putzke, J.D., Woodruff, B.K., Ghetti, B. *et al.* (2005) The effect of tau genotype on clinical features in FTDP-17. *Parkinsonism Relat. Disord.*, **11**, 205–208.
26. Morris, H.R., Khan, M.N., Janssen, J.C., Brown, J.M., Perez-Tur, J., Baker, M., Ozansoy, M., Hardy, J., Hutton, M., Wood, N.W. *et al.* (2001) The genetic and pathological classification of familial frontotemporal dementia. *Arch. Neurol.*, **58**, 1813–1816.
27. Mackenzie, I.R., Baker, M., West, G., Woulfe, J., Qadi, N., Gass, J., Cannon, A., Adamson, J., Feldman, H., Lindholm, C. *et al.* (2006) A family with tau-negative frontotemporal dementia and neuronal intranuclear inclusions linked to chromosome 17. *Brain*, **129**, 853–867.
28. Rademakers, R., Cruts, M., Dermaut, B., Sleegers, K., Rosso, S.M., Van den Broeck, M., Backhovens, H., van Swieten, J., van Duijn, C.M. and Van Broeckhoven, C. (2002) Tau negative frontal lobe dementia at 17q21: significant finemapping of the candidate region to a 4.8 cM interval. *Mol. Psychiatry*, **7**, 1064–1074.
29. van der Zee, J., Rademakers, R., Engelborghs, S., Gijselinck, I., Bogaerts, V., Vandenberghe, R., Santens, P., Caekebeke, J., De Pooter, T., Peeters, K. *et al.* (2006) A Belgian ancestral haplotype harbours a highly prevalent mutation for 17q21-linked tau-negative FTLD. *Brain*, **129**, 841–852.
30. Cruts, M., Rademakers, R., Gijselinck, I., van der Zee, J., Dermaut, B., de Pooter, T., de Rijk, P., Del-Favero, J. and van Broeckhoven, C. (2005) Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. *Hum. Mol. Genet.*, **14**, 1753–1762.
31. Baker, M., Mackenzie, I.R., Pickering-Brown, S.M., Gass, J., Rademakers, R., Lindholm, C., Snowden, J., Adamson, J., Dessa Sadovnick, A., Rollinson, S. *et al.* (2006) Mutations in *Progranulin* cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature*, doi:10.1038/nature05016.
32. Cruts, M., Gijselinck, I., van der Zee, J., Engelborghs, S., Wils, H., Pirici, D., Rademakers, R., Vandenberghe, S., Dermaut, B., Martin, J.-J. *et al.* (2006) Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature*, doi:10.1038/nature05017.
33. 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.
34. Di Maria, E., Tabaton, M., Vigo, T., Abbruzzese, G., Bellone, E., Donati, C., Frasson, E., Marchese, R., Montagna, P., Munoz, D.G. *et al.* (2000) Corticobasal degeneration shares a common genetic background with progressive supranuclear palsy. *Ann. Neurol.*, **47**, 374–377.
35. 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.
36. Bennett, P., Bonifati, V., Bonuccelli, U., Colosimo, C., De Mari, M., Fabbrini, G., Marconi, R., Meco, G., Nicholl, D.J., Stocchi, F. *et al.* (1998) Direct genetic evidence for involvement of tau in progressive supranuclear palsy. European Study Group on Atypical Parkinsonism Consortium. *Neurology*, **51**, 982–985.
37. Higgins, J.J., Golbe, L.I., De Biase, A., Jankovic, J., Factor, S.A. and Adler, R.L. (2000) An extended 5′-tau susceptibility haplotype in progressive supranuclear palsy. *Neurology*, **55**, 1364–1367.
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. Psychiatry*, **66**, 665–667.
39. Oliva, R., Tolosa, E., Ezquerra, M., Molinuevo, J.L., Valldeoriola, F., Burguera, J., Calopa, M., Villa, M. and Ballesta, F. (1998) Significant changes in the tau A0 and A3 alleles in progressive supranuclear palsy and improved genotyping by silver detection. *Arch. Neurol.*, **55**, 1122–1124.
40. 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.
41. Healy, D.G., Abou-Sleiman, P.M., Lees, A.J., Casas, J.P., Quinn, N., Bhatia, K., Hingorani, A.D. and Wood, N.W. (2004) Tau gene and Parkinson's disease: a case–control study and meta-analysis. *J. Neurol. Neurosurg. Psychiatry*, **75**, 962–965.
42. Pittman, A.M., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A. and de Silva, R. (2004) The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum. Mol. Genet.*, **13**, 1267–1274.
43. Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A. and Cox, D.R. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science*, **307**, 1072–1079.
44. 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.
45. 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.
46. 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.
47. Ezquerra, M., Campdelacreu, J., Munoz, E., Oliva, R. and Tolosa, E. (2004) Sequence analysis of tau 3′untranslated region and saitohin gene in sporadic progressive supranuclear palsy. *J. Neurol. Neurosurg. Psychiatry*, **75**, 155–157.
48. Fung, H.C., Evans, J., Evans, W., Duckworth, J., Pittman, A., de Silva, R., Myers, A. and Hardy, J. (2005) The architecture of the tau haplotype block in different ethnicities. *Neurosci. Lett.*, **377**, 81–84.
49. Hardy, J., Pittman, A., Myers, A., Gwinn-Hardy, K., Fung, H.C., de Silva, R., Hutton, M. and Duckworth, J. (2005) Evidence suggesting that *Homo neanderthalensis* contributed the H2 MAPT haplotype to *Homo sapiens*. *Biochem. Soc. Trans.*, **33**, 582–585.
50. Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G. *et al.* (2005) A common inversion under selection in Europeans. *Nat. Genet.*, **37**, 129–137.
51. Shaw-Smith, C., Pittman, A.M., Willatt, L., Martin, H., Rickman, L., Gribble, S., Curley, R., Cumming, S., Dunn, C., Kalaitzopoulos, D. *et al.* (2006) Microdeletion encompassing *MAPT* at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat. Genet.*, doi:10.1038/ng1858.
52. Varela, M.C., Krepischi-Santos, A.C., Paz, J.A., Knijnenburg, J., Szuhai, K., Rosenberg, C. and Koiffmann, C.P. (2006) A 17q21.31 microdeletion encompassing the MAPT gene in a mentally impaired patient. *Cytogenet. Genome Res.*, **114**, 89–92.
53. Evans, W., Fung, H.C., Steele, J., Eerola, J., Tienari, P., Pittman, A., Silva, R., Myers, A., Vrieze, F.W., Singleton, A. *et al.* (2004) The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci. Lett.*, **369**, 183–185.

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 H1/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.

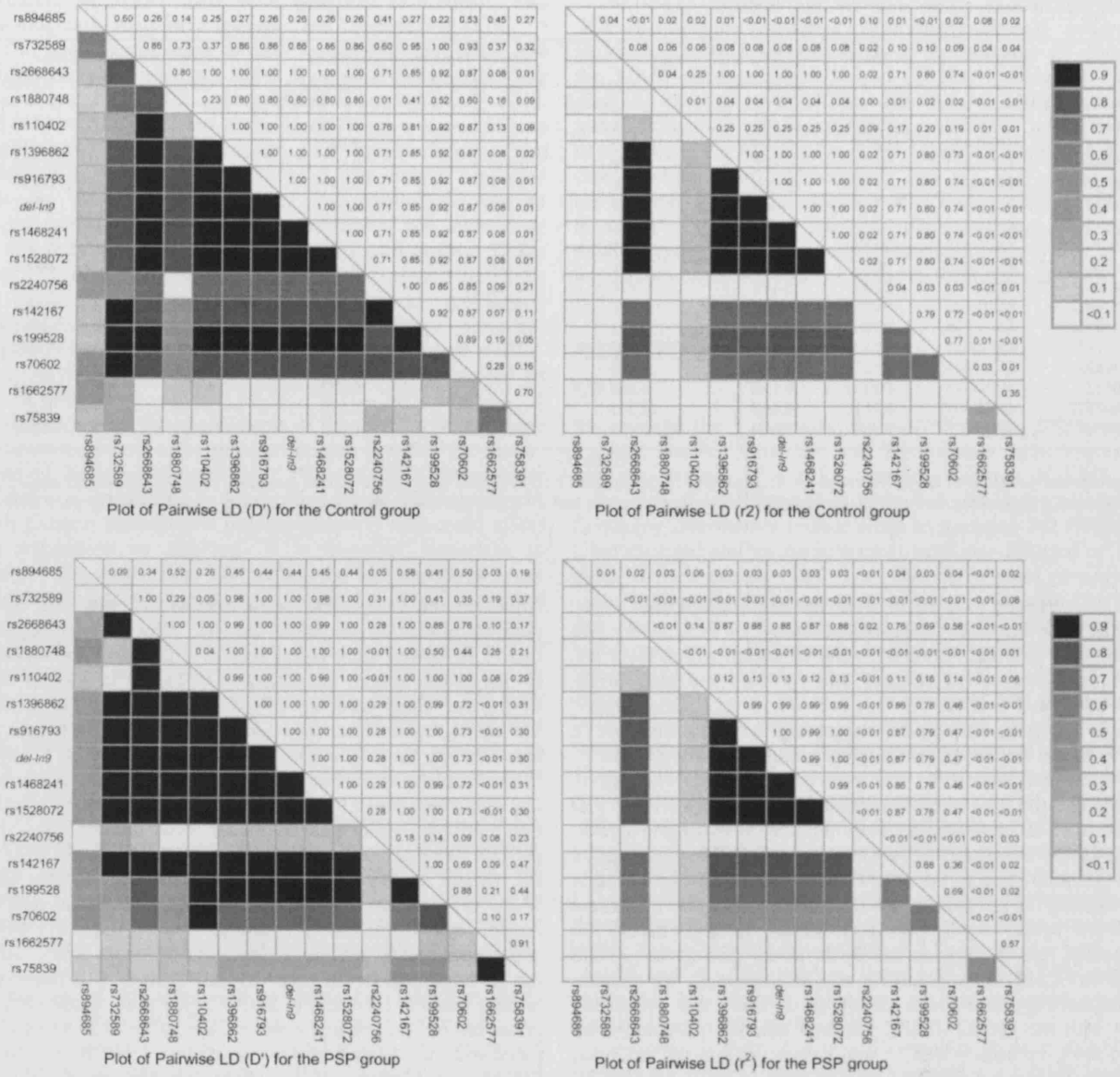

Figure 1. Pair-wise r^2 and D' LD analysis of all of the SNPs for the control and the PSP populations. The blocks are shaded corresponding to the values which were obtained from the LD analysis programme TagIT.

(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

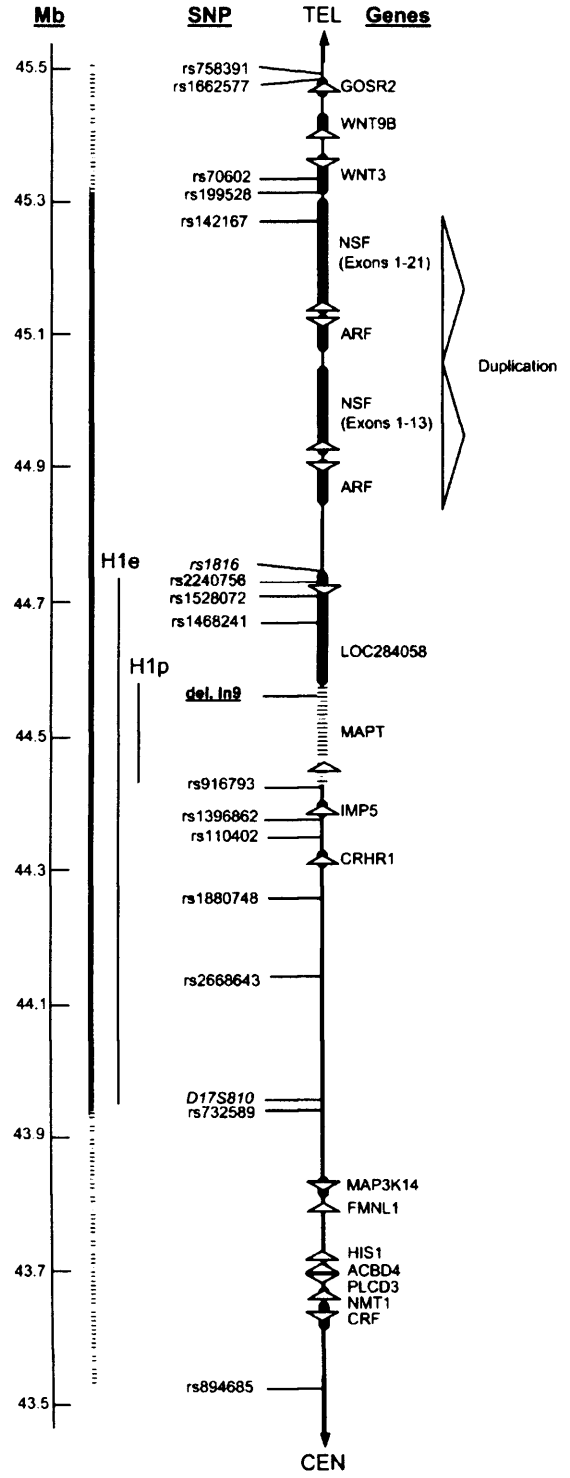

Figure 3. Region of chromosome 17q21.31 containing the extended *MAPT* haplotype block. The chromosomal co-ordinates (Mb; million base pairs) are indicated on the left-hand axis. They are based on the July 2003 draft of the human genome sequence (http://genome.ucsc.edu). Relative positions of SNPs and 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

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.

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	F1†	F2†	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.

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

ID*	Haplotype†																									Frequency (%)	
																										EM‡	Rs
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	1	0	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 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	G	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	A	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).[20] 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[20] 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%.

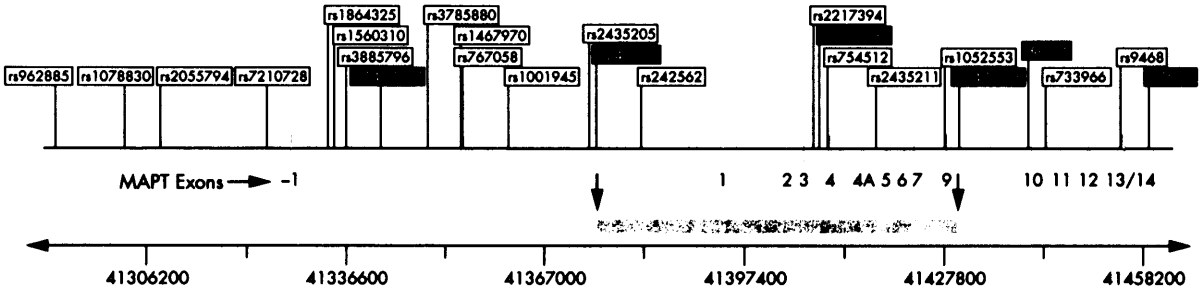

Figure 3 Genomic organisation of *MAPT* and distribution of the 25 markers used in the linkage disequilibrium and haplotype analysis in the CEPH trios. Relative positions of the promoter, coding exons, and genetic markers in the HapMap CEPH trios are to scale. Chromosome coordinates (base pairs) are 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.[32 33] In a more recent study,[34] 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]

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.

Table 1 Clinical features of affected individuals with the 17q21 microdeletion

	Individual 1	Individual 2	Individual 3
Current age	20 years	13 years	3 years
Gender	M	F	F
Gestation	Term	38 weeks	38 weeks
Birth weight (centile)	2.72 kg (2^{nd}–9th)	2.35 kg (0.4^{th}–2^{nd})	2.56 kg (0.4^{th}–2^{nd})
Perinatal	Hypotonic postnatally	Hypotonic postnatally. Extended breech presentation. Bilateral subluxation of hips	Hypotonic postnatally
Feeding and speech	Poor feeding; required nasogastric tube feeding from 4 weeks of age Oral dyspraxia	Slow feeding with weak suck Oral dyspraxia	Slow feeding with poor suck Excessive chewing/mouthing
Age at sitting	16 months	8 months	8 months
Age at walking	24 months	21 months	19 months
Learning disability	Special school for children with moderate to severe learning disability	Special school for children with moderate learning disability	Motor and speech delay
Behavior	Unremarkable	Unremarkable	Frequent laughing
Head circumference (percentile)	25^{th}–50^{th} at 17 years	9^{th} –25^{th} at 13 years	25^{th}–50^{th} at 2 years
Height (percentile)	$<0.4^{th}$ at 17 years	9^{th}–25^{th} at 13 years	75^{th}–90^{th} at 2 years
Weight (percentile)	0.4^{th}–2^{nd} at 17 years	9^{th}–25^{th} at 13 years	50^{th}–75^{th} at 2 years
Dysmorphic craniofacial features	Long face, deep-set eyes, narrow nose with bulbous tip. High palate	Long face. Submucous cleft palate and tongue tie repair	Wide mouth, short philtrum
Other findings	Pectus excavatum, two café au lait patches, mild contractures of elbows and knees; reduced visual acuity in each eye with static mild pigmentary macular epithelial changes; no cataract. Bilateral undescended testes. Petit mal seizures in early childhood	Pectus excavatum, small joint hypermobility, bilateral cataract extraction and lens implantation at 10 years. Mild bilateral hydronephrosis Mild bilateral hearing loss Deep sacral dimple	Single clonic seizure at 13 months, congenital dislocation of left hip, scoliosis, strabismus, hypermetropia

(**Supplementary Table 1** online). The centromeric breakpoint was within clone RP11-707O23, which gave a diminished signal on one chromosome 17 homolog (normal flanking clone RP11-798G7; deleted flanking clone CTD-3070M1). The telomeric breakpoint was between clone RP5-843B9 (no signal on one chromosome 17 homolog) and clone RP11-259 G18 (normal signals on both chromosome 17 homologs). We subsequently undertook FISH studies with clones from the RP11 library that had previously been shown to be specific to the H1 and H2 haplotypes[5] (**Fig. 2**; details of the FISH probes used in this analysis are given in **Supplementary Table 2** online, and examples of FISH images are given in **Supplementary Fig. 1** online) The H1 and H2 clones on either side of the deleted region showed signals in 17q24 as well as 17q21.3, consistent with the known presence of segmental duplications within this region[6] (**Supplementary Table 1**) and with the LCRs indicated in **Figure 2**.

The previously described 900-kb inversion[5] provides an explanation for the absence of recombination in this region and for the occurrence of two ancestral haplotypes, which have been named H1 and H2. We built on existing data[5,6,12] in order to construct a detailed assembly of the H2 haplotype in relation to a reference H1 haplotype (University of California Santa Cruz (UCSC) Genome Browser, May 2004 assembly), depicting SNP marker and LCR order and orientation. We used the available sequence and positioning of the available H2 BAC clones relative to H1 for annotation of the architectural features (**Fig. 2**). The differences between the H1 and H2 assemblies are far from that of a simple inversion: there are also substantial differences between H1 and H2 in both the number and the orientation of segmental duplications. For example, LCR subunit 4 lies in inverted orientation on the H1 haplotype, but it has undergone an inversion event on the H2 haplotype such that the LCRs now lie in direct orientation on this background. SNP marker position and orientation differ on the H1 and H2 haplotypes. The H1 centromeric markers rs2696425 and rs2668643 are duplicated on both H1 and H2, both flanking the deletion. Some SNP markers are present at one copy on H1 (rs1528072 and rs2532418) but are duplicated on H2, and others are present at one copy in both haplotypes (rs241041 and rs199528). Markers that fall within the confines of the inversion are in reverse order on H2 relative to H1 (rs1396862, rs916793, H1/H2 238-bp insertion/deletion polymorphism[13] and rs1468241).

We performed detailed haplotyping with respect to H1 and H2 alleles for each member of each trio (**Fig. 3**). In trio 1, the father is an H1/H2 heterozygote, and the mother an H2/H2 homozygote. The parental origin of the *MAPT* deletion is the H1/H2 father. Of particular interest, the chromosome carrying the deletion in trio 1 is composed of both H1 and H2 alleles, suggestive of an H1/H2 recombination event. To our knowledge, this is the first documented example of recombination between H1 and H2 haplotypes. In trio 2, the father is an H2/H2 homozygote and the mother an H1/H2 heterozygote. Genotypes of all markers in the affected individual are consistent with H2. We were unable to determine the parental origin of the deletion in this trio despite typing a panel of repeat markers across the deleted region. In trio 3, the father is an H1/H2 heterozygote and the mother an H1/H1 homozygote. Haplotype analysis excluded the maternal origin of the deletion. On the deleted chromosome, both the telomeric and centromeric markers are of type H2, as seen in trio 2.

We noted that variations in genomic architecture associated with the H1 and H2 haplotypes at 17q21.3 present possibilities for the

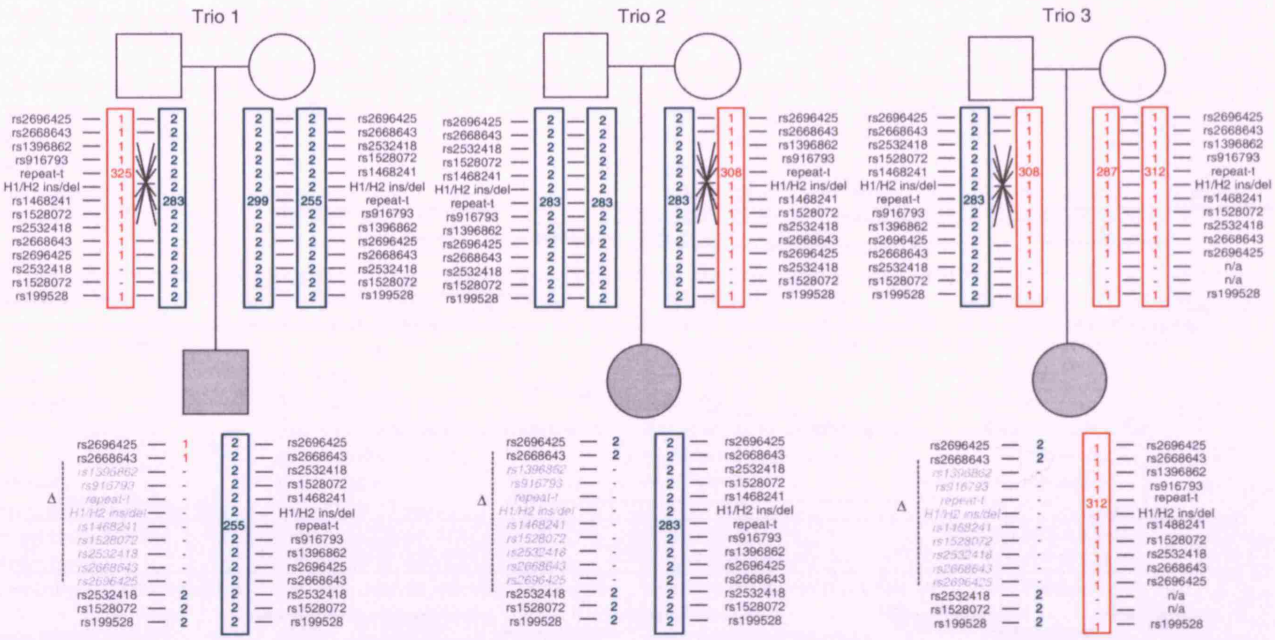

Figure 3 Extended *MAPT* haplotypes in the three trios. The figure is a depiction of the genotyping data in **Supplementary Table 4** online and the FISH data in **Supplementary Table 1**. The phased parental haplotypes (boxed) have been constructed in accordance with strict H1/H2 ruling from the SNP marker genotypes, and the deletion haplotypes of the affected offspring have been inferred (unboxed) from the genotype data in **Supplementary Table 2**. Markers in faint italics cannot reliably be used to infer haplotypes of the deleted chromosome because they lie within the confines of the deleted region. Dotted line (Δ) corresponds to the deleted region established by FISH. Included in the haplotype analysis are genotypes of the size of a mixed tetranucleotide repeat (repeat-t) situated in *MAPT*. Repeat-t distinguishes between the parental chromosomes in trio 1 and trio 3 and excludes a maternal origin.

microdeletions are recognized to make a greater contribution than mutations in the *NSD1* gene to the pathogenesis of Sotos syndrome in Japanese relative to Western cohorts[19], it might be of interest, by analogy with the present study, to examine the world-wide population incidence of the genomic inversion in the Sotos syndrome region and its impact on the ratio of cases with a microdeletion versus those with an *NSD1* mutation.

The deletion encompasses the gene encoding microtubule-associated protein tau (*MAPT*), and, to the best of our knowledge, these cases represent the first reports of haploinsufficiency of *MAPT* in humans. Gain-of-function mutations in *MAPT* cause frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17)[20,21], and common genetic variation of *MAPT* is associated with progressive supranuclear palsy[22] and corticobasal degeneration and Alzheimer's disease[23]. The possible role of *MAPT* in development has been studied in mice lacking the *tau* gene ($tau^{-/-}$)[24]. These mice are phenotypically and developmentally normal. The sole identified histological abnormality is a reduction in microtubule density and stability in some small-caliber axons. More detailed phenotypic analysis of these mice has subsequently been reported[25]. Muscle weakness in the wire-hanging test, hyperactivity in a novel environment and impairment of contextual fear conditioning were found, suggesting that a possible role for *MAPT* haploinsufficiency in learning disability in humans merits further investigation. The deleted region contains other genes: corticotrophin releasing hormone receptor 1 (*CRHR1*), intramembrane protease 5 (*NM_175882*, also known as *IMP5*) and the predicted genes *NP_689679.1*, *NP_787078.1* and *KIAA1267*. Of these, there is some evidence for a role for *CHRH1* in central nervous system development[26], although no data are available on the possible impact of haploinsufficiency of this gene on that process.

Microscopic chromosomal deletions involving 17q21.3 have been reported before[27], but these have been much larger than the submicroscopic deletion reported here, and the clinical phenotypes have been quite different, with major congenital malformations such as esophageal atresia and congenital heart defect. The DECIPHER database (http://decipher.sanger.ac.uk) was set up in order to facilitate and expedite the identification of new disorders in clinical cytogenetics and has proved instrumental in identifying the overlapping genotypic and phenotypic features of the three cases reported in the present study. Low birth weight (0.4th–9th centile), neonatal hypotonia, poor feeding in infancy and oromotor dyspraxia together with mental retardation seem to be common features of deletion of 17q21.3. When occurring together, these features should prompt consideration of this diagnosis. Further cases are needed in order to define the full phenotype of this disorder and its evolution with age. For genetic counseling, the presence of an inversion in the transmitting parent might be considered a risk factor for deletion in the offspring. This inversion may prove to be a necessary factor for deletion to occur, but it is unlikely to be sufficient, given the relatively high frequency of the inversion in the populations in which it occurs compared with the much lower frequency of the deletion. A prediction from the present study is that if H1/H2 heterozygotes and possibly H2/H2 homozygotes are predisposed to undergo deletions at 17q21.3, then the incidence of these deletions should be broadly restricted to those populations in which the H2 haplotype occurs, namely those of Iceland, Europe and the Middle East[5,28]. Again, investigation of this possibility will require analysis of additional cases. Furthermore, a follow-up study in which the parents of individuals with this new microdeletion syndrome are genotyped with respect to H1/H2 would clearly be worthwhile.

23. Myers, A.J. *et al.* The H1c haplotype at the *MAPT* locus is associated with Alzheimer's disease. *Hum. Mol. Genet.* **14**, 2399–2404 (2005).
24. Harada, A. *et al.* Altered microtubule organization in small-calibre axons of mice lacking tau protein. *Nature* **369**, 488–491 (1994).
25. Ikegami, S., Harada, A. & Hirokawa, N. Muscle weakness, hyperactivity and impairment in fear conditioning in tau-deficient mice. *Neurosci. Lett.* **279**, 129–132 (2000).
26. Bishop, G.A. & King, J.S. Corticotropin releasing factor in the embryonic mouse cerebellum. *Exp. Neurol.* **160**, 489–499 (1999).
27. Khalifa, M.M., MacLeod, P.M. & Duncan, A.M. Additional case of *de novo* interstitial deletion del(17)(q21.3q23) and expansion of the phenotype. *Clin. Genet.* **44**, 258–261 (1993).
28. Evans, W. *et al.* The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci. Lett.* **369**, 183–185 (2004).
29. Pittman, A.M. *et al.* The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum. Mol. Genet.* **13**, 1267–1274 (2004).

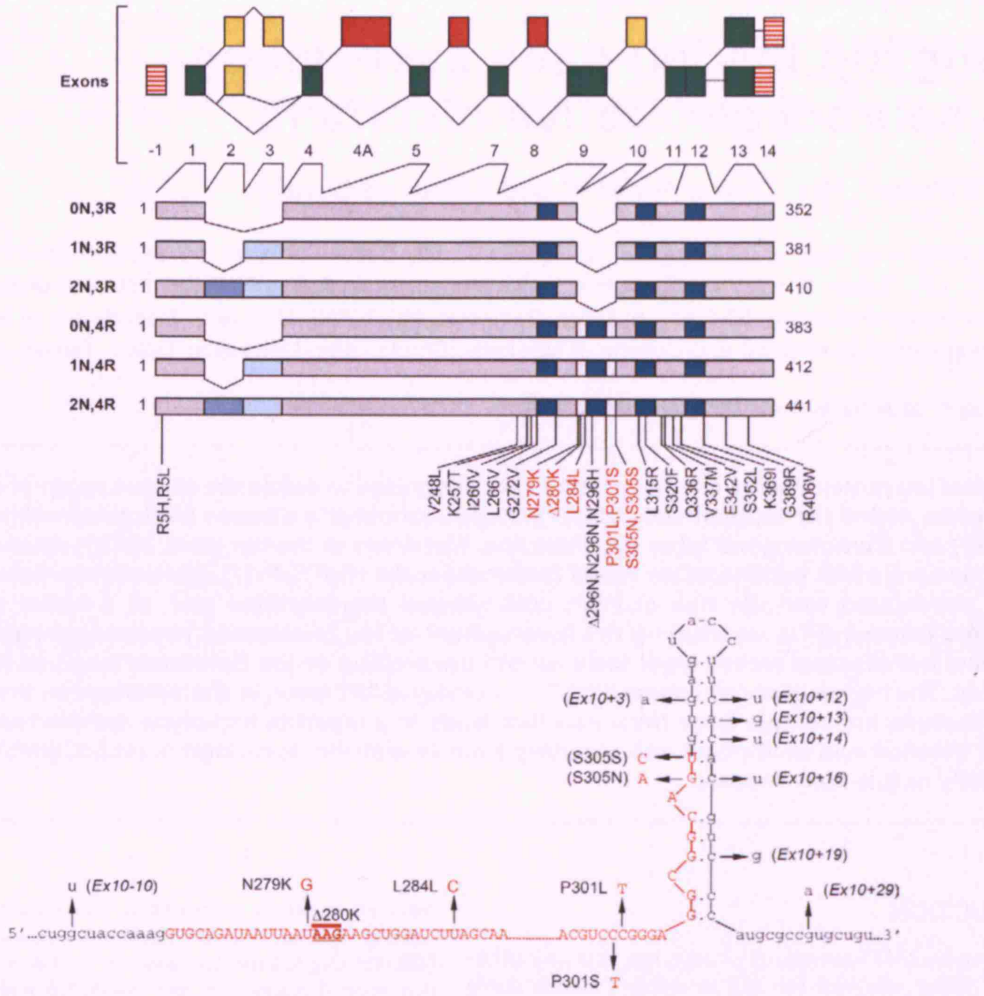

Figure 1. (Top) Tau in the central nervous system (CNS) exists as six isoforms due to the alternative splicing of exons 2, 3 and 10 (yellow boxes). Exons 4A, 7 and 8 (red boxes) are absent in the CNS and exon 4A is included in peripheral nervous system tau. Exons 2 and 3 code for N-terminal inserts, alternative splicing leads to tau isoforms with 2, 1 or no N-terminal inserts (2N, 1N or 0N). Exon 10 codes for one of four MT binding domains—alternative splicing results in tau with three or four MT binding repeat domains (3R, 4R). FTDP-17 missense and silent mutations and deletions are indicated with numbering relating to the longest 441 residue 2N,4R isoform. Mutations in red affect the alternative splicing of exon 10. Proportions are not to scale. (Bottom) FTDP-17 mutations affecting the splicing of exon 10. The majority of these mutations disrupt a predicted pre-mRNA stem-loop structure, inducing increased incorporation of exon 10. Partial sequence of exon 10 in red. Intronic sequence in black. Proportions are not to scale. Modified from Goedert (21).

tau dysfunction as central in the aetiopathogenesis of these disorders. The mutations affecting *MAPT* in FTDP-17 can be divided into two classes, the missense and the exon 10 splicing mutations (Fig. 1) (recently reviewed in 20–22). The chief effect of the former is in the biochemistry of the tau protein, where in most cases, the MT-binding capacity is reduced, leading to a greater abundance of unbound tau species that could lead to aggregation and the formation of the insoluble tau inclusions that characterize these disorders. The exon 10 splicing mutations in the main cause increased incorporation of exon 10, thereby increasing the relative levels of 4R-tau isoforms. Although classified under the rubric of FTDP-17, these disorders have diverse clinical phenotypes including movement disorders such as PSP and CBD, memory dysfunction similar to that found in AD and typical clinical pictures observed in PiD/frontotemporal lobar degeneration (23). In fact, there are reports of the same mutation, P301L causing either CBD or FTDP in the same family (24). It is therefore clear that the manifestations of each mutation are subject to potential epigenetic and perhaps environmental influences and the question of genotype–phenotype correlation is complex (25,26).

In addition to FTDP-17, another group of familial frontotemporal dementias (FTDs) with intranuclear ubiquitin inclusions, but lacking any tau pathology, was linked to the same region of chromosome 17q21 (FTDU-17) (27–29). This region is defined by the markers D17S1787–D17S806 and contains *MAPT* (27). Extensive sequencing of *MAPT*

Table 1. Association studies of the *MAPT* locus and identified risk alleles in PSP

Study	Risk genotype and allele frequency (%)						
	PSP			Controls			OR
Tau intron 9 dinucleotide repeat	*n*	a0/a0	a0	*n*	a0/a0	a0	
Conrad *et al.* (33)	22	95.5	97.7	61	57.4	74.6	15.6
Conrad *et al.* (82)	31	100	100	67	97	98	n.a
Oliva *et al.* (39)	30	86.7	93.3	50	52	73	6.5
Morris *et al.* (38)	53	84.3	90.6	75	53	72.7	n.a
Baker *et al.* (40)	64	79.7	89.8	139	51.1	70.5	n.a
Bennett *et al.* (36)	30	86.7	93.3	36	61.1	76.4	4.1
Higgins *et al.* (83)	24	83.3	91.6	86	53.4	75.6	4.4
H1 haplotype		H1/H1	H1		H1/H1	H1	
Baker *et al.* (40)	64	87.5	93.7	145	62.8	78.4	4.1
Morris *et al.* (84)	50	98	99	71	63	79	n.a
Tau promoter haplotypes		C/C	C		C/C	C	
Ezquerra *et al.* (85)	35	91.4	94.2	195	49.7	72.5	11.8
		H1p/H1p	H1p		H1p/H1p	H1p	
de Silva *et al.* (71)	42	97.6	98.8	70	61.4	78.6	n.a
5′ tau haplotype		Hap A/A	Hap A		Hap A/A	Hap A	
Higgins *et al.* (86)	52	98	99	54	48	77.7	n.a
Extended 17q21 haplotypes		H1E/H1E	H1E		H1E/H1E	H1E	
Pastor *et al.* (62)	45	84.4	87.7	45	33.3	51.1	10.9
Pittman *et al.* (42)	60	88	93	63	62	80	4.7
H1c *MAPT* haplotype		H1c/H1c	H1c		H1c/H1c	H1c	
Pittman *et al.* (54), US	238	n/a	24	131	n.a	7.8	n.a
Pittman *et al.* (54), UK	83	n/a	24.3	169	n.a	11.3	n.a
H1-specific SNP rs242557		A/A	A		A/A	A	
Rademakers *et al.* (55)	244	n/a	56.7	396	n.a	45.1	1.6

All studies found a significant association of H1 haplotype in populations from Western Europe, Spain and North America. n.a, data not available or not applicable; *n*, number of cases–controls in study; OR, odds ratio for the risk genotype.

more similar to each other than the chimpanzee sequence, they do not follow a predictable relationship; at some sequences, the chimpanzee sequence is similar to H1, and at others, it is similar to H2 (56). Thus, the H1 and H2 sequences do not follow a precursor–product relationship and one cannot be derived from the other, rather both must have been derived independently from a more distant precursor predating the inversion (30,49). Although the estimated divergence of the H1 and H2 clades is ~3 million years ago, it is estimated that the H2 haplotype was re-introduced into the European population only about 10 000–30 000 years ago, overlapping with a period when *Homo sapiens* and *Homo neanderthalensis* co-existed in Europe. We have therefore suggested that the invariant H2 haplotype could have been derived from the neanderthals (49) with subsequent positive selection as shown in the Icelandic population (50). However, it is also possible that H2 represents a rare and early chromosomal change in Africa that rapidly expanded in European populations from a few founder chromosomes (50).

MAPT HAPLOTYPE DIVERSITY AND DISEASE

Although by definition all genes within the region of LD encompassing the *MAPT* locus are also associated with PSP, the hallmark tau pathology of this disorder strongly implicates *MAPT* itself. Several recent studies have analysed the population-wide haplotypic diversity of *MAPT* for case–control comparisons (30,54,55). This is with the ultimate goal of identifying the underlying mechanistic basis of the genetic association.

We utilized validated high-density genotype data from the International HapMap Project, which provides the useful basis for analysing this diversity in a representative Caucasian population (CEPH-Utah collection). A single H2 haplotype and multiple H1 variant haplotypes were identified (54). Using a minimal number of haplotype-tagging SNPs and the bi-allelic *del-In9* marker for H1/H2 assignment (40,54), we determined that only one of the common (frequency >0.1) sub-haplotypes in the H1 clade, namely, H1c to be highly

and exon 10-containing *MAPT* mRNA transcript levels are significantly higher in H1c when compared with the other H1 and H2 haplotypes (70). A possible basis for the increased expression is the H1-specific SNP, rs242557, which is located within intron 0 of the gene in an evolutionarily conserved island (>75% conservation between mouse and human), about 47 kb downstream from the *MAPT* core promoter at exon −1. Cellular reporter assays showed that this conserved domain is a putative control element for the *MAPT* core promoter and that the allelic variants of rs242557 alter this capacity (55). Cellular assays for transcriptional activity now show that the rs242557 allele A (as found in haplotype H1c) in conjunction with both the H1p and H2p variants of the *MAPT* core promoter (exon −1) (71,72) has significantly higher (2–4-fold, respectively) expression than the allele G (73).

CONCLUSION

The *MAPT* gene is an interesting paradigm in the study of the genetic causes of human disease. Like amyloid pathology in AD and the subsequent identification of its originating protein and gene and mutations that cause the autosomal dominant familial forms of AD (74), tau pathology consisting of insoluble, fibrillar aggregates of tau has for long been the pathological hallmark of the tauopathies. The recent identification of *MAPT* mutations established the central importance of tau dysfunction in the aetiology of these disorders. However, like in AD and PD, the large majority of the cases are sporadic, caused by a combination of genetic and environmental risk factors and ageing. A common theme in many of the neurodegenerative disorders is the abnormal deposition of insoluble proteins, coded by the very genes that are mutated in a small number of the familial cases of these disorders. In the absence of protein defects caused by mutations, there are several other mechanisms that could lead to accumulation, misfolding and/or mislocalization of these proteins and the production of toxic aggregation intermediates. These could include oxidative damage, abnormal post-translational modifications (hyperphosphorylation, glycosylation, sumoylation, etc.) and defective ubiquitin–proteasome degradation. However, recent findings also implicate changes in expression levels of normal proteins caused by common genetic variability (75). For example, genetic variation in α-synuclein associated with PD modulates the disease risk by influencing protein expression (76). The over-production of protein as a result of gene multiplications such as α-synuclein in PD (77–79) and APP in AD (80) causes the disease. The *MAPT* haplotype association with sporadic tauopathies provides another clear example of a gene identified as a Mendelian pathogenic locus (in familial FTDP-17), contributing to risk for sporadic tauopathies by influencing either the expression of the protein or alternative splicing (75). This would be a potentially important therapeutic target for developing strategies to selectively reduce expression levels of pathogenic proteins.

ACKNOWLEDGEMENTS

This work was supported by the Reta Lila Weston Trust for Medical Research (A.M.P. and R.d.S.) and the PSP (Europe) Association and Society for PSP, grants from the Bogue Fellowship of UCL and the Chang Gung Memorial Hospital Biomedical Scholarship to H.C.F. R.d.S. is funded by a research grant from the Medical Research Council (MRC), UK.

Conflict of Interest statement. None of the authors have declared any conflict of interest.

REFERENCES

1. Witman, G.B., Cleveland, D.W., Weingarten, M.D. and Kirschner, M.W. (1976) Tubulin requires tau for growth onto microtubule initiating sites. *Proc. Natl Acad. Sci. USA*, **73**, 4070–4074.
2. Binder, L.I., Frankfurter, A. and Rebhun, L.I. (1985) The distribution of tau in the mammalian central nervous system. *J. Cell Biol.*, **101**, 1371–1378.
3. Gu, Y., Oyama, F. and Ihara, Y. (1996) Tau is widely expressed in rat tissues. *J. Neurochem.*, **67**, 1235–1244.
4. LoPresti, P., Szuchet, S., Papasozomenos, S.C., Zinkowski, R.P. and Binder, L.I. (1995) Functional implications for the microtubule-associated protein tau: localization in oligodendrocytes. *Proc. Natl Acad. Sci. USA*, **92**, 10369–10373.
5. Avila, J. (2006) Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett.*, **580**, 2922–2927.
6. Kanemaru, K., Takio, K., Miura, R., Titani, K. and Ihara, Y. (1992) Fetal-type phosphorylation of the tau in paired helical filaments. *J. Neurochem.*, **58**, 1667–1675.
7. Andreadis, A., Brown, W.M. and Kosik, K.S. (1992) Structure and novel exons of the human tau gene. *Biochemistry*, **31**, 10626–10633.
8. Goedert, M., Spillantini, M.G., Potier, M.C., Ulrich, J. and Crowther, R.A. (1989) Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J.*, **8**, 393–399.
9. Goedert, M., Wischik, C.M., Crowther, R.A., Walker, J.E. and Klug, A. (1988) Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau. *Proc. Natl Acad. Sci. USA*, **85**, 4051–4055.
10. Goedert, M., Spillantini, M.G., Jakes, R., Rutherford, D. and Crowther, R.A. (1989) Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron*, **3**, 519–526.
11. Delacourte, A., Sergeant, N., Wattez, A., Gauvreau, D. and Robitaille, Y. (1998) Vulnerable neuronal subsets in Alzheimer's and Pick's disease are distinguished by their tau isoform distribution and phosphorylation. *Ann. Neurol.*, **43**, 193–204.
12. de Silva, R., Lashley, T., Strand, C., Shiarli, A.M., Shi, J., Tian, J., Bailey, K.L., Davies, P., Bigio, E.H., Arima, K. *et al.* (2006) An immunohistochemical study of cases of sporadic and inherited frontotemporal lobar degeneration using 3R- and 4R-specific tau monoclonal antibodies. *Acta Neuropathol. (Berl.)*, **111**, 329–340.
13. Arai, T., Ikeda, K., Akiyama, H., Shikamoto, Y., Tsuchiya, K., Yagishita, S., Beach, T., Rogers, J., Schwab, C. and McGeer, P.L. (2001) Distinct isoforms of tau aggregated in neurons and glial cells in brains of patients with Pick's disease, corticobasal degeneration and progressive supranuclear palsy. *Acta Neuropathol. (Berl.)*, **101**, 167–173.
14. Sergeant, N., Wattez, A. and Delacourte, A. (1999) Neurofibrillary degeneration in progressive supranuclear palsy and corticobasal degeneration: tau pathologies with exclusively 'exon 10' isoforms. *J. Neurochem.*, **72**, 1243–1249.
15. Togo, T., Sahara, N., Yen, S.H., Cookson, N., Ishizawa, T., Hutton, M., de Silva, R., Lees, A. and Dickson, D.W. (2002) Argyrophilic grain disease is a sporadic 4-repeat tauopathy. *J. Neuropathol. Exp. Neurol.*, **61**, 547–556.
16. de Silva, R., Lashley, T., Gibb, G., Hanger, D., Hope, A., Reid, A., Bandopadhyay, R., Utton, M., Strand, C., Jowett, T. *et al.* (2003) Pathological inclusion bodies in tauopathies contain distinct complements of tau with three or four microtubule-binding repeat domains as demonstrated by new specific monoclonal antibodies. *Neuropathol. Appl. Neurobiol.*, **29**, 288–302.

54. Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L. *et al.* (2005) Linkage disequilibrium fine-mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J. Med. Genet.*, **42**, 837–846.
55. Rademakers, R., Melquist, S., Cruts, M., Theuns, J., Del-Favero, J., Poorkaj, P., Baker, M., Sleegers, K., Crook, R., De Pooter, T. *et al.* (2005) High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. *Hum. Mol. Genet.*, **14**, 3281–3292.
56. Conrad, C., Vianna, C., Schultz, C., Thal, D.R., Ghebremedhin, E., Lenz, J., Braak, H. and Davies, P. (2004) Molecular evolution and genetics of the Saitohin gene and tau haplotype in Alzheimer's disease and argyrophilic grain disease. *J. Neurochem.*, **89**, 179–188.
57. Baker, M., Graff-Radford, D., Wavrant DeVrieze, F., Graff-Radford, N., Petersen, R.C., Kokmen, E., Boeve, B., Myllykangas, L., Polvikoski, T., Sulkava, R. *et al.* (2000) No association between TAU haplotype and Alzheimer's disease in population or clinic based series or in familial disease. *Neurosci. Lett.*, **285**, 147–149.
58. Crawford, F., Freeman, M., Town, T., Fallin, D., Gold, M., Duara, R. and Mullan, M. (1999) No genetic association between polymorphisms in the Tau gene and Alzheimer's disease in clinic or population based samples. *Neurosci. Lett.*, **266**, 193–196.
59. Green, E.K., Thaker, U., McDonagh, A.M., Iwatsubo, T., Lambert, J.C., Chartier-Harlin, M.C., Harris, J.M., Pickering-Brown, S.M., Lendon, C.L. and Mann, D.M. (2002) A polymorphism within intron 11 of the tau gene is not increased in frequency in patients with sporadic Alzheimer's disease, nor does it influence the extent of tau pathology in the brain. *Neurosci. Lett.*, **324**, 113–116.
60. Russ, C., Powell, J.F., Zhao, J., Baker, M., Hutton, M., Crawford, F., Mullan, M., Roks, G., Cruts, M. and Lovestone, S. (2001) The microtubule associated protein Tau gene and Alzheimer's disease—an association study and meta-analysis. *Neurosci. Lett.*, **314**, 92–96.
61. Myers, A.J., Kaleem, M., Marlowe, L., Pittman, A.M., Lees, A.J., Fung, H.C., Duckworth, J., Leung, D., Gibson, A., Morris, C.M. *et al.* (2005) The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Hum. Mol. Genet.*, **14**, 2399–2404.
62. Pastor, P., Ezquerra, M., Munoz, E., Marti, M.J., Blesa, R., Tolosa, E. and Oliva, R. (2000) Significant association between the tau gene A0 A0 genotype and Parkinson's disease. *Ann. Neurol.*, **47**, 242–245.
63. de Silva, R., Hardy, J., Crook, J., Khan, N., Graham, E.A., Morris, C.M., Wood, N.W. and Lees, A.J. (2002) The tau locus is not significantly associated with pathologically confirmed sporadic Parkinson's disease. *Neurosci. Lett.*, **330**, 201–203.
64. Farrer, M., Skipper, L., Berg, M., Bisceglio, G., Hanson, M., Hardy, J., Adam, A., Gwinn-Hardy, K. and Aasly, J. (2002) The tau H1 haplotype is associated with Parkinson's disease in the Norwegian population. *Neurosci. Lett.*, **322**, 83–86.
65. Hoenicka, J., Perez, M., Perez-Tur, J., Barabash, A., Godoy, M., Vidal, L., Astarloa, R., Avila, J., Nygaard, T. and de Yebenes, J.G. (1999) The tau gene A0 allele and progressive supranuclear palsy. *Neurology*, **53**, 1219–1225.
66. Maraganore, D.M., Hernandez, D.G., Singleton, A.B., Farrer, M.J., McDonnell, S.K., Hutton, M.L., Hardy, J.A. and Rocca, W.A. (2001) Case–control study of the extended tau gene haplotype in Parkinson's disease. *Ann. Neurol.*, **50**, 658–661.
67. Rademakers, R., Sleegers, K., Pals, P., Nuytemans, K., Lohman, E., Durr, A., Engelborghs, S., de Pooter, T., Van den Broeck, M., Pickut, B. *et al.* (2005) *MAPT* H1 subhaplotyping in European PD association samples implicates tau splicing in early-onset Parkinson's disease susceptibility. *Alzh. Dement.*, **2**, S420–S421.
68. Skipper, L., Wilkes, K., Toft, M., Baker, M., Lincoln, S., Hulihan, M., Ross, O.A., Hutton, M., Aasly, J. and Farrer, M. (2004) Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am. J. Hum. Genet.*, **75**, 669–677.
69. Poorkaj, P., Kas, A., D'Souza, I., Zhou, Y., Pham, Q., Stone, M., Olson, M.V. and Schellenberg, G.D. (2001) A genomic sequence analysis of the mouse and human microtubule-associated protein tau. *Mamm. Genome*, **12**, 700–712.
70. Myers, A.J., Pittman, A.M., Rohrer, K., Zhao, A., Leung, D., Bryden, L., Kaleem, M., Marlowe, L., Fung, H.C., Lees, A.J. *et al.* (2006) The H1c risk haplotype of the MAPT gene is over-expressed in human temporal cortex relative to other common alleles of MAPT. *Alzh. Dement.*, **2**, S32.
71. 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.
72. Andreadis, A., Wagner, B.K., Broderick, J.A. and Kosik, K.S. (1996) A tau promoter region without neuronal specificity. *J. Neurochem.*, **66**, 2257–2263.
73. Pittman, A., Myers, A., Hardy, J., Wood, N., Lees, A. and de Silva, R. (2006) Increased MAPT expression is associated with progressive supranuclear palsy. *Alzh. Dement.*, **2**, S421.
74. Hardy, J. and Selkoe, D.J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **297**, 353–356.
75. Singleton, A., Myers, A. and Hardy, J. (2004) The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases. *Hum. Mol. Genet.*, **13** (Spec no. 1), R123–R126.
76. Chiba-Falek, O. and Nussbaum, R.L. (2001) Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. *Hum. Mol. Genet.*, **10**, 3101–3109.
77. Chartier-Harlin, M.C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M. *et al.* (2004) Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet*, **364**, 1167–1169.
78. Ibanez, P., Bonnet, A.M., Debarges, B., Lohmann, E., Tison, F., Pollak, P., Agid, Y., Durr, A. and Brice, A. (2004) Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet*, **364**, 1169–1171.
79. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R. *et al.* (2003) Alpha-synuclein locus triplication causes Parkinson's disease. *Science*, **302**, 841.
80. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M. *et al.* (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.*, **38**, 24–26.
81. Shaw-Smith, C., Redon, R., Rickman, L., Rio, M., Willatt, L., Fiegler, H., Firth, H., Sanlaville, D., Winter, R., Colleaux, L. *et al.* (2004) Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J. Med. Genet.*, **41**, 241–248.
82. Conrad, C., Amano, N., Andreadis, A., Xia, Y., Namekata, K., Oyama, F., Ikeda, K., Wakabayashi, K., Takahashi, H., Thal, L.J. *et al.* (1998) Differences in a dinucleotide repeat polymorphism in the tau gene between Caucasian and Japanese populations: implication for progressive supranuclear palsy. *Neurosci. Lett.*, **250**, 135–137.
83. Higgins, J.J., Litvan, I., Pho, L.T., Li, W. and Nee, L.E. (1998) Progressive supranuclear gaze palsy is in linkage disequilibrium with the tau and not the alpha-synuclein gene. *Neurology*, **50**, 270–273.
84. Morris, H.R., Vaughan, J.R., Datta, S.R., Bandopadhyay, R., de Silva, R., Schrag, A., Cairns, N.J., Burn, D., Nath, U., Lantos, P.L. *et al.* (2000) Multiple system atrophy/progressive supranuclear palsy: alpha-Synuclein, synphilin, tau, and APOE. *Neurology*, **55**, 1918–1920.
85. 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.
86. Higgins, J.J., Adler, R.L. and Loveless, J.M. (1999) Mutational analysis of the tau gene in progressive supranuclear palsy. *Neurology*, **53**, 1421–1424.