Genetic Characterisation of Neurodegenerative disorders

Thesis submitted in fulfillment of the degree
of Doctor of Philosophy

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

October 2007

Hon Chung, Fung

I, Hon Chung, Fung, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

Firstly I would like to express my gratitude to all the patients, healthy controls and their families for their understanding and their participation in this research, without their generous support we would not have been to make progress in the research of neurodegenerative diseases.

I give my deepest thanks to my supervisor and director of my Institute, Andrew Lees, for providing an inspiring and enjoyable environment for my research, and my principal supervisor Rohan de Silva at the Reta Lila Weston Institute of Neurological Studies, UCL for his encouragement and for providing an exciting research project. I must also express my special thanks to my supervisor and friend, John Hardy, for giving me the confidence and support to begin my doctoral program. I also give thanks to Alan Pittman, Andrew Singleton and Amanda Myers for help with genetics throughout this journey.

I am grateful to all the staff and lab colleagues at the Reta Lila Weston Institute and Laboratory of Neurogenetics, NIA at the National Institutes of Health for guidance and assistance: Rina Bandopadhyay, Yvonne Mwelwa, Joan Ward, Kimberly Singleton, Katrina Gwinn-Hardy, Fabienne Wavrant-DeVrieze, Jaime Duckworth, Sampath Arepalli, Whitney Evans, Dena Hernandez, Sonja Scholz, Mar Matarin, and Bryan Traynor. Thanks are given to my colleagues of Department of Neurology, Chang Gung Memorial Hospital: Wen-Chuin Hsu, Sien-Tsong Chen, Chuin-Mei Chen, Rong-Kuo Lyu, Hong-Shiu Chang, Long-Sun Ro, Lok-Ming Tang, Tsong-Hai Lee. In particular I would like to thank Yih-Ru Wu for her great support and encouragement. I also express my thanks to Guey-Jen Lee-Chen for her support in the early days of my research.

I would also like to gratefully acknowledge the support of some very special individuals, Roman Cheung, Matthew Chu, Stephen Chow and Denise Chan. They helped me immensely by giving me encouragement and friendship.

Thanks are given to the Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taiwan and the UCL Bogue fellowship for financial support.

Finally, I express my warmest gratitude to my Dad and Mom, my family Thomas and Sarah. Your support and prayer is the key of my success.

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

The cases and control samples from Guam in this study were obtained from Micronesian Health Study II, 515 Alupang Cove, Tamuning, Guam 96913.

The control samples from Finland were obtained from Department of Neurology, University of Helsinki, Biomedicum-Helsinki, Haarmaninkatu 8, FIN-02900, Helsinki, Finland with thanks to Dr. Johanna Eerola.

Chapter 4

The pathologically confirmed PSP cases from the UK in this study were all obtained from the Sara Koe PSP Research Centre, 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.

Chapter 5

The Taiwanese PD cases and controls samples were from the Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taiwan with thanks to Drs. Yih-Ru, Wu and Chiung-Mei, Chen.

The Finnish PD cases and control were obtained from Department of Neurology, Helsinki University Central Hospital and University of Helsinki, Biomedicum-Helsinki, Haarmaninkatu 8, FIN-02900, Helsinki, Finland and Department of Neurology, Seinäjoki Central Hospital, Seinäjoki.

The Greek PD case and control samples were obtained from Neurogenetic Unit, Department of Neurology, Medical School, University of Thessaly, Larissa, Greece with special thanks to Dr. Georgia Xiromerisiou.

Chapter 6

The pathologically confirmed control samples from the US were obtained from the Laboratory of Neurogenetics NIA, NIH, Bethesda from Dr. Amanda Myers. Many data and biomaterials were collected from several NIA-NACC funded institutes with thanks to Drs. Marcelle Morrison-Bogorad, Tony Phelps and Walter Kukull for helping to coordinate this collection.

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 Taiwanese AD cases and controls samples were from the Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taiwan with thanks to Drs Wen-Chuin Hsu and Yih-Ru Wu for collecting the samples, and special thanks to Dr. Guey-Jen Lee-Chen from Department of Life Science, National Taiwan Normal University for preparation of the biomaterial samples.

Publications

Published papers either as a direct result from (Appendix 10.4) or through collaborative work during this thesis:

Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, Britton A, Gibbs JR, Langefeld C, Stiegert ML, Schymick J, Okun MS, Mandel RJ, Fernandez HH, Foote KD, Rodriguez RL, Peckham E, De Vrieze FW, Gwinn-Hardy K, Hardy JA, Singleton A. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol. 2006 Nov;5(11):911-6.

Fung HC, Xiromerisiou G, Gibbs JR, Wu YR, Eerola J, Gourbali V, Hellstrom O, Chen CM, Duckworth J, Papadimitriou A, Tienari PJ, Hadjigeorgiou GM, Hardy J, Singleton AB. Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts. Neurodegener Dis. 2006;3(6):327-33.

Fung HC, Evans J, Evans W, Duckworth J, Pittman A, de Silva R, Myers A, Hardy J. The architecture of the tau haplotype block in different ethnicities. Neurosci Lett. 2005 Mar 29;377(2):81-4.

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 Oct 21;369(3):183-5.

Pittman AM, **Fung HC**, de Silva R. Untangling the tau gene association with neurodegenerative disorders. Hum Mol Genet. 2006 Oct 15;15 Special No 2: R188-95.

Simon-Sanchez J, Scholz S, **Fung HC**, Matarin M, Hernandez D, Gibbs JR, Britton A, de Vrieze FW, Peckham E, Gwinn-Hardy K, Crawley A, Keen JC, Nash J, Borgaonkar D, Hardy J, Singleton A. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet. 2007 Jan 1;16(1):1-14.

Scholz SW, Xiromerisiou G, **Fung HC**, Eerola J, Hellstrom O, Papadimitriou A, Hadjigeorgiou GM, Tienari PJ, Fernandez HH, Mandel R, Okun MS, Gwinn-Hardy K, Singleton AB. The human prion gene M129V polymorphism is not associated with idiopathic Parkinson's disease in three distinct populations. Neurosci Lett. 2006 Mar 13;395(3):227-9.

Hardy J, Pittman A, Myers A, **Fung HC**, de Silva R, Duckworth J. Tangle diseases and the tau haplotypes. Alzheimer Dis Assoc Disord. 2006 Jan-Mar;20(1):60-2.

Pittman AM, Myers AJ, Abou-Sleiman P, **Fung HC**, Kaleem M, Marlowe L, Duckworth J, Leung D, Williams D, Kilford L, Thomas N, Morris CM, Dickson D, Wood NW, Hardy J, Lees AJ, de Silva R. Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. J Med Genet. 2005 Nov;42(11):837-46.

Hardy J, Pittman A, Myers A, Gwinn-Hardy K, **Fung HC**, de Silva R, Hutton M, Duckworth J. Evidence suggesting that *Homo neanderthalensis* contributed the H2 MAPT haplotype to *Homo sapiens*. Biochem Soc Trans. 2005 Aug;33(Pt 4):582-5.

Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, **Fung HC**, Duckworth J, Leung D, Gibson A, Morris CM, de Silva R, Hardy J. The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. Hum Mol Genet. 2005 Aug 15;14(16):2399-404.

Abbreviations

Aβ	β-amyloid
AD	Alzheimer’s disease
AGD	argyrophilic grain disease
ALS	amyotrophic lateral sclerosis
CEPH	Centre d’Etude du Polymorphisme Humain
CI	confidence interval
del-in9	238 bp haplotyping-defining insertion/deletion polymorphism
df	degree of freedom
EM	expectation-maximization
EOFAD	early-onset familial AD
FTD	frontotemporal dementia
FTDP-17	FTD with parkinsonism linked to chromosome 17
HapMap	The International Haplotype Map Project
HD	Huntington’s disease
htSNPs	haplotype tagging single nucleotide polymorphisms
HWE	Hardy-Weinberg equilibrium
LCR	low-copy repeat
LD	linkage disequilibrium
LOAD	late-onset AD
LRT	likelihood ratio test
MT	microtubule
NFTs	neurofibrillary tangles
OR	odds ratios
PD	Parkinson’s disease
PDC	Parkinson dementia complex of Guam
PiD	Pick’s disease
PL-EM	partition ligation-expextation maximization
PSP	progressive supranuclear palsy
SNP	single nucleotide polymorphism
3R	3 MT-binding repeats
4R	4 MT-binding repeats

Gene Abbreviations

APOE	apolipoprotein E gene
APP	amyloid-β precursor protein
ATP13A2	ATPase type 13A2
CRHR1	corticotrophin releasing hormone receptor
DJ1	DJ-1
IMP5	intramembrane protease 5
LRRK2	leucine-rich repeat kinase 2
MAPT	microtubule associated protein tau
NR4A2	nuclear receptor subfamily 4, group A, member 2
NSF	N-ethylmaleimide sensitive factor
PINK1	PTEN-induced putative kinase I
PRKN	parkin
PSEN1	presenilin 1
PSEN2	presenilin 2
SNCA	α-synuclein
STH	saitohin
UCHL1	ubiquitin carboxy-terminal hydrolase L1

Abstract

Our global population is ageing and an ever increasing number of elderly are affected with neurodegenerative diseases, including the subjects of the studies in this work, Alzheimer's disease (AD), Parkinson's disease (PD), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD).

On strong evidence that several genes may influence the development of sporadic neurodegenerative diseases, the genetic association approach was used in the work of this thesis to identify the multiple variants of small effect that may modulate susceptibility to common, complex neurodegenerative diseases. It has been shown that the common genetic variation of one of these susceptibility genes, *MAPT*, that of the microtubule associated protein, tau, is an important genetic risk factor for neurodegenerative diseases. There are two major *MAPT* haplotypes at 17q21.31 designated as H1 and H2.

In order to dissect the relationship between *MAPT* variants and the pathogenesis of neurodegenerative diseases, the architecture and distribution the major haplotypes of *MAPT* have been assessed. The distribution of H2 haplotype is almost exclusively in the Caucasian population, with other populations having H2 allele frequencies of essentially zero.

A series of association studies of common variation of *MAPT* in PSP, CBD, AD and PD in different populations were performed in this work with the hypothesis that common molecular pathways are involved in these disorders. Multiple common variants of the H1 haplotypes were identified and one common haplotype, H1c, showed preferential association with PSP and AD.

A whole-genome association study of PD was also undertaken in this study in order to detect if common genetic variability exerts a large effect in risk for disease in idiopathic PD. Twenty six candidate loci have been found in this whole-genome association study and they provide the basis for our investigation of disease causing genetic variants in idiopathic PD.

Contents

Tables

Figures

Chapter 1 Introduction

1.1. Overview

1.1.1 The impact of neurodegenerative diseases in the ageing global population.

The population of our world is ageing and an ever-increasing number of elderly are affected by neurodegenerative diseases. In the developed world, about 2% of the population is afflicted at any time [1]. These neurodegenerative diseases include Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), Huntington's disease (HD), frontotemporal dementia (FTD), Pick's Disease (PiD) and prion diseases. Recently, the "Global Market and Future Outlook for Neurodegenerative Disorder Therapies 2007" forecasts that the overall neurodegenerative disease drug market will grow from around US$9 billion in 2005 to more than US$17 billion by 2010 (Piribo.com). The overall neurodegenerative diseases market is growing at over 12% each year. With an increase in life expectancy and the number of old people, along with advances in treatments of neurodegenerative diseases, the increases look set to continue.

Neurodegenerative diseases cost the U.S. economy billions of dollars each year in direct health care costs and lost opportunities; it is estimated that $100 billion per years is spent on AD alone. The market for AD therapy is expected to rise from US$4 billion in 2005 to US$7 billion in 2010 with the US market continuing to dominate. In addition to the financial costs, there is an immense emotional burden on patients and their caregivers. As the number of elderly citizens increases, these costs to society will also increase [2].

1.1.2 The environmental risk factors of neurodegenerative diseases

Neurodegenerative diseases are characterized by progressive nervous system dysfunction. These disorders result from the gradual and progressive loss of neural cells, leading to nervous system dysfunction. Neurodegenerative diseases can affect abstract thinking, skilled movements, emotional feelings, cognition, memory and other abilities [3].

Known risk factors for neurodegenerative diseases include certain genetic polymorphisms and increasing age. Other possible causes may include gender, poor education, endocrine conditions, oxidative stress, inflammation, stroke, hypertension, diabetes, smoking, head trauma, depression, infection, tumors, vitamin deficiencies, immune and metabolic conditions, and chemical exposure. Because the pathogenesis of many of these diseases remains unknown, the role of environmental factors in these diseases has to be considered [2]. Although there is a growing body of evidence suggesting that a large proportion of the non-familial cases are significantly influenced by genetic factors, it is likely like the environmental risk factors play a part in the etiology of some of these conditions. MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) is a good though rare example of an environmental agent that caused an epidemic of Parkinsonism among drug misusers [4]. MPTP is a simple chemical, and a viable hypothesis is that autointoxication by similar molecules may cause sporadic diseases [5]. A good example is the epidemic of a variety of degenerative diseases on Guam, where a possible environmental agent has not been discovered [6].

Despite the enormous effort put into searching for the environmental factors for these diseases, epidemiological evidence for an association between environmental agents and neurodegenerative diseases remains inconclusive [2].

On the other hand, evidence from the familial, twin and association studies for neurodegenerative diseases, such as AD, PD, FTD, suggesting it is promising to apply molecular genetics to reveal the pathogenesis of those diseases with understanding their underlying genetics. The genetics of various neurodegenerative diseases are briefly reviewed in the following sections.

1.1.3 The genetics of neurodegenerative diseases

Until recently, it was considered impossible to find a common molecular mechanism for neurodegeneration. However, despite their diverse clinical manifestations and disease progression, these disorders share some common characteristics: all these diseases (except Huntington disease) have both sporadic and inherited types, the onset of all these diseases is usually after the fourth or fifth decade, and their pathology involves neuronal loss and protein aggregation. For instance, a normal soluble cellular protein is converted into an abnormal insoluble aggregated protein rich in β-sheets that is toxic such as β-amyloid in AD. Emerging evidence for a causal role of the conformational changes of proteins in neurodegenerative diseases has become clearer recently from genetic studies [7].

Mutations in the genes that encode the protein components of fibrillar aggregates are genetically associated with the inherited form of neurodegenerative diseases

[3]. These include mutations in genes for the β-amyloid (Aβ) precursor protein, causing AD; α-synuclein, causing PD; or in microtubule-associated protein tau, causing FTD with parkinsonism. Though these mutations cause mainly the familial type of neurodegenerative diseases, the pathology of both familial and sporadic forms is either identical or very similar [8]. Investigation of the molecular genetics of the familial forms has provided a window into the molecular biology of both the familial and the sporadic forms of neurodegenerative diseases. Indeed, there is a growing body of evidence suggesting that a large proportion of the sporadic cases are also significantly influenced by genetic factors. These risk genes are likely to be numerous, displaying intricate patterns of interaction with each other as well as with non-genetic variables, and — unlike classical Mendelian ("simplex") disorders — exhibit no simple or single mode of inheritance. Hence, the genetics of these diseases has been labelled "complex" [9].

A conception regarding the genetic makeup of complex diseases is the "common disease/common variant" (CD/CV) hypothesis [10]. According to this theory common disorders are also governed by common DNA variants, such as single nucleotide polymorphisms. These variants significantly influence disease risk but are insufficient to actually cause a specific disorder. Current empirical and theoretical data support this hypothesis, although there remains great uncertainty as to the number of the underlying risk factors and their specific effect sizes. It is believed that there is a risk spectrum predisposing to common disease, such as AD, as one continuum [10]. The continuum extends from the most clearly defined genetic forms to cases influenced by genetic susceptibility factors, and further extending to a less well defined area of cases that may be caused by genes of low penetrance and/or non-genetic factors. In AD, for instance, rare, fully penetrant autosomal dominant mutations in 3 genes (i.e. *APP*, *PSEN1*, and *PSEN2*) have been shown to cause the disease, while a common, incompletely penetrant susceptibility variant, ε4 in the apolipoprotein E gene (*APOE*) significantly increases the risk for AD.

Similar genetic susceptibility traits such as *APOE* in AD are also found in other neurodegenerative diseases, such as *MAPT* in PSP. Therefore, genetic analysis is having a substantial impact in the attempts to dissect the etiology and pathogenesis of the neurodegenerative diseases, making it increasingly clearer how disease may be initiated. In the following sections, the genetics of various neurodegenerative diseases and the genetic approach for studying of these diseases will be reviewed.

1.2 Neurodegenerative Diseases

1.2.1 The genetics of Alzheimer's disease

AD is the most common neurodegenerative disease and afflicts ~5% of those over 65 years. It is one of the most serious health problems in the industrialized world. It is an insidious and progressive neurodegenerative disorder that accounts for the vast majority of age-related dementia and is characterized by global cognitive decline and the accumulation of Aβ deposits and neurofibrillary tangles in the brain. Family history is the second-greatest risk factor for the disease after age, and the growing understanding of AD genetics has been central to the knowledge of the pathogenic mechanisms leading to the disease. Genetically, AD is complex and heterogenous and appears to follow an age-related dichotomy: rare and highly penetrant early-onset familial AD (EOFAD) mutations in different genes are transmitted in an autosomal dominant fashion, while late-onset AD (LOAD) without obvious familial segregation is thought to be explained by the CD/CV hypothesis [11].

The EOFAD represents only a small fraction of all AD cases (≤ 5%) and typically presents with onset ages younger than 65 years, showing autosomal dominant transmission within affected families. These forms of the disease are caused by mutations in the amyloid-β precursor protein (APP) on chromosome 21 [12] and the presenilin 1 gene (*PSEN1*) on chromosome 14 [13] and presenilin 2 gene (*PSEN2*) on chromosome 1 [14-16]. To date, more than 160 mutations in these 3 genes have been reported to cause EOFAD. An up-to-date overview of disease-causing mutations in these genes can be found at the Alzheimer Disease & Frontotemporal Dementia Mutation Database [17]. The most frequently mutated gene, *PSEN1*, accounts for the majority of AD cases with onset prior to age 50. While these AD-causing mutations occur in 3 different genes located on 3 different chromosomes, they all share a common biochemical pathway, i.e., the altered production of Aβ leading to an overabundance of the Aβ42 species relative to the Aβ40 species, which eventually results in neuronal cell death and dementia. The mutations in the APP gene occur at the Aβ cleavage sites, thereby altering APP processing such that more Aβ42 is produced [18]. The presenilins are a central component of γ-secretase, the enzyme responsible for liberating the C-terminal fragment of APP, and mutations in the presenilins also alter APP processing such that more Aβ42 is produced [19]. These genetic data are the intellectual basis for the amyloid hypothesis of the disorder, suggests that Aβ42 is the initiating molecule in Alzheimer's disease.

LOAD, on the other hand, is classically defined as AD with onset at age 65 years or older and represents the vast majority of all AD cases. While segregation and twin studies conclusively suggest a major role of genetic factors in this form of AD [20], to date, only one such factor has been established, the ε4 allele of the *APOE* gene on chromosome 19q13 [21;22]. In contrast to all other association-based findings in AD, the risk effect of *APOE*-ε4 has been consistently replicated in a large number of studies across many ethnic groups, yielding odds ratios (OR) between approximately 3 and approximately 15 for heterozygous and homozygous carriers, respectively, of the ε4 allele in white individuals [23]. In addition to the increased risk exerted by the ε4-allele, a weak, albeit significant, protective effect for the minor allele, ε2, has also been reported in several studies [9]. Unlike the mutations in the known EOFAD genes, *APOE*-ε4 is neither necessary nor sufficient to cause AD but instead operates as a genetic risk modifier by decreasing the age of onset in a dose-dependent manner. Despite its long-known and well-established genetic association, the biochemical consequences of APOE-ε4 in AD pathogenesis are not yet fully understood but likely encompass Aβ-aggregation/clearance and/or cholesterol homeostasis. Several lines of evidence suggest that numerous additional LOAD loci [24] — and probably also EOFAD loci [25;26] — remain to be identified, since the 4 known genes together account for probably less than 50% of the genetic variance of AD. As outlined above, it is currently unclear how many of these loci will be proved to be risk factors as opposed to causative variants. As candidates for the former, more than 3 dozen genes have been significantly associated with AD in the past [27;28]. Despite the more than 500 independent association studies, however, no single gene has been shown to be a risk factor with the same degree of replication or consistency that has been shown for *APOE-ε4*. One of the conclusions to be drawn from currently available data, as well as from the few independently performed meta-analyses on putative AD risk factors is that even if some of the published associations were genuine, their overall effect size is likely to be only minor, i.e. with OR not exceeding 2.

1.2.2 The genetics of Parkinson's disease

PD is the second most common neurodegenerative disease of adult onset. Histopathologically, it is characterized by a severe loss of dopaminergic neurons in the substantia nigra and cytoplasmic inclusions consisting of insoluble protein aggregates, which lead to a progressive movement disorder including the classic triad of tremor, bradykinesia, and rigidity, with an average onset age between 50 and 60 years. As in AD, there appears to be an age-dependent dichotomy: the majority of individuals with an early or even juvenile onset show typical Mendelian inheritance. However, unlike in AD, these cases show a predominantly autosomal-recessive mode of inheritance, and there is an ongoing debate as to whether genetic factors play any substantial role in contributing to disease risk in cases with onset beyond approximately 50 years [29;30]. Notwithstanding these uncertainties, there are a plethora of genetic studies on both forms of the disease, and mutations in at least 6 genes have now been shown to cause familial early-onset parkinsonism (α-synuclein, *SNCA*) [31]; parkin (*PRKN*) [32]; DJ-1 (*DJ1*)

[33]; PTEN-induced putative kinase I (*PINK1*) [34]; and leucine-rich repeat kinase 2 or dardarin (*LRRK2*) [35] and *ATP13A2* [36], with several other linkage regions pending characterization and/or replication. As was the case in the study of AD, the first locus to be characterized – *SNCA*, on chromosome 4q21 – which codes for α-synuclein the protein that is the major constituent of the Lewy body (LB), one of the classic neuropathological hallmarks of the disease [31], which can be found at the core of LBs. While the exact mechanisms underlying α-synuclein toxicity currently remains only incompletely understood, recent evidence suggests that some *SNCA* mutations may change normal protein function quantitatively rather than qualitatively, via duplication or triplication of the *SNCA* gene [37;38]. Mutations in a second gene, *LRRK2* with dominant inheritance have been identified by several different laboratories [35]. While the functional consequences of *LRRK2* mutations are still unknown, it was suggested that at least some mutations could interfere with the protein's kinase activity [39]. While changes in *SNCA* and *LRRK2* are the leading causes of autosomal-dominant forms of PD, the majority of affected pedigrees actually show a recessive mode of inheritance. The most frequently involved gene in recessive parkinsonism is parkin (*PRKN*) on chromosome 6q25 [32;40], which causes nearly half of all early-onset PD cases. Parkin is an ubiquitin ligase that is involved in the ubiquitination of proteins targeted for degradation by the proteasomal system. The spectrum of parkin mutations ranges from amino acid-changing single base mutations to complex genomic rearrangements and exon deletions, which probably result in a loss of protein function. It has been speculated that this may trigger cell death by rendering neurons more vulnerable to cytotoxic insults, e.g., the accumulation of glycosylated α-synuclein [41]. In addition to parkin mutations,

genetic analyses of two non-parkin early-onset, autosomal-recessive PD pedigrees revealed two independent, homozygous mutations in *DJ1*[33] on chromosome 1p36 [42]. Both mutations result in a loss-of-function of DJ-1, a protein that is suggested to be involved in oxidative stress response. While several studies have independently confirmed the presence of *DJ1* mutations in other PD cases, the frequency of disease-causing variants in this gene is estimated to be low (~1%) [43]. Less than 13 Mb toward the long arm of the same chromosome, additional PD-causing mutations were subsequently discovered in *PINK1* [34] following positive linkage evidence to this region [44]. *PINK1* codes an enzyme that is expressed at particularly high levels in brain, and the first two mutations identified (G309D and W437ter) were predicted to lead to a loss-of-function that may render neurons more vulnerable to cellular stress, similar to the effects of *PRKN* mutations. While Lewy bodies are typically not found in brains of patients bearing *PRKN* mutations, it is currently unclear whether these are present in PD cases with mutations in *DJ1* and *PINK1*. At least six additional candidate PD loci have been described, including putative disease-causing mutations in the ubiquitin carboxy-terminal hydrolase L1 (*UCHL1*) on chromosome 4p14 [45], and in a nuclear receptor of subfamily 4 (*NR4A2*, or *NURR1*) [46] located on 2q22. However, and unlike the previously outlined PD genes, neither of these maps to known PD linkage regions, nor were they independently confirmed beyond the initial reports. However, polymorphisms in both genes have been, albeit inconsistently, associated with PD in some case-control studies. A recent meta-analysis of the S18Y polymorphism in *UCHL1* showed a modest but significant protective effect of the Y allele [47], which suggests that this gene may actually be a susceptibility factor rather than a causal PD gene. Unlike early-onset PD, the heritability of late-onset PD is probably low [29]. Despite this caveat, while a number of whole genome screens across several late-onset PD family samples have been performed, only a few overlapping genomic intervals have been identified. One of the more extensively studied regions is 17q21, containing the gene encoding the microtubule-associated protein tau (*MAPT*) [48]. Previously, it had been shown that rare missense mutations in *MAPT* lead to a syndrome of FTD with parkinsonism linked to chromosome 17 (FTDP-17), but to date no mutation has been identified as causing parkinsonism without frontotemporal degeneration. However, haplotype analyses of the tau gene have revealed evidence of genetic association of the H1 haplotype with both PD [49;50] and PSP [51]. Despite the lack of evidence for genetic linkage to chromosome 19q13, variants in *APOE* have also been tested for a role in PD. Across the nearly three dozen different studies available to date, some authors report a significant risk effect of *APOE*-ε4 for PD, while others only see association with certain PD phenotypes or even a risk effect of the ε2 allele, which is protective in AD (see above). A recent meta-analysis on the effects of *APOE* in PD concluded that only the ε2-related increase in PD risk remains significant when all published studies are considered jointly [52]. Finally, and in addition to the findings in autosomal-dominant familial PD, there is also some support for a potential role of *SNCA* variants in the risk for late-onset PD [53].

1.2.3 The genetics of progressive supranuclear palsy (PSP)

PSP is the second most frequent cause of degenerative parkinsonism after PD [54]. In addition to parkinsonism, the clinical symptoms include early postural instability and supranuclear gaze palsy [55]. Neuropathologically, PSP is characterized by abundant neurofibrillary tangles and neurophil threads consisting largely of four repeat tau [56]. Robust genetic association of PSP with *MAPT* and rare reports of families with more than one affected member indicated that genetic factors could play a role in PSP [57;58].

In Europeans, the *MAPT* gene has an unusual genetic structure, two distinct and inverted haplotypes have been found and designated as H1 and H2. The H2 haplotype has an allele frequency of approximately 25% [59;60]. This has made the genetic analysis of PSP and CBD easy in this population and has shown a robust association between the H1 haplotype of the *MAPT* locus and both PSP and CBD [61]. It is likely that there will be a *MAPT* association with these diseases in other populations, but the relevant analyses are more difficult to perform in these other populations because of the absence of the H2 haplotype. A haplotypic association, in the absence of coding changes, implies that the biological effect could be mediated either by differences in expression or differences in splicing between haplotypes. The fact that the disturbances in the splicing of the *MAPT* gene is one of the causes of FTDP-17, and the fact that the tangle deposits consist almost exclusively, of four-repeat tau, suggests that either or both of the above are equally likely explanations. A more detailed analysis of the structure of the H1 haplotype revealed that it has considerable complexity, and that in fact, the haplotypic association between H1 and PSP and CBD is driven by a variant of the haplotype named H1c, which defines a region from the promoter to intron 10 of the gene. Analysis of this haplotype has not yet led to the determination of whether it is a particularly high-expressing haplotype, one that particularly expresses the exon-10 containing transcript, or a mixture of both. (see section

1.2.5.1 and **Figure 1.1**) This haplotypic association is an example of the general principle that genetic variability at the loci causing autosomal dominant disease (in this case FTDP-17) is part of the genetic contribution to the sporadic diseases (in this case PSP and CBD) [62].

1.2.4 The genetics of corticobasal degeneration (CBD)

Corticobasal degeneration (CBD) is a progressive neurological disorder characterised by atrophy of multiple brain areas including the cerebral cortex and the basal ganglia [63]. Initial symptoms, such as poor coordination, akinesia, rigidity, impaired balance and limb dystonia which typically appear at the age of around 60, are similar to those found in Parkinson’s disease. Other symptoms such as cognitive and visuo-spatial impairments, apraxia, hesitant and halting speech, myoclonus (muscular jerks), and dysphagia may also occur.

Neuropathologically, CBD is distinguished from PSP and other dementias by several important features. Most pathology in CBD is in the cerebrum, whereas the basal ganglia, diencephalon, and brainstem are mainly the targets of PSP. Histologically, there are ballooned neurons, astrocytosis, and four repeat 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 [64;65].

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 [66] that was subsequently independently replicated [67].

1.2.5 The tauopathies

The above neurodegenerative diseases AD, PSP and CBD are collectively belong to a group of disorders known as the tauopathies, as they all have pathological fibrillar aggregates of tau in the brain. The characteristics of tau protein and the tauopathies are reviewed below.

1.2.5.1 Microtubule associated protein tau

The microtubule associated protein, tau was first identified as a “factor essential for microtubule (MT) assembly”, a heat stable protein that induced the assembly of MTs from purified tubulin and belonging to the family of MT-associated proteins [68]. Tau is abundantly expressed in the both the peripheral and central nervous system [69], where it is enriched in the axons of mature and growing neurones and, low levels of tau are also present in oligodendrocytes and astrocytes [70;71]. Tau is a phosphoprotein with developmentally regulated phosphorylation profiles at up to 38 phosphorylation sites [72]. The level of protein phosphorylation is highly elevated in foetal tau and pathological tau found within the insoluble, fibrillar inclusions that define tauopathies, when compared to normal adult brain tau [73]. 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 [74-76] **(Figure 1.1)**

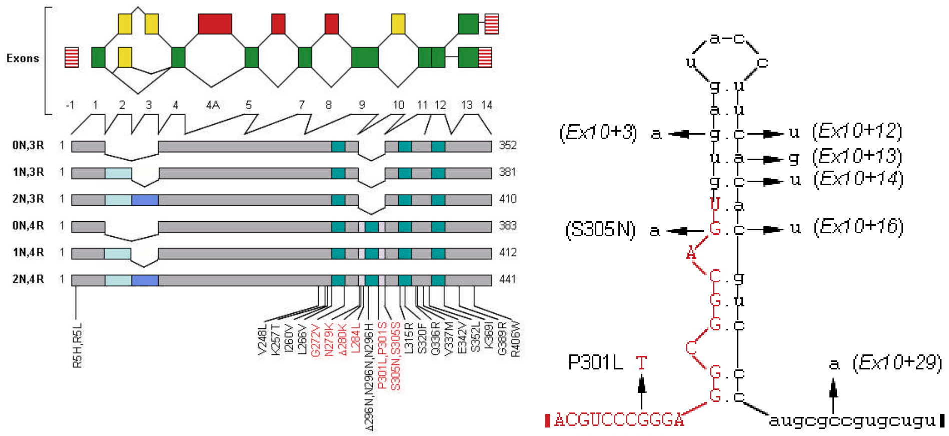


Figure 1.1 MAPT structure and the FTDP-17 mutation spectrum

Left: 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 tau. Exons 2 and 3 code for 29 residue 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). 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. Right: 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 increase incorporation of exon 10. Partial sequence of 3'-end of exon 10 in red. intronic sequence in black. Proportions are not to scale. (Modified from Goedert [77])

In the healthy adult human brain, tau protein exists as six major isoforms produced by the alternative splicing of exons 2, 3 and 10 [78]. (**Figure 1.1**) 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 recognised 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 [79;80], whereas in PSP, CBD and argyrophilic grain disease (AGD), both neuronal and glial inclusions contain mostly 4R-tau isoforms [79;81-84], and roughly equal amounts of 3R- and 4R-tau make up the paired helical filaments and straight filaments observed in AD [84;85].

1.2.5.2 The tauopathies

The tauopathies are a group of neurodegenerative disorders that are characterized pathologically by the presence of fibrillar aggregates of tau in the brain [76;86] (**Table1.1**). The most common tauopathy, AD, is characterized 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 biomarkers that include the loss of choline acetyl transferase and synaptophysin reactivity. In addition to abundant extracellular amyloid Aβ deposits, the senile plaques, hyperphosphorylated tau neurofibrillary tangles (NFTs) constitute the pathological lesions [87]. Aβ and NFTs also coexist in some other tauopathies like Down's syndrome [88;89] however NFTs occur alone in argyrophilic grain disease (AGD) [82], PiD [90], CBD, PSP [91], FTDP-17 [92], some ALS [93], Niemann-Pick disease type C [94] and subacute sclerosing panencephalitis. These disorders are classified as primary tauopathies, since pathological aggregates of neurofibrillary tau are their main defining characteristic (**Table 1.1**). 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-17
Gerstmann-Sträussler-Scheinker disease
Hallervorden-Spatz disease
Myotonic dystrophy
Niemann-Pick disease
Non-Guamanian motor neuron disease with neurofibrillary tangles
Pick’s disease
Postencephalitic parkinsonism
Prion protein cerebral amyloid angiopathy
Progressive subcortical gliosis
Progressive supranuclear palsy
Supacute sclerosing panencephalitis
Tangle only dementia

Table 1.1 The Tauopathies
Primary tauopathies are shaded grey; secondary tauopathies are shaded white.

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 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 NFTs and in distal dendrites as neuropil threads. Ultrastructurally, these lesions consist of paired helical filaments and straight filaments [76;94]. The tau inclusions in the different tauopathies have characteristic morphologies and distributions.

The pathological tau filaments are insoluble but can be isolated for biochemical analysis as the detergent sarkosyl-insoluble fractions of brain homogenates [95]. 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 different compositions 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).

1.3 Genetic Approach to Study Neurodegenerative Diseases

In the following sections, the different genetic approaches, which have been employed in the work of this thesis, for studying of neurodegenerative diseases will be reviewed.

1.3.1 Genetic epidemiology

Genetic epidemiology is a discipline closely related to traditional epidemiology that focuses on the familial and the population, towards identification of genetic determinants of disease and the joint effects of genes and non-genetic determinants such as the environment [96]. 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 most neurodegenerative diseases, diabetes mellitus and cancer. These diseases are thought to be caused by several interacting genetic and environmental determinants [97] and require quite different genetic epidemiological study design and interpretation compared to the traditional genetic linkage studies in monogenic Mendelian disorders [96].

1.3.2 Genetic Mapping of common complex disease genes

1.3.2.1 The rationale of the 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 populations 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 correlation between nearby variants such that the alleles at neighbouring markers (observed on the same chromosome) are associated within a population more than if they were expected by chance.

Various methods of marker pairwise LD measures have been proposed [98] that are usually based upon Lewontin’s D’ [99]; 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 pair-wise 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 locus 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 locus B [100].

1.3.2.2 The design of population genetic association studies

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 [101] 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 individual representatives 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 [102]. 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 [103]. 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. 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, which have a high relative risk.

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

1.3.2.3 Bias due to population stratification

In a population-based association study involving hundreds of thousand markers, minimizing false positives is essential. Sources of false-positives association can be divided into three main categories: statistical fluctuations that arise by chance and result in low p-values; technical artefacts; and underlying systematic bias due to study design. The issue from multiple hypothesis testing is best addressed using robust criteria for declaring significant associations. While technical artefacts would probably be avoided if cases and controls are genotyped in an identical manner, because genotyping errors or missing data should affect cases and controls equally.

The population stratification remains the major bias which the researchers have to consider from the beginning of the sample collection. Population stratification bias is a systematic bias which occurs in the studies of genoytype-disease associations if the component populations have different genotypic distribution. Population stratification is the presence of multiple sub-groups within a population that differ in disease prevalence (or average trait value, for quantitative traits). This is most commonly due to ethnic admixture, which is defined as combining two or more populations into a single group 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 [104] 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 [105].

1.3.3 Whole-genome association approach for neurodegenerative diseases

1.3.3.1 Whole-genome association study

A genome-wide association approach is an association study that surveys most of the genome for predisposing genetic variants. Because no assumption is made about the genomic location of the causal variants, this approach could exploit the strengths of association studies without having to guess the identity of the causal genes. The genome-wide association approach therefore represents an unbiased yet fairly comprehensive option that can be attempted even in the absence of convincing evidence regarding the function or location of the genes [106].

Genome-wide association studies require knowledge about common genetic variation and the ability to genotype a sufficiently comprehensive set of variants in a large patient sample. The dbSNP database now contains nearly 9 million SNPs, including most of the ~11million SNPs with minor allele frequencies of 1% or greater that are estimated to exist in the human genome [107]. Importantly, genotyping technology has considerably improved and become cheaper in recent years. One recent review of SNP genotyping technology cited ‘large-scale’ studies that involved nearly a hundred thousand genotypes [108]. By contrast, the HapMap project (discussed in more detail below) plans to include information on ~300 million genotypes.

Another crucial advance towards enabling efficient genome-wide studies is the determination of LD patterns on a genome-wide scale through the HapMap project, which will be particularly useful for methods that use markers selected on the basis of LD.

1.3.3.2 The International HapMap project

The HapMap is a catalog of common genetic variants that occur in human beings. It describes what these variants are, where they occur in our DNA, and how they are distributed among people within populations and among populations in different parts of the world. The International HapMap Project is not using the information in the HapMap to establish connections between particular genetic variants and diseases. Rather, the project is designed to provide information that other researchers can use to link genetic variants to the risk for specific illnesses, which will lead to new methods of preventing, diagnosing, and treating disease. This large and ambitious project (http://www.hapmap.org/) aims to construct genome-wide maps of LD patterns, at a density of at least 1 SNP per kb, in samples collected in the USA (Caucasians of western European origin), Nigeria, China and Japan 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 [109]. For the purposes of candidate gene association studies, HapMap project data can be analysed to identify haplotype-tagging SNPs for more efficient and economical genotyping in case-control cohorts.

1.3.3.3 Markers for genome-wide association studies

Useful markers for a genome-wide association studies must either be the causal allele or highly correlated (in LD) with the causal allele [110;111]. Most of the genome falls into segments of strong LD, within which variants are strongly correlated with each other, and most chromosomes carry one of only a few common haplotypes [112-114]. Recently, several large genomic regions (of ~500 kb) have been comprehensively examined as part of the “Encyclopedia of DNA Elements” (ENCODE) project. This project involved the resequencing of 96 chromosomes to ascertain all common variants, and the genotyping of all SNPs that are either in the dbSNP database or that were identified by resequencing. (www.genome.gov). These studies have shown that most of the roughly 11 million common SNPs in the genome have groups of neighbours that are all nearly perfectly correlated with each other — the genotype of one SNP perfectly predicts those of correlated neighbouring SNPs. One SNP can thereby serve as a proxy for many others in an association screen. Once the patterns of LD are known for a given region, a few such tag SNPs can be chosen such that, individually or in multimarker combinations (haplotypes), they capture most of the common variation within the region [114;115] (**Figure1.2**). A proportionally higher density of variants must be typed to comprehensively survey the fraction of the genome that shows low LD.

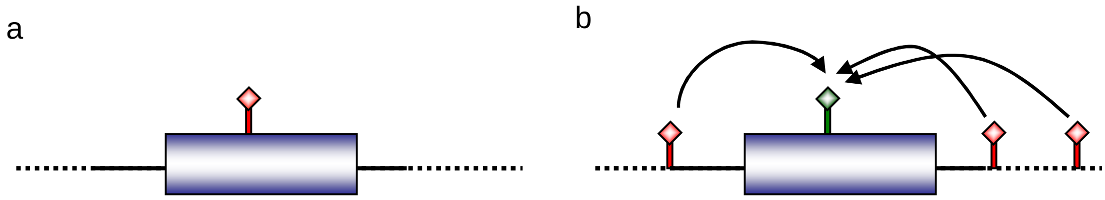


Figure 1.2 Testing SNPs for association by direct and indirect methods.

(a) The candidate SNP (red) which to be genotyped is located within the causal gene. A direct association with the disease/phenotype is tested. (b) The SNPs to be genotyped (red) are chosen on the basis of linkage disequilibrium (LD) patterns to provide information about as many other SNPs as possible. In this case, the SNP shown in green is tested for association indirectly, as it is in LD with the other three SNPs [9].

On the basis of previous studies [112-114;116] and initial HapMap data, a few hundred thousand well-chosen SNPs should be adequate to provide information about most of the common variation in the genome; a larger number of tag SNPs is likely to be required in African populations (and those with very recent origins in Africa), because these populations generally contain more variation and less LD [114;117]. The precise number of tag SNPs needed is yet to be determined, and will depend on the methods used to select SNPs, the degree of long-range LD between blocks and the efficiency with which SNPs in regions of low LD can be tagged [118;119]. Various algorithms have been proposed for selecting tag SNPs [115;120-125]; the optimal method will depend partly on which of the many methods for searching for associations is employed (using haplotypes, single markers, multiple markers and so on).

1.4 Thesis aims and objectives

In this chapter, the rationale and strategies behind the genetic association approach for neurodegenerative diseases were reviewed. As a common pathological finding, tau protein inclusions have long been recognized to define one of the diverse categories of neurodegenerative diseases, that is, tauopathies. Tau protein dysfunction in those neurodegenerative diseases has been firmly established as there is growing evidence from two independent lines of research. First, the biochemical study of the neuropathological lesions that defines these diseases led to the identification of their molecular components. Second, the study of familial forms of disease led to the identification of gene defects that cause the inherited variants of the different diseases. For example, the association of the H1 *MAPT* haplotype implies tau dysfunction is related to the pathogenesis of PSP and this is also supported by the pathognomic tau pathology found in the disease post-mortem. Though the exact mechanism of the formation of pathological tau inclusion may vary in different neurodegenerative diseases, the abnormal expression of *MAPT* gene still remains as the possible culprit causing neuronal death among various neurodegenerative diseases. Most of the previous association studies were performed on a single population from a solitary ethnicity. The impact of this plausible disease candidate gene among different ethnic groups has not yet been studied.

At the onset of this work, though two distinct *MAPT* H1, H2 non-recombinant haplotype blocks had been defined, the characteristics of these unusual haplotype blocks in different populations had not been established. The first task of the work in this thesis was to investigate the distribution of these *MAPT* haplotypes and the variation, if any, of these blocks among different populations worldwide. With these results, we could gain an understanding of the origins of these haplotypes and their possible effects on neurodegeneration in different populations.

This work is based on a general hypothesis that the genetic variations at 17q21.31 affecting *MAPT* gene expression, splicing or mRNA stability modulate pathways that lead to the death of neurons in different neurodegenerative diseases. Identification of mutations in the *MAPT* that directly led to neurodegeneration in FTDP-17 confirm that tau dysfunction is an important part of the neurodegenerative sequence [126;127]. Furthermore, the overlap in pathological findings of deposition of tau tangles in various neurodegenerative diseases suggest that understanding tangle formation in other diseases besides FTDP-17 would be critical to understanding the pathogenesis of cell loss. The main aim of the work in this thesis was to investigate the genetic association of the *MAPT* haplotype in various types of neurodegenerative diseases, including PSP, AD and PD.

At the onset of this study, the genetic diversity of the *MAPT* gene was defined in terms of the H1 and H2 haplotypes but there was evidence of much greater diversity.

Using population genetic methods, the underlying LD structure and haplotype diversity of the *MAPT* gene was examined. Expanding on this, a Taiwanese 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 tagged SNPs (htSNPs) for more streamlined and efficient genotyping of the different cohorts of neurodegenerative diseases and healthy controls from various populations. These htSNPs were genotyped in those cohorts to determine if the *MAPT* association, if any, is the same in different cohorts. The allele, genotype and haplotype frequencies of the htSNPs were statistically assessed for differences between cases and controls among different populations.

On strong evidence that several genes may influence the development of sporadic neurodegenerative diseases, an effective and powerful approach to identify the multiple variants of small effect that modulate susceptibility to common, complex diseases is the key to detect the small genetic effects on diseases susceptibility [128]. As neurodegenerative diseases, which studied in the context of this thesis are a collection of common complex diseases under the influence of genetic risk factors, they would be good candidates for whole-genome association study. The emerging high-throughput whole genome genotyping gave us a feasible technique to detect the genetic variants and susceptibility loci that affect the risk of developing the neurodegenerative diseases. The first whole-genome association study of a Parkinson's disease cohort has been included in the work of this thesis. This was carried out in order to determine if there is any common genetic variability exerting a large effect in those risks for Parkinson's disease in a population cohort.

Chapter 2 Methods and Materials

2.1 Methods

2.1.1 DNA sample extraction from tissue

DNA was routinely extracted by hand from fresh frozen brain material or blood as required. In the laboratory, two methods were used, the DNeasy Tissue Kit (Qiagen) or by Proteinase K/phenol-chloroform extraction.

For the the DNeasy Tissue Kit, 100μl anticoagulated blood was mixed with 20 μl proteinase K in a 2 ml microcentrifuge tube. The volume was adjusted to 220 μl with PBS. The solution was incubated at 56^{o}C for 10 minutes after 200μl Buffer AL (Qiagen) was added. 200 μl ethanol (96%-100%) was added to the sample followed by vortexing to mix the solution thoroughly. The mixture was pipeted into the DNeasy Mini spin column place in a collection tube. The whole set of column was centrifuged at 8000 rpm for 1 minutes. After adding 500 μl of Buffer AW1 (Qiagen), the column was centrifuged again at 8000 rpm. The content in the DNeasy mini spin column was washed with 500 μl Buffer AW2 (Qiagen) and was centrifuged at 14,000 rpm for 3 minutes to dry the DNeasy membrane. The elution of the DNA sample was carried out as adding 200 μl Buffer AE (Qiagen) directly onto the DNeasy membrane, followed by incubation for 1 minute at room temperature. Finally, the eluate with DNA sample was collected with centrifuged for 1 minute at 8000 rpm.

In the latter protocol the blood cells or the frozen brain tissue 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^{o}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^{o}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^{o}C.

2.1.2 DNA quantification

DNA extraction quantity and quality was monitored by UV spectrophotometer absorption (ND-1000, Nanodrop). 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 absorption coefficient for double-stranded DNA) for the final concentration value in ng/µl. 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 a common method of creating and amplifying copies of specific fragments of DNA. PCR rapidly amplifies a single DNA molecule into many billions of molecules. 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 selectively PCR-amplify a specific DNA fragment, suitable primers need to be designed and synthesized. Some prior DNA sequence information from the target sequence is required for designing two oligonucleotide primers which are specific for the target (designed in the computer program Primer Express version 2.0 (Applied Biosystems) and Primer3 [http://fokker.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi]). Primers are short oligonucleotides, that is, chemically synthesized, single-stranded DNA fragments — often not more than 50 and usually only 18 to 25 base pairs long — containing nucleotides that are complementary to the nucleotides at both ends of the DNA fragment to be amplified. These complementary bases in primer and DNA template facilitate annealing of the primer to the DNA template to which the DNA polymerase can bind and begin with the synthesis of a new DNA strand that is complementary to the DNA template. 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.

PCR was performed using the Taq DNA polymerase core kit (Qiagen) and FastStart PCR Master (Roche). For the former protocol, 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. While in the latter protocol, 20ul reactions contained 2x reaction reagent (containing the dNTP and Taq DNA polymerase), 0.1-0.5 uM of each forward and reverse primers, and 25 ng of genomic DNA template. 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 Hybaid Multiblock System. 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 Agarose gel electrophoresis

Agarose gels, used for analysing DNA fragment sizes and quality were made by melting agarose powder (American Bioanalytical) 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 visualized with AlphaEase FC software version 3.2.1 (Alphainnotech) under UV illumination.

2.1.5 Genotyping

2.1.5.1 Restriction fragment length polymorphism

Most of the nucleotide variations within the genome of a specific species are 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. The generation or destruction of restriction sites after the point mutation allows the rapid detection of point mutations after the genomic sequences are amplified by the PCR. 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 Gene Runner version 3.05.

Restriction endonucleases are enzymes that cleave DNA molecules at specific nucleotide sequences depending on the particular enzyme used. Enzyme recognition sites are usually 4 to 6 base pairs in length. If molecules differ in nucleotide sequence, fragments of different sizes may be generated. The fragments can be separated by gel electrophoresis.

Typically, 15 μl of raw PCR products were incubated with 1 unit of the corresponding restriction endonuclease (New England Biolabs and Fermentas) in a reaction volume of 20 μl at the recommended temperature for at least four hours. Digests were separated on a 3~4% agarose gel, depending on the sizes of the predicted fragmented PCR products. The fragments were visualized with ethidium bromide staining. The images were captured with AlphaEase FC software version 3.2.1 (Alphainnotech) under ultraviolet illumination for genotype scoring.

2.1.5.2 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, 15 µl of the PCR product was first immobilised on Streptavidin-SepharoseTM HP (Amersham Pharmacia Biotech): 2 µl of Streptavidin-SepharoseTM HP is re-suspended in 38 µl of binding buffer (10 mM Tris-HCl, 2 M NaCl, 1mM EDTA, 0.1 % Tween-20) and 20 µl of water. 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), then immersed for 10 seconds in ethanol followed by denaturing buffer (0.2 M NaOH) and finally wash buffer (20 mM Tris-acetate, pH 7.6, 2 mM magnesium acetate). 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 (Biotage) 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.

2.1.5.3 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. The images were captured with AlphaEase FC software version 3.2.1 (Alphainnotech) under UV illumination for genotype scoring.

2.1.6 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 ddNTPs 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 Multiscreen PCR μ96 filter plates (Millipore). Typically, 85 μl of TE buffer was added to the PCR product (15 μl). The mixture was transferred to the filter plate. The filter plate was placed on the vacuum system (Millipore) and applied vacuum at 20 inch Hg for 7-12 minutes. The PCR samples were washed with using 25 μl of water. Then the samples were dissolved in 20 μl of water by vigorously mixing before these mixtures were used to set up the sequencing reaction. For the sequencing reaction, 1.5 μl of treated PCR product was made up to a volume of 10 μl with molecular biology grade H_2O. The mixture including 1 μl (3.2 pmol) of primer (forward or reverse), 2.0 μl of 5x sequencing buffer (Applied Biosystems) and 0.5 μl of BigDye Mix. The sequencing reaction involved an initial denaturing step at 96^{o}C for 3 minutes, followed by 50 cycles of denaturation (95^{o}C, 30 seconds), primer annealing (variable, depending on the annealing temperature of the primers, 15 seconds) and extension (60^{o}C, 4 minutes) and a refrigeration hold at 4^{o}C until sample collection. The resulting reactions were purified using the Montage SEQ Sequencing Reaction Cleanup Kit (Millipore). The resulting products were subjected to capillary electrophoresis on an ABI 3100 capillary sequencer (Applied Biosystems). Sequence traces (ABI chromatograms) were viewed and analysed using Sequencher software.

2.1.7 Whole Genome Scanning

Genetic association studies offer a potentially powerful approach for mapping causal genes with modest effects, but are limited because only a small number of genes can be studied at a time. With the completion of phases I and II of the International Haplotype Map (HapMap) project, together with the development of efficient, affordable, high-density SNP genotyping technology, the whole genome scanning is an increasingly used to identify genetic risk factors for complex diseases. The Infinium II Whole-Genome Genotyping Assay is designed to interrogate a large number of SNPs at unlimited levels of loci multiplexing. Using a single bead type and dual colour channel approach, this panel genotypes from 240,000 to 550,000 SNPs per sample.

The infinium II whole genome genotyping consists of four modular steps: (1) whole genome amplification, (2) target capture to 50-mer probe array, the probes have been immobilized on beads which have been plated on the BeadChips, (3) array-based primer extension SNP scoring and (4) signal amplification and staining.

Typically, the 750 ng of each of DNA samples (Coriell institute for Medical Research, Philadelphia, PA, USA) was denatured with 0.1 M NaOH and then neutralized by 270μl of WG#-MP1 reagent (Illumina). The amplification was proceeded with 300 μl of WG#AMM (Amplification Master Mix) at 37^{o}C for 20 hours. The amplified samples were enzymatically fragmented with WG#FRG (Illumina) for 1 hour at 37^{o}C. The DNA samples were precipitated with 300 μl of 2-propanol and 100 μl WG#PA1 (Illumina). The mixtures were incubated at 4^{o}C for 30 minutes and centrifuged at 3000x g for 20 minutes at the same temperature.

42 μl of WG#RA1 (Illumina) was added to resuspend the precipitated DNA samples in hybridization buffer.

The fragmented, resuspended DNA samples were loaded on the humanhap 240, 330 or 550 BeadChips, which contains 25,789,740 tagged SNPs, 33,847,060 tagged SNPs or 59,341,039 tagged SNPs respectively for hybridization (16 hours, 48^{o}C). Following hybridization, WG#RA1 (Illumina) reagent was used to wash away unhybridized and non-specifically hybridized DNA sample. While WG#XB1 (Illumina) and WG#XB2 (Illumina) were added for preparing the Beadchips for the extension reaction. WG#EMM (Illumina) reagent was dispensed to extend primers hybridized to DNA on the BeadChips, incorporating labelled nucleotides. NaOH was used to remove the unhybridized DNA. After neutralization using the WG#XB3 (Illumina) reagent, the labelled extended primers underwent a multi-layer staining process. Finally, the Beadchips were washed in the WG#PB1 reagent (Illumina), then dried for 1 hours before they were transferred to the BeadArray Reader (Illumina).

The BeadArray Reader used a laser to excite the fluorescence of the allele-specifically extended product on the beads of the BeadChip sections. Light emissions from these fluorescent products were then recorded in high-resolution images of the BeadChip sections. Data from these images were analyzed to determined SNP genotypes using Beadstudio, genotyping software package (Illumina).

2.1.8 Statistical analysis in population genetic association studies

2.1.8.1 Hardy-Weinberg equilibrium

In population genetics, the Hardy–Weinberg equilibrium (HWE), states that, under certain conditions, after one generation of random mating, the genotype frequencies at a single gene locus will become fixed at a particular equilibrium value. It also specifies that those equilibrium frequencies can be represented as a simple function of the allele frequencies at that locus. In the simplest case of a single locus with two alleles **A** and **a** with allele frequencies of *p* and *q*, respectively, the HWE predicts that the genotypic frequencies for the **AA** homozygote to be p^2, the **Aa** heterozygote to be 2*pq* and the other **aa** homozygote to be q^2. 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.

2.1.8.2 Genetic association studies

2.1.8.2.1 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.

2.1.8.2.2 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 lies not below 1 and vice 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 stated otherwise.

2.1.8.2.3 Haplotype Analysis

A “haplotype” is a DNA sequence that has been inherited from one parent. Each person possesses two haplotypes for most regions of the genome. The most common type of variation among haplotypes possessed by individuals in a population is the SNP, in which different alleles are present at a given locus. Almost always, there are only two alleles at a SNP site among the individuals in a population. Given the likely complexity of trait determination, it is widely assumed that the genetic basis (if any) of important traits (e.g. diseases) can be best understood by assessing the association between the occurrence of particular haplotypes and particular traits.

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 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 or SHEsis 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 (degree of freedom) 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.

2.1.8.3 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.

2.1.9 Bioinformatics/Web resources

2.1.9.1 The National Centre for Biotechnology Information (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.

2.1.9.2 The University of California Santa Cruz genome browser (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.

2.1.9.3 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.

2.1.9.4 Ensembl genome browser

Ensembl is a joint project between European Molecular Biology Laboratory - European Bioinformatics Institute and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes. This project maintains a shared web-based program Ensembl 'BLASTView' which provides access to the WU-BLAST and SSAHA sequence similarity search algorithms via a single interface. It allows for simultaneous searches with up to 30 query sequences against multiple target species. Throughout the work in this thesis, we retrieve sequences homologous to the one of interest and this program was used routinely.

2.1.9.5 SHEsis

SHEsis (www.nhgg.org/analysis/) is a software plateform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. In haplotype analysis, this platform uses a Full-Precise Iteration algorithm, which could reconstruct ambiguous haplotypes and estimate haplotype frequencies in the given random sample set. For estimation of linkage disequilibrium (LD): Lewontin's D' (|D'|) and r^2 were calculated between each pair of genetic markers.

The SHEsis platform estimates haplotype frequency individually in controls and in cases to give the results of both single haplotype and a global data automatically.

2.2 Materials

2.2.1 PCR reagents

Taq DNA polymerase kit (Qiagen)
FastStart PCR Master (Roche)

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 and Fermentas.

2.2.2.3 Pyrosequencing reagents

Pyro Gold reagent (Biotage)

Binding Buffer

10mM Tris-HCl

2M NaCl

1mM EDTA

0.1% Tween20

1 X Annealing Buffer

20 mM Tris-Acetate

2mM MgAc2

Denaturation Solution

0.2 M NaOH

Washing Buffer

10 mM Tris-Acetate

2.2.2.4 Sequencing reagents

Big Dye Mix (Applied Biosystems)

BigDye® v1.1/3.1 Sequencing Buffer (5X) (Applied Biosystems)

2.2.3 Whole Genome Scanning Reagents (Illumina)

0.1 N NaOH

WG#MP1 (Neutralizing Reagent)

WG#AMM (Amplification Master Mix)

WG#FRG (Fragmentation Reagent)

WG#PA1 (Precipitation Reagent)

WG#RA1 (Resuspension Reagent)

100% Ethanol (American Bioanalytical)

100% Formamide (American Bioanalytical)

Iso-Propanol (American Bioanalytical)

Staining Reagents:

WG#XC1

WG#XC2

WG#XC3

WG#XC4

WG#TEM

WG#ATM

WG#LTM

2.2.4 Molecular biology reagents

TBE Buffer

Tris-borate pH 8.0 (Mediatech)

TE Buffer:

Tris-HCl pH 7.4

2.3 Suppliers

Suppliers of materials and services used throughout this study:

Alpha Innotech Corporation, 2401 Merced Street, San Leandro, CA 94577, USA.

American Bioanalytical, 15 Erie Drive, Natick MA 01760, USA.

Amersham Pharmacia Biotech, 800 Centennial Avenue, Piscataway, NJ 08855 USA.

Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Biotage, 1725 Discovery Drive, Charlottesville, VA 22911, USA.

Fermentas, 7520 Connelley Drive, Hanover, MD 21076, USA.

Illumina, 9885 Towne Centre Drive, San Diego, CA 92121, USA

Mediatech, 13884 Park Center Road, Herndon, VA 20717, USA

Millipore, 290 Concord Road, Billerica, MA 01821, USA.

Nanodrop, 3411 Silverside Rd, Bancroft Building, Wilmington, DE 19810, USA.

New England Biolabs, 240 County Road, Ipswich, MA 01938, USA.

Qiagen, 27220 Turnberry Lane, Valencia, CA 91355, USA.

Roche, 9115 Hague Road, Indianapolis, IN 46250, USA.

Chapter 3 The architecture of the tau haplotype

3.1 Overview

The microtubule associated protein, tau is the major component of the fibrillar aggregates which are found in a number of neurodegenerative disorders, including AD, PSP, CBD and FTDP-17 [129]. That tau dysfunction plays an important role in neurodegeneration is affirmed by the discovery of mutations in the *MAPT* gene causing autosomal dominant disease [127].

In our study, the *MAPT* locus is found to be very unusual. It appears as two distinct haplotype clades, H1 and H2, over a region of approximately 1.8Mb. These two haplotype clades H1 and H2 were found only in European/Caucasian populations [130;131]. In other populations, only the H1 occurs and shows a normal pattern of recombination [59;132]. The H2 haplotype shows remarkably little genetic variation and differs from the H1 haplotype in both sequence and in terms of the orientation of several elements of the locus. Presumably, these differences prevent recombination between the heterologous clades [60;133].

Understanding the architecture and distribution of these haplotypes is important, both for an understanding of population genetics and history and to develop an understanding of the pathogenesis of neurodegenerative disorders, such as PD and AD. Therefore we have assessed the distribution of the *MAPT* H1/H2 haplotype in different racial groups worldwide and the pattern of the extended haplotype block over the *MAPT* gene in different ethnicities.

3.2 Background

The *MAPT* locus is unusual in that there appear to be two distinct haplotype clades covering the tau gene, *MAPT*, and the surrounding genetic material. This locus contains several other genes besides *MAPT* [131] (**Figure 3.1**). The H1 haplotype clade is the most common, having an allele frequency of about >70% in European populations [130]. There appears to be no recombination between H1 and H2 haplotype clades over a region of ~1.8Mb although it is likely that recombination occurs between different H1 haplotypes and possibly also between different H2 haplotypes [131]. The 238 bp deletion between exons 9 and 10 was found in the H2 haplotype exclusively; this insertion/deletion polymorphism was denoted as *del-In9* and used as a haplotype-defining marker. [130].

We have been interested in the *MAPT* locus because it is a susceptibility locus for diseases with tau pathology such as neurofibrillary tangles, including PSP [130] and CBD [67], and possibly also including Parkinson dementia complex of Guam (PDC) [134], a devastating epidemic tangle disease which, at one time, was the major cause of death in South Guam, but has now virtually disappeared [135-137].

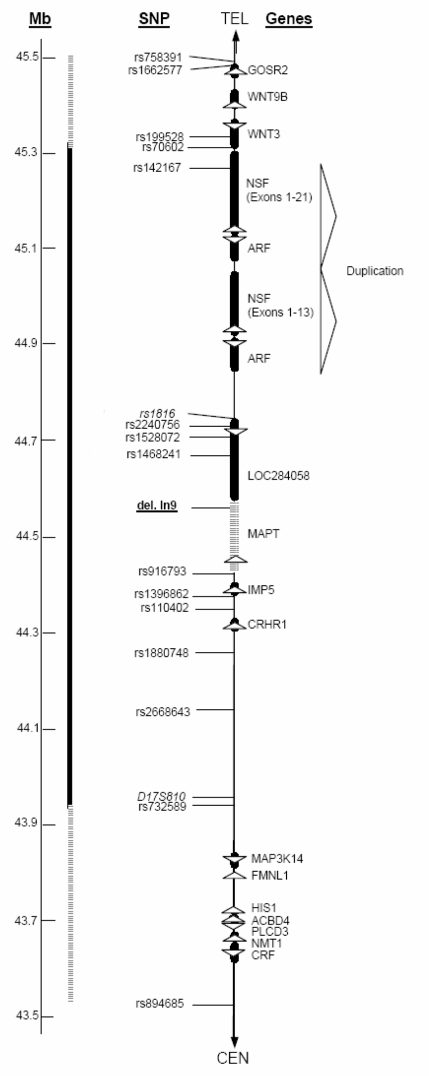


Figure 3.1 The extended haplotype block at 17q21.31
The region of chromosome 17q21.31 containing the extended *MAPT* haplotype block. The chromosomal coordinates (Mb; million base pairs) are indicated on the left hand axis. They are based upon the July 2003 draft of the human genome sequence. Relative positions of the SNPs and confirmed genes are indicated. Arrowheads on genes indicate the direction of transcription. CEN,centromeric; TEL,telomeric.

With this background, we genotyped *MAPT* for the haplotype-defining insertion/deletion (*del-In9*) polymorphism in the *MAPT* region, five SNPs flanking this particular *MAPT* haplotype block in the CEPH diversity panel and eleven SNPs which differentiate the *MAPT* haplotype H1 from that of H2 in the primate panel. The aim of this study was the understanding of the architecture and distribution of these haplotypes and an understanding of population genetics and evolution to reveal the pathogenesis of these neurodegenerative diseases.

3.3 Methods and Materials

3.3.1. Samples

In this study, the DNAs used were from the CEPH panels that have been previously described [138], a panel of 30 controls from Guam who were age-matched for our Guam Parkinson Dementia Complex study (mean age 75 years) and 150 controls from Finland who were age-matched for the PD study and genotype described in Chapter 5. (mean age 70 years) [139;140]

A set of primate DNAs including Chimpanzee, Gorilla, Gibbon, Marmoset, Orangutang, Owl Monkey, Cynomolous Monkey and African Green Monkey from European Collection of Cell Cultures (http://www.ecacc.org.uk) were also used in this part of the study.

3.3.2. Methods

3.3.2.1 Genotyping of the H1/H2 haplotype-defining insertion/deletion polymorphism in *MAPT* intron 9 (*del-In9)*

For the insertion/deletion polymorphism, genotyping was carried out by running PCR products using primers flanking the deletion, on an agarose gel, visualized with ethidium bromide staining and photographed with a Polaroid camera for genotype scoring.

3.3.2.2 Genotyping of the SNPs flanking the *MAPT* haplotype block and the differentiating SNPs of the *MAPT* haplotype H1/H2

Five SNPs, rs758391, rs1662577, rs70602, rs2668643 and rs894685 were chosen to flank the telomeric and centromeric ends of the *MAPT* haplotype block. Seven SNPs, including rs1801353, rs1047833, rs393152, rs1052553, rs7687, rs2240758 and rs199533 were selected to differentiate the *MAPT* H1 and H2 haplotypes in comparison between human and primate genomes. The genotyping of these SNPs were analyzed by restriction enzyme digestion or Pyrosquencing as described in the Methods section (Chapter 2). Furthermore, eleven SNPs, including rs1864325, rs1560310, rs1984937, rs767058, rs754512, rs733966, rs1078830, rs2055794, rs2217394, *STH Q7R* [158] and rs9468 were chosen to compare the *MAPT* H1 and H2 haplotypes between human and primate genomes from MIT Chimp Assembly (November 2003) (www.ncbi.nlm.nih.gov).

3.3.2.2.1 Pyrosequencing

The SNPs rs70602, rs1801353, rs1047833, rs393152, rs1052553, rs7687, rs2240758 and rs199533 were analysed by Pyrosequencing. Samples were analysed using a PSQ 96 System together with SNP Software and SNP reagent kits (Pyrosequencing Inc., Biotage, Charlottesville, VA) following the manufacturer's instructions (Section 2.1.5.2).

3.3.2.2.2 Restriction fragment length polymorphism

For SNP analysis by RFLP, 15 μl of PCR product was digested by 1 unit of the corresponding restriction endonuclease (rs758391 [Hph I (A)]; rs1662577 [BsrG I (C)]; rs2668643 [Apo I (A)] and rs894685 [Acc I (T)]). 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.

3.3.2.3. Determination of Linkage Disequilibrium

Linkage disequilibrium (LD) was determined by the "UNPHASED" program. The LD was calculated pair-wise, using the statistical LD calculations for *D*' and r^2 from the expectation-maximization (EM)-derived haplotypes.

3.4 Results

The geographical distribution of *the MAPT* haplotyoes are described and illustrated graphically in **Table 3.1 and Figure 3.2**. It was surprising to note that the H2 haplotype is found almost exclusively in populations with European/Caucasian ancestry, with middle eastern and European populations having H2 allele frequencies of ~25%, central Asian (including Finnish) populations ~5% and other populations (African, East Asian and native American) having H2 allele frequencies of essentially zero.

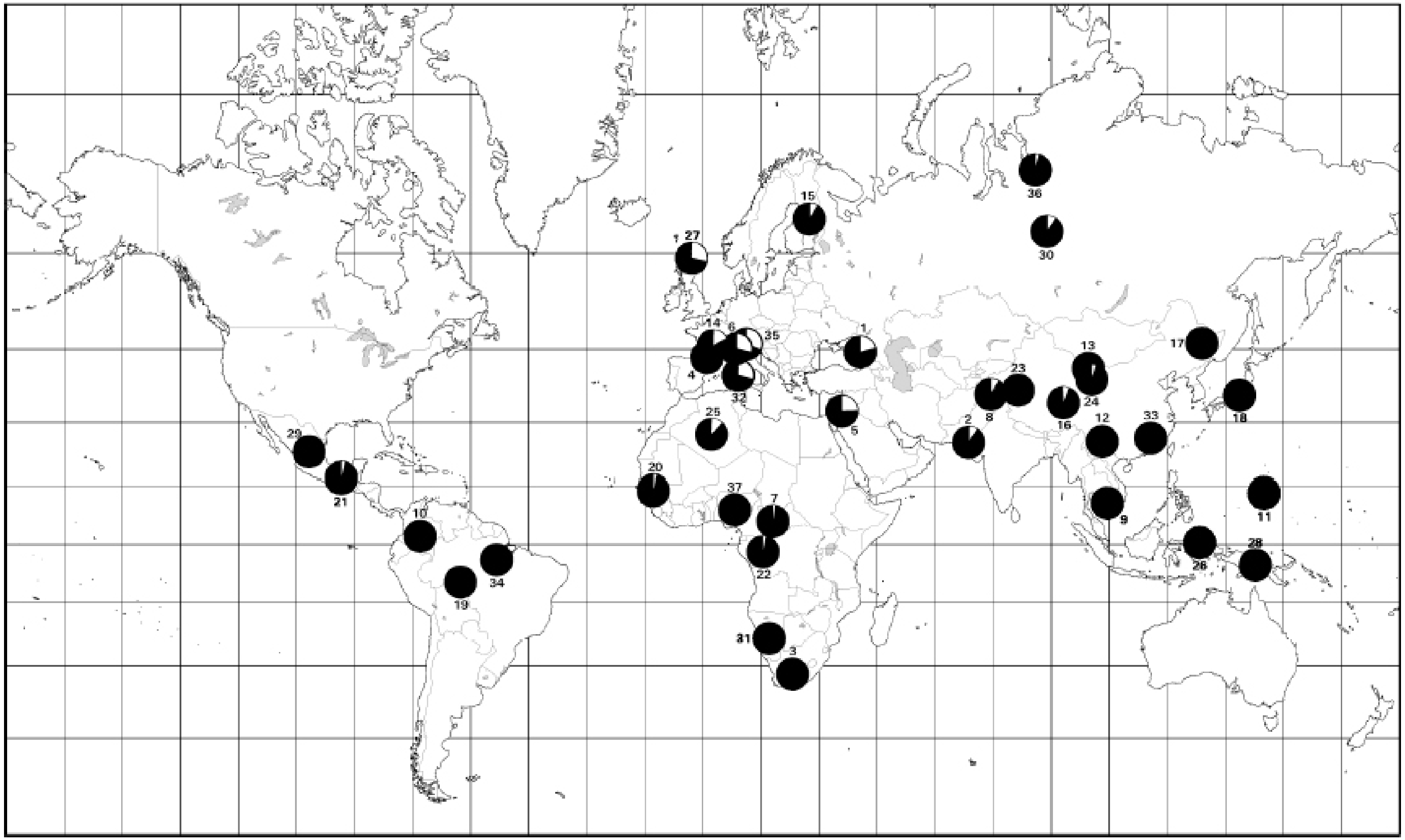


Figure 3.2 Worldwide Tau H1/H2 Distribution

Diagram showing geographic distribution of the tau locus. H1 is black and H2 is white segment.

No.	Population (number of subjects)	H1	H2
1	Adygei(17)	79	21
2	Balochi(25),Brahui(36),Makrani(25),Sindhi(21)	90	10
3	Bantu(20)	100	0
4	Basque(24)	73	27
5	Bedouin(49),Druze(10),Palestinian(51)	76	24
6	Bergamo(29)	71	29
7	BiakaPygmy(14)	97	3
8	Burusho(25),Hazara(25),Kalash(25),Pathan(25)	93	7
9	Cambodian(25)	100	0
10	Colombian(11)	100	0
11	Chamorro(85)	94	6
12	Dai(13),Lahu(10),Naxi(10)	100	0
13	Daur(10)	100	0
14	French(48)	83	17
15	Finnish(138)	92	8
16	Han(45)	94	6
17	Hezhen(10),Oroqen(10),Tu(10)	100	0
18	Japanese(31)	100	0
19	Karitiana(24)	100	0
20	Mandenka(24)	96	4
21	Maya(25)	96	4
22	Mbuti Pygmy(15)	97	3
23	Miaozu(10),Uygur(9),Xibo(25)	100	0
24	Mongola(10)	95	5
25	Mozabite(30)	87	13
26	NAN Melanesian(22)	100	0
27	Orcadian(16)	72	28
28	Papuan(17)	100	0
29	Pima(25)	100	0
30	Russian(7)	90	10
31	San(28)	100	0
32	Sardinian(10)	71	29
33	She(25),Tujia(8),Yizu(25)	100	0
34	Surui(10)	100	0
35	Tuscan(10)	69	31
36	Yakut(10)	96	4
37	Yoruba(25)	100	0

Table 3.1 Worldwide *MAPT* H1/H2 distribution
The *MAPT* H1/H2 haplotype distribution in different populations worldwide is shown. The number shown is the percentage of the total number of samples tested in each population.

The region of complete LD was noted in five of the ethnically different populations, including Italians, Pakistani, French, Orkney Islanders, and Russians in the CEPH panel and the British population in our previous report [131]. In those populations, the block of LD extended from the chromosomal coordinate of 45334515 (rs70602) telomerically to that of 44146521 (rs2668643) centromerically relative to *MAPT*. (the coordinates are based on the July 2003 draft of the human genome sequence which was lastest version when this study was carried out.) The coordinates and the allele frequencies of these SNPs are illustrated in **Table 3.2** and **Figure 3.3**. Meanwhile, SNPs rs758391, rs1662577, and rs894685 showed very low D’ with *del-In9* indicating they fall outside the limits of the haplotype block (**Figure 3.4**). At the telomeric end of the haplotype block, we observed only 12 out of 186 individuals who were recombinant. It is noteworthy the recombinants are almost exclusively found in the H2 haplotype block where the mechanism of this rare event is under investigation. This phenomenon is noted in the Italians, Pakistani, British, and French.

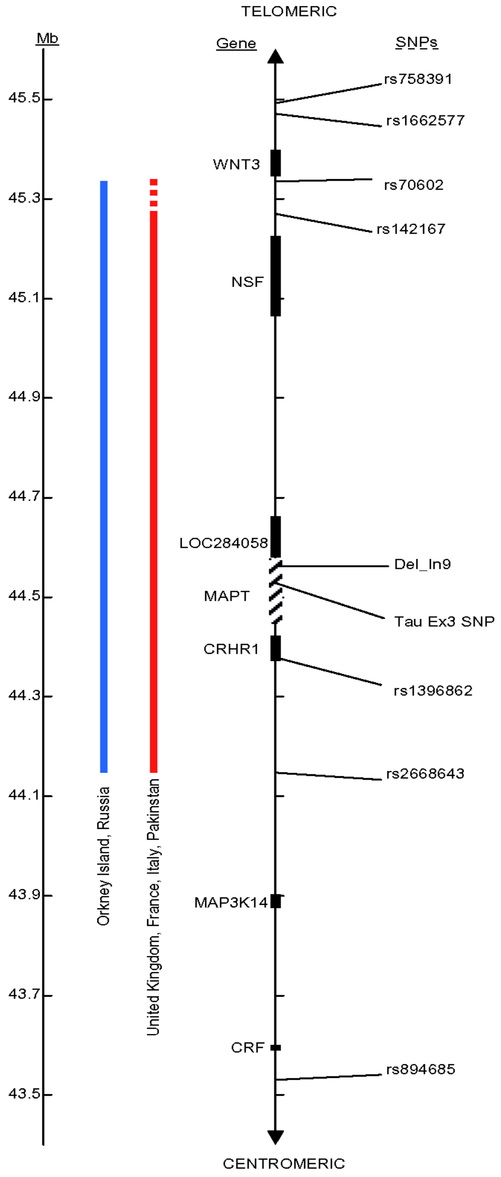
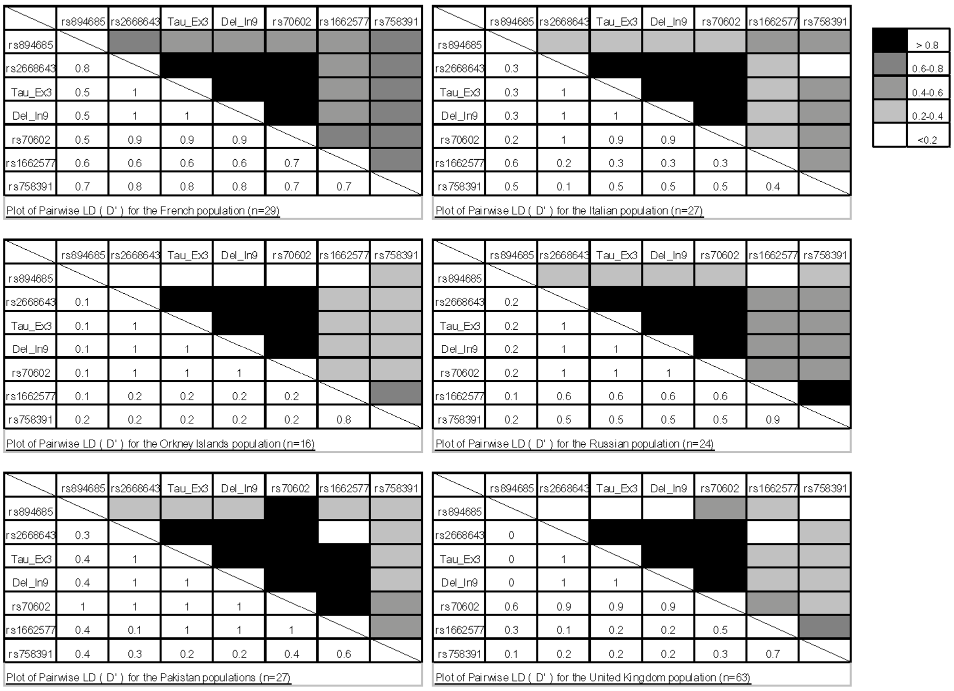


Figure 3.3 The architecture of the tau haplotype block in different ethnicities The haplotype block of *MAPT* region from rs2668643 (centromeric) to rs70602 (telomeric) in six different ethnic groups. The solid bar (red and blue) shows the complete LD throughout the region. The red dashed bar shows the region that recombination may have occurred in given populations. The chromosomal coordinations (Mb: million base pairs) are indicated on the left-hand axis.

			Minor Allele Frequency of SNP						
Ethnicity	Population	Number of Subjects	rs894685	rs2668643	Tau Ex3 SNP	del_In9	rs70602	rs1662577	rs758391
Italy	Sardinian	27	0.19	0.30	0.28	0.28	0.24	0.33	0.39
Orkney Island	Orcadian	16	0.31	0.28	0.28	0.28	0.28	0.44	0.41
Russia	Russian	24	0.19	0.10	0.10	0.10	0.10	0.40	0.35
Pakistan	Makrani	27	0.57	0.15	0.11	0.11	0.07	0.35	0.37
France	French	29	0.17	0.25	0.21	0.21	0.17	0.48	0.50
United Kingdom	British	63	0.19	0.19	0.20	0.20	0.17	0.42	0.50

Table 3.2 The allele frequencies of the SNPs used in the analysis of the *MAPT* haplotype block in different ethnicities.

Comparison between primate and human genomes over the *MAPT* region revealed that according to the haplotype-defining insertion/deletion polymorphism marker (*del-In9),* and SNPs rs1801353, rs1047833, rs1864325, rs1560310, rs767058, rs754512, rs733966, rs2240758 and rs199533, the chimpanzee has the H1 variant. On the other hand, with the rs393152, rs1078830, rs2055794, rs2217394, rs1052553, *STH Q7R*, rs9468 and rs7687, the primate sequence corresponds to the human H2 sequence (November 2003 MIT Chimp Assembly) suggesting that the evolution of this locus has been complex.



	rs894685	rs2668643	Tau_Ex3	Del_In9	rs70602	rs1662577	rs758391
rs894685		▓	▒	▒	▒	▒	▓
rs2668643	0.8		■	■	■	▒	▓
Tau_Ex3	0.5	1		■	■	▒	▓
Del_In9	0.5	1	1		■	▒	▓
rs70602	0.5	0.9	0.9	0.9		▓	▓
rs1662577	0.6	0.6	0.6	0.6	0.7		▓
rs758391	0.7	0.8	0.8	0.8	0.7	0.7	

Plot of Pairwise LD (D') for the French population (n=29)

	rs894685	rs2668643	Tau_Ex3	Del_In9	rs70602	rs1662577	rs758391
rs894685		░	░	░	░	▒	▒
rs2668643	0.3		■	■	■	░	
Tau_Ex3	0.3	1		■	■	░	▒
Del_In9	0.3	1	1		■	░	▒
rs70602	0.2	1	0.9	0.9		░	▒
rs1662577	0.6	0.2	0.3	0.3	0.3		▒
rs758391	0.5	0.1	0.5	0.5	0.5	0.4	

Plot of Pairwise LD (D') for the Italian population (n=27)

■	> 0.8
▓	0.6-0.8
▒	0.4-0.6
░	0.2-0.4
	<0.2

	rs894685	rs2668643	Tau_Ex3	Del_In9	rs70602	rs1662577	rs758391
rs894685							░
rs2668643	0.1		■	■	■	░	░
Tau_Ex3	0.1	1		■	■	░	░
Del_In9	0.1	1	1		■	░	░
rs70602	0.1	1	1	1		░	░
rs1662577	0.1	0.2	0.2	0.2	0.2		▓
rs758391	0.2	0.2	0.2	0.2	0.2	0.8	

Plot of Pairwise LD (D') for the Orkney Islands population (n=16)

	rs894685	rs2668643	Tau_Ex3	Del_In9	rs70602	rs1662577	rs758391
rs894685		░	░	░	░		░
rs2668643	0.2		■	■	■	▒	▒
Tau_Ex3	0.2	1		■	■	▒	▒
Del_In9	0.2	1	1		■	▒	▒
rs70602	0.2	1	1	1		▒	▒
rs1662577	0.1	0.6	0.6	0.6	0.6		■
rs758391	0.2	0.5	0.5	0.5	0.5	0.9	

Plot of Pairwise LD (D') for the Russian population (n=24)

	rs894685	rs2668643	Tau_Ex3	Del_In9	rs70602	rs1662577	rs758391
rs894685		░	░	░	■	░	░
rs2668643	0.3		■	■	■		░
Tau_Ex3	0.4	1		■	■	■	░
Del_In9	0.4	1	1		■	■	░
rs70602	1	1	1	1		■	▒
rs1662577	0.4	0.1	1	1	1		▒
rs758391	0.4	0.3	0.2	0.2	0.4	0.6	

Plot of Pairwise LD (D') for the Pakistan populations (n=27)

	rs894685	rs2668643	Tau_Ex3	Del_In9	rs70602	rs1662577	rs758391
rs894685					▒	░	
rs2668643	0		■	■	■		░
Tau_Ex3	0	1		■	■	░	░
Del_In9	0	1	1		■	░	░
rs70602	0.6	0.9	0.9	0.9		▒	░
rs1662577	0.3	0.1	0.2	0.2	0.5		▓
rs758391	0.1	0.2	0.2	0.2	0.3	0.7	

Plot of Pairwise LD (D') for the United Kingdom population (n=63)

Figure 3.4 Pair-wise D' LD analysis of the different ethnic populations.
The blocks are shaded corresponding to the values which were obtained from the LD analysis program UNPHASED.

Position	dbSNP ID	Chimp	H1	H2
43795192	rs1801353	C	C	T
43817563	rs1047833	G	G	C
44194565	rs393152	G	A	G
44421540	rs1078830	C	T	C
44427146	rs2055794	A	G	A
44453262	rs1864325	C	C	T
44453969	rs1560310	G	G	A
44455965	rs1984937	C	T	G
44474234	rs767058	C	C	G
44528923	rs2217394	G	A	G
44531122	rs754512	T	T	A
44549365	rs1052553	G	A	G
44552141	*STH Q7R*	G	A	G
44562127	*del-In9*	+	+	-
44565039	rs733966	C	C	T
44577047	rs9468	C	T	C
44578781	rs7687	C	T	C
44723915	rs2240758	C	C	G
45103737	rs199533	C	C	T

Table 3.3 Polymorphisms in the extended *MAPT* locus that differentiate H1 clades from H2 and comparison with the chimp assembly.

del-In9 is the H1/H2 haplotype-defining insertion deletion polymorphism in intron 9 of the *MAPT* gene and *STH Q7R* is the saitohin gene polymorphism that is in complete LD with *del-In9*. All positions are given relative to Build 34 (July 2003) of the Human Genome.

3.5 Discussion

The distribution of H2 haplotype exclusively in Caucasians is of interest for three reasons; first, it raises the question of the origin of the H2 haplotype, since it appears to be so divergent from the H1 haplotype and that recombination does not occur with the H1 haplotype over ~1.8Mb [82;131]; second, it suggests that the *MAPT* haplotype can be used as an approximate population marker for Caucasian ancestry in admixed populations and third, it leaves the possibility open that populations with more H1 individuals could have a higher incidence of H1-associated diseases, especially, PSP and CBD.

The relative constancy of the H2 allele frequency in Caucasian populations from the Middle East to the Orkneys suggests that its origin in European populations is ancient and coincides with the colonization of Europe. The lower frequency in the Finnish population is consistent with previous genetic data showing that this group has a substantial genetic contribution from Asian populations [141]. It is difficult to envisage the origin of the H2 haplotype, either as a mutational event or as a result of admixture with an earlier human population: the divergence of the two haplotypes is extensive and suggests considerable genetic separation. This could reflect either a mutation preventing recombination, and thus maintaining the separate integrity of the haplotype, or the ancient genetic isolation of a group of humans in which this haplotype occurred.

Determining whether the difference in H2 haplotype frequency might lead to a difference in the incidence of neurodegenerative disease is fraught with difficulty at two levels: first cross-cultural comparisons of incidence of late-onset neurodegenerative diseases is notoriously difficult and second, it is not yet clear whether the H1 haplotype *in toto*, or some variant of it, is responsible for the association; recent data would suggest the latter [142] and, if this were the case it would make straightforward predictions concerning incidence, difficult until the pathogenic variant(s) are precisely resolved. Nevertheless, a direct prediction of disease incidence for these tauopathies based on H1 allele frequencies would be that non-Caucasians would have an incidence approximately double that of Caucasians since those populations would have nearly twice as many H1 homozygotes.

In the study of the *MAPT* haplotype block in different ethnicities, the complete LD region was noted in five of the ethnically different populations, including Italians, Pakistanis, French, Orkney Islanders, and Russians, in the CEPH panel and the British population. The same LD pattern over *MAPT* region is shown by different ethnic groups in the diversity panel confirms that this particular LD block is shared between populations indicating that haplotype structure in human is ancient, predating the separation of Caucasians.

This pattern of LD strongly suggests that the formation of the H2 haplotype was a single event either indicating a chromosomal rearrangement [59;131;133] or limited intermixing with a predating population [143]. Given this, the high prevalence (~25%) of the H2 haplotype in Caucasians is surprising and may suggest a strong selection for the H2 haplotype. Of course, the H2 haplotype is protective against both PSP and CBD [67;130], but these diseases are far too rare, and too late in lifespan to have had significant impact on allele frequencies. However, there are occasions when tauopathies could become major causes of mortality: on Guam, in Umatac, Parkinson–dementia complex was the major cause of death in the 1930 to 1950s [129;135], the Spanish flu epidemic in 1919 led to an epidemic of postencephalitic Parkinsonism: von Economo’s syndrome [129;144;145] and subacute sclerosing panencephalitis is a, now rare, but frequently fatal complication of measles infection [129;146;147]. In the latter two diseases, there has been no study of the *MAPT* haplotype, and in Guam the H2 haplotype is so rare that there is not sufficient evidence to show any association between PDC and the H2 haplotype of *MAPT*. [59;134;140].

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, while both H1 and H2 sequences are more similar to each other than to the chimp sequence, they do not follow a predictable relationship: at some sequences, the chimp sequence is similar to H1 and at others, it is similar toH2 (**Table 3.3**, and also [148;149]). Thus the H1 and H2 sequences do not follow a precursor–product relationship and one cannot be derived directly from the other, rather both must have been derived independently from a more distant precursor. Logically, therefore their relationship could be as illustrated in **Figure 3.5**.

During the time of the writing of this thesis, Stefansson et al, from the deCODE group, has found that a ~900 kb inversion polymorphism at the region 17q21.31 in H2 chromosomes with respect to the H1 haplotype [60]. Jaime Duckworth from our group also used the bioinformatic data to show similar findings on the structures in these distinct haplotypes. Chromosomes with the inverted segment in different orientations represent two distinct lineages, H1 and H2, that have diverged for as much as 3 million years and show no evidence of having recombined [60]. This size of this inversion, 900-kb, is smaller than the linkage disequilibrium block, that is ~1.8 Mb. The mechanism that gives rise to this discrepancy of the sizes had not yet been established. The proposed hypothesis for this inconsistency of the sizes is the inverted, non-complementary segments, from H1 and H2 clades, over the *MAPT* not only makes the recombination and crossing over between two haplotype clades impossible; but also extends beyond the ends of this haplotype clade as the non-complementary segment may repel each other and a "recombination bleb" may be formed to hinder the exchange of genetic materials between two chromatids. A number of low-copy repeat (LCR) sequences identified in this region, and their complex architecture also suggest there could be different break points within or beyond the break-points which results in different sizes of the chromosomal rearrangement including deletion and/or reciprocal duplication.

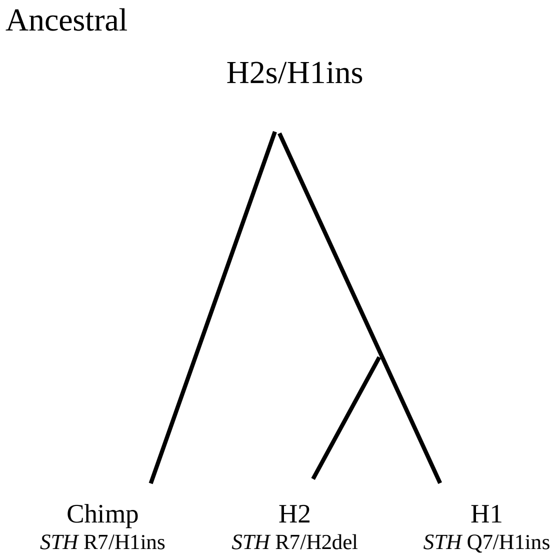


Figure 3.5 Parsimony Tree of Relationships between Chimp MAPT Locus and H1 and H2 Haplotypes.

Parsimony tree showing relationship between the saitohin (*STH*-Q7R) and the *del-In9* polymorphisms in the *MAPT* locus indicating that the H1 and H2 variants of these are more likely to have derived from a common founder than that either H1 or H2 is the predecessor of the other. H1 haplotypes carry *STH*-Q7 alleles and H1 insertion, H2 haplotypes carry *STH* R7 alleles H2 deletion. The same diagrams could be drawn for the other polymorphisms in Table 3.3. (H1ins: H1 insertion; H2del: H2 deletion)

Chapter 4 Genetic Association of *MAPT* haplotypes with progressive supranuclear palsy and corticobasal degeneration

4.1 Overview

The haplotype H1 of the tau gene, *MAPT*, was found to be highly associated with progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) [130]. In order to investigate the pathogenic basis of this association, the association of *MAPT* with PSP and CBD based on the underlying haplotype architecture of *MAPT* was refined. The common haplotype structure of *MAPT* and associations with these related tauopathies were also explored.

Detailed linkage disequilibrium (LD) architecture and common haplotype structure of *MAPT* were examined in 27 CEPH-trio individuals. Based on this, 5 htSNPs were identified that capture 95% of the common haplotype diversity of the region. These, together with the del-In9 polymorphism to define the H1/H2 division, were used to genotype well characterised PSP and CBD case-control cohorts.

Two common haplotypes defined by the htSNPs and del-In9 were identified to be associated with PSP, 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. The sole H2-derived haplotype is under-represented and, one of the common H1-derived haplotypes is highly associated, with a similar trend observed in CBD. We also observed particularly powerful and highly significant associations with PSP and CBD of haplotypes formed by 3 H1-specific SNPs. These findings 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.

4.2 Background

PSP 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 [55;150;151]. CBD is an atypical parkinsonian condition occurring much less frequently than PSP and 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 [57;58] 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 [51]. The overrepresentation of the commoner allele (a0) in PSP and also later in CBD was then confirmed by a number of groups [66;67]. This suggests that either 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-17 cause a clinical picture closely resembling PSP [152-154], no pathogenic variations of *MAPT* have yet been identified in clinically and pathologically diagnosed sporadic and familial PSP [155].

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). In approximately 200 unrelated Caucasians, these polymorphisms were in complete LD, forming two extended haplotypes H1 and H2. The study demonstrated that the more common haplotype, H1, with which the a0 allele segregated, was significantly over-represented in PSP [130]. Follow up 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 [156;157]. This was then extended extended to a ~1.8Mb haplotype which is in near complete LD [131]. This region associated with PSP includes several other genes in addition to *MAPT*, including saitohin [158;159] (*STH;* situated within intron 9 of *MAPT*), *NSF* (*N*-ethylmaleimide sensitive factor), *IMP5* (intramembrane protease 5, a presenilin homologue) [160], *CRHR1* (corticotrophin releasing hormone receptor) and LOC284058, an unknown gene just adjacent to *MAPT*. (**Figure 3.1**) 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 multi-gene haplotype block are associated with PSP and CBD, the hallmark tau pathology of these disorders strongly implicates *MAPT* itself.

The objective of the work in this chapter was therefore to exhaustively analyse the *MAPT* haplotype association with PSP and CBD in order to identify non-coding variants that could affect *MAPT* gene expression, splicing or processing, leading to tau pathology and selective neuronal loss.

The findings in this section of the study were based on the results from a close collaborative work with Alan Pittman. The findings have been published in Journal of Medical Genetics (2005) [177]. With the framework of the association study between H1c clade of *MAPT* and PSP and CBD, we went further to investigate whether these genetic traits also associated with other neurodegenerative diseases. The important findings in this section are an inevitably important foundation of the whole project. Therefore, the details of the study have been put in here for a complete picture of the study.

4.3 Analysis of MAPT haplotype structure in the CEPH-trios

SNP data for the region of the *MAPT* locus in 27 CEPH trios (Coriell Institute for Medical Research; http://locus.umdnj.edu/nigms/) from the International HapMap project (HapMap) web site (http://www.hapmap.org/), was downloaded for genetic analysis of the *MAPT*. The raw SNP genotype data was 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) [123].

Initially, any SNPs that had a minor allele frequency of less than 5% were removed from the HapMap data. The inconsistencies in the data through the parental-offspring relationship in the CEPH-trios were also checked. A resulting set of 24 SNPs and the *del-In9* (**Table 4.1**) was used, they cover 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) ($\sum = 1x10^{-6}$) algorithm specifically for CEPH-trio material (EM-trio) [123]. The average density of the markers was one SNP every 6.7 kb. For convenience, the bi-allelic (+/-) intron 9 deletion-insertion polymorphism (*del-In9*) was designated as a SNP. A total of 34 haplotypes were resolved from parental chromosomes. The pair-wise 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 firstly by estimating pair-wise haplotype frequencies through EM-trio, then assessing the statistical strength of association via 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*(A1B1) – *f*(A1)*f*(B1) [99]. A and B refer to two genetic markers and *f* is their frequency. D' is obtained from D/DMAX 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 viewed to be the preferred measure of LD for association based studies [123].

SNP	Position[a]	dbSNP ID	Alleles	Ancestral	F1[b]	F2[b]	p-value[c]
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	41375573	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

Table 4.1 The 24 single nucleotide polymorphisms and del-ln9 used for the linkage disequilibrium and haplotype structure analysis of MAPT in the CEPH trios

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 available genotype data for these SNPs 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).

[a] SNP position on chromosome [b] allelic frequencies in the CEPH-trios
[c] *p* values for test to fit Hardy-Weinberg equilibrium

Allelic and genotype frequencies followed by statistical assessment of Hardy-Weinberg equilibrium (HWE) 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. Six tagging SNPs (*del-In9*, rs1467967, rs242557, rs3785883, rs2471738 and 7521) were selected, then, using TagIT, their performance was assessed on the CEPH-trios. The tagging approach was focused on the coefficient of determination (i.e. haplotype, r^2) in a linear regression, which uses the haplotypes defined by the htSNPs to predict the state of the tagged-SNPs. 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: 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 data set 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.

4.4 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 all pathologically confirmed. The majority of these cases have been used in previous studies [131;155;157;158;161]. Pathological confirmation of the diagnosis of PSP was made following standardized criteria [161]. 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.

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

4.5 Genotyping

The htSNPs (dbSNP numbers: rs1467967, rs242557, rs3785883, rs2471738 and rs7521 and *del-In*9) were genotyped in the PSP case-control cohorts as follows: The 238bp MAPT *del-In*9 was genotyped as in previous chapter (Chapter 3). PCR primer pairs 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: 10µl reactions, which contained one unit of DNA polymerase (Qiagen, Crawley, West Sussex), 10x PCR reaction buffer, 5x Q solution (Qiagen), 10 pmoles of each oligonucleotide primer pair and 25ng of sample template genomic DNA.

Genotyping of the SNPs, rs1467967, rs242557, rs3785883, rs2471738 and rs7521 were conducted by Pyrosequencing (Biotage AB) or by restriction digest (RFLP);

the following restriction endonucleases cutting the PCR product once at the (N) allele, respectively: Dra I (A), ApaL I (A), BsaH I (G), BstE II (T) and Pst I (A) (New England Biolabs and Fermentas). 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 visualized with ethidium bromide staining.

Genotyping accuracy was assessed by re-typing 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/).

4.6 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 HWE were made using TagIT. Case-control single-locus htSNP allelic and genotypic association was calculated statistically in CLUMP software [162]. 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. The heterogeneity between the H1/H1 homozygote populations versus the whole population was tested using a standard Pearson's χ^2 test.

Distribution 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 a SNP haplotype analysis suite that performs a regression-based haplotype association test through a LRT, which is χ^2 and n-1 degrees of freedom to derive the associated p-value, where n is the number of haplotypes observed for the data. This test was used to give an initial assessment of haplotype association (an omnibus test) and then individual haplotype tests (haplotype-specific tests) of association were performed again through a LRT (d.f = 1) and by also obtaining empirical p-values by Monte-Carlo methods (20,000 simulations used). To test the effect of the H1-specific htSNPs whilst controlling for the extended H1/H2 haplotype, a set of equality constraints was imposed under the null across the haplotypes identical at the *del-In9* and single-locus and haplotype analysis was performed. The p-values were corrected according to the number of tests performed where appropriate by the Bonferroni correction, the significance of which is discussed throughout the text.

4.7 Results

4.7.1 Linkage disequilibrium and haplotype structure of *MAPT*

The average density of the markers is one SNP every 6.7 kilobases (kb). None of the polymorphisms deviated from Hardy-Weinberg equilibrium (HWE). The details of all SNPs analysed in the CEPH-trios was summarized in Table 4.1. Pairwise LD was evaluated across *MAPT* for all 24 selected SNPs and *del-In9* in the 27 CEPH-trios both by D' and r^2, calculated from the Expectation–Maximisation-trio (EM-trio) inferred haplotypes. By pairwise LD analysis of the 25 SNPs in CEPH-trios, a greater diversity was identified than reflected by the description of the two extended H1 and H2 haplotypes alone. The entire *MAPT* gene is featured by significant LD as is particularly evident by the measure of D' (**Figure 4.1**). However, when LD was assessed by the more stringent measure of r^2 (that 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 define haplotypes within the H1 clade. LD correlation by D' between many of the described H1-specific SNPs is relatively low, 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 with smaller blocks in the Taiwanese population, in which the extended H2 haplotype is absent (result is shown in Chapter 6).

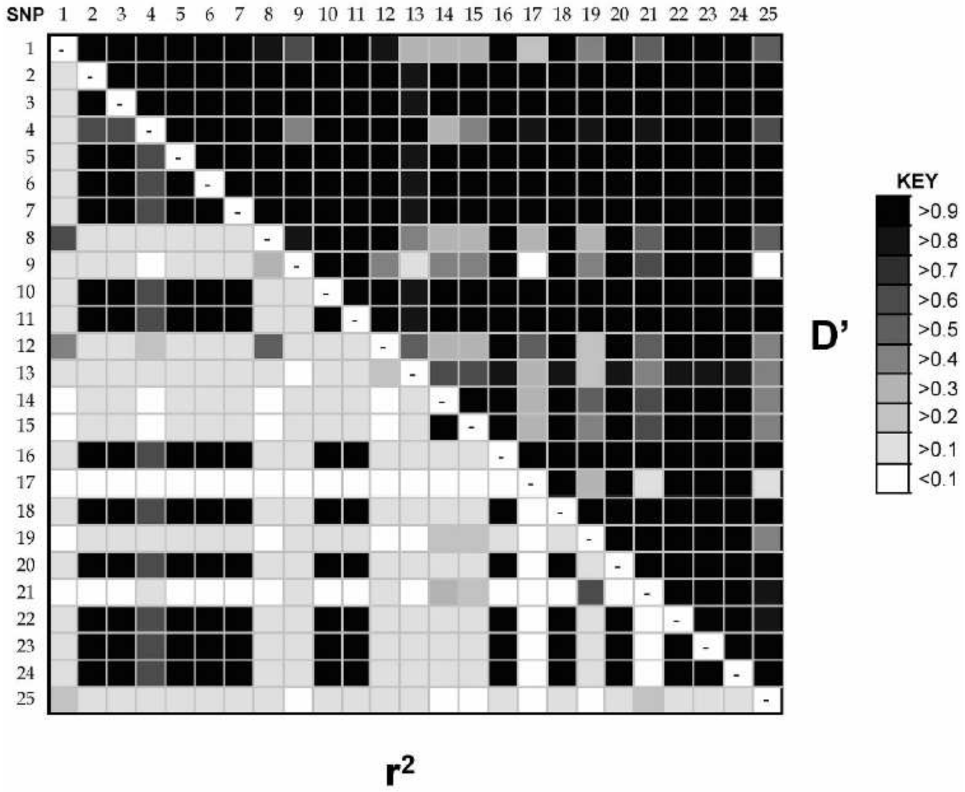


Figure 4.1 Linkage Disequilibrium (LD) across the *MAPT* gene. Numerical LD is presented by grey-scale, pair-wise between each SNP by both D' (upper right) and and the more stringent measure r2 (bottom left). The darker the shading indicates a higher extent of LD between the SNPs. SNPs are numbered as in Table 4.1.

The EM-inferred *MAPT* haplotypes and their respective frequencies were obtained by using the EM estimation algorithm specifically tailored to deal with trio data (EM trio) as structured in the CEPH-trios [123]. The phased haplotypes were also obtained (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%. 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 4.2**).

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 that differed from haplotype A by SNP 13 (**Table 4.2**). However, this haplotype was not predicted by EM-trio for output as a significant frequency in the population and represents 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. The ancestral (chimpanzee) haplotype was also constructed based upon the alleles of the 24 SNPs and the *del-In9*. This appears not to resemble any haplotype present in the CEPH-trios, though its closest relative (but different by ten loci) would appear to be that of the extended H2 (CEPH-trio haplotype A, from **Table 4.2**). The other ancestral SNP loci are either consistent with the H1 haplotype family (SNPs rs962885, rs1864325, rs1560310, rs1467970, rs1001945, rs754512 and rs733966), including the presence of the 238 bp insertion sequence (*del-In9*), or the allele is not observed in *Homo sapiens* (dbSNPs rs3885796 and rs 767058).

																										Frequency (%)	
ID[a]	Haplotype[b]																									EM[c]	R[d]
A	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	1	1	0	1	18.1	17.6
B	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0	17.2	23.5
C	0	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1	1	14.3	23.5
D	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	3.8	…
E	0	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0	1	0	0	1	1	1.9	2.9
F	0	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	1.9	2.9
G	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	…	2.9
H	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	…	2.9
I	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	…	2.9
J	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	1	…	2.9
K	0	1	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	1	1	0	1	0	0	1	1	…	2.9
L	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	…	2.9
M	1	1	1	1	0	1	1	0	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	1	0	…	2.9
N	1	1	1	1	0	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	…	2.9
O	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1	…	2.9
P	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0	…	2.9
Q	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0	1	1	0	1	0	0	1	1	1.9	…
R	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1.9	…
S	0	1	1	1	0	1	1	0	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	1	0	1.9	…
T	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0	1.9	…
U	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1.9	…
V	0	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	0	1.9	…
W	0	1	1	1	0	1	1	0	0	1	0	0	1	0	0	0	1	1	1	0	1	0	0	1	1	1.9	…
X	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	1.9	…
	1	0	0	0	0	1	-	0	1	1	-	1	0	1	1	1	1	1	0	1	0	0	0	0	1	Ancestral	

Table 4.2 The haplotype structure of the *MAPT* gene in CEPH-trios.

The haplotype structure is 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%. Additionally presented is the build of the ancestral genotype (Chimpanzee). a:haplotype identity. b: binary representation. c: infered frequency by Expectation-Maximization (all data). d: resolved haplotype frequency

4.7.2 Selection, performance assessment and association analysis of *MAPT* haplotype-tagging SNPs in PSP and CBD

Using an association-based criterion (criterion 5 in TagIT, haplotype r^2), the htSNPs were selected [123]. Six htSNPs (rs1467967, rs242557, rs3785883, rs2471738 and rs7521 and the *del-In*9) 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 i.e. 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 [130].

In CEPH-trios, 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 5 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% [123] .

The *MAPT* htSNPs were genotyped 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 4.3. In all the groups, there were no significant deviations from HWE 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=1.14x10^{-5}, 4.021x10^{-8}, respectively; **Table 4.3**). The same trend was observed in CBD but the difference was not significant, possibly due to a small sample size. No evidence of association was found for htSNPs rs1467967, rs3785883 and rs7521 in the studies, except in the US CBD study where htSNP rs3785883 is moderately associated (*p*=0.019, allelic). The OR and their 95% confidence intervals were calculated. The values for all 6 htSNPs by comparison of each minor allele verses each major allele was present (**Table 4.3**).

		Frequency (F1%)		Association (p)		Odds Ratio (MA)	
	dbSNP ID	Cases	Controls	Allelic	Genotypic	OR	95% CI
US PSP							
	rs1467967	62.8	62.6	0.963	1	0.965	0.703~1.325
	rs242557	54.4	31	2.91ex-9	2.29ex-8	2.356	1.706~3.255
	rs3785883	17	22.4	0.072	*0.168	0.713	0.487~1.044
	rs2471738	67	81.5	1.87ex-5	*1.15ex-4	2.224	1.535~3.222
	del-In9	91.6	77.1	4.02ex-8	*1.00ex-5	0.298	0.193~0.462
	rs7521	43.2	44.5	0.456	0.671	1.124	0.827~1.526
UK PSP							
	rs1467967	67.9	64.6	0.993	0.77	0.998	0.639~1.56
	rs242557	47.9	35.7	0.012	0.016	1.815	1.209~2.726
	rs3785883	25.5	20.6	0.365	0.68	1.227	0.762~1.974
	rs2471738	66	80.1	0.001	0.005	2.142	1.368~3.355
	del-In9	93.2	76.6	1.14ex-5	5.31ex-5	0.215	0.099~0.466
	rs7521	51.2	45.7	0.546	0.814	0.773	0.505~1.183
US CBD							
	rs1467967	61.9	62.6	0.909	*0.870	1.03	0.619~1.713
	rs242557	50	31	0.002	0.01	2.231	1.322~3.764
	rs3785883	33.3	22.4	0.019	0.022	1.047	0.586~1.872
	rs2471738	67	81.5	0.005	0.011	2.165	1.254~3.736
	del-In9	86.4	77.1	0.063	**	0.532	0.271~1.043
	rs7521	43.2	44.5	0.826	0.464	0.807	0.494~1.32

Table 4.3 Allele frequencies (*F*1) and *p*-values of single-locus association in the three studies.

Allele frequencies (*F*1) and *p*-values of single-locus association in the three studies. The *p*-values were derived by standard Pearson's χ2 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 performed for the *del-In9* in intron 9 in the CBD series, since there were no rare homozygotes in the CBD cases, thus preventing us from performing a valid test. Significant single-locus association 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.

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 light of the lack of association of *del-In9* in this particular study.

There is potentially the greater power to detect the contribution to association of causal variants by performing tests of association for the htSNP-defined haplotypes rather than individual htSNPs themselves. The six htSNPs were identified to capture 95% of the common haplotypic diversity of *MAPT*. An omnibus test of haplotype frequency differences estimated by EM between cases and controls in both the UK and US PSP groups was performed. The haplotype distribution (all haplotypes >1.0%) was found to be highly significant 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 CBD (p=0.120, d.f = 17). In addition to the global significance of the haplotype-wide comparison, individual haplotype tests (d.f = 1) were undertaken for significance through LRT and empirical *p*-values were derived through Monte-Carlo methods (20,000 simulations, data not shown). Two common haplotypes, A and C, which were strongly associated with both UK and US PSP were identified (**Table 4.4**). Haplotype A, which derives from the *del-In9*-defined H2 haplotype was the most common haplotype in the controls and was significantly under-represented 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, 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 rs242557 and rs2471738 which, in addition the *del-In9,* also show powerful single-locus effects. Haplotypes A and C do not differ by htSNPs rs1467967 and rs7521, 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.

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	Case	p(pcorrected)		Control	Case	p(pcorrected)		Control	Case	p(pcorrected)	
A	A	G	G	C	H2	G	20.7	6.3	1.46ex-5	(2.77ex-4)	22	6.3	9.55ex-9	(2.01ex-7)	22	8.2	0.02	(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	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	7.9	0.077	(1.000)	4	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	1	0.291	(1.000)	1.4	0		…	1.400	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	3.8	0.404	(1.000)	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	0.61	(1.000)
J	A	G	G	C	H1	G	2.4	0	0.033	(0.621)	0	3	0.055	(1.000)	0	3.4	0.237	(1.000)
K	A	A	A	C	H1	G	2.2	0.9	0.378	(1.000)	0	0	…		0.000	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	0.759	(1.000)
M	G	A	G	C	H1	G	2	2.6	0.744	(1.000)	3.5	3.4	0.930	(1.000)	3.5	5	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	0.018	(0.322)
O	A	A	A	C	H1	A	0	3.6	0.070	(1.000)	3.4	1.3	0.350	(1.000)	3.4	5	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	…	
Q	A	A	G	T	H1	A	0.7	2.8	0.040	(0.760)	0	1.6	0.003	(0.073)	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	0.920	(1.000)	2.6	0	0.621	(1.000)
T	A	G	A	T	H1	G	0.3	0	…		1.100	0		…	1.100	7	0.713	(1.000)
U	A	A	G	C	H1	G	1.1	0	…		1.100	1.7	0.270	(1.000)	1.1	3.5	0.17	(1.000)
v	G	G	A	T	H1	G	1.3	0	…		1.900	1	0.207	(1.000)	1.9	2.8	0.699	(1.000)
w	G	G	G	C	H2	G	0	0	…		0.000	0		…	0.000	2.9	0.326	(1.000)
x	G	A	A	T	H1	G	0	0	…		2.700	0.5	0.205	(1.000)	2.7	0	0.174	(1.000)

Table 4.4 Association of common MAPT haplotypes with progressive supranuclear palsy and corticobasal degeneration
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.

Haplotype				Frequency (%) and association (LRT) of haplotype											
				UK PSP				US PSP				US CBD			
ID	rs242557	rs3785883	rs2471738	Control	PSP	p-value	p (corrected)	Control	PSP	p-value	p (corrected)	Control	CBD	p-value	p (corrected)
I	G	G	C	50	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	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)	13.90	22.1	0.145	(1.000)
IV	G	A	C	10	16.6	0.316	(1.000)	10.2	7.1	**0.008**	(0.064)	10.2	0	**0.034**	(0.275)
V	A	A	C	6.9	9	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	3	0.611	(1.000)	4	4.4	0.603	(1.000)
VII	A	A	T	3.2	0	0.907	(1.000)	2.9	1.6	0.751	(1.000)	2.9	0	0.321	(1.000)
VIII	G	A	T	2.4	0	**0.045**	(0.356)	3.4	1.4	0.103	(1.000)	3.4	10.7	0.186	(1.000)

Table 4.5 Association of the subset of htSNP haplotypes with progressive supranuclear palsy and corticobasal degeneration
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 studey are significant. CBD, corticobasal degeneration; LRT, likelihood ratio test; PSP, progressive supranuclear palsy.

4.7.3 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, are independent to that of *del-In9*, each htSNP was incorporated as additional explanatory factors to the logistic regression model of the *del-In9* that serves to define the extended H1 and H2 haplotype status. Significant associations of single locus htSNPs rs242557, rs3785883 and rs2471738 were found (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 htSNPs 14 and 21 (p = 0.0183 and 0.0436, respectively) for the CBD cases. The effects of haplotypes were probed on sub-sets of htSNPs again entering the extended haplotype (H1 and H2 status, defined by the *del-In9*) as an explanatory factor. Highly significant differences were found in the distribution of haplotypes defined by three htSNPs rs242557, rs3785883 and rs2471738 in the UK and US PSP and to a lesser extent, the CBD cases (p= 9.34x10-4, 9.31x10-5, 0.0292, respectively). This was significant (p =2.49x10-5, 1.44x10-8, 0.006) in UK and US PSP and CBD, respectively, when the extended haplotype was excluded as an explanatory factor (**Table 4.4**). The haplotypes those SNPs define are associated with PSP and CBD after consideration of the *del-in9*, suggestive 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) under-represented (**Table 4.5**). The SNPs rs242557, rs3785883 and rs2471738 are H1-specific SNPs in *MAPT*, i.e. variable only on the H1 background though the haplotype I allelic combination is fixed and representative of H2 in addition to H1-derived variants.

The htSNP data were re-analysed, after removing all individuals with a H2 chromosome, thus leaving us with a biased H1H1 homozygote population. A significant (*p* <0.05) heterogeneity was found 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 Caucasian cohorts. For this purpose, it would be important to extend this study in an H1-only population such as the Japanese and Taiwanese [59].

4.8 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, the haplotype tagging approach was applied in this study. This provides a substantially streamlined and economical protocol by using a minimal set of tagging SNPs to study the LD and common haplotypic diversity of the entire gene or locus.

The underlying LD and haplotype structure of *MAPT* were first assessed 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, the *del-In9* status at the MAPT locus were included to define the H1 and H2 haplotypes [130]. 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 [163]. The strict H1/H2 dichotomy and H1 diversity across *MAPT* and beyond has also been demonstrated in other studies [49;164]. In a more recent study [60], the lack of recombination between H1 and H2 has been shown to be due to inversion of the chromosomal region on 17q21.31 corresponding to the extended *MAPT* H1/H2 haplotype block which described in previous chapter [131].

Then, an association-based criterion was used to assign a set of five haplotype-tagging SNPs (htSNPs) that, together with *del-In9* as a sixth bi-allelic tagging polymorphism, capture 95% of the common haplotype diversity in *MAPT*. The six htSNPs were genotyped 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, very strong associations of two common haplotypes were clearly demonstrated. Firstly, the significant under-representation of the 'classical' H2 (haplotype A, **Table 4.4**) and secondly, strong over-representation of an H1-derived haplotype (haplotype C, **Table 4.4**). 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 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 under-representation of H2 (Haplotype A) and overrepresentation of the H1-derived haplotype C (**Table 4.4**). However, they were not significant, possibly due to the smaller number of CBD cases. Assuming that these findings can be confirmed in a larger CBD cohort, they suggest that causative variant(s) 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 1.14Mb around *MAPT* that is associated with PSP and CBD. 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 [165]. The haplotype structure and its associations of the *MAPT* gene alone are refined in this study. A particular H1-derived haplotype in *MAPT* has been demonstrated to be highly associated with PSP.

In an attempt to further minimise the candidate pathogenic domain of *MAPT*, a strong association with PSP and CBD of three-locus haplotypes were identified based on the sub-set of H1-specific htSNPs, rs242557, rs3785883 and rs2471738. These associations are independent of the extended H1 and H2 haplotypes, defined by *del-In9*. Haplotypes derived from these SNPs span a minimal region from dbSNPs rs242557 to rs2471738 on the H1 haplotype background in *MAPT*.

This minimal region incorporates ~56.3kb of sequence, from upstream of exon 1 downstream to intron 9 that could harbour potential causal variant(s) that are in LD with these SNPs. Skipper and colleagues defined a similar associated candidate region in the 5'-half of *MAPT* in Norwegian PD cases, thereby proposing genetic variability that could influence the alternative splicing of *MAPT* exons 2 and 3 or, expression levels of *MAPT*. However, they carried out their analysis only on H1 homozygous individuals, having removed all H2 carriers [49]. For this reason, we cannot compare findings from both studies. As explained previously in section 4.7.3, 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 a *MAPT* promoter haplotype in PD based not only on allelic association of the previously defined extended H1 haplotype but also on differences in transcriptional activity [142]. In future studies, it would be important to compare LD and association of the *MAPT* locus in PSP, CBD and PD 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 [130]. It is more plausible that the associated SNPs in our study that confer greatest risk (SNPs rs242557 and rs2471738, **Table 4.3**), or protection (*del-In9* and associated SNPs through LD; **Table 4.1** and **Figure 4.1**) are in LD with variants that could cause subtle changes either in the alternative splicing or overall expression levels. It is possible that each neuronal sub-group is dependent on a particular tau isoform profile and expression level. Aberrations in this homeostasis, could affect one neuronal sub-group more than another and lead to the selective and disease-specific neuronal death and tau pathology [166].

Chapter 5 Association of tau haplotype-tagging polymorphisms with Parkinson disease in diverse ethnic cohorts.

5.1 Overviews

The genetic variation of *MAPT*, has not only been found to be associated with tauopathies, including Alzheimer's disease[167], progressive supranuclear palsy and corticobasal degeneration[67;130;131;165] as described in Chapters 4 and 6. Several *MAPT* polymorphisms that define the tau H1 haplotype have been investigated for an association with PD with conflicting results[133;142;168]. In order to demonstrate the association of *MAPT* with PD, a systematic framework of genetic analysis was devised to examine the possible genetic variations for genetic study in PD case-control cohorts from three ethnically diverse populations: Taiwanese, Greek and Finnish.

A moderate association at SNP rs3785883 in the *MAPT* region in the Greek cohort as well as for SNP rs7521 and rs242557 (*p*=0.01 genotypic *p*=0.04 allelic) in the Finnish population were found. There were no significant differences in genotype or allele distribution between cases and controls in the Taiwanese cohort. There is therefore no consistent association between the *MAPT* H1 haplotype and PD in three ethnically diverse populations; however, the sub-haplotypes of *MAPT* H1 may confer susceptibility to PD.

5.2 Background

5.2.1 Overlaps in the clinical and pathological features of tauopathies and synucleinopathies

Parkinson disease (PD) is the second most common chronic neurodegenerative disease, characterized by tremor, rigidity, postural instability and bradykinesia. Epidemiological studies have estimated a cumulative prevalence of PD of greater than 1 per thousand. PD belongs to a group of diseases termed ‘synucleinopathies’ based on the strong immunostaining of its pathological hallmark, Lewy bodies, for α-synuclein.[169] Increasing evidence indicates that there is an overlap of the clinical and pathological features of tauopathies and synucleinopathies, thereby re-enforcing the notion that these disorders might be linked mechanistically. This observation raises the possibility that tau protein may be important in PD pathogenesis.[170;171]

5.2.2 Genetic risk factors of Parkinson’s disease

One of the strongest risk factors of PD is a positive family history. The estimated genetic risk ratio for PD is approximately 1.7 (70% increase risk for PD if a sibling has PD) for all ages, and increases over 7-fold for those under age 66 years. Growing evidence shows that genetic abnormalities play a major role in the aetiopathogenesis of PD. Several loci for familial PD have been reported, including α-synuclein (*SNCA)*, parkin, PTEN-induced kinase 1 (*PINK1*), ubiquitin-C terminal hydrolase-L1 (*UCH-L1*), DJ-1 and leucine-rich repeat kinase 2 (*LRRK2*). (**Table 5.1**) *SNCA* was the first genetic factor linked to familial PD. In 1996, PD within the Contursi kindred was linked to chromosome 4q21-23 and in

1997 the A53T mutation in the *SNCA* was identified as the causative mutation [31;172]. Other than point mutations in the causal genetic loci, the alteration of gene dosage could also confer the risk of PD in particular affected families. In 2003, Singleton et al. discovered a triplication of *SNCA* in an autosomal dominant PD family known as the Iowa kindred. Their result provides evidence that *SNCA* behaves differently from the wild-type protein in a quantitative rather than qualitative manner could be the cause of PD [37].

Despite the discovery of the gene loci for familial PD, the role of genetic factors in sporadic PD remains unclear.

Locus	Chromosome Position	Gene	Phenotype	Inheritance
PARK1 and 4	4q21-23	α-synuclein	Earlier onset and features of common DLB	AD
PARK 2	6q25.2-q27	Parkin	Earlier onset with slow progression. Usually no Lewy bodies	AR
PARK 3	2q13	unknown	Classical PD with Lewy bodies and dementia	AD
PARK 5	4q14	UCH-L1	Classical PD	AD
PARK 6	1p35-36	PINK1	Earlier onset with slow progression	AR
PARK 7	1p36	DJ-1	Earlier onset with slow progression	AR
PARK 8	12p11.2-q13.1	LRRK2	Classical PD with and without Lewy bodies	AD
PARK 10	1p32	unknown	Classical PD	Unclear
PARK 11	2q36-37	unknown	Classical PD	AD/Unclear

Table 5.1 Parkinson's disease genetic loci
AD: autosomal dominant; AR: autosomal recessive

One of the genetic susceptibility factors that have been found to be associated with sporadic PD is *MAPT*. In spite of two meta-analyses supporting an association between *MAPT* haplotype H1 and PD[49;173], several *MAPT* polymorphisms that define the H1 haplotype have been investigated for an association with PD with conflicting results.

The aim of the work described in this chapter was to determine the possible genetic role of *MAPT* in PD in three ethnically different PD populations, using htSNPs and to demonstrate if there is a consistent association between the *MAPT* H1 haplotype (delineated by *del-In9*) and PD in different populations worldwide.

5.3 Case-control samples

In this study, 508 patients and 611 healthy controls were recruited from Greek, Finnish and Taiwanese populations (**Table 5.2**). All patients were sporadic, based on pedigree analysis. The PD diagnoses were verified according to PD Society Brain Bank criteria[174]. Patients with dementia or patients who reported first-degree relatives with parkinsonism were excluded. Patients with evidence of secondary parkinsonism or with atypical features such as early dementia, ophthalmoplegia, early autonomic failure, and pyramidal signs were not included in this study. Age-at-onset was defined as the age at which the patient noticed the first symptom indicative of PD. Controls were age-matched normal subjects living in the same geographical area as the patients who visited our outpatient clinic and finally were found free of any neurological disease (PD included). After approval from the hospital internal ethics and scientific boards and written informed consent, blood samples were drawn for DNA extraction from patients and controls. Certified neurologists who were blind to genotyping results performed all clinical assessments.

5.3.1 Taiwanese Series

One hundred and nineteen patients with sporadic idiopathic PD (47.1% female, mean age of onset: 61.7±10.9 years, range: 24-91 years) were selected from the neurology clinic of Chang-Gung Memorial Hospital[174;175]. Two hundred and sixty unrelated healthy adult volunteers (43.8% female, mean age at exam: 59.0±10.1 years, range: 31-84 years) were recruited. These normal control individuals are genetically unrelated to the individuals in the patient group. All patients were followed up for at least 1 year and up to 3 years.

5.3.2 Greek Series

The Greek cohort consisted of 242 patients with sporadic idiopathic PD (41.3% female, mean age of onset: 63.3±9.6 years, range: 30-88 years) were residents of Thessaly (Central Greece) and were identified prospectively during a 3-year period (2001-2004) in the outpatient clinic for movement disorders in Larissa University Hospital[174;176]. Two hundred and fifteen unrelated healthy adult volunteers (46% female, mean age at exam: 59.6±9.6 years, range: 38-86 years) were recruited in this study.

5.3.3 Finnish Series

The Finnish series is composed of 147 sporadic PD patients (mean age of onset: 61.5 years, range 40 to 87, 41% female) and 136 neurologically normal healthy control subjects (mean age 66.4 years, range 38 to 88, 63% female). Eleven patients had an onset before 45 years old. All affected individuals were recruited from the Neurological outpatient clinics of the Helsinki University Central Hospital and Seinäjoki Central Hospital. Patients were either followed for at least 4 years or, alternatively, clinical follow-up for at least 2 years plus ^{123}I-β-CIT-SPECT findings supporting idiopathic PD. Patients with dementia or patients who reported first-degree relatives with parkinsonism were excluded. Written informed consent was obtained from all participants.

	Greek		Finnish		Taiwanese	
	Patients	Controls	Patients	Controls	Patients	Controls
Gender (F:M)	100(41.3):142(58.7)	94(43.7):121(56.3)	60(40.8):87(59.2)	86(63.2):50(36.8)	56(47.1):63(52.9)	114(43.8):146(56.2)
age	69.8+/-8.7(44-95)	68.3+/-12.8(32-93)	67.2+/-8.4(37-87)	66.4+/-9.2(38-88)	68.9+/-10.3(41-96)	59.0+/-10.1(31-84)
age at onset	63.3+/-9.6(30-88)		61.5+/-8.8(40-87)		61.7+/-10.9(24-91)	

Table 5.2 Demographic description of the diverse ethnic PD cohorts

5.4 Genotyping of htSNPs

5.4.1 Selection of haplotype-tagging SNP markers

Six htSNPs that function as a minimal set of highly informative SNP markers and capture 95% of the common haplotype diversity of *MAPT* in Caucasians[167;177]. These SNPs were selected after analysis of the LD architecture of *MAPT*, using validated HapMap genotype data (http://www.hapmap.org/) for 24 SNPs spanning *MAPT* in 27 CEPH-trios individuals. (http://locus.umdnj.edu/nigms/)[177]. The 6 htSNPs were previously analysed and found a good coverage of this region with capturing > 95% of the diversity of the MAPT haplotype in European populations. (See section 4.3 and section 6.4.1) However, the performance of these htSNPs in Asian populations was unknown. A further analysis of *MAPT* on Taiwanese healthy controls was later performed and described in Chapter 6.

5.4.2 Pyrosequencing Assay

The htSNPs (db SNPs numbers: rs1467967, rs242557, rs3785883, rs2471738, rs7521 and the intron 9 deletion-insertion (*del-In9)*; (**Table 5.3**) were genotyped in the Parkinson's disease case-control cohorts as described in previous chapter.

For each SNP, the allele and genotype distributions were compared between cases and controls in each population. Case-control single locus htSNP allelic and genotypic association was calculated statistically in De Finetti software (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl). The square of the correlation coefficient (r^2) and D' for LD was calculated pair-wise between each using the genetics software program TagIT (http://popgen.biol.ucl.ac.uk/software.html). Haplotype predictions were made using an Expectation–Maximisation (EM) algorithm using TagIT. Haplotypes with frequencies <1% were included in the analysis only when they were observed in one of the three populations with higher frequency for comparison reasons. SHEsis software (www.nhgg.org/analysis) was used to compare the distribution of haplotypes defined by the htSNPs in the Parkinson's disease cases and controls [178]. This is a SNP haplotype analysis based on a Full-Precise-Iteration algorithm, which could reconstruct ambiguous haplotypes and estimate haplotype frequencies in the given random sample set [178].

SNP	Position[a]	dbSNPID	Alleles
1	41342006	rs1467967	A/G
2	41375573	rs242557	A/G
3	41410268	rs3785883	A/G
4	41431900	rs2471738	C/T
5	41442488	*del-In9*	+/-
6	41461242	rs7521	A/G

Table 5.3 List of the htSNPs

[a]SNP position on chromosome. (Position on chromosome (in bp) is based on May 2004 build of Human Genome Sequence (http://genome.ucsc.edu).)

5.5 Results

The single locus associations are illustrated in Table 5.5. In all groups, there were no deviations from Hardy-Weinberg Equilibrium (HWE) at any of the htSNPs. Testing disease association with the *del-In9* polymorphism, which delineates the H1 from H2 haplotype were negative either looking at contingency tables or by examining the odds ratios. However rs242557 is moderately associated with the disease in the Finnish population (*p*=0.01 genotypic *p*=0.04 allelic). A moderate association was identified at SNP rs3785883 in the Greek cohort for both allele and genotype frequency (*p*=0.01, *p*=0.05 respectively) as well as for SNP rs7521 in the Finnish cohort (genotype *p*=0.02). There were no significant differences in genotype or allele distribution between cases and controls in the Taiwanese cohort (**Table 5.5**).

A global test of haplotype frequency was performed and the differences were estimated by EM between cases and controls in Greek, Finnish and Taiwanese series but no difference was found at a *p* of <0.05 (**Table 5.4**). Individual haplotype tests (df = 1) identified haplotype associations in all three cohorts (Greek cohort haplotype I; Finnish cohort haplotypes H, I, M and U; Taiwanese cohort haplotypes Q and S). Although an association of the haplotype clade I was identified in two cohorts the direction of effect was opposite, being protective in the Greek cohort and a risk factor in the Finnish cohort.

The H2 allele frequency in Greek population (21%) is consistent with other Caucasian populations but the frequency is significantly lower in the Finnish population (8%)[59]. There is no Haplotype A (H2) found in the Taiwanese cohort, in accordance with the reported low frequency of H2 allele in Asian populations[59]. As expected, no difference was noted in the frequency of haplotype A (H2) between cases and controls in either the Greek or Finnish series. These results were consistent with the single-locus analysis that showed that the *del-In9* polymorphism of *MAPT* is not associated with Parkinson’s disease.

5.6 Discussion

In the analysis of a positional and/or functional candidate gene or region, the optimal experimental design would be to contiguously screen the entire genomic sequence to detect all the common variants that exist. This would ensure that all the common haplotypes are defined. However the approach that we followed succeeds to scan the common variation of a gene sensitively and comprehensively and provides key fine-mapping data within regions of strong LD particularly in the Greek and Finish population. In the Taiwanese population, the 6 htSNPs only capture 87% of common genetic diversity at this locus. Therefore, the haplotype association of the Taiwanese cohort lacks power to represent all those haplotypes over the *MAPT* region. We have found that in the later association study of *MAPT* and Taiwanese Alzheimer's disease, 9 htSNPs to capture 95% of the diversity of the known CEPH variants (see section 5.5.2). Unfortunately, in this Taiwanese PD series, there is a lack of material for further investigation with those additional 4 htSNPs.

The association of the polymorphic variants, and constructed haplotypes, with susceptibility to PD in this study does not reflect a universal or consistent effect in the populations studied here. These data suggest several possibilities; first, that genetic variation in *MAPT* has a different degree of influence on disease in populations with different ethnic backgrounds, perhaps modulated by other genetic factors or environmental influences; second, that the variants assessed here are not directly causal or protective but are in linkage disequilibrium with other variability and that this variability is different or the LD structure is different, between populations; third, the positive results identified here are simply type I errors. The application of a correction for multiple testing, such as Bonferroni's would remove the statistical significance of the observed results, however this particularly stringent procedure would mask association in these series for all effects except those with particularly high odds ratios. In an attempt to avoid this type of correction we have utilized independent cohorts to perform the analysis, this study design is optimal for identifying a risk gene, or for analyzing a risk variant, but it will often give converse associations for alleles in LD with the risk allele. We aim to replicate the individual associations here in larger series from these ethnic groups.

In summary, our study failed to demonstrate a consistent association between the *MAPT* H1 and H2 haplotypes (delineated by intron 9 *del-In9*) and PD in three ethically diverse populations. However, the data presented here suggests that sub-haplotypes of haplotype H1 may confer susceptibility to PD, and that either allelic heterogeneity or divergent haplotype composition explain the divergent haplotype results.

							Greek					Finnish					Taiwanese				
Haplotype ID	rs1467967	rs242557	rs3785883	rs2471738	INDEL	rs7521	% Case	% Control	Chi-square	p-value	Odd ratio (95%CI)	% Case	% Control	Chi-square	p-value	Odd ratio (95%CI)	% Case	% Control	Chi-square	p-value	Odd ratio (95%CI)
A	A	G	G	C	H2	G	0.19	0.18	0.05	0.81	1.04(0.72-1.51)	0.06	0.06	0.00	0.95	0.97(0.46-2.07)	0.00	0.00			
B	G	G	G	C	H1	A	0.21	0.18	1.29	0.25	1.23(0.85-1.77)	0.14	0.11	1.48	0.22	1.39(0.81-2.36)	0.01	0.00	1.74	0.18	119(0-114223)
C	A	A	G	C	H1	A	0.08	0.09	0.13	0.71	0.90(0.54-1.52	0.08	0.13	3.53	0.06	0.57(0.31-1.03)	0.02	0.00	3.57	0.06	9995.36(0-141596)
D	A	G	G	C	H1	A	0.10	0.08	0.56	0.45	1.21(0.73-1.99)	0.09	0.09	0.01	0.94	1.02(0.56-1.86)	0.01	0.02	0.21	0.64	0.71(0.17-2.93)
F	A	A	G	T	H1	G	0.06	0.08	0.36	0.54	0.83(0.47-1.48)	0.12	0.12	0.00	0.98	0.99(0.57-1.71)	0.00	0.01	1.58	0.20	0.11(0.002-6.95)
G	G	A	G	C	H1	A	0.04	0.06	1.92	0.16	0.62(0.31-1.22)	0.05	0.05	0.05	0.82	0.91(0.41-2.02)	0.05	0.06	0.31	0.57	0.82(0.41-1.64)
H	A	G	A	C	H1	A	0.07	0.04	2.18	0.13	1.62(0.84-3.10)	**0.00**	**0.05**	**12.46**	**0.00**	**0.013(0-1.48)**	0.00	0.00			
I	A	A	G	C	H1	G	**0.02**	**0.04**	**3.65**	**0.05**	**0.39(0.14-1.05)**	**0.08**	**0.02**	**10.09**	**0.00**	**4.60(1.65-12.85)**	0.10	0.11	0.08	0.77	0.92(0.54-1.56)
J	A	G	G	C	H1	G	0.03	0.03	0.00	0.97	1.01(0.41-2.45)	0.04	0.04	0.02	0.89	0.93(0.37-2.34)	0.11	0.07	1.83	0.17	1.49(0.83-2.62)
K	A	G	A	C	H1	G	0.03	0.02	0.80	0.37	1.49(0.61-3.59)	0.00	0.00				0.06	0.05	0.12	0.72	1.13(0.54-2.36)
L	G	A	G	C	H1	G	0.02	0.02	0.03	0.86	0.91(0.32-2.57)	0.00	0.02	3.54	0.06	0.02(0-25)	0.32	0.39	3.37	0.06	0.73(0.52-1.02)
M	A	A	A	T	H1	G	0.03	0.01	2.86	0.09	2.57(0.82-7.99)	**0.08**	**0.03**	**7.15**	**0.01**	**3.19(1.30-7.82)**	0.00	0.00	2.62	0.60	2.5(0.06-95.9)
N	A	A	A	C	H1	A	0.00	0.10	2.40	0.14	0.14(0.0007-2.5)	0.04	0.02	0.78	0.37	1.59(0.56-4.50)	0.00	0.00	1.26	0.26	
O	G	G	A	C	H1	A	0.01	0.01	1.18	0.27	2.34(0.48-11.32)	0.04	0.03	0.34	0.55	1.31(0.52-3.31)	0.00	0.00			
P	G	A	A	T	H1	G	0.02	0.01	0.31	0.57	1.40(0.42-4.61)	0.01	0.03	2.14	0.14	0.36(0.09-1.47)	0.00	0.01	1.76	0.18	0.21(0.0019-2.56)
Q	G	G	G	C	H1	G	0.01	0.01	0.23	0.63	1.37(0.36-5.12)	0.01	0.03	3.71	0.05	0.23(0.04-1.16)	**0.08**	**0.04**	**4.36**	**0.04**	**2.19(1.03-4.65)**
R	G	A	A	C	H1	G	0.00	0.01	3.97	0.04		0.01	0.00	2.34	0.13	66(0.002-2334636)	0.04	0.04	0.00	0.96	1.021(0.46-2.26)
S	G	A	G	T	H1	G	0.00	0.14	5.17	0.02		0.01	0.00	2.63	0.10	11.2(0.28-440.7)	**0.04**	**0.01**	**5.74**	**0.02**	**5.72(1.14-28.57)**
T	G	G	G	T	H1	G	0.01	0.00	0.43	0.48	2.37(0.19-28.8)	0.01	0.01	0.29	0.59	0.63(0.11-3.37)	0.03	0.02	0.06	0.80	1.13(0.39-3.27)
U	A	G	G	T	H1	G	0.01	0.01	0.24	0.60		**0.03**	**0.07**	**5.02**	**0.03**	**0.36(0.146-0.911)**	0.08	0.11	1.58	0.20	0.7(0.4-1.21)
							Global x^2 (df=63)		p(Fisher's)			Global x^2 (df=63)		p(Fisher's)			Global x^2 (df=63)		p(Fisher's)		
							60.83		0.55			80.14		0.05			15.5		0.07		

Table 5.4 Association of common *MAPT* haplotypes with PD in the three diverse ethnic populations.

		Controls			Cases		X^2	P value	Odds Ratio[95%CI]
rs1467967, 5' of exon1									
Genotypic	C/C	C/T	T/T	C/C	C/T	T/T			
Greek	28(0.133)	95(0.452)	87(0.414)	27(0.127)	87(0.412)	97(0.459)	0.03	0.87	0.95(0.54-1.68)
Finnish	15(0.104)	60(0.419)	68(0.475)	10(0.074)	67(0.496)	68(0.429)	0.81	0.36	0.68(0.29-1.57)
Taiwanese	95(0.37)	115(0.45)	43(0.16)	42(0.36)	58(0.50)	15(0.130)	0.29	0.547	0.87(0.54-1.41)
Allelic	C	T		C	T				
Greek	151(0.359)	269(0.6400		141(0.334)	281(0.665)		0.60	0.43	0.89(0.67-1.18)
Finnish	90(0.314)	196(0.685)		87(0.322)	183(0.677)		0.04	0.84	1.03(0.72-1.47)
Taiwanese	305(0.60)	201(0.39)		142(0.61)	88(0.38)		0.14	0.70	1.06(0.77-1.46)
rs242557, 5' of exon1									
Genotypic	A/A	A/G	G/G	A/A	A/G	G/G			
Greek	18(0.088)	96(0.472)	89(0.438)	20(0.098)	115(0.566)	68(0.334)	0.12	0.73	1.12(0.57-2.19)
Finnish	35(0.243)	80(0.555)	29(0.201)	27(0.203)	62(0.466)	44(0.330)	5.97	0.01	1.96(1.13-3.37)
Taiwanese	84(0.34)	125(0.51)	36(0.14)	41(0.36)	63(0.56)	8(0.07)	1.78	0.13	1.36(0.8-1.71)
Allelic	A	G		A	G				
Greek	132(0.325)	274(0.674)		155(0.381)	251(0.618)		2.85	0.09	1.28(0.96-1.71)
Finnish	150(0.520)	138(0.479)		116(0.436)	150(0.563)		3.98	0.04	1.4(1.00-1.96)
Taiwanese	293(0.59)	197(0.40)		145(0.64)	79(0.35)		1.5	0.2	1.23(0.88-1.71)
rs2471738, Intron 9									
Genotypic	G/G	G/A	A/A	G/G	G/A	A/A			
Greek	133(0.627)	73(0.344)	6(0.028)	154(0.733)	49(0.233)	7(0.033)	0.00	0.95	1.01(0.66-1.54)
Finnish	74(0.50)	61(0.414)	11(0.075)	60(0.444)	70(0.518)	5(0.0370	1.09	0.29	0.77(0.48-1.24)
Taiwanese	172(0.66)	82(0.31)	4(0.01)	72(0.63)	39(0.34)	3(0.026)	0.18	0.71	1.1(0.69-1.74)
Allelic	G	A		G	A				
Greek	354(0.842)	66(0.157)		344(0.839)	66(0.160)		0.02	0.88	0.97(0.67-1.41)
Finnish	209(0.715)	83(0.214)		190(0.703)	80(0.296)		0.10	0.75	0.94(0.65-1.35)
Taiwanese	426(0.82)	90(0.17)		183(0.80)	45(0.19)		0.46	0.45	0.86(0.56-1.32)
rs3785883, Intron 3									
Genotypic	G/G	G/A	A/A	G/G	G/A	A/A			
Greek	133(0.627)	73(0.3440)	6(0.0280)	154(0.733)	49(0.233)	7(0.033)	5.45	0.01	1.63(1.08-2.47)
Finnish	86(0.605)	45(0.316)	11(0.077)	76(0.584)	46(0.353)	8(0.061)	0.12	0.72	0.91(0.56-1.48)
Taiwanese	191(0.73)	63(0.24)	5(0.019)	86(0.72)	28(0.23)	5(0.04)	0.44	0.50	1.23(0.75-1.77)
Allelic	G	A		G	A				
Greek	339(0.779)	85(0.200)		357(0.85)	63(0.15)		3.72	0.05	1.42(0.99-2.03)
Finnish	217(0.764)	67(0.235)		198(0.76)	62(0.238)		0.00	0.94	0.98(0.66-1.46)
Taiwanese	445(0.85)	77(0.14)		200(0.84)	38(0.18)		0.46	0.49	1.15(0.75-1.77)
Indel, Intron 9									
Genotypic	H1/H1	H1/H2	H2/H2	H1/H1	H1/H2	H2/H2			
Greek	134(0.623)	73(0.339)	8(0.037)	128(0.571)	83(0.37)	13(0.058)	1.22	0.26	0.80(0.55-1.18)
Finnish	117(0.835)	21(0.15)	2(0.014)	114(0.850)	19(0.14)	1(0.007)	0.12	0.73	1.12(0.58-2.15)
Taiwanese									
Allelic	H1	H2		H1	H2				
Greek	341(0.793)	89(0.206)		339(0.756)	109(0.243)		1.66	0.19	0.81(0.59-1.11)
Finnish	255(0.910)	25(0.089)		247(0.921)	21(0.078)		0.21	0.64	1.15(0.62-2.11)
Taiwanese									
rs7521, 3'of exon14									
Genotypic	A/A	A/G	G/G	A/A	A/G	G/G			
Greek	71(0.334)	91(0.429)	50(0.235)	55(0.264)	98(0.471)	55(0.264)	2.48	0.11	0.71(0.46-1.08)
Finnish	32(0.223)	67(0.468)	44(0.307)	27(0.2)	82(0.607)	26(0.192)	4.88	0.02	0.53(0.30-0.93)
Taiwanese	206(0.80)	47(0.18)	2(0.007)	93(0.78)	24(0.20)	1(0.008)	0.01	0.852	0.903(0.08-10.08)
Allelic	A	G		A	G				
Greek	233(0.54)	191(0.450)		208(0.5)	208(0.5)		2.07	0.15	0.82(0.62-1.05)
Finnish	131(0.45)	155(0.541)		136(0.503)	134(0.496)		1.16	0.28	0.83(0.59-1.16)
Taiwanese	459(0.9)	51(0.1)		210(0.88)	26(0.11)		0.18	0.67	0.89(0.54-1.47)

Table 5.5 Allele frequencies and p-values of single locus association in the three diverse ethnic populations

Chapter 6 Association Study of MAPT gene haplotypes with Alzheimer's disease

6.1 Overview

Although it is clear that microtubule associated protein tau is involved in Alzheimer's disease (AD) pathology, it has not been clear whether it is involved genetically. Several studies on variability within *MAPT* and the occurrence of AD have been published, with inconclusive, though largely negative results [179-184] (Table 6.1). As stated in the chapter 4, a variant of *MAPT* subhaplotype, H1c is associated with the sporadic tauopathies, such as PSP and CBD. Our study was extended to examine the association of the PSP-associated haplotype, H1c, with autopsy confirmed, late-onset AD (LOAD) in US and UK populations and clinical confirmed Taiwanese AD patients.

Study	Population	Case		Control		H1/H1 vs All Others	
		N	H1/H1	N	H1/H1	OR	95% CI
Baker et al. [218]	Finland	110	0.83	174	0.88	0.77	0.38-1.55
Baker et al.[218]	USA	271	0.62	419	0.63	0.95	0.69-1.32
Bullido et al. [181]	Spain	167	0.44	194	0.50	0.76	0.49-1.18
Crawford et al. [179]	USA	138	0.64	117	0.62	1.09	0.67-1.77
Lilius et al. [180]	Sweden	184	0.78	62	0.68	1.66	0.84-3.28
Roks et al. [183]	Netherlands	101	0.60	116	0.60	1.00	0.56-1.79
Present study	UK	179	0.78	121	0.78	1.01	0.69-1.46
Present study	USA	181	0.74	131	0.74	1.02	0.70-1.50

Table 6.1 Summary of studies on MAPT H1/H1 diplotype as risk factor for Alzheimer's disease.
The odds ratios (ORs) and 95% confidence intervals (CIs) indicated were taken from the original publication.

Two of the tagging variants (rs242557 and rs2471738) have significant associations with AD in the US series and when both the US and UK series were collapsed. One of these variants (rs242557) also showed a significant p-value in Taiwanese AD series and a trend towards association (allele p-value = 0.094, genotypic p-value = 0.061) within the UK population.

6.2 Background

AD is the most common cause of dementia in the elderly. It is characterized clinically by a gradual onset and progression of memory loss, and characterized postmortem by the presence of two types of neuropathological inclusions: neurofibrillary tangles and senile plaques [185].

The neurofibrillary tangles, comprised of paired helical filaments of phosphorylated tau protein are a pathognomic feature of AD. Using immunohistochemical and biochemical means, it has been demonstrated that tau is modified in AD. One of the diversified MAPT H1 subhaplotypes, H1c has been shown to be largely responsible for the association between the H1 clade and the sporadic tauopathies [131;133;164;165;177]. The inconclusive association between AD and the variants of *MAPT* has been studied by several groups [179-184]. These reports have compared alleles that discriminated between the H1 and H2 clades, but did not assess whether variability on the H1 background showed association with disease.

The objective of this work was to examine whether the PSP-associated *MAPT* haplotype is responsible for the disease pathogenesis with late-onset AD in autopsy confirmed, late-onset AD (LOAD) and clinical confirmed Taiwanese AD patients.

6.3 Case-control samples

6.3.1 US and UK series

In UK series, there were 179 cases (66% female, mean age of death: 81 years, range: 65-96 years) and 121 controls (51% female, mean age of death: 78 years, range: 65-

100 years). While in US series, there were 181 cases (55% female, mean age of death: 81 years, range: 66-97 years) and 131 controls (51% female, mean age of death: 81 years, range: 65-99 years). All samples were pathologically confirmed. Controls were free of neuropathology at autopsy.

All samples were of Caucasian origin obtained from either the Newcastle Brain Bank, Newcastle-upon-Tyne, UK (UK series) or various brain banks throughout the United States, including National institute on Aging, Johns Hopkins Alzheimer's Disease Research Center, University of California, the Kathleen Price Bryan Brain Bank, Duke University Medical Center, Stanford University, New York Brain Bank, Taub Institute, Columbia University, Massachusetts General Hospital, University of Michigan, University of Kentucky, Mayo Clinic Jacksonville, University Southern California, Washington University, St Louis Alzheimer's Disease Research Center, University of Washington, Seattle.

6.3.2 Taiwanese Series

Patients were recruited from the dementia outpatient clinic of Chang Gung Medical Center, Linkou, Taiwan. The diagnosis of AD (n = 110; 60% female, 76.9±6.5 years) was made by consensus according to the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) for probable AD [186]. Subjects without stroke and cognitive impairment represented the control group (n = 117; 46.2% female, aged 59.0±10.1 years). Each subject was informed of the aim of the study, and all gave their consents to study. Patients with previous clinical history of neurological, psychiatric, somatic, or toxic causes for dementia were excluded. Evaluation included general physical and neurological assessment, the Mini-Mental State Examination (MMSE) [187], and Hachinski ischaemia score [188]. Laboratory studies included complete blood cell count, biochemistry analysis, erythrocyte sedimentation rate, vitamin B12 and folic acid levels, thyroid stimulating hormone level and syphilis serological testing. Each patient underwent brain computerized tomography (CT) scan. All the patients were examined by at least 2 neurologists and confirmed to fulfil the DSM-IV criteria for dementia.

6.4 Selection of haplotype-tagging SNPs

6.4.1 Markers for the US and UK series

The haplotype structure of *MAPT* was previously analysed using markers from the CEPH database (http://www.hapmap.org) (see section 4.3). Using the program TagIT (http://popgen.biol.ucl.ac.uk/software), a minimum of five SNPs were found in order to capture the haplotype diversity at the *MAPT* locus. The performance of our five tagging SNPs and the *del-In9* against each individual SNP typed were tested within the CEPH-trios, using criterion 5 (an association based criterion using haplotype r^2) in the program TagIT. On average, this set of tag SNPs captures 95% of the diversity of the known CEPH variants and scores 90% for most individual loci. Performance plots of the tagging variants as well as the *del-In9* polymorphism on its own are shown in Figure 6.1.

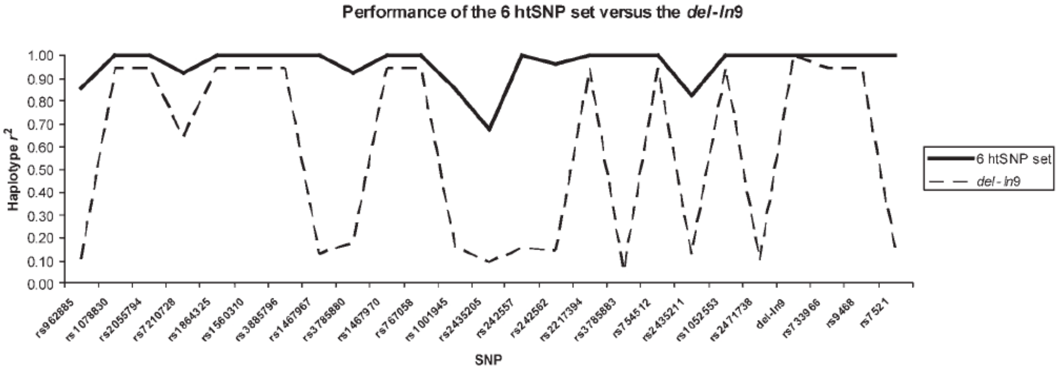


Figure 6.1 MAPT tagging markers capture the diversity of MAPT. Solid line: Performance plot of the six MAPT SNP tag markers using the data available from CEPH in the hapmap project (http://www.hapmap.org. The plot is a row of vector performance values (as measured by haplotype r^2 using criterion 5 from TagIt) for each of our tag SNPs against each of the SNP loci typed in the CEPH trios. High r^2 values indicate good performance, because r^2 is a measure of linkage disequilibrium. If there is perfect linkage disequilibrium between two markers, r^2 will approach 1, indicating that the two markers are segregating together and thus are genetically equivalent. On average, this set of tagging SNPs captures 95% (average $r^2 = 0.95$) of the diversity of the known SNPs as a whole and predominantly scores > 90% against individual SNP loci. Broken line: In contrast, examining just the *del-In9* variant's performance demonstrates that it performs well for some loci ($r^2 > 80\%$), whereas it performs poorly ($r^2 < 50\%$) for several loci. This is because there are many variants of the H1 clade. As the del-In9 variant only distinguishes H1 and H2, it is a reasonable marker to tag the variants that occur on the H1 and H2 backgrounds; however, it will perform poorly if used to tag loci that define sub-haplotypes of the H1 clade. This is important for the current study, as all previous studies examining AD risk and MAPT genotypes only looked at variants that defined the H1 or H2 clades and variants of the H1 clade.

6.4.2 Markers for the Taiwanese series

6.4.2.1 SNP Selection

For the Taiwanese population, there was no genotyping information available on the known SNPs in the extended *MAPT* region that covers 45-kb upstream and downstream of the gene. We selected all the Japanese SNPs with frequency information published on the JSNP database at http://snp.ims.u-tokyo.ac.jp in the region. We then tried to use any SNP(s) from dbSNP to fill gaps, when two adjacent JSNPs were more than 14kb apart. This led us to the SNPs: rs962885, rs2301689, rs3744457, rs2280004, rs1467967, rs3785880, rs1001945, rs242557, rs242562, rs2303867, rs3785882, rs3785883, rs3785885, rs2258689, rs2471738, rs916896, rs7521, rs2074432, rs2277613, rs876944 and rs2301732. (**Table 6.2**)

SNP	Position[a]	dbSNP ID	Alleles	Ancestral
1	41291420	rs962885	C/T	C
2	41291627	rs2301689	A/G	G
3	41328711	rs3744457	C/T	C
4	41340959	rs2280004	C/T	T
5	41342006	rs1467967	A/G	G
6	41349204	rs3785880	G/T	T
7	41361649	rs1001945	C/G	G
8	41375573	rs242557	A/G	A
9	41382599	rs242562	A/G	A
10	41395691	rs2303867	A/G	A
11	41397030	rs3785882	C/G	G
12	41410269	rs3785883	A/G	G
13	41414477	rs3785885	A/G	G
14	41423219	rs2258689	C/T	C
15	41431900	rs2471738	C/T	C
16	41454642	rs916896	A/G	G
17	41461242	rs7521	A/G	G
18	41465690	rs2074432[b]	A/G	A
19	41472982	rs2277613	C/T	C
20	41490227	rs876944	G/T	G
21	41500036	rs2301732	A/G	A

Table 6.2 The 21 single nucleotide polymorphisms used for the linkage disequilibrium and haplotype structure analysis of *MAPT* in the Taiwanese cohort

[a]SNP position on chromosome. (Position on chromosome (in bp) is based on May 2004 build of Human Genome Sequence (http://genome.ucsc.edu).) [b]The dbSNP ID rs2074432 was merged into rs1078997 (in May 2004 Human Assembly).

6.4.2.2 Selection, performance, assessment, and association analysis of *MAPT* haplotype tagging SNPs in Taiwanese cohort

An association based criterion (haplotype r^2) as previous described was selected for the criterion to tagging SNPs. Nine haplotype tagged SNPs (rs2277613, rs962885, rs3785880, rs3744457, rs1467967, rs242557, rs3785883, rs2471738, rs7521) were required to represent all the SNPs in our Taiwanese cohort. The bi-allelic *del-In9* marker was used to unambiguously confirm all the Taiwaneses in our study are of the H1 clade [130]. The performance value for the 9 htSNPs was interpreted at an average haplotype r^2 value of 0.95 (95%). (**Figure 6.2**)

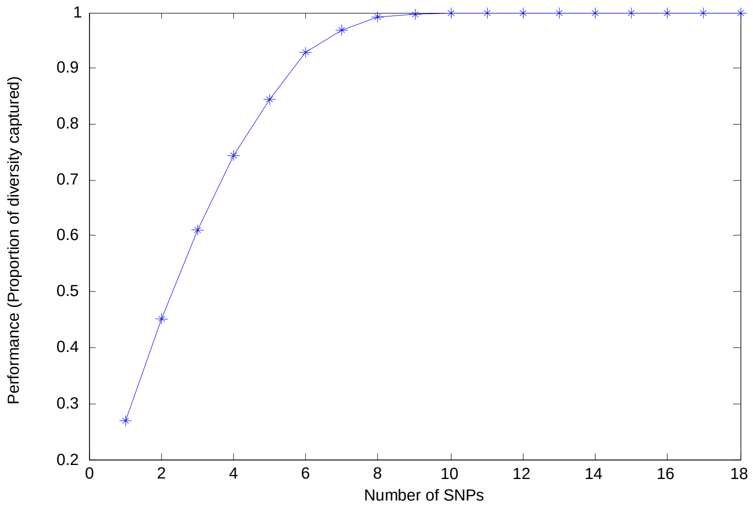


Figure 6.2 Analysis of performance of different number of htSNPs in Taiwanese cohort.

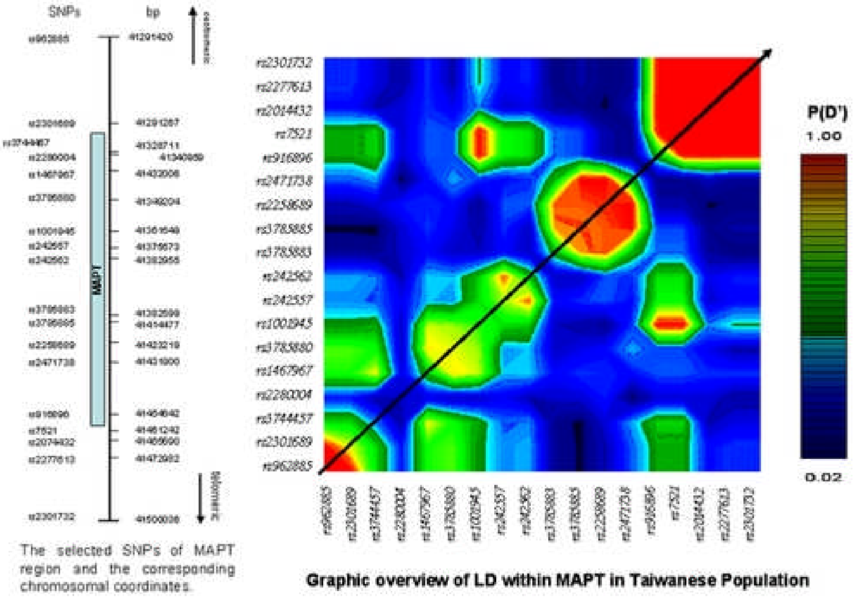
This analysis was calculated with the program TagIT. A minimum of 8 htSNPs were required to capture more than 95% of the diversity of all haplotypes in the Taiwanese cohort.

Oligonucleotide primer pairs were designed to specifically amplify by PCR, the *MAPT* haplotype SNPs of interest (rs2277613, rs962885, rs3785880, rs3744457, rs1467967, rs242557, rs3785883, rs2471738, rs7521). Genotyping of the SNPs was carried out either by pyrosequencing (Biotage, Pyrosequencing Inc., Charlottesville, VA, USA) or by RFLP. PCR and pyrosequencing primer sequences are shown in Appendices 10.3. The SNPs rs962885, rs2301689, rs3744457, rs2280004, rs1467967, rs3785880, rs1001945, rs242562, rs2303867, rs3785882, rs3785885, rs2258689, rs916896, rs2074432, rs2277613, rs876944 and rs2301732 were analysed with Pyrosequencing. (See section 2.1.5.2). For SNP analysis by RFLP, 15 µl of PCR product—SNPs: (rs242557, rs3785883, rs2471738, rs7521)—were digested by 1 U of the corresponding restriction endonuclease [New England Biolabs, Hertfordshire, UK (rs242557 [*Apa*L I(G)]; rs3785883 [*Bsa*H I(G)]; rs2471738 [*Bst*E II(T)]; rs7521[*Pst* I(A)]) in a reaction volume of 20 µl for 4 h. The PCR products are all cleaved by the corresponding enzyme once at the indicated (N) allele. Digests were run on a 4%

agarose gel for analysis. Genotype scoring was carried out blindly by two individuals. Any discrepancies between the two were resolved by repeating the assay.

6.4.2.3LD and statistical analysis

For each SNP, the allele and genotype distributions in the group of patients were compared with those in the control group. Statistical assessments for the allele and genotype frequencies and Hardy-Weinberg equilibrium (HWE) were made using the genetics software program TagIT (http://popgen.biol.ucl.ac.uk/software.html) [123]. The square of the correlation coefficient (r^2) and D' for LD was calculated pair-wise between each pair of SNP. Haplotype predictions were made using an EM algorithm using TagIT. Case–control locus-by-locus association was calculated statistically using a χ^2 distribution and the significance was calculated using a Monte–Carlo approach as implemented by CLUMP software [162]. The LD pattern of Taiwanese normal control population was illustrated with the software package, Graphical Overview of Linkage Disequilibrium [189]. (**Figure 6.3**)



The selected SNPs of MAPT region and the corresponding chromosomal coordinates.

Graphic overview of LD within MAPT in Taiwanese Population

Figure 6.3 Linkage disequilibrium (LD) across the *MAPT* in Taiwanese population.
The chromosomal coordinates of all single nucleotide polymorphisms (SNPs) used in this study with respect to the *MAPT* gene (pale blue rectangle on left) were shown in the left panel. Distribution of LD in Taiwanese is shown in the right part of this figure. Red, green and blue represent strong (D' ≥ 0.8), moderate (D' ≥ 0.5) and weak (D' ≥ 0.2) LD, respectively. The chromosomal coordinates (in bp) are based on May 2004 build of Human Genome Sequence (http://genome.ucsc.edu).

6.4.3 SNP amplification and genotyping

6.4.3.1 ApoE genotyping

APOE genotyping was performed as previously described [190]. Briefly, the SNP-containing region was amplified by PCR:

(Forward: 5'- GACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCGACGCAGG CCCGGCTGGACGCGGACATGGAGGA-3' and Reverse: 5'- AGGCCACGCTCG ACGCCCTCGCGGGCCCCGGCCTGGTACACT-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 was GC rich and required a modified PCR protocol: this included a longer initial denaturing step of 10 minutes at 94^{o}C and an additional 10 seconds in the denaturing steps in each cycle. A primer annealing temperature of 56^{o}C is used. Digested PCR products were run on a 4% low-melting agarose gel and stained with ethidium bromide to visualize the allele discriminating bands.

6.5 Results

6.5.1 The association of the H1c haplotype at the *MAPT* locus with AD in UK and US series

6.5.1.1 Single locus analysis

6.5.1.1.1 Single locus analysis

We genotyped six polymorphisms which we have shown previously tag the haplotype diversity in *MAPT* in Caucasians in two series of autopsy-confirmed controls and autopsy confirmed LOAD cases obtained from brain banks in the US and UK. (see Figure 6.1 for information on the usefulness of these SNPs to tag diversity in these populations). The single locus associations with AD are shown in Table 6.3. Tests of association of the intron 9 insertion–deletion (*del-In9*) polymorphism, which has been used to define the H1/H2 clades, were negative as we and others have previously reported [179-184]. Two of the tagging variants (rs242557 and rs2471738) had significant P-values in both the US series and when both the US and UK series were collapsed (**Table 6.3**). One of these tag variants (rs242557) also showed a trend towards association (allelic P-value = 0.094, genotypic P-value = 0.061) within the UK population.

	Location in		Major allele frequency		P-value					
Variants	MAPT	Major Allele	CO	AD	Allelelic		Genotypic		Risk Allele	Odds Ratio
US series										
APOE	N/A	ε4[a]	12	45	**3.47E-18**	**[2.09E-06]**	**1.36E-17**	**[3.46E-14]**	ε4	8.313
rs1467967	5' of ex 1	A	62	68	0.199	[0.763]	0.334	[0.077]	A	1.261
rs242557	5' of ex 1	G	70	62	**0.038**	**[0.043]**	0.107	[0.753]	A	1.456
rs3785883	Intron 3	G	78	83	0.1	[0.013]	0.195	[0.137]	G	1.403
rs2471738	Intron 9	C	82	74	**0.024**	**[0.032]**	**0.02**	**[0.075]**	T	1.598
del-In9	Intron 9	H1	78	78	1	[0.119]	0.248	[0.059]	H1	1.020
rs7521	3' of ex 14	G	56	55	0.87	[0.930]	0.315	[0.025]	A	1.035
UK Series										
APOE	N/A	ε4[a]	13	38	**1.28E-11**	**[7.66E-6]**	**3.77E-10**	**[1.24E-8]**	ε4	4.375
rs1467967	5' of ex 1	A	67	68	0.856	[0.848]	0.821	[0.053]	A	1.048
rs242557	5' of ex 1	G	65	58	0.094	[0.616]	0.061	[0.931]	A	1.351
rs3785883	Intron 3	G	85	87	0.465	[0.575]	0.293	[0.659]	G	1.222
rs2471738	Intron 9	C	82	76	0.081	[0.217]	0.229	[0.942]	T	1.449
del-In9	Intron 9	H1	74	74	1	[0.478]	0.763	[0.361]	H1	1.005
rs7521	3' of ex 14	G	57	57	1	[0.557]	0.515	[0.490]	A	1.006
UK and US series										
APOE	N/A	ε4[a]	13	41	**2.45E-28**	**[7.07E-11]**	**5.13E-27**	**[1.46E-22]**	ε4	5.938
rs1467967	5' of ex 1	A	65	68	0.262	[0.003]	0.469	[0.7]	A	1.157
rs242557	5' of ex 1	G	68	60	**0.007**	**[2.49E-07]**	0.013	[0.057]	A	1.408
rs3785883	Intron 3	G	81	85	0.071	[0.988]	0.114	[0.046]	G	1.325
rs2471738	Intron 9	C	82	75	**0.004**	**[0.046]**	**0.009**	**[0.021]**	T	1.524
del-In9	Intron 9	H1	76	76	0.946	[0.018]	0.224	[0.106]	H1	1.016
rs7521	3' of ex 14	G	57	56	0.906	[0.002]	0.168	[0.676]	A	1.020

Table 6.3 Single locus association in UK series and US series
ex, exon; CO, control; AD, Alzheimer's disease; CI, confidence interval. The relative locations, major allele frequencies, P-values, odds ratios and confidence intervals for all tagging loci tested in each series are shown. All values were calculated using SPSS software v.11. Significant values are highlighted in bold. P-values corrected for age, gender and *APOE* effects are given in brackets.
[a] For APOE, the ε 4 allele frequencies are listed, since this is the risk allele for LOAD, even though ε 4 is not the major allele.

6.5.1.1.2 Single locus analysis: APOE sub-analysis

We noted that when the single locus P-values were adjusted for age, sex and *APOE* status, many of the single locus P-values became more significant (see **Table 6.4** for both series and after collapsed; the P-value of rs242557 prior to age, sex, *APOE* adjustment = 0.007, after adjustment = 2.49E-07). We examined whether this effect was mainly due to age differences, gender differences or differences due to *APOE* background and found that the most robust interaction for each marker was with *APOE* status. To further examine this interaction, we divided the entire sample including both US and UK samples into two sub-series on the basis of *APOE*-ε4 genotype; cases and controls that possessed at least one ε4 allele were analysed separately from cases and controls that had no ε4 alleles. We found significant single locus P-values only in the sub-series where neither cases nor controls had any *APOE*-

ε4 alleles, suggesting that the single tag variant association is driven by those individuals who do not possess *APOE*-ε4 alleles.

			Major allele frequency		P-value				
Variants	Location in MAPT	Major Allele	CO	AD	Allelic	Genotypic	Risk Allele	Odds Ratio	CI (95%)
APOE-ε4 negative US and UK series collapsed									
rs1467967	5' of ex 1	A	63	67	0.355	0.533	A	1.177	0.833-1.667
rs242557	5' of ex 1	G	68	58	0.01	0.029	A	1.1577	1.112–2.096
rs3785883	Intron 3	G	79	83	0.249	0.159	G	1.288	0.837–1.980
rs2471738	Intron 9	C	83	75	0.017	0.048	T	1.637	1.092–2.454
del-In9	Intron 9	H1	75	78	0.541	0.006	H1	1.129	0.765–1.667
rs7521	3' of ex 14	G	57	54	0.594	0.66	A	1.094	0.787–1.520
APOE-ε4 positive series US and UK series collapsed									
rs1467967	5' of ex 1	A	71	68	0.649	0.751	G[a]	1.103	0.723–1.681
rs242557	5' of ex 1	G	67	61	0.206	0.182	A[a]	1.307	0.863–1.980
rs3785883	Intron 3	G	86	86	0.83	0.653	G[a]	1.064	0.607–1.862
rs2471738	Intron 9	C	81	75	0.207	0.308	T[a]	1.365	0.841–2.217
del-In9	Intron 9	H1	78	75	0.592	0.842	H2[a]	1.135	0.714–1.803
rs7521	3' of ex 14	G	56	57	0.845	0.119	G[a]	1.04	0.702–1.539

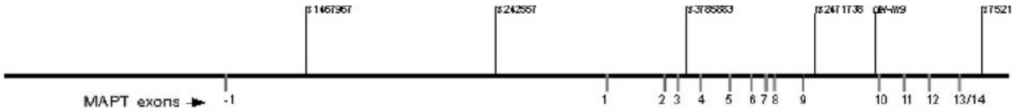
Table 6.4 Single locus associations within *APOE*-ε4 positive and negative subsets of the combined US and UK sample
The major allele frequencies, P-values, odds ratios and confidence intervals for all tagging loci tested in the subset of the entire sample that was *APOE*-ε4 negative and the subset of the series that was *APOE*-ε4 positive are shown. The *APOE*-ε4 negative subset contained all individuals from both the US and UK series that did not possess any *APOE*-ε4 alleles (genotypes 23 and 33). The *APOE*-ε4 positive subset contained all individuals from both the US and UK series that possessed at least one ε4 allele (genotypes 24, 34, 44). Abbreviations are the same as given in Table 6.3.

[a] Note that in the *APOE*-ε4 positive series the risk alleles are flipped for loci rs1467967, del-In9 and rs7521. This probably reflects the smaller sample size of the controls in this series (n=66).

6.5.1.2 Haplotype analysis

6.5.1.2.1 Haplotype analysis

The location of all tag SNPs with respect to *MAPT* exon structure as well as major haplotype frequencies and description of the alleles at each tag SNP for the major *MAPT* haplotypes is shown in Figure 6.4. Haplotype frequencies were obtained from the program SNPHAP (Clayton, D: http://www-gene.cimr.cam.ac.uk/clayton/software/snphap). As expected, we found no difference in the frequency of haplotype A (H2) between cases and controls in the UK or US series or when both samples are combined (UK LOAD frequency = 24.41%, control frequency = 25.44%, US LOAD frequency = 22.10%, US control frequency = 22.52%, combined series LOAD frequency = 23.39%, control frequency = 23.48%). This is in contrast to the clear negative association of this haplotype with PSP [177; see chapter 4]. Thus, these results are consistent with our single-locus analysis and with previous reports [177;182;184] that the *del-In9* polymorphism of *MAPT* is not associated with LOAD. As we had previously shown that the H1c variant of the *MAPT* locus was involved in risk for PSP [177], we decided to perform an analysis on this haplotype alone, thus minimizing multiple testing confounds. Using the constrain flag in the program Whap (http://www.genome.wi.mit.edu/~shaun/whap), we found a significant result testing the H1c variant against all other haplotypes in the UK series (empirical P-value = 0.045, 500 permutations of likelihood ratio test to obtain P-value), which replicated in the US series (empirical P-value = 0.018, 500 permutations of likelihood ratio test to obtain P-value). When both series were combined, likelihood ratio tests of haplotype H1c gave a P-value of 0.004. Examining the frequencies as predicted by SNPHAP, it appeared that the association was due to an over-representation of the H1c haplotype in LOAD (UK LOAD frequency = 15.11%, control frequency = 9.29%, US LOAD frequency = 13.17%, US control frequency = 8.05%, combined series LOAD frequency = 13.91%, control frequency = 8.51%). This association is in the same direction as, although smaller, that seen in our analysis of PSP where the H1c frequency is ~24%.



Haplotype	Frequency Control	LOAD	Alleles:				
H2a	24%	23%	A	G	G	C	*del*	G
H1b	13%	16%	G	G	G	C	*ins*	A
H1c	9%	14%	A	A	G	T	*ins*	G
H1d	5%	7%	A	A	G	C	*ins*	A
H1e	12%	7%	A	G	G	C	*ins*	A

Figure 6.4 MAPT haplotypes
The locations of all tag SNPs used in this study with respect to the exon structure of MAPT are shown. In addition, the five major (frequency > 5%) MAPT haplotypes are listed along with the frequency in controls, LOAD. Under the location of each tag SNP, the allele for that particular SNP is shown for each haplotype.

6.5.1.2.2 Haplotype analysis: *APOE* sub-analysis

In the light of the putative interaction with *APOE* genotype that we found examining single locus associations, we decided to examine whether *APOE* status had any influence on H1c haplotype association that we observed. We first tested whether there was an *APOE* interaction with *MAPT* using the full dataset and the '–gxe' and '–alt-gxe' flags in Whap. These two flags test whether there is a significant haplotype effect while adjusting for epistatic effects of another locus (–gxe) and whether there is a significant interaction between the genotypes of one locus and the haplotypes of the other (–alt-gxe). When haplotype H1c was tested for association controlling for the effect of the *APOE* locus, the P-value increased moderately from 0.004 to 0.005 (500 permutations of LRT to obtain P-value). When *APOE* was included in the model, using the '–alt-gxe' flag, the P-value decreased considerably (P-value with *APOE* = 5.16E-22, 500 permutations to obtain P-value), indicating a significant interaction between *APOE* genotype and *MAPT* haplotype. We then stratified our sample as in our single locus analysis by splitting the combined series into the subset of cases and controls that possessed *APOE*-ε4 alleles and into the cases and controls that had no *APOE*-ε4 alleles. As in the single locus analysis, the association between haplotype H1c and disease risk was only seen in those individuals who had no *APOE*-ε4 alleles (H1c P-value in subset of entire series where individuals had at least one ε4 allele = 0.238 and H1c P-value in subset of entire series where individuals had no ε4 allele = 0.008, 500 permutations to obtain P-values for both tests).

6.5.2 The association of the H1 haplotype at the *MAPT* locus with Taiwanese series

6.5.2.1 Linkage disequilibrium pattern of MAPT in Taiwanese

For the LD analysis of *MAPT* gene, we used the above 21 SNPs to genotype 21 normal Taiwanese individuals. We discarded SNPs that had a minor allele frequency of less than 5%. The average density of the markers is one SNP every 20 kb. None of the polymorphisms deviated from HWE. Additionally, the *del-In9* marker which defines the extended H1 and H2 clades was genotyped. As expected, all the Taiwanese individuals including cases and control were H1 homozygotes.

We evaluated pairwise LD across *MAPT* for all 18 SNPs in the 21 normal Taiwanese individuals both by D' and r^2, calculated from the EM inferred haplotypes. The *MAPT* region is featured by LD as is particularly evident by the measure D' (**Figure 6.3**). This pattern of LD across the extended *MAPT* region shows that the LD blocks in Taiwanese population have four blocks of D' LD. This structure, consisting of smaller blocks is consistent with LD throughout the genome and is in contrast with the unusually large region of LD in the Caucasian population. This fragmented haplotype structure would suggest that it would be easier to identify the region of the *MAPT* locus in which genetic variability contributes to disease risk in this population compared to the entire region of LD in the Caucasian population.

6.5.2.2 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). Nine htSNPs are required to represent all the SNPs in our Taiwanese cohort. The performance value for the 9 htSNPs was interpreted at an average haplotype r^2 value of 0.99 (99%). None of the 9 Taiwanese htSNPs deviated from HWE in any of the populations tested. The single locus association results are summarized in Table 6.5. APOE ε4 had a significant association with AD in this Taiwanese control (allelic p-value = 9.05e-4, genotypic p-value = 5.84e-4, OR = 2.91, CI = 1.56-5.43) and one of the tagging variants (rs242557) had a significant p-value (allellic p-value =1.93e-4, genotypic p-value =0.041 OR =2.17, CI=1.02-4.64) in our series.

Variant	CHROMOSOMAL LOCATION	Major Allele	Major Allele Frequency		p-value		Odds Ratio	CI (95%)		
			CONTROLS	AD	Allelic	Genotypic				
rs962885	41291420	T	77.4	75.7	0.674	0.589	1.10	0.71	to	1.70
rs3744457	41328711	C	63.0	67.3	0.352	0.651	1.21	0.81	to	1.80
rs1467967	41342006	G	61.9	66.5	0.316	0.188	1.22	0.83	to	1.80
rs3785880	41349204	T	67.1	72.3	0.237	0.337	0.78	0.52	to	1.18
rs242557	41375573	A	65.9	54.6	*0.015*	*0.021*	*0.62*	*0.42*	*to*	*0.91*
rs3785883	41410269	G	84.2	85.8	0.624	0.587	1.14	0.68	to	1.92
rs2471738	41431900	C	80.4	76.6	0.337	0.601	1.25	0.79	to	1.97
rs7521	41461242	G	88.9	89.9	0.725	0.619	1.11	0.61	to	2.03
rs2277613	41472982	C	61.5	56.4	0.279	0.207	1.24	0.84	to	1.82
APOE	n/a	ε4	9.0	21.6	*9.05E-04*	*5.84E-04*	*2.91*	*1.56*	*to*	*5.43*

Table 6.5 Single locus association analysis of *MAPT* with Taiwanese AD cohort.
The relative locations, major allele frequencies, p-values, odd ratios, and confidence intervals for all tagging loci tested in Taiwanese cases and control groups.
AD = Alzheimer's disease, CI: Confidence Interval. Significant values are shown in bold italics.

6.6 Discussion

The dominant hypothesis for the aetiology and pathogenesis of AD has been the amyloid hypothesis. This hypothesis is based on the observation that all the autosomal dominant mutations in either the APP or presenilin genes that cause AD, do so through their effect on APP/Aβ metabolism. However, experiments with mice with APP mutations (eg APPSwe (KM670/671NL)) that have been crossed with those with *MAPT* mutations (eg, P301L) have shown that the major pathway by which Aβ kills neurons involves tau biology and tangle formation [18;191;192]. Furthermore, tau expression appears to be needed for Aβ toxicity in *ex vivo* experiments [193]. We believe our data are most consistent with the view that, in the presence of an amyloid load, those individuals with *MAPT* variants that are either highly expressing or prone to express a more pathogenic species of tau through alternate splicing [194], are more prone to disease. Our single locus results would indicate that, as we found with PSP, the most likely region for a pathogenic variant(s) to occurs from just upstream of exon 1 to just within intron 9, as only rs242557 (5’ of exon 1) and rs2471738 (within intron 9) give significant single locus associations. The SNP rs242557 is within an 181 bp region that is conserved in human, chimp, mouse, dog and rat (ch17: 41375547 – 41375728, UCSC genome browser build 35, May 2004), and is ~19 kb upstream of the first coding exon of *MAPT*. The SNP rs2471738 does not lie within a conserved region of intron 9 and does not appear to interrupt the donor or acceptor splice sites, as it is ~2 kb away from the nearest intron–exon junction. This suggests that perhaps this variant is not functional, but is associated with risk because it is in linkage disequilibrium with another variant. In our analysis, it appears as though there might be a stronger association between the H1c variant of *MAPT* and risk in individuals who do not possess *APOE*-ε4 alleles. However, it should be noted that our *APOE* sub-analyses are underpowered because control individuals possessing *APOE*-ε4 alleles are fairly rare. Analysis in larger populations will need to be performed to confirm these initially interesting results. Irrespective of *APOE* status, we obtained significant associations of the H1c haplotype with both of our pathological AD series; the same haplotype shows robust association with PSP. This suggests that modulation of *MAPT* gene expression is a worthwhile approach to consider for the treatment of AD [62]. The *MAPT* gene has four blocks of high LD in the Taiwanese population (**Figure 6.3**). This haplotype structure is not unusual for the human genome but differs from the haplotype structure in European populations which is distorted by the occurrence of the non-recombining and inverted H2 haplotype [131].

Our study showed that rs242557 is associated with AD. This SNP is 5' of exon 1 (intron 0) and is the same SNP that we have previously reported to show a strong association with AD in European populations [167]. We have found that both the European and Taiwanese Alzheimer populations show the association with this locus but in different directions. Meanwhile, the A-allele is the major allele of this polymorphism in the Taiwanese population (~65%) while it is the minor allele in the European population (~35%) [167]. We propose that risk variants of *MAPT* associated with AD should be in the region within linkage disequilibrium region containing rs242557. This suggests that the variability that leads to predisposition is within a short distance of this SNP in intron 0 of *MAPT*. Furthermore, previously, in Chapter 4, this same SNP also showed a robust association with PSP and CBD. Together, these results indicate that this risk domain is involved in a common pathway in developing several of these neurodegenerative disorders. The SNP rs242557 lies within a ~181bp region that is conserved between human, chimp, mouse, dog and rat (ch17:41375547-41375728, USC Human Genome build 35). Sequencing of this 181bp region in all of our Taiwanese AD cases failed to reveal additional variability. This suggests either that rs242557 itself or other variants that lie outside this conserved region but within this LD block are likely to contribute the risk of development of AD and other related neurodegenerative disorders. Given these data and the position of this SNP, the most plausible explanation from this work and previous studies is that genetic variability in *MAPT* expression contributes to the risk of developing AD. This interpretation would be favoured either by the confirmation of this association with AD in another Asian population, or, indirectly, through the analysis of PSP or CBD in Taiwanese population; since in Caucasians, these share the same haplotypic association, but have a higher relative risk. Further investigation of this region with functional approaches should help to elucidate the pathogenesis of these neurodegenerative disorders.

Chapter 7 First-Stage Whole-Genome Association Study of Parkinson's disease

7.1 Overview

The previous decade has witnessed considerable progress in the identification of genes underlying rare monogenic forms of Parkinson's disease (PD). Despite evidence supporting a role for genetics in sporadic PD, no common genetic variants have been unequivocally linked to this disorder. In this study, a whole genome single nucleotide polymorphism genotyping was performed in a cohort of publicly available PD cases and neurologically normal controls using >408,000 unique SNPs from the Illumina Infinium I and Infinium II assays. These experiments were performed with the primary aim to detect if there is common genetic variability that exerts a large effect in risk for disease in our cohort.

Approximately 220 million genotypes were produced in 539 subjects. For the 408,803 SNPs studied, the genotype call rate was >95% for each of 396,591 SNPs. A total of 219,577,497 unique genotype calls were made. Analyzing the current data, seven loci were shown with a *p*-value less than $1\text{x}10^{-6}$, with OR ranging from 0.24 to 0.37 and 2.59 to 3.08

7.2 Background

Parkinson's disease (PD) is a chronic neurodegenerative disease with a cumulative prevalence greater than 1 per thousand [195]. The estimated sibling risk ratio (λs) for

PD is approximately 1.7 (70% increase risk for PD if a sibling has PD) for all ages, and increases by more than seven times for those younger than 66 years [196]. These data are consistent with a significant genetic contribution to disease risk.

While attempts to define the underlying lesions in monogenic forms of PD have been particularly successful [31-36], traditional testing of candidate-gene associations has been less successful. Few common variants have shown repeatable association with risk for Parkinson’s disease, the notable exception being common variation in *SNCA*, a gene originally implicated by results from family-based studies.

The completion of stages I and II of the International HapMap Project [197;198] in concert with the arrival of efficient, affordable high density SNP typing methods, promises to provide an approach with which to define the role of common genetic variation in risk for disease. This approach, much like traditional linkage methods, provides researchers with the ability to test variation in the genome in a relatively unbiased global manner, and thus does not rely on *a priori* hypotheses regarding mechanistic underpinnings of disease.

The International HapMap Project has provided a resource with which to calculate a minimum set of SNPs, often called tagging SNPs (tSNPs), which act a proxy markers for neighbouring genetic variation (also discussed in Chapter 2). Thus, a well-chosen set of several hundred thousand tSNPs will provide information about several million common genetic variants throughout the genome.

To begin to address the role of common genetic variation in idiopathic PD, a genome-wide SNP typing was performed using more than 408,000 unique SNPs across the genome. By using Illumina Infinium I and HumanHap300 assays, a genome-wide association study in 276 patients with Parkinson's disease and 276 neurologically normal controls was done.

7.3 Case-Control Samples

The samples used for this study were derived from the National Institute of Neurological Disorders and Stroke (NINDS) funded Neurogenetics repository, which includes collections of patients with PD, cerebrovascular disease, epilepsy and amyotrophic lateral sclerosis, in addition to neurologically normal controls (http://ccr.coriell.org/ninds/).

7.3.1 Subject Collection

Samples were derived from the NINDS Neurogenetics repository hosted by Coriell Institute for Medical research (NJ, USA) (http://ccr.coriell.org/ninds/). All subjects gave written informed consent to participate in the study. Six pre-compiled panels each consisting of 92 cases or controls were selected for the analysis. The panels that containing samples from patients with PD were *NDPT001*, *NDPT005* and *NDPT007*; these included DNA from 273 unique participants and three replicate samples. The panels that contained samples from neurologically normal controls were *NDPT002*, *NDPT006* and *NDPT008*; these comprised DNA from 275 unique subjects and one replicate sample. For the control population utilized in these experiments, blood samples were drawn from neurologically normal, unrelated, white individuals at many different sites within the USA. Each participant underwent a detailed medical history interview. None had a history of the following neurological diseases: Alzheimer's disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, brain aneurysm, dementia, dystonia, or PD. Folstein mini-mental state examination scores ranged from 26-30. All participants were interviewed for family history in detail. None have any first-degree relative with a known primary neurological disorder including amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson’s disease and schizophrenia. The mean age at collection was 68 years (range 55-88 years).

For the PD cohort, blood was drawn from unique and unrelated white individuals with idiopathic PD. The age of patients at onset of the disease ranged from 55 years to 84 years. Disease onset was defined as the time when symptoms of the disease were first noted, including at least one of the following: resting tremor, rigidity, bradykinesia, gait disorder or postural instability. All patients were queried about family history of parkinsonism, dementia, tremor, gait disorders, and other neurological dysfunction. Both those with and without a reported family history of PD were included on this panel. None were included who had three or more relatives with parkinsonism, nor with apparent Mendelian inheritance of PD.

7.3.2 Sample Preparation

DNA for the genotyping experiments was extracted using a salting out procedure from the Epstein-Barr virus immortalised lymphocyte cell lines. The average passage number for each line was five (range five to seven). Epstein-Barr virus immortalisation was undertaken as previously described [199]. At the same time, DNA was extracted from 0.5mL of blood from all participants for subsequent quality control steps in the cell-banking process.

7.4 Genotyping

All samples were assayed with the Illumina Infinium I and Infinium HumanHap300 SNP chips (Illumina Inc, San Diego, CA). These products assay 109,365 gene-centric SNPs (Infinium I) and 317,511 haplotype tagging SNPs derived from phase I of the international HapMap project (HumanMap300). There are 18,073 SNPs in common between the two arrays; thus the assays combined provide data on 408,803 unique SNPs. Any samples with a call rate below 95% were repeated on a fresh DNA aliquot, if the call rate persisted below this level the sample was excluded from the analysis. Low quality genotyping led us to repeat eleven individual samples; of which seven were ultimately excluded from the analysis.

7.5 Data Analysis

For each SNP, a series of estimates and tests was computed using a program developed at Wake Forest University called Snpgwa. Each SNP was tested for departures from HWE. Five tests of genotypic association were computed: 2 degree of freedom overall test for 2x3 tables, dominant model, additive model (Cochran-Armitage trend test), recessive model and lack of fit to an additive model (LOF). The odd ratios (OR), 95% CIs and *p* values were calculated for each of the association models. The computational program Dandelion which ran within Snpgwa was employed to perform two-marker and three-marker moving-window haplotype association analysis for those SNPs that were consistent with HWE in controls. For all p values with an uncorrected significance of less than 0.05 we did permutation tests within Snpgwa using a variable number of permutations based on the p value of the test. For each permutation, Snpgwa permutes the affection status (case or control) of the entire sample represented in the input file while preserving the total number of cases and total number of controls in each permutation. The permutation is done using a Wichman-Hill random number generator.

In an attempt to detect the presence of significant population sub-structure or ethnically mismatched individuals. 267 random, unlinked SNPs were selected throughout the genome and ran the program STRUCTURE on these data from all genotyped individuals [200].

7.6 Results

Two hundred and seventy six samples from patients with PD and 276 from unrelated population controls were genotyped. In the PD cohort there were 273 unique individuals, and in the control cohort there were 275 unique individuals. Genotyping of the four replicate samples with the Infinium I assay gave genotype concordance rates of greater than 99.99%. Analysis of the 18,073 SNPs that overlap between the Infinium I and HumanHap300 products revealed genotype concordance rates of 99.94% between the assays across 537 samples. Four samples were dropped from the control cohort due to low-quality genotyping; further analysis revealed that two of these samples (*ND01630* and *ND01666*) were contaminated; the other two samples (*ND03447* and *ND03704*) failed to meet the genotype quality threshold (95% call rate) after repeated assay. Thus, the total number of fully genotyped samples in the control cohort was 271. Six samples were dropped from the PD cohort, this included three young-onset samples that were erroneously included in panel *NDPT007* (*ND05074*, *ND05416* and *ND05841*). Samples *ND01500*, *ND04424* and *ND04744* were excluded from analysis because of genotype call rates below 95% after being assayed twice.

For the 408 803 SNPs studied, the genotype call rate was greater than 99% for each of 395,275 SNPs (96.6%) and greater than 95% for 406,312 SNPs (99.4%). The Hardy-Weinberg equilibrium p-value was higher than 0·001 for 395,493 SNPs and higher than 0.05 for 375,527 SNPs. The average minor allele frequency in autosomes was 26.47%. A total of 219,577,497 unique genotype calls were made and the average call rate across all samples was 99.6%.

Statistical analysis of association was done for all genotypes, irrespective of Hardy-Weinberg disequilibrium or minor allele frequency. The most significantly associated SNPs are shown in **Table 7.1**

Chr. Location	dbSNP ID	Location bp (genome build 36.1)	No. Geno	Gene	Putative function	HWE *p*-value	*p*-value 2DF	Empirical *p*-value 2DF	*p*-value Dom/Add/Rec	OR (95% CI)	Empirical *p*-value
11q14	rs10501570	84095494	536	*DLG2*	a member of the membrane-associated guanylate kinase family, may interact at postsynaptic sites	0.396	7.3E-6	2.0E-06	5.3E-4^{R}	0.2 (0.0-0.5)	4.9E-4^{R}
17p11.2	rs281357	19683106	537	*ULK2*	similar to a serine/threonine kinase in *C. elegans* which is involved in axonal elongation	0.852	9.8E-6	4.0E-06	0.0002^{R}	0.4 (0.2-0.6)	1.5E-5^{R}
4q13.2	rs2242330^{+}	68129844	537	*BRDG1*	docking protein acting downstream of Tec tyrosine kinase in B cell antigen receptor signaling	0.708	1.7E-5	1.2E-05	2.9E-6^{A}	0.5 (0.4-0.7)	<1E-6^{A}
10q11.21	rs1480597*	44481115	525	Intergenic		1.000	1.9E-5	7.0E-06	3.2E-6^{D}	0.4 (0.3-0.6)	2.0E-6^{D}
4q13.2	rs6826751^{+}	68116450	536	*BRDG1*	as above	0.024	2.0E-5	1.8E-05	3.5E-6^{A}	0.6 (0.4-0.7)	5.0E-6^{A}
16q23.1	rs4888984	78066835	537	Intergenic		1.000	2.7E-5	1.1E-05	4.6E-6^{A}	0.5 (0.3-0.7)	3.0E-6^{A}
4q35.2	rs4862792	188438344	511	Intergenic		0.358	3.5E-5	8.0E-06	6.8E-6^{D}	2.9 (1.8-4.6)	7.0E-6^{D}
4q13.2	rs3775866^{+}	68126775	537	*BRDG1*	as above	0.911	4.6E-5	3.3E-05	7.8E-6^{A}	0.5 (0.4-0.7)	8.0E-6^{A}
20q13.13	rs2235617‡	47988384	530	*ZNF313*	metal ion binding, protein binding, zinc ion binding, involved in cell differentiation and spermatogenesis	0.034	4.7E-5	4.7E-05	8.8E-6^{D}	0.4 (0.3-0.6)	1.2E-5^{D}
1p31	rs988421	72322424	536	*NEGR1*	Neuronal growth regulator	0.667	4.9E-5	4.3E-05	7.0E-4^{R}	2.0 (1.3-3.0)	8.2E-4^{R}
10q11.21	rs7097094*	44530696	537	Intergenic		0.294	5.0E-5	2.7E-05	8.9E-6^{D}	0.5 (0.3-0.7)	8.0E-6^{D}
10q11.21	rs999473*	44502322	537	Intergenic		0.294	5.0E-5	3.8E-05	8.9E-6^{D}	2.2 (1.5-3.1)	8.0E-6^{D}
11q11	rs1912373	56240441	537	Intergenic		0.375	5.6E-5	6.1E-05	9.7E-6^{D}	2.2 (1.6-3.2)	1.2E-5^{D}
1q25	rs1887279$^{\#}$	182176783	537	*GLT25D2*	Glycosyltransferase 25 domain containing 2	0.424	5.7E-5	3.5E-05	1.2E-5^{A}	0.5 (0.4-0.7)	6.0E-6^{A}
1q25	rs2986574$^{\#}$	182173237	536	*GLT25D2*	Glycosyltransferase 25 domain containing 2	0.350	6.3E-5	2.4E-05	1.3E-5^{A}	2.0 (1.4-2.7)	6.0E-6^{A}
22q13	rs11090762	46133989	536	Intergenic		0.730	6.3E-5	4.2E-05	1.2E-5^{D}	0.4 (0.3-0.6)	8.0E-6^{D}
20q13.13	rs6125829‡	48002336	509	*ZNF313*	metal ion binding, protein binding, zinc ion binding, involved in cell differentiation and spermatogenesis	0.004	6.6E-5	7.2E-05	1.4E-5^{D}	2.2 (1.6-3.2)	1.8E-5^{D}
7p12	rs7796855	49627992	537	Intergenic		0.931	6.6E-5	7.2E-05	1.3E-5^{D}	0.4 (0.3-0.6)	1.2E-5^{D}
4q13.2	rs355477^{+}	68079120	533	*BRDG1*	as above	0.207	7.9E-5	7.4E-05	1.5E-5^{A}	0.6 (0.5-0.8)	1.7E-5^{A}
1q25	rs3010040$^{\#}$	182174845	537	*GLT25D2*	Glycosyltransferase 25 domain containing 2	0.421	8.0E-5	6.2E-05	1.6E-5^{A}	0.5 (0.4-0.7)	1.2E-5^{A}
1q25	rs2296713$^{\#}$	182176340	537	*GLT25D2*	Glycosyltransferase 25 domain containing 2	0.421	8.0E-5	6.2E-05	1.6E-5^{A}	2.0 (1.4-2.7)	1.2E-5^{A}
4q13.2	rs355461^{+}	68063319	537	*BRDG1*	as above	0.150	8.3E-5	6.0E-05	1.6E-5^{A}	1.7 (1.3-2.2)	1.9E-5^{A}
4q13.2	rs355506^{+}	68068677	537	*BRDG1*	as above	0.150	8.3E-5	6.0E-05	1.6E-5^{A}	1.7 (1.3-2.2)	1.9E-5^{A}
4q13.2	rs355464^{+}	68061719	531	*BRDG1*	as above	0.086	8.9E-5	9.3E-05	1.7E-5^{A}	1.7 (1.3-2.2)	2.1E-5^{A}
4q13.2	rs1497430^{+}	68040409	535	*BRDG1*	as above	0.150	9.7E-5	8.0E-05	1.8E-5^{A}	1.7 (1.3-2.2)	1.9E-5^{A}
4q13.2	rs11946612^{+}	68018566	535	*BRDG1*	as above	0.150	9.7E-5	8.5E-05	1.8E-5^{A}	0.6 (0.5-0.8)	2.1E-5^{A}

Table 7.1 Summary of the most significant associated SNPs in Genome-wide genotyping. p-values with uncorrected significant > 0.0001 for SNPs that gave successful genotypes in > 95% of samples. HWE=Hardy Weinberg equilibrium. D=dominant. R=recessive. A=additive. No. geno=number of successful genotypes generated. +,#,‡,* : Closely associated SNPs.

Analysis with STRUCTURE [200] showed that there is no discernable difference in the population sub-structure between cases and controls (**Figure 7.1**). Furthermore, comparison of the cases and controls pooled together versus genotypes from a cohort of 173 non-white participants showed clear separation of the PD and control group from the non-white group, with the exception of a single patient from the former cohort, who, based on these analyses, had significant non-white genetic background. This individual was removed from the association analysis.

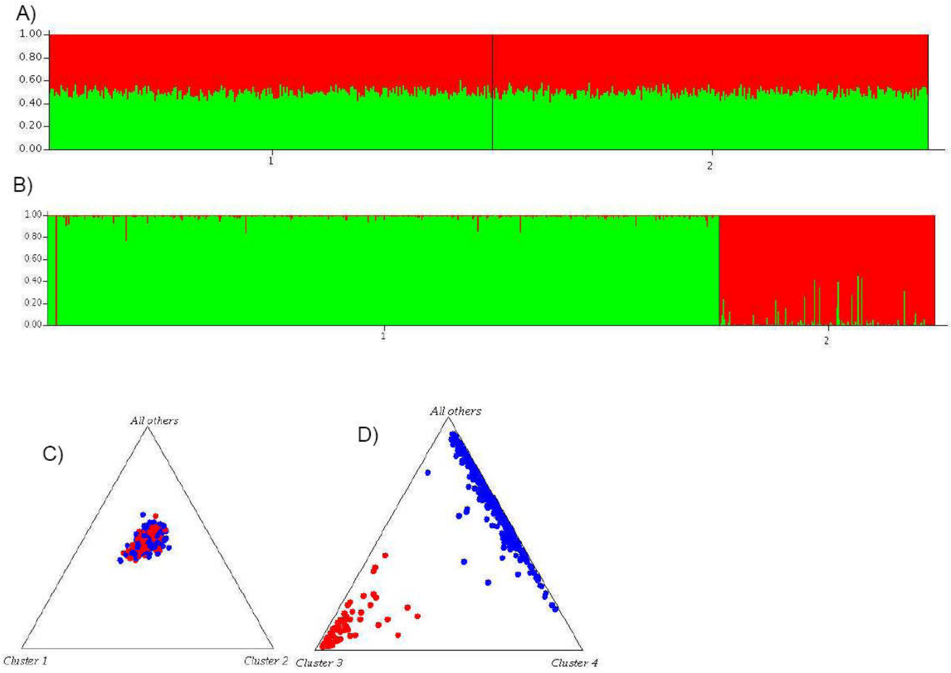


Figure 7.1 Bar and triangle plots from STRUCTURE using 267 random autosomal SNPs
A) Bar plot for K=2, sorted by putative population 1 consists of 271 white controls and population 2 consists of 267 patients with sporadic PD. B) Bar plot for K=2, sorted by putative population using the same set of 267 SNPs where population 1 consists of 538 whites (sporadic PD case/control series) and population 2 consists of 173 non-white participants. C) Triangle plot with same putative population as bar plot A) but with K=4, where blue dots are population 1 (controls) and red dots are population 2 (PD). D) Triangle plot with same putative population as bar plot B) but with K=4, where blue dots are population 1 (white sporadic PD patients and controls) and red dots are population 2 (non-white participants). The non-white population are self-idenitified African American subjects from the NIA sponsored study Health Aging in Neighbourhoods of Diversity across the Life Span (http://handis.nih.gov/)

7.7 Discussion

The aim of the present experiments was two-fold; first, to generate publicly available genotype data for Parkinson's disease patients and controls so that these data could be mined and augmented by other researchers; second, to perform a preliminary analysis in an attempt to localize common genetic variation exerting a large influence on risk for PD in a white north American cohort. These are the first genome-wide SNP genotype data, outside of the International HapMap Project, to be made publicly available.

Our data provides 80% power to detect an allelic association with an odds ratio of more than 2.09 and less than 0.40 at an uncorrected significance level of p=0.000001. This calculation is based on the average observed minor allele frequency of 26% and assumes that either the causal variant is typed or that there is complete and efficient tagging of common variation by the genotyped tSNPs. Although the sample size here is of limited power there is precedence for the use of small cohorts to identify genes of large effect by gene-wide association studies; the analysis of around 100,000 SNPs in only 96 cases with age-related macular degeneration and 50 controls led to the identification of variability within the gene encoding complement factor H as a risk factor for disease [212]. These data on macular degeneration draw attention to the use of genome-wide association studies in localisation of common genetic variability associated with disease, although the size of effect in that study was much higher than would generally be expected in most complex diseases (in macular degeneration the OR for homozygous carriers was 7.4). Illustrating the size of effects expected in complex disorders, the locus most robustly associated with risk of PD is the *SNCA* gene. A significant association was not identified at this particular locus; however, given that the OR associated with this locus is estimated at 1.4 it is not surprising that we were unable to identify an association.

Analysis of our data showed 26 loci with a two-degree of freedom p value less than 0.0001 (**Table 7.1**), with ORs ranging from 0.2 (95% CI 0.04–0.5) to 0.6 (0.5–0.8) and from 1.7 (1.3–2.2) to 2.2 (1.6–3.2). A stringent Bonferroni correction based on 408,803 independent tests means that a precorrection p value of less than 1.2×10^{-7} would be needed to provide a corrected significant p value of less than 0.05. Thus, none of the values listed were significant after correction. Although speculation on the plausibility and biological significance of these candidate loci is tempting, we regard these data as hypothesis generating. Furthermore, given the inevitably high false-positive rate of genome-wide association studies, the next step in these analyses should involve genotyping in additional sample series. In the first instance, this work should be done in a cohort comprising patients and controls of similar demographic characteristics to reduce the confounds of allelic and genetic heterogeneity between ethnic groups. This approach would involve continued whole-genome SNP genotyping in the additional PD cases and controls available from the Coriell Neurogenetics repository. However, a more cost-effective measure would be to do follow-up genotyping of several thousand of the most significantly associated SNPs in additional cases and controls. The release of genotype data and not just allele frequency data means that genotype data from additional samples can be added easily to the current set allowing investigators to undertake joint analysis rather than replication-based analysis. The former approach is more powerful than the latter in identifying common genetic risk factors [201]. The control samples in the current study have been specifically obtained so that they can be used for other neurological disorders, including but not restricted to stroke and amyotrophic lateral sclerosis, so these data will also be of use to other researchers outside of the PD speciality.

A genome-wide association analysis of Parkinson’s disease has been done with a two-tiered design with slightly fewer than 200 000 SNPs [215]. Although this study used fewer than half of the SNPs used in our study, the multistage design added substantial power and sensitivity to the results. The authors of these experiments suggested that their data revealed 13 SNPs associated with risk for Parkinson’s disease. We, and others, have not been able to confirm these findings in independent cohorts [216,217]. Side-by-side comparison of the current data and the most significantly associated SNPs, published by Maraganore and colleagues, did not show a replication of any of these published associations [215]. One plausible approach is to combine or compare odds ratios of physically close SNPs, although data compared between studies and across platforms should be viewed with appropriate caution.

Our data suggest that there are no common genetic variants that exert an effect of greater than an OR of 4 in PD. From the standpoint of experimental design this information is very useful. However, there are important drawbacks to this interpretation. First, these results can strictly only be applied to the current population. Second, analysis of young-onset PD cases, where a genetic effect is thought to be stronger, could reveal genetic variants with an effect of this size [202]. Third, this statement is reliant on either genotyping the causal variant or efficient and complete tagging of the causal variant.

In summary, the generation and release of genotype data derived from publicly available PD and neurologically normal control samples were presented. These data suggest that there is no common genetic variant that exerts a large genetic risk for late-onset PD in white north Americans. These data are now available for future mining and augmentation to identify common genetic variability that results in minor and moderate risk for disease.

Chapter 8 Discussion

8.1 Summary

Under the framework of this thesis, the geographical distribution worldwide of *MAPT* H1 and H2 haplotype were studied. There is an almost complete association between the H2 haplotype and Caucasian ancestry as this H2 locus is not found in other populations. This extended haplotype block has not just shown a unique distribution, but is also the longest region of complete linkage disequilibrium, which spans over ~1.8 Mb, in the genome. The pattern of this LD over the *MAPT* region was also found to be shared by different ethnic groups, including French, Orkney Islanders, Italian, Russian, Pakistani and United Kingdom populations, but not in the Japanese or Taiwanese populations, which have only H1 haplotype. A series of association studies between this *MAPT* and different neurodegenerative diseases, PSP, CBD, AD and PD were conducted with different populations globally. Locus-by-locus association analysis revealed the defined *MAPT* haplotype block was associated with PSP. In this study, 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) for Caucasian populations (United Kingdom, Finnish, Greek, United States) and 9 htSNPs for Taiwanese population were selected that together captured almost all (> 95%) the genetic diversity of the gene. Association analysis revealed that two common haplotypes were associated with PSP, AD and the same trend in CBD. In 2007, Myers et al has revealed that the H2 haplotype as defined by *del-In9* is a significant protective factor while there is an increased expression of 4 microtuble binding repeat containing *MAPT* transcripts driven by the variant of H1 (H1c) in AD.

These findings also gave support to the hypothesis of our studies in this thesis [213]. In PSP association study, the strongest associated H1-specific SNPs 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. In AD, the single locus analysis revealed that the same rs242557 and rs2471738 were associated with AD in the US and UK populations, while only the rs242557 in Taiwanese population. Significant associations were obtained in both US and UK of the pathological AD series with the H1c haplotypes. In the PD series, the rs242557 and rs7521 were found to be associated in Finnish cohort, while rs3785883 was moderately associated with the disease in the Finnish population. No significant differences in genotype or allele distribution were identified between cases and controls in the Taiwanese cohort. To address the role of common genetic variation, besides *MAPT* in idiopathic PD (Chapter 5), a whole-genome association analysis was performed with 26 loci with a two degree of freedom p-value less than 0.0001, with odds ratios ranging from 0.2 (95% CI 0.04-0.5) to 0.6 (0.5-0.8) and 1.7 (1.3-2.2) to 2.2 (1.6-3.2).

8.2 General Discussion

8.2.1 A general association between common *MAPT* haplotypes and neurodegenerative diseases

Testing for association between the *MAPT* locus and PSP, CBD, AD and PD was facilitated by the ease and consistency by which one could test H1 and H2 haplotypes in European populations. In this work, PSP, CBD and AD showed robust associations, but PD did not show a consistent association. Furthermore, the absence of the H2 haplotype from Asian populations meant it was not possible to test simple H1/H2 associations in those populations. However, with the delineation of further variability on the H1 background, the assessments were done in this work. The results in this thesis have suggested that the H1 variant that showed the strongest association to PSP (H1c) also showed associations with AD and CBD. Furthermore, the H1-specific SNP rs242557, showed associations with the AD, PSP and CBD in all the studied populations in this study. These observations raise the possibility, which is perhaps not surprising, that genetic variability in *MAPT* expression and/or splicing contributes to the risk of developing all tangle diseases.

Recently, Myers et al further extended their study from confirming the association between H1c clade and AD in an autopsy confirmed series to found that the H1 clade increases the expression of total MAPT transcript and especially of 4 microtubule binding repeat containing transcripts [213]. Caffrey et al independently showed that the protective *MAPT* H2 haplotype significantly expresses two-fold more *MAPT* transcripts with both exon 2 and 3 than the disease-associated H1 haplotype in both cell culture and post-mortem brain tissue. Caffrey et al suggested in the report that inclusion of exon 3 in MAPT transcripts may contribute to protecting H2 carrier from neurodegeneration [214]. Though these findings may imply that different mechanisms lead to the ultimate neurodegeneration, they support the notion that the basis of genetic associations could be by control of expression or splicing. From the perspective of the pathogenesis of AD, these data suggest that the up-regulatinon of production of the disease related, four-repeat isoform of the tau protein (4R-tau) is the major drive of pathogenesis. Its role in tangle pathogenesis may be analogous to that of Aβ42 in plaque pathogenesis. This speculation ostensibly has therapeutic implications.

In PD, several case-control association studies have evaluated the association of *MAPT* [49;161;163;203-207]. However, these studies are largely underpowered and failed to give a definite answer. In 2004, Healy et al reported 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 [49]. However, this association was weaker. This could be because PD is a synucleinopathy, rather than a sole tauopathy. One of the plausible explanations is the changes in expression of *MAPT* caused by the genetic variations could contribute to increase vulnerability but without the concomitant production of tangle pathology; and the risks conferred by those genetic variations to PD is less than those to the tauopathies.

This observation supports the hypothesis that the neurodegenerative diseases could be classified by molecular variations in the genes. Molecular genetic analysis is having an impact on this approach to disease by two ways. First, it is offering a window on the aetiology and pathogenesis of disease, making clear how disease may be initiated. Second, it is showing that the boundaries of diseases are not where they might have been expected to be [208] .

8.3 Future Research

8.3.1 Population-based genetic association studies

The diverse clinical manifestations of the neurodegenerative diseases, with similar pathological findings in different populations, suggests that the underlying genetic variations could be a key to understand the pathogenesis of neurodegenerative diseases. Current and developing genetic techniques make it possible to study genomic variation between population groups and offer the opportunity to test the hypothesis that diseases have divergent clinical features between populations. There is growing evidence that race and ethnicity modulate disease via genetic background, although it is difficult to consider this separately from the influence of environment on the phenotype, since ethnicity may correlate with geographic and behavioural differences.

The present study identified the association of the H1c haplotype and H1-specific SNP (rs242557) with PSP, CBD and AD in different populations. In order to recognise the genetic risk factors of the neurodegenerative diseases, it would be important to extend this study in two directions: First, by genotyping the same set of htSNPs in other neurodegenerative diseases; second, by performing the association studies of other candidate genes, such as alpha-synuclein (*SNCA)* and *APOE,* with other neurodegenerative diseases in different ethnic groups. These investigations of neurodegenerative diseases in different populations are required to further clarify the phenotypes differ within as well as across different racial and ethnic groups. Such investigations would also have a significant impact on the appropriate diagnosis and treatment of neurodegenerative diseases globally. Once the diagnosis and treatment for neurodegenerative diseases are based on an understanding of pathogenesis, these differences in the phenotype will be of even greater clinical relevance.[209]

8.3.2 Whole-genome association studies in neurodegenerative diseases

The many possible approaches to mapping the genes that underlie common disease and quantitative traits fall broadly into two categories: candidate gene studies, which use either association or re-sequencing approaches, and genome-wide studies, which include linkage mapping and genome-wide association study [106].

In this study, the genome-wide association study in PD was carried out. This genome wide association approach is an association study that surveys most of the genome for causal genetic variants. Because no assumption is made about the genomic location of the causal variants, this approach could exploit the strengths of association studies without having to guess or be biased by the identity of the causal genes. Thus, the genome-wide association approach represents an unbiased yet fairly comprehensive option that can be attempted even in the absence of convincing evidence regarding the function or location of the causal gene. The genome-wide genotyping is an emerging powerful tool for detecting the genetic variants which significantly increase disease risk but insufficient to actually cause a specific disorder. Association studies will be carried out in this setting more cost-effectively to find out the common variants which cause the neurodegenerative diseases with complex genetic traits. [106;210;211]

9 Reference List

[1] Hardy,J., Orr,H., The genetics of neurodegenerative diseases, J. Neurochem., 97 (2006) 1690-1699.

[2] Brown,R.C., Lockwood,A.H., Sonawane,B.R., Neurodegenerative diseases: an overview of environmental risk factors, Environ. Health Perspect., 113 (2005) 1250-1256.

[3] Soto,C., Unfolding the role of protein misfolding in neurodegenerative diseases, Nat. Rev. Neurosci., 4 (2003) 49-60.

[4] Langston,J.W., Ballard,P., Tetrud,J.W., Irwin,I., Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis, Science, 219 (1983) 979-980.

[5] Parsons,R.B., Smith,M.L., Williams,A.C., Waring,R.H., Ramsden,D.B., Expression of nicotinamide N-methyltransferase (E.C. 2.1.1.1) in the Parkinsonian brain, J. Neuropathol. Exp. Neurol., 61 (2002) 111-124.

[6] Garruto,R.M., Gajdusek,D.C., Chen,K.M., Amyotrophic lateral sclerosis and parkinsonism-dementia among Filipino migrants to Guam, Ann. Neurol., 10 (1981) 341-350.

[7] Hardy,J., Gwinn-Hardy,K., Genetic classification of primary neurodegenerative disease, Science, 282 (1998) 1075-1079.

[8] Cummings JL, Neurodegenerative Disorders as Proteinopathies: Phenotypic Relationships. In Genotype-Proteotype-Phenotype Relationships in Neurodegenerative Diseases Springer-Verlag Berlin Heidelberg New York, New York, 2005, pp. 1-10.

[9] Bertram,L., Tanzi,R.E., The genetic epidemiology of neurodegenerative

disease, J. Clin. Invest, 115 (2005) 1449-1457.

[10] Lander,E.S., The new genomics: global views of biology, Science, 274 (1996) 536-539.

[11] Tanzi,R.E., A genetic dichotomy model for the inheritance of Alzheimer's disease and common age-related disorders, J. Clin. Invest, 104 (1999) 1175-1179.

[12] Goate,A., Chartier-Harlin,M.C., Mullan,M., Brown,J., Crawford,F., Fidani,L., Giuffra,L., Haynes,A., Irving,N., James,L., ., Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease, Nature, 349 (1991) 704-706.

[13] Sherrington,R., Rogaev,E.I., Liang,Y., Rogaeva,E.A., Levesque,G., Ikeda,M., Chi,H., Lin,C., Li,G., Holman,K., ., Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease, Nature, 375 (1995) 754-760.

[14] Rogaev,E.I., Sherrington,R., Rogaeva,E.A., Levesque,G., Ikeda,M., Liang,Y., Chi,H., Lin,C., Holman,K., Tsuda,T., ., Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene, Nature, 376 (1995) 775-778.

[15] Levy-Lahad,E., Wasco,W., Poorkaj,P., Romano,D.M., Oshima,J., Pettingell,W.H., Yu,C.E., Jondro,P.D., Schmidt,S.D., Wang,K., ., Candidate gene for the chromosome 1 familial Alzheimer's disease locus, Science, 269 (1995) 973-977.

[16] Rogaeva,E., The solved and unsolved mysteries of the genetics of early-onset Alzheimer's disease, Neuromolecular. Med., 2 (2002) 1-10.

[17] Alzheimer Disease & Frontotemporal Dementia Mutation Database 2007

(URL: http://www.molgen.ua.ac.be/ADMutations/)

[18] Hardy,J., Selkoe,D.J., The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics, Science, 297 (2002) 353-356.

[19] Wilquet,V., De Strooper,B., Amyloid-beta precursor protein processing in neurodegeneration, Curr. Opin. Neurobiol., 14 (2004) 582-588.

[20] Mayeux,R., Sano,M., Chen,J., Tatemichi,T., Stern,Y., Risk of dementia in first-degree relatives of patients with Alzheimer's disease and related disorders, Arch. Neurol., 48 (1991) 269-273.

[21] Strittmatter,W.J., Saunders,A.M., Schmechel,D., Pericak-Vance,M., Enghild,J., Salvesen,G.S., Roses,A.D., Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease, Proc. Natl. Acad. Sci. U. S. A, 90 (1993) 1977-1981.

[22] Schmechel,D.E., Saunders,A.M., Strittmatter,W.J., Crain,B.J., Hulette,C.M., Joo,S.H., Pericak-Vance,M.A., Goldgaber,D., Roses,A.D., Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease, Proc. Natl. Acad. Sci. U. S. A, 90 (1993) 9649-9653.

[23] Farrer,L.A., Cupples,L.A., Haines,J.L., Hyman,B., Kukull,W.A., Mayeux,R., Myers,R.H., Pericak-Vance,M.A., Risch,N., van Duijn,C.M., Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium, JAMA, 278 (1997) 1349-1356.

[24] Daw,E.W., Payami,H., Nemens,E.J., Nochlin,D., Bird,T.D., Schellenberg,G.D., Wijsman,E.M., The number of trait loci in late-onset

Alzheimer disease, Am. J. Hum. Genet., 66 (2000) 196-204.

[25] Cruts,M., van Duijn,C.M., Backhovens,H., Van den,B.M., Wehnert,A., Serneels,S., Sherrington,R., Hutton,M., Hardy,J., George-Hyslop,P.H., Hofman,A., Van Broeckhoven,C., Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease, Hum. Mol. Genet., 7 (1998) 43-51.

[26] Campion,D., Dumanchin,C., Hannequin,D., Dubois,B., Belliard,S., Puel,M., Thomas-Anterion,C., Michon,A., Martin,C., Charbonnier,F., Raux,G., Camuzat,A., Penet,C., Mesnage,V., Martinez,M., Clerget-Darpoux,F., Brice,A., Frebourg,T., Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum, Am. J. Hum. Genet., 65 (1999) 664-670.

[27] Finckh,U., van Hadeln,K., Muller-Thomsen,T., Alberici,A., Binetti,G., Hock,C., Nitsch,R.M., Stoppe,G., Reiss,J., Gal,A., Association of late-onset Alzheimer disease with a genotype of PLAU, the gene encoding urokinase-type plasminogen activator on chromosome 10q22.2, Neurogenetics., 4 (2003) 213-217.

[28] Tanzi,R.E., Bertram,L., New frontiers in Alzheimer's disease genetics, Neuron, 32 (2001) 181-184.

[29] Tanner,C.M., Ottman,R., Goldman,S.M., Ellenberg,J., Chan,P., Mayeux,R., Langston,J.W., Parkinson disease in twins: an etiologic study, JAMA, 281 (1999) 341-346.

[30] Maher,N.E., Golbe,L.I., Lazzarini,A.M., Mark,M.H., Currie,L.J., Wooten,G.F., Saint-Hilaire,M., Wilk,J.B., Volcjak,J., Maher,J.E., Feldman,R.G., Guttman,M., Lew,M., Waters,C.H., Schuman,S., Suchowersky,O., Lafontaine,A.L., Labelle,N., Vieregge,P., Pramstaller,P.P., Klein,C., Hubble,J.,

Reider,C., Growdon,J., Watts,R., Montgomery,E., Baker,K., Singer,C., Stacy,M., Myers,R.H., Epidemiologic study of 203 sibling pairs with Parkinson's disease: the GenePD study, Neurology, 58 (2002) 79-84.

[31] Polymeropoulos,M.H., Lavedan,C., Leroy,E., Ide,S.E., Dehejia,A., Dutra,A., Pike,B., Root,H., Rubenstein,J., Boyer,R., Stenroos,E.S., Chandrasekharappa,S., Athanassiadou,A., Papapetropoulos,T., Johnson,W.G., Lazzarini,A.M., Duvoisin,R.C., Di Iorio,G., Golbe,L.I., Nussbaum,R.L., Mutation in the alpha-synuclein gene identified in families with Parkinson's disease, Science, 276 (1997) 2045-2047.

[32] Kitada,T., Asakawa,S., Hattori,N., Matsumine,H., Yamamura,Y., Minoshima,S., Yokochi,M., Mizuno,Y., Shimizu,N., Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism, Nature, 392 (1998) 605-608.

[33] Bonifati,V., Rizzu,P., van Baren,M.J., Schaap,O., Breedveld,G.J., Krieger,E., Dekker,M.C., Squitieri,F., Ibanez,P., Joosse,M., van Dongen,J.W., Vanacore,N., van Swieten,J.C., Brice,A., Meco,G., van Duijn,C.M., Oostra,B.A., Heutink,P., Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism, Science, 299 (2003) 256-259.

[34] Valente,E.M., Abou-Sleiman,P.M., Caputo,V., Muqit,M.M., Harvey,K., Gispert,S., Ali,Z., Del Turco,D., Bentivoglio,A.R., Healy,D.G., Albanese,A., Nussbaum,R., Gonzalez-Maldonado,R., Deller,T., Salvi,S., Cortelli,P., Gilks,W.P., Latchman,D.S., Harvey,R.J., Dallapiccola,B., Auburger,G., Wood,N.W., Hereditary early-onset Parkinson's disease caused by mutations in PINK1, Science, 304 (2004) 1158-1160.

[35] Paisan-Ruiz,C., Jain,S., Evans,E.W., Gilks,W.P., Simon,J., van der,B.M., Lopez,d.M., Aparicio,S., Gil,A.M., Khan,N., Johnson,J., Martinez,J.R.,

Nicholl,D., Carrera,I.M., Pena,A.S., de Silva,R., Lees,A., Marti-Masso,J.F., Perez-Tur,J., Wood,N.W., Singleton,A.B., Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease, Neuron, 44 (2004) 595-600.

[36] Ramirez,A., Heimbach,A., Gründemann,J., Stiller,B., Hampshire,D., Cid,L.P., Goebel,I., Mubaidin,A.F., Wriekat,A.L., Roeper,J., Al-Din,A., Hillmer,A.M., Karsak,M., Liss,B., Woods,C.G., Behrens,M.I., Kubisch,C., Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase, Nat Genet, 38 (2006) 1184-1191.

[37] Singleton,A.B., Farrer,M., Johnson,J., Singleton,A., Hague,S., Kachergus,J., Hulihan,M., Peuralinna,T., Dutra,A., Nussbaum,R., Lincoln,S., Crawley,A., Hanson,M., Maraganore,D., Adler,C., Cookson,M.R., Muenter,M., Baptista,M., Miller,D., Blancato,J., Hardy,J., Gwinn-Hardy,K., alpha-Synuclein locus triplication causes Parkinson's disease, Science, 302 (2003) 841.

[38] Chartier-Harlin,M.C., Kachergus,J., Roumier,C., Mouroux,V., Douay,X., Lincoln,S., Levecque,C., Larvor,L., Andrieux,J., Hulihan,M., Waucquier,N., Defebvre,L., Amouyel,P., Farrer,M., Destee,A., Alpha-synuclein locus duplication as a cause of familial Parkinson's disease, Lancet, 364 (2004) 1167-1169.

[39] Albrecht,M., LRRK2 mutations and Parkinsonism, Lancet, 365 (2005) 1230.

[40] Matsumine,H., Saito,M., Shimoda-Matsubayashi,S., Tanaka,H., Ishikawa,A., Nakagawa-Hattori,Y., Yokochi,M., Kobayashi,T., Igarashi,S., Takano,H., Sanpei,K., Koike,R., Mori,H., Kondo,T., Mizutani,Y., Schaffer,A.A., Yamamura,Y., Nakamura,S., Kuzuhara,S., Tsuji,S., Mizuno,Y., Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to

chromosome 6q25.2-27, Am. J. Hum. Genet., 60 (1997) 588-596.

[41] Petrucelli,L., O'Farrell,C., Lockhart,P.J., Baptista,M., Kehoe,K., Vink,L., Choi,P., Wolozin,B., Farrer,M., Hardy,J., Cookson,M.R., Parkin protects against the toxicity associated with mutant alpha-synuclein: proteasome dysfunction selectively affects catecholaminergic neurons, Neuron, 36 (2002) 1007-1019.

[42] van Duijn,C.M., Dekker,M.C., Bonifati,V., Galjaard,R.J., Houwing-Duistermaat,J.J., Snijders,P.J., Testers,L., Breedveld,G.J., Horstink,M., Sandkuijl,L.A., van Swieten,J.C., Oostra,B.A., Heutink,P., Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36, Am. J. Hum. Genet., 69 (2001) 629-634.

[43] Abou-Sleiman,P.M., Healy,D.G., Quinn,N., Lees,A.J., Wood,N.W., The role of pathogenic DJ-1 mutations in Parkinson's disease, Ann. Neurol., 54 (2003) 283-286.

[44] Valente,E.M., Bentivoglio,A.R., Dixon,P.H., Ferraris,A., Ialongo,T., Frontali,M., Albanese,A., Wood,N.W., Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36, Am. J. Hum. Genet., 68 (2001) 895-900.

[45] Leroy,E., Boyer,R., Auburger,G., Leube,B., Ulm,G., Mezey,E., Harta,G., Brownstein,M.J., Jonnalagada,S., Chernova,T., Dehejia,A., Lavedan,C., Gasser,T., Steinbach,P.J., Wilkinson,K.D., Polymeropoulos,M.H., The ubiquitin pathway in Parkinson's disease, Nature, 395 (1998) 451-452.

[46] Le,W.D., Xu,P., Jankovic,J., Jiang,H., Appel,S.H., Smith,R.G., Vassilatis,D.K., Mutations in NR4A2 associated with familial Parkinson disease, Nat. Genet., 33 (2003) 85-89.

[47] Maraganore,D.M., Lesnick,T.G., Elbaz,A., Chartier-Harlin,M.C., Gasser,T., Kruger,R., Hattori,N., Mellick,G.D., Quattrone,A., Satoh,J., Toda,T., Wang,J., Ioannidis,J.P., de Andrade,M., Rocca,W.A., UCHL1 is a Parkinson's disease susceptibility gene, Ann. Neurol., 55 (2004) 512-521.

[48] Scott,W.K., Nance,M.A., Watts,R.L., Hubble,J.P., Koller,W.C., Lyons,K., Pahwa,R., Stern,M.B., Colcher,A., Hiner,B.C., Jankovic,J., Ondo,W.G., Allen,F.H., Jr., Goetz,C.G., Small,G.W., Masterman,D., Mastaglia,F., Laing,N.G., Stajich,J.M., Slotterbeck,B., Booze,M.W., Ribble,R.C., Rampersaud,E., West,S.G., Gibson,R.A., Middleton,L.T., Roses,A.D., Haines,J.L., Scott,B.L., Vance,J.M., Pericak-Vance,M.A., Complete genomic screen in Parkinson disease: evidence for multiple genes, JAMA, 286 (2001) 2239-2244.

[49] Healy,D.G., Abou-Sleiman,P.M., Lees,A.J., Casas,J.P., Quinn,N., Bhatia,K., Hingorani,A.D., Wood,N.W., Tau gene and Parkinson's disease: a case-control study and meta-analysis, J. Neurol. Neurosurg. Psychiatry, 75 (2004) 962-965.

[50] Martin,E.R., Scott,W.K., Nance,M.A., Watts,R.L., Hubble,J.P., Koller,W.C., Lyons,K., Pahwa,R., Stern,M.B., Colcher,A., Hiner,B.C., Jankovic,J., Ondo,W.G., Allen,F.H., Jr., Goetz,C.G., Small,G.W., Masterman,D., Mastaglia,F., Laing,N.G., Stajich,J.M., Ribble,R.C., Booze,M.W., Rogala,A., Hauser,M.A., Zhang,F., Gibson,R.A., Middleton,L.T., Roses,A.D., Haines,J.L., Scott,B.L., Pericak-Vance,M.A., Vance,J.M., Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease, JAMA, 286 (2001) 2245-2250.

[51] Conrad,C., Andreadis,A., Trojanowski,J.Q., Dickson,D.W., Kang,D., Chen,X., Wiederholt,W., Hansen,L., Masliah,E., Thal,L.J., Katzman,R., Xia,Y., Saitoh,T., Genetic evidence for the involvement of tau in progressive supranuclear palsy, Ann. Neurol., 41 (1997) 277-281.

[52] Huang,X., Chen,P.C., Poole,C., APOE-[epsilon]2 allele associated with higher prevalence of sporadic Parkinson disease, Neurology, 62 (2004) 2198-2202.

[53] Farrer,M., Maraganore,D.M., Lockhart,P., Singleton,A., Lesnick,T.G., de Andrade,M., West,A., de Silva,R., Hardy,J., Hernandez,D., alpha-Synuclein gene haplotypes are associated with Parkinson's disease, Hum. Mol. Genet., 10 (2001) 1847-1851.

[54] Bower,J.H., Maraganore,D.M., McDonnell,S.K., Rocca,W.A., Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990, Neurology, 49 (1997) 1284-1288.

[55] Litvan,I., Agid,Y., Calne,D., Campbell,G., Dubois,B., Duvoisin,R.C., Goetz,C.G., Golbe,L.I., Grafman,J., Growdon,J.H., Hallett,M., Jankovic,J., Quinn,N.P., Tolosa,E., Zee,D.S., Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop, Neurology, 47 (1996) 1-9.

[56] Sergeant,N., Delacourte,A., Buee,L., Tau protein as a differential biomarker of tauopathies, Biochim. Biophys. Acta, 1739 (2005) 179-197.

[57] de Yebenes,J.G., Sarasa,J.L., Daniel,S.E., Lees,A.J., Familial progressive supranuclear palsy. Description of a pedigree and review of the literature, Brain, 118 (Pt 5) (1995) 1095-1103.

[58] Rojo,A., Pernaute,R.S., Fontan,A., Ruiz,P.G., 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,A.C., Hanson,A., de Yebenes,J.G., Clinical genetics of familial progressive supranuclear palsy, Brain, 122 (Pt 7) (1999) 1233-1245.

[59] Evans,W., Fung,H.C., Steele,J., Eerola,J., Tienari,P., Pittman,A., Silva,R.,

Myers,A., Vrieze,F.W., Singleton,A., Hardy,J., The tau H2 haplotype is almost exclusively Caucasian in origin, Neurosci. Lett., 369 (2004) 183-185.

[60] Stefansson,H., Helgason,A., Thorleifsson,G., Steinthorsdottir,V., Masson,G., Barnard,J., Baker,A., Jonasdottir,A., Ingason,A., Gudnadottir,V.G., Desnica,N., Hicks,A., Gylfason,A., Gudbjartsson,D.F., Jonsdottir,G.M., Sainz,J., Agnarsson,K., Birgisdottir,B., Ghosh,S., Olafsdottir,A., Cazier,J.B., Kristjansson,K., Frigge,M.L., Thorgeirsson,T.E., Gulcher,J.R., Kong,A., Stefansson,K., A common inversion under selection in Europeans, Nat. Genet., 37 (2005) 129-137.

[61] Hutton,M., Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms, Neurology, 56 (2001) S21-S25.

[62] Singleton,A., Myers,A., Hardy,J., 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 (2004) R123-R126.

[63] Gibb,W.R., Luthert,P.J., Marsden,C.D., Clinical and pathological features of corticobasal degeneration, Adv. Neurol., 53 (1990) 51-54.

[64] Komori,T., Tau-positive glial inclusions in progressive supranuclear palsy, corticobasal degeneration and Pick's disease, Brain Pathol., 9 (1999) 663-679.

[65] Baba,Y., Putzke,J.D., Tsuboi,Y., Josephs,K.A., Thomas,N., Wszolek,Z.K., Dickson,D.W., Effect of MAPT and APOE on prognosis of progressive supranuclear palsy, Neurosci. Lett., 405 (2006) 116-119.

[66] Di Maria,E., Tabaton,M., Vigo,T., Abbruzzese,G., Bellone,E., Donati,C., Frasson,E., Marchese,R., Montagna,P., Munoz,D.G., Pramstaller,P.P., Zanusso,G., Ajmar,F., Mandich,P., Corticobasal degeneration shares a

common genetic background with progressive supranuclear palsy, Ann. Neurol., 47 (2000) 374-377.

[67] 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., Khan,M.N., Hardy,J., Lantos,P.L., George-Hyslop,P., Munoz,D.G., Mann,D., Lang,A.E., Bergeron,C., Bigio,E.H., Litvan,I., Bhatia,K.P., Dickson,D., Wood,N.W., Hutton,M., Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype, Neurology, 56 (2001) 1702-1706.

[68] Witman,G.B., Cleveland,D.W., Weingarten,M.D., Kirschner,M.W., Tubulin requires tau for growth onto microtubule initiating sites, Proc. Natl. Acad. Sci. U. S. A, 73 (1976) 4070-4074.

[69] Binder,L.I., Frankfurter,A., Rebhun,L.I., The distribution of tau in the mammalian central nervous system, J. Cell Biol., 101 (1985) 1371-1378.

[70] Gu,Y., Oyama,F., Ihara,Y., Tau is widely expressed in rat tissues, J. Neurochem., 67 (1996) 1235-1244.

[71] LoPresti,P., Szuchet,S., Papasozomenos,S.C., Zinkowski,R.P., Binder,L.I., Functional implications for the microtubule-associated protein tau: localization in oligodendrocytes, Proc. Natl. Acad. Sci. U. S. A, 92 (1995) 10369-10373.

[72] Avila,J., Tau phosphorylation and aggregation in Alzheimer's disease pathology, FEBS Lett., 580 (2006) 2922-2927.

[73] Kanemaru,K., Takio,K., Miura,R., Titani,K., Ihara,Y., Fetal-type phosphorylation of the tau in paired helical filaments, J. Neurochem., 58 (1992) 1667-1675.

[74] Andreadis,A., Brown,W.M., Kosik,K.S., Structure and novel exons of the

human tau gene, Biochemistry, 31 (1992) 10626-10633.

[75] Goedert,M., Spillantini,M.G., Potier,M.C., Ulrich,J., Crowther,R.A., 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 (1989) 393-399.

[76] Goedert,M., Wischik,C.M., Crowther,R.A., Walker,J.E., Klug,A., 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, 85 (1988) 4051-4055.

[77] Goedert,M., Tau gene mutations and their effects, Mov Disord., 20 Suppl 12 (2005) S45-S52.

[78] Goedert,M., Spillantini,M.G., Jakes,R., Rutherford,D., Crowther,R.A., Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease, Neuron, 3 (1989) 519-526.

[79] de Silva,R., Lashley,T., Strand,C., Shiarli,A.M., Shi,J., Tian,J., Bailey,K.L., Davies,P., Bigio,E.H., Arima,K., Iseki,E., Murayama,S., Kretzschmar,H., Neumann,M., Lippa,C., Halliday,G., MacKenzie,J., Ravid,R., Dickson,D., Wszolek,Z., Iwatsubo,T., Pickering-Brown,S.M., Holton,J., Lees,A., Revesz,T., Mann,D.M., An immunohistochemical study of cases of sporadic and inherited frontotemporal lobar degeneration using 3R- and 4R-specific tau monoclonal antibodies, Acta Neuropathol. (Berl), 111 (2006) 329-340.

[80] Delacourte,A., Sergeant,N., Wattez,A., Gauvreau,D., Robitaille,Y., Vulnerable neuronal subsets in Alzheimer's and Pick's disease are distinguished by their tau isoform distribution and phosphorylation, Ann. Neurol., 43 (1998) 193-204.

[81] Arai,T., Ikeda,K., Akiyama,H., Shikamoto,Y., Tsuchiya,K., Yagishita,S., Beach,T., Rogers,J., Schwab,C., McGeer,P.L., 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 (2001) 167-173.

[82] Togo,T., Sahara,N., Yen,S.H., Cookson,N., Ishizawa,T., Hutton,M., de Silva,R., Lees,A., Dickson,D.W., Argyrophilic grain disease is a sporadic 4-repeat tauopathy, J. Neuropathol. Exp. Neurol., 61 (2002) 547-556.

[83] Sergeant,N., Wattez,A., Delacourte,A., Neurofibrillary degeneration in progressive supranuclear palsy and corticobasal degeneration: tau pathologies with exclusively "exon 10" isoforms, J. Neurochem., 72 (1999) 1243-1249.

[84] de Silva,R., Lashley,T., Gibb,G., Hanger,D., Hope,A., Reid,A., Bandopadhyay,R., Utton,M., Strand,C., Jowett,T., Khan,N., Anderton,B., Wood,N., Holton,J., Revesz,T., Lees,A., 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 (2003) 288-302.

[85] Sergeant,N., David,J.P., Goedert,M., Jakes,R., Vermersch,P., Buee,L., Lefranc,D., Wattez,A., Delacourte,A., 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 (1997) 834-844.

[86] Golbe,L.I., Neurodegeneration in the age of molecular biology, BMJ, 324 (2002) 1467-1468.

[87] Goedert,M., Sisodia,S.S., Price,D.L., Neurofibrillary tangles and beta-amyloid deposits in Alzheimer's disease, Curr. Opin. Neurobiol., 1 (1991) 441-447.

[88] Ikeda,S., Tokuda,T., Yanagisawa,N., Kametani,F., Ohshima,T., Allsop,D., Variability of beta-amyloid protein deposited lesions in Down's syndrome brains, Tohoku J. Exp. Med., 174 (1994) 189-198.

[89] Yoshimura,N., Kubota,S., Fukushima,Y., Kudo,H., Ishigaki,H., Yoshida,Y., Down's syndrome in middle age. Topographical distribution and immunoreactivity of brain lesions in an autopsied patient, Acta Pathol. Jpn., 40 (1990) 735-743.

[90] Pearce,J.M., Pick's disease, J. Neurol. Neurosurg. Psychiatry, 74 (2003) 169.

[91] Burn,D.J., Lees,A.J., Progressive supranuclear palsy: where are we now?, Lancet Neurol., 1 (2002) 359-369.

[92] Rademakers,R., Cruts,M., Van Broeckhoven,C., The role of tau (MAPT) in frontotemporal dementia and related tauopathies, Hum. Mutat., 24 (2004) 277-295.

[93] Nirmalananthan,N., Greensmith,L., Amyotrophic lateral sclerosis: recent advances and future therapies, Curr. Opin. Neurol., 18 (2005) 712-719.

[94] Vanier,M.T., Millat,G., Niemann-Pick disease type C, Clin. Genet., 64 (2003) 269-281.

[95] Greenberg,S.G., Davies,P., A preparation of Alzheimer paired helical filaments that displays distinct tau proteins by polyacrylamide gel electrophoresis, Proc. Natl. Acad. Sci. U. S. A, 87 (1990) 5827-5831.

[96] Weeks,D.E., Lathrop,G.M., Polygenic disease: methods for mapping complex disease traits, Trends Genet., 11 (1995) 513-519.

[97] Elston,R.C., The genetic dissection of multifactorial traits, Clin. Exp. Allergy, 25 Suppl 2 (1995) 103-106.

[98] Devlin,B., Risch,N., A comparison of linkage disequilibrium measures for fine-scale mapping, Genomics, 29 (1995) 311-322.

[99] LEWONTIN,R.C., THE INTERACTION OF SELECTION AND LINKAGE. II. OPTIMUM MODELS, Genetics, 50 (1964) 757-782.

[100] Cordell,H.J., Clayton,D.G., Genetic association studies, Lancet, 366 (2005) 1121-1131.

[101] Hirose,S., Mitsudome,A., Okada,M., Kaneko,S., Genetics of idiopathic epilepsies, Epilepsia, 46 Suppl 1 (2005) 38-43.

[102] Hattersley,A.T., McCarthy,M.I., What makes a good genetic association study?, Lancet, 366 (2005) 1315-1323.

[103] Ioannidis,J.P., Trikalinos,T.A., Khoury,M.J., Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases, Am. J. Epidemiol., 164 (2006) 609-614.

[104] Cardon,L.R., Palmer,L.J., Population stratification and spurious allelic association, Lancet, 361 (2003) 598-604.

[105] Pritchard,J.K., Donnelly,P., Case-control studies of association in structured or admixed populations, Theor. Popul. Biol., 60 (2001) 227-237.

[106] Hirschhorn,J.N., Daly,M.J., Genome-wide association studies for common diseases and complex traits, Nat. Rev. Genet., 6 (2005) 95-108.

[107] Kruglyak,L., Nickerson,D.A., Variation is the spice of life, Nat. Genet., 27 (2001) 234-236.

[108] Syvanen,A.C., Accessing genetic variation: genotyping single nucleotide polymorphisms, Nat. Rev. Genet., 2 (2001) 930-942.

[109] A haplotype map of the human genome, Nature, 437 (2005) 1299-1320.

[110] Kruglyak,L., Prospects for whole-genome linkage disequilibrium mapping of common disease genes, Nat. Genet., 22 (1999) 139-144.

[111] Jorde,L.B., Linkage disequilibrium and the search for complex disease genes, Genome Res., 10 (2000) 1435-1444.

[112] Daly,M.J., Rioux,J.D., Schaffner,S.F., Hudson,T.J., Lander,E.S., High-resolution haplotype structure in the human genome, Nat. Genet., 29 (2001) 229-232.

[113] Patil,N., Berno,A.J., Hinds,D.A., Barrett,W.A., Doshi,J.M., Hacker,C.R., Kautzer,C.R., Lee,D.H., Marjoribanks,C., McDonough,D.P., Nguyen,B.T., Norris,M.C., Sheehan,J.B., Shen,N., Stern,D., Stokowski,R.P., Thomas,D.J., Trulson,M.O., Vyas,K.R., Frazer,K.A., Fodor,S.P., Cox,D.R., Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21, Science, 294 (2001) 1719-1723.

[114] Gabriel,S.B., Schaffner,S.F., Nguyen,H., Moore,J.M., Roy,J., Blumenstiel,B., Higgins,J., DeFelice,M., Lochner,A., Faggart,M., Liu-Cordero,S.N., Rotimi,C., Adeyemo,A., Cooper,R., Ward,R., Lander,E.S., Daly,M.J., Altshuler,D., The structure of haplotype blocks in the human genome, Science, 296 (2002) 2225-2229.

[115] Johnson,G.C., Esposito,L., Barratt,B.J., Smith,A.N., Heward,J., Di Genova,G., Ueda,H., Cordell,H.J., Eaves,I.A., Dudbridge,F., Twells,R.C., Payne,F., Hughes,W., Nutland,S., Stevens,H., Carr,P., Tuomilehto-Wolf,E., Tuomilehto,J., Gough,S.C., Clayton,D.G., Todd,J.A., Haplotype tagging for the identification of common disease genes, Nat. Genet., 29 (2001) 233-237.

[116] Dawson,E., Abecasis,G.R., Bumpstead,S., Chen,Y., Hunt,S., Beare,D.M.,

Pabial,J., Dibling,T., Tinsley,E., Kirby,S., Carter,D., Papaspyridonos,M., Livingstone,S., Ganske,R., Lohmussaar,E., Zernant,J., Tonisson,N., Remm,M., Magi,R., Puurand,T., Vilo,J., Kurg,A., Rice,K., Deloukas,P., Mott,R., Metspalu,A., Bentley,D.R., Cardon,L.R., Dunham,I., A first-generation linkage disequilibrium map of human chromosome 22, Nature, 418 (2002) 544-548.

[117] Crawford,D.C., Carlson,C.S., Rieder,M.J., Carrington,D.P., Yi,Q., Smith,J.D., Eberle,M.A., Kruglyak,L., Nickerson,D.A., Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations, Am. J. Hum. Genet., 74 (2004) 610-622.

[118] Carlson,C.S., Eberle,M.A., Kruglyak,L., Nickerson,D.A., Mapping complex disease loci in whole-genome association studies, Nature, 429 (2004) 446-452.

[119] Goldstein,D.B., Ahmadi,K.R., Weale,M.E., Wood,N.W., Genome scans and candidate gene approaches in the study of common diseases and variable drug responses, Trends Genet., 19 (2003) 615-622.

[120] Zhang,K., Deng,M., Chen,T., Waterman,M.S., Sun,F., A dynamic programming algorithm for haplotype block partitioning, Proc. Natl. Acad. Sci. U. S. A, 99 (2002) 7335-7339.

[121] Stram,D.O., Haiman,C.A., Hirschhorn,J.N., Altshuler,D., Kolonel,L.N., Henderson,B.E., Pike,M.C., Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study, Hum. Hered., 55 (2003) 27-36.

[122] Ke,X., Cardon,L.R., Efficient selective screening of haplotype tag SNPs, Bioinformatics., 19 (2003) 287-288.

[123] Weale,M.E., Depondt,C., Macdonald,S.J., Smith,A., Lai,P.S., Shorvon,S.D.,

Wood,N.W., Goldstein,D.B., Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping, Am. J. Hum. Genet., 73 (2003) 551-565.

[124] Carlson,C.S., Eberle,M.A., Rieder,M.J., Yi,Q., Kruglyak,L., Nickerson,D.A., Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium, Am. J. Hum. Genet., 74 (2004) 106-120.

[125] Halldorsson,B.V., Bafna,V., Lippert,R., Schwartz,R., De La Vega,F.M., Clark,A.G., Istrail,S., Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies, Genome Res., 14 (2004) 1633-1640.

[126] Poorkaj,P., Bird,T.D., Wijsman,E., Nemens,E., Garruto,R.M., Anderson,L., Andreadis,A., Wiederholt,W.C., Raskind,M., Schellenberg,G.D., Tau is a candidate gene for chromosome 17 frontotemporal dementia, Ann. Neurol., 43 (1998) 815-825.

[127] Hutton,M., Lendon,C.L., 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,R.C., Stevens,M., de Graaff,E., Wauters,E., van Baren,J., Hillebrand,M., Joosse,M., Kwon,J.M., Nowotny,P., Che,L.K., Norton,J., Morris,J.C., Reed,L.A., Trojanowski,J., Basun,H., Lannfelt,L., Neystat,M., Fahn,S., Dark,F., Tannenberg,T., Dodd,P.R., Hayward,N., Kwok,J.B., Schofield,P.R., Andreadis,A., Snowden,J., Craufurd,D., Neary,D., Owen,F., Oostra,B.A., 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, 393 (1998) 702-705.

[128] Abou-Sleiman,P.M., Hanna,M.G., Wood,N.W., Genetic association studies of

complex neurological diseases, J. Neurol. Neurosurg. Psychiatry, 77 (2006) 1302-1304.

[129] Spillantini,M.G., Goedert,M., Tau protein pathology in neurodegenerative diseases, Trends Neurosci., 21 (1998) 428-433.

[130] 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., 8 (1999) 711-715.

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

[132] Fung,H.C., Evans,J., Evans,W., Duckworth,J., Pittman,A., de Silva,R., Myers,A., Hardy,J., The architecture of the tau haplotype block in different ethnicities, Neurosci. Lett., 377 (2005) 81-84.

[133] Skipper,L., Wilkes,K., Toft,M., Baker,M., Lincoln,S., Hulihan,M., Ross,O.A., Hutton,M., Aasly,J., Farrer,M., Linkage disequilibrium and association of MAPT H1 in Parkinson disease, Am. J. Hum. Genet., 75 (2004) 669-677.

[134] Poorkaj,P., Tsuang,D., Wijsman,E., Steinbart,E., Garruto,R.M., Craig,U.K., Chapman,N.H., Anderson,L., Bird,T.D., Plato,C.C., Perl,D.P., Weiderholt,W., Galasko,D., Schellenberg,G.D., TAU as a susceptibility gene for amyotropic lateral sclerosis-parkinsonism dementia complex of Guam, Arch. Neurol., 58 (2001) 1871-1878.

[135] Mulder,D.W., Kurland,L.T., Iriarte,L.L., Neurologic diseases on the island of Guam, U. S. Armed. Forces. Med. J., 5 (1954) 1724-1739.

[136] Plato,C.C., Garruto,R.M., Galasko,D., Craig,U.K., Plato,M., Gamst,A., Torres,J.M., Wiederholt,W., Amyotrophic lateral sclerosis and parkinsonism-dementia complex of Guam: changing incidence rates during the past 60 years, Am. J. Epidemiol., 157 (2003) 149-157.

[137] Morris,H.R., Steele,J.C., Crook,R., Wavrant-De Vrieze,F., Onstead-Cardinale,L., Gwinn-Hardy,K., Wood,N.W., Farrer,M., Lees,A.J., McGeer,P.L., Siddique,T., Hardy,J., Perez-Tur,J., Genome-wide analysis of the parkinsonism-dementia complex of Guam, Arch. Neurol., 61 (2004) 1889-1897.

[138] Cann,H.M., de Toma,C., Cazes,L., Legrand,M.F., Morel,V., Piouffre,L., Bodmer,J., Bodmer,W.F., Bonne-Tamir,B., Cambon-Thomsen,A., Chen,Z., Chu,J., Carcassi,C., Contu,L., Du,R., Excoffier,L., Ferrara,G.B., Friedlaender,J.S., Groot,H., Gurwitz,D., Jenkins,T., Herrera,R.J., Huang,X., Kidd,J., Kidd,K.K., Langaney,A., Lin,A.A., Mehdi,S.Q., Parham,P., Piazza,A., Pistillo,M.P., Qian,Y., Shu,Q., Xu,J., Zhu,S., Weber,J.L., Greely,H.T., Feldman,M.W., Thomas,G., Dausset,J., Cavalli-Sforza,L.L., A human genome diversity cell line panel, Science, 296 (2002) 261-262.

[139] Eerola,J., Hernandez,D., Launes,J., Hellstrom,O., Hague,S., Gulick,C., Johnson,J., Peuralinna,T., Hardy,J., Tienari,P.J., Singleton,A.B., Assessment of a DJ-1 (PARK7) polymorphism in Finnish PD, Neurology, 61 (2003) 1000-1002.

[140] Perez-Tur,J., Buee,L., Morris,H.R., Waring,S.C., Onstead,L., Wavrant-De Vrieze,F., Crook,R., Buee-Scherrer,V., Hof,P.R., Petersen,R.C., McGeer,P.L., Delacourte,A., Hutton,M., Siddique,T., Ahlskog,J.E., Hardy,J., Steele,J.C., Neurodegenerative diseases of Guam: analysis of TAU, Neurology, 53 (1999) 411-413.

[141] Zerjal,T., Dashnyam,B., Pandya,A., Kayser,M., Roewer,L., Santos,F.R., Schiefenhovel,W., Fretwell,N., Jobling,M.A., Harihara,S., Shimizu,K., Semjidmaa,D., Sajantila,A., Salo,P., Crawford,M.H., Ginter,E.K., Evgrafov,O.V., Tyler-Smith,C., Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis, Am. J. Hum. Genet., 60 (1997) 1174-1183.

[142] Kwok,J.B., Teber,E.T., Loy,C., Hallupp,M., Nicholson,G., Mellick,G.D., Buchanan,D.D., Silburn,P.A., Schofield,P.R., Tau haplotypes regulate transcription and are associated with Parkinson's disease, Ann. Neurol., 55 (2004) 329-334.

[143] Zietkiewicz,E., Modern human origins and prehistoric demography of Europe in light of the present-day genetic diversity, J. Appl. Genet., 42 (2001) 509-530.

[144] Dickman,M.S., von Economo encephalitis, Arch. Neurol., 58 (2001) 1696-1698.

[145] Ishii,T., Takeyasu,K., The C-terminal 165 amino acids of the plasma membrane Ca(2+)-ATPase confer Ca2+/calmodulin sensitivity on the Na+,K(+)-ATPase alpha-subunit, EMBO J., 14 (1995) 58-67.

[146] Garg,R.K., Subacute sclerosing panencephalitis, Postgrad. Med. J., 78 (2002) 63-70.

[147] Ikeda,K., Akiyama,H., Kondo,H., Arai,T., Arai,N., Yagishita,S., Numerous glial fibrillary tangles in oligodendroglia in cases of subacute sclerosing panencephalitis with neurofibrillary tangles, Neurosci. Lett., 194 (1995) 133-135.

[148] Holzer,M., Craxton,M., Jakes,R., Arendt,T., Goedert,M., Tau gene (MAPT)

sequence variation among primates, Gene, 341 (2004) 313-322.

[149] Conrad,C., Vianna,C., Schultz,C., Thal,D.R., Ghebremedhin,E., Lenz,J., Braak,H., Davies,P., Molecular evolution and genetics of the Saitohin gene and tau haplotype in Alzheimer's disease and argyrophilic grain disease, J. Neurochem., 89 (2004) 179-188.

[150] Maher,E.R., Lees,A.J., The clinical features and natural history of the Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy), Neurology, 36 (1986) 1005-1008.

[151] Daniel,S.E., de Bruin,V.M., Lees,A.J., The clinical and pathological spectrum of Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy): a reappraisal, Brain, 118 (Pt 3) (1995) 759-770.

[152] Wszolek,Z.K., Tsuboi,Y., Uitti,R.J., Reed,L., Hutton,M.L., Dickson,D.W., Progressive supranuclear palsy as a disease phenotype caused by the S305S tau gene mutation, Brain, 124 (2001) 1666-1670.

[153] 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., Schellenberg,G.D., An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype, Ann. Neurol., 52 (2002) 511-516.

[154] Morris,H.R., Osaki,Y., Holton,J., Lees,A.J., Wood,N.W., Revesz,T., Quinn,N., Tau exon 10 +16 mutation FTDP-17 presenting clinically as sporadic young onset PSP, Neurology, 61 (2003) 102-104.

[155] 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., Lees,A.J., Sequence analysis of tau in familial and sporadic progressive supranuclear palsy, J. Neurol. Neurosurg. Psychiatry, 72 (2002) 388-390.

[156] Ezquerra,M., Pastor,P., Valldeoriola,F., Molinuevo,J.L., 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., 275 (1999) 183-186.

[157] de Silva,R., Weiler,M., Morris,H.R., Martin,E.R., Wood,N.W., Lees,A.J., Strong association of a novel Tau promoter haplotype in progressive supranuclear palsy, Neurosci. Lett., 311 (2001) 145-148.

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

[159] 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, 99 (2002) 7751-7756.

[160] Ponting,C.P., Hutton,M., Nyborg,A., Baker,M., Jansen,K., Golde,T.E., Identification of a novel family of presenilin homologues, Hum. Mol. Genet., 11 (2002) 1037-1044.

[161] Morris,H.R., Janssen,J.C., Bandmann,O., Daniel,S.E., Rossor,M.N., Lees,A.J., Wood,N.W., The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases, J. Neurol. Neurosurg. Psychiatry, 66 (1999) 665-667.

[162] Sham,P.C., Curtis,D., Monte Carlo tests for associations between disease and alleles at highly polymorphic loci, Ann. Hum. Genet., 59 (1995) 97-105.

[163] Golbe,L.I., Lazzarini,A.M., Spychala,J.R., Johnson,W.G., Stenroos,E.S., Mark,M.H., Sage,J.I., The tau A0 allele in Parkinson's disease, Mov Disord., 16 (2001) 442-447.

[164] Oliveira,S.A., Scott,W.K., Zhang,F., Stajich,J.M., Fujiwara,K., Hauser,M., Scott,B.L., Pericak-Vance,M.A., Vance,J.M., Martin,E.R., Linkage disequilibrium and haplotype tagging polymorphisms in the Tau H1 haplotype, Neurogenetics., 5 (2004) 147-155.

[165] Pastor,P., Ezquerra,M., Perez,J.C., Chakraverty,S., Norton,J., Racette,B.A., McKeel,D., Perlmutter,J.S., Tolosa,E., Goate,A.M., Novel haplotypes in 17q21 are associated with progressive supranuclear palsy, Ann. Neurol., 56 (2004) 249-258.

[166] Buee,L., Delacourte,A., Comparative biochemistry of tau in progressive supranuclear palsy, corticobasal degeneration, FTDP-17 and Pick's disease, Brain Pathol., 9 (1999) 681-693.

[167] 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., The H1c haplotype at the MAPT locus is associated with Alzheimer's disease, Hum. Mol. Genet., 14 (2005) 2399-2404.

[168] Martin,E.R., Scott,W.K., Nance,M.A., Watts,R.L., Hubble,J.P., Koller,W.C., Lyons,K., Pahwa,R., Stern,M.B., Colcher,A., Hiner,B.C., Jankovic,J., Ondo,W.G., Allen,F.H., Jr., Goetz,C.G., Small,G.W., Masterman,D., Mastaglia,F., Laing,N.G., Stajich,J.M., Ribble,R.C., Booze,M.W., Rogala,A., Hauser,M.A., Zhang,F., Gibson,R.A., Middleton,L.T., Roses,A.D., Haines,J.L., Scott,B.L., Pericak-Vance,M.A., Vance,J.M., Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease, JAMA, 286 (2001) 2245-2250.

[169] Spillantini,M.G., Schmidt,M.L., Lee,V.M., Trojanowski,J.Q., Jakes,R., Goedert,M., Alpha-synuclein in Lewy bodies, Nature, 388 (1997) 839-840.

[170] Jensen,P.H., Hager,H., Nielsen,M.S., Hojrup,P., Gliemann,J., Jakes,R., alpha-

synuclein binds to Tau and stimulates the protein kinase A-catalyzed tau phosphorylation of serine residues 262 and 356, J. Biol. Chem., 274 (1999) 25481-25489.

[171] Lee,V.M., Giasson,B.I., Trojanowski,J.Q., More than just two peas in a pod: common amyloidogenic properties of tau and alpha-synuclein in neurodegenerative diseases, Trends Neurosci., 27 (2004) 129-134.

[172] Polymeropoulos,M.H., Higgins,J.J., Golbe,L.I., Johnson,W.G., Ide,S.E., Di Iorio,G., Sanges,G., Stenroos,E.S., Pho,L.T., Schaffer,A.A., Lazzarini,A.M., Nussbaum,R.L., Duvoisin,R.C., Mapping of a gene for Parkinson's disease to chromosome 4q21-q23, Science, 274 (1996) 1197-1199.

[173] Zhang,J., Song,Y., Chen,H., Fan,D., The tau gene haplotype h1 confers a susceptibility to Parkinson's disease, Eur. Neurol., 53 (2005) 15-21.

[174] Hughes,A.J., Daniel,S.E., Kilford,L., Lees,A.J., Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases, J. Neurol. Neurosurg. Psychiatry, 55 (1992) 181-184.

[175] Fung,H.C., Chen,C.M., Hardy,J., Singleton,A.B., Lee-Chen,G.J., Wu,Y.R., Analysis of the PINK1 gene in a cohort of patients with sporadic early-onset parkinsonism in Taiwan, Neurosci. Lett., 394 (2006) 33-36.

[176] Clarimon,J., Xiromerisiou,G., Eerola,J., Gourbali,V., Hellstrom,O., Dardiotis,E., Peuralinna,T., Papadimitriou,A., Hadjigeorgiou,G.M., Tienari,P.J., Singleton,A.B., Lack of evidence for a genetic association between FGF20 and Parkinson's disease in Finnish and Greek patients, BMC. Neurol., 5 (2005) 11.

[177] 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., Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration, J. Med. Genet., 42 (2005) 837-846.

[178] Shi,Y.Y., He,L., SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci, Cell Res., 15 (2005) 97-98.

[179] Crawford,F., Freeman,M., Town,T., Fallin,D., Gold,M., Duara,R., Mullan,M., No genetic association between polymorphisms in the Tau gene and Alzheimer's disease in clinic or population based samples, Neurosci. Lett., 266 (1999) 193-196.

[180] Lilius,L., Froelich,F.S., Basun,H., Forsell,C., Axelman,K., Mattila,K., Andreadis,A., Viitanen,M., Winblad,B., Fratiglioni,L., Lannfelt,L., Tau gene polymorphisms and apolipoprotein E epsilon4 may interact to increase risk for Alzheimer's disease, Neurosci. Lett., 277 (1999) 29-32.

[181] Bullido,M.J., Aldudo,J., Frank,A., Coria,F., Avila,J., Valdivieso,F., A polymorphism in the tau gene associated with risk for Alzheimer's disease, Neurosci. Lett., 278 (2000) 49-52.

[182] 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., Mann,D.M., 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 (2002) 113-116.

[183] Roks,G., Dermaut,B., Heutink,P., Julliams,A., Backhovens,H., Van de,B.M., Serneels,S., Hofman,A., Van Broeckhoven,C., van Duijn,C.M., Cruts,M., Mutation screening of the tau gene in patients with early-onset Alzheimer's

disease, Neurosci. Lett., 277 (1999) 137-139.

[184] Russ,C., Powell,J.F., Zhao,J., Baker,M., Hutton,M., Crawford,F., Mullan,M., Roks,G., Cruts,M., Lovestone,S., The microtubule associated protein Tau gene and Alzheimer's disease--an association study and meta-analysis, Neurosci. Lett., 314 (2001) 92-96.

[185] Myers,A.J., Goate,A.M., The genetics of late-onset Alzheimer's disease, Curr. Opin. Neurol., 14 (2001) 433-440.

[186] McKhann,G., Drachman,D., Folstein,M., Katzman,R., Price,D., Stadlan,E.M., Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, Neurology, 34 (1984) 939-944.

[187] Folstein,M.F., Folstein,S.E., McHugh,P.R., "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician, J. Psychiatr. Res., 12 (1975) 189-198.

[188] Rosen,W.G., Terry,R.D., Fuld,P.A., Katzman,R., Peck,A., Pathological verification of ischemic score in differentiation of dementias, Ann. Neurol., 7 (1980) 486-488.

[189] Abecasis,G.R., Cookson,W.O., GOLD--graphical overview of linkage disequilibrium, Bioinformatics., 16 (2000) 182-183.

[190] Hixson,J.E., Vernier,D.T., Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI, J. Lipid Res., 31 (1990) 545-548.

[191] Lewis,J., Dickson,D.W., Lin,W.L., Chisholm,L., Corral,A., Jones,G., Yen,S.H., Sahara,N., Skipper,L., Yager,D., Eckman,C., Hardy,J., Hutton,M., McGowan,E., Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP, Science, 293 (2001) 1487-1491.

[192] Hardy,J., Duff,K., Hardy,K.G., Perez-Tur,J., Hutton,M., Genetic dissection of Alzheimer's disease and related dementias: amyloid and its relationship to tau, Nat. Neurosci., 1 (1998) 355-358.

[193] Oddo,S., Caccamo,A., Shepherd,J.D., Murphy,M.P., Golde,T.E., Kayed,R., Metherate,R., Mattson,M.P., Akbari,Y., LaFerla,F.M., Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction, Neuron, 39 (2003) 409-421.

[194] Rapoport,M., Dawson,H.N., Binder,L.I., Vitek,M.P., Ferreira,A., Tau is essential to beta -amyloid-induced neurotoxicity, Proc. Natl. Acad. Sci. U. S. A, 99 (2002) 6364-6369.

[195] Kuopio,A.M., Marttila,R.J., Helenius,H., Rinne,U.K., Changing epidemiology of Parkinson's disease in southwestern Finland, Neurology, 52 (1999) 302-308.

[196] de Rijk,M.C., Breteler,M.M., Graveland,G.A., Ott,A., Grobbee,D.E., van der Meche,F.G., Hofman,A., Prevalence of Parkinson's disease in the elderly: the Rotterdam Study, Neurology, 45 (1995) 2143-2146.

[197] Maraganore,D.M., de Andrade,M., Elbaz,A., Farrer,M.J., Ioannidis,J.P., Kruger,R., Rocca,W.A., Schneider,N.K., Lesnick,T.G., Lincoln,S.J., Hulihan,M.M., Aasly,J.O., Ashizawa,T., Chartier-Harlin,M.C., Checkoway,H., Ferrarese,C., Hadjigeorgiou,G., Hattori,N., Kawakami,H., Lambert,J.C., Lynch,T., Mellick,G.D., Papapetropoulos,S., Parsian,A., Quattrone,A., Riess,O., Tan,E.K., Van Broeckhoven,C., Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease, JAMA, 296 (2006) 661-670.

[198] The International HapMap Project, Nature, 426 (2003) 789-796.

[199] Miller,G., Shope,T., Lisco,H., Stitt,D., Lipman,M., Epstein-Barr virus:

transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes, Proc. Natl. Acad. Sci. U. S. A, 69 (1972) 383-387.

[200] Falush,D., Stephens,M., Pritchard,J.K., Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, Genetics, 164 (2003) 1567-1587.

[201] Skol,A.D., Scott,L.J., Abecasis,G.R., Boehnke,M., Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies, Nat. Genet., 38 (2006) 209-213.

[202] Payami,H., Zareparsi,S., James,D., Nutt,J., Familial aggregation of Parkinson disease: a comparative study of early-onset and late-onset disease, Arch. Neurol., 59 (2002) 848-850.

[203] de Silva,R., Hardy,J., Crook,J., Khan,N., Graham,E.A., Morris,C.M., Wood,N.W., Lees,A.J., The tau locus is not significantly associated with pathologically confirmed sporadic Parkinson's disease, Neurosci. Lett., 330 (2002) 201-203.

[204] Farrer,M., Skipper,L., Berg,M., Bisceglio,G., Hanson,M., Hardy,J., Adam,A., Gwinn-Hardy,K., Aasly,J., The tau H1 haplotype is associated with Parkinson's disease in the Norwegian population, Neurosci. Lett., 322 (2002) 83-86.

[205] Hoenicka,J., Perez,M., Perez-Tur,J., Barabash,A., Godoy,M., Vidal,L., Astarloa,R., Avila,J., Nygaard,T., de Yebenes,J.G., The tau gene A0 allele and progressive supranuclear palsy, Neurology, 53 (1999) 1219-1225.

[206] Maraganore,D.M., Hernandez,D.G., Singleton,A.B., Farrer,M.J., McDonnell,S.K., Hutton,M.L., Hardy,J.A., Rocca,W.A., Case-Control study of the extended tau gene haplotype in Parkinson's disease, Ann. Neurol., 50

(2001) 658-661.

[207] Pastor,P., Ezquerra,M., Munoz,E., Marti,M.J., Blesa,R., Tolosa,E., Oliva,R., Significant association between the tau gene A0/A0 genotype and Parkinson's disease, Ann. Neurol., 47 (2000) 242-245.

[208] Hardy,J., Gwinn-Hardy,K., Neurodegenerative disease: a different view of diagnosis, Mol. Med. Today, 5 (1999) 514-517.

[209] Katrina Gwinn-Hardy, Racial and Ethnic influences on the Expression of the Genotype in Neurodegenerative diseases. In Genotype-Proteotype-Phenotype Relationships in Neurodegenerative Diseases Springer-Verlag Berlin Heidelberg New York, New York, 2005, pp. 25-36.

[210] Fung,H.C., Scholz,S., Matarin,M., Simon-Sanchez,J., Hernandez,D., Britton,A., Gibbs,J.R., Langefeld,C., Stiegert,M.L., Schymick,J., Okun,M.S., Mandel,R.J., Fernandez,H.H., Foote,K.D., Rodriguez,R.L., Peckham,E., de Vrieze,F.W., Gwinn-Hardy,K., Hardy,J.A., Singleton,A., Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data, Lancet Neurol., 5 (2006) 911-916.

[211] Gunderson,K.L., Kuhn,K.M., Steemers,F.J., Ng,P., Murray,S.S., Shen,R., Whole-genome genotyping of haplotype tag single nucleotide polymorphisms, Pharmacogenomics., 7 (2006) 641-648.

[212] Klein,R.J., Zeiss,C., Chew,E.Y., Tsai,J.Y., Sackler,R.S., Haynes,C., Henning,A.K., SanGiovanni,J.P., Mane,S.M., Mayne,S.T., Bracken,M.B., Ferris,F.L., Ott,J., Barnstable,C., Hoh,J., Complement factor H polymorphism in age-related macular degeneration, Science., 308 (2005) 385-389.

[213] Myers,A.J., Pittman,A.M., Zhao,A.S., Rohrer,K., Kaleem,M., Marlowe,L., Lees,A., Leung,D., McKeith,I.G., Perry,R.H., Morris,C.M., Trojanowski,J.Q.,

Clark, C., Karlawish,J., Arnold,S., Forman,M.S., Van Deerlin,V., de Silva,R., Hardy, J., The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. Neurobiol Dis., 25 (2007) 561-570.

[214] Caffrey,T.M., Joachim,C., Wade-Martins,R.,Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus, Neurobiol Aging., 2007 [doi:10.1016/j.neurobiolaging.2007.05.002]

[215] Maraganore DM, de Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 2005; 77: 685–93.

[216] Farrer MJ, Haugarvoll K, Ross OA, et al. Genomewide association, Parkinson disease, and PARK10. Am J Hum Genet 2006; 78: 1084–88.

[217] Goris A, Williams-Gray CH, Foltynie T, Compston DA, Barker RA, Sawcer SJ. No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am J Hum Genet 2006; 78: 1088–90.

[218] Baker M, Graff-Radford D, Wavrant DeVrieze F, Graff-Radford N, Petersen RC, Kokmen E, Boeve B, Myllykangas L, Polvikoski T, Sulkava R, Verkoniemmi A, Tienari P, Haltia M, Hardy J, Hutton M, Perez-Tur J: No association between TAU haplotype and Alzheimer's disease in population or clinic based series or in familial disease. Neurosci Lett 2000, 285(2):147-9.

10. Appendices

10.1 Chapter 3 Genotyping assays:

dbSNP ID	Forward	Reverse	Size of Amplicon (bp)
rs758391	ACCGTGGAGACATCTGTAGT	TGGTAGGCCTGTGGTAAA	228
rs1662577	TCCAAGTGGTTTAGCCATA	ACGTATTTAGGCCTTCCTCT	108
rs70602	GCGGAAATCTAACCATCTGTGC	GAACGGCTTCTTGACCTAAGTGG	419
rs2668643	AGAACCAAAGATGGAATCCT	AAGCGAAAACCCTAAGACA	395
rs894685	AAGTCTCCCCAAACAACAG	TTGCCTGTCTGTCCATCT	119
del_In9	GGAAGACGTTCTCACTGATCTG	AGGAGTCTGGCTTCAGTCTCTC	484
rs1801353	TCCAGACTAAGTTCCGAATG	CAGCACCTTCTCATCATTG	101
rs1047833	GTTTGCTCCACCCTTAGAT	GCTGAGCCTGTGTTTCAG	266
rs393152	ATCAGGTGACTCCCAAGAA	TTAGCATCAAGGGTAGATCC	121
rs1052553	GGTGAACCTCCAAAATCAG	GGACTTGACATTCTTCAGGT	219
rs7687	ATAGTATCAGCCCTCCACAC	AACTCAACAGGGTGCAGAT	201
rs2240758	ATTACAAAAGCGCTTACAGG	TCTGAAGTGGTGGTCTCACT	195
rs199533	CATACGGAGAACGAAGTACC	AACCAATGTTGATGTGTTCA	197

Table 10.1 Sequences of PCR primer pairs used for genotying in Chapter 3

dbSNP ID	Pyrosequencing primer	Enzyme for RFLP assay
rs758391	---	***Hph I (A)***
rs1662577	---	***BsrG I (C)***
rs70602	TCTCCTGTGGTCATTTT	
rs2668643	---	***ApoI (A)***
rs894685	---	***Acc I (T)***
del_In9	---	
rs1801353	TCTGGCTGGGTTTCA	
rs1047833	GCAGCCTTCAGCTTG	
rs393152	GCTGTGGCTCTTTCC	
rs1052553	GGAGTACGGACCAC	
rs7687	CCTTGGAAATGGTTCTTT	
rs2240758	GCCAAACTTGGAATC	
rs199533	CAAGTCAAAGGGAAGAA	

Table 10.2 Genotyping assays for the SNP in Chapter 3

Genotyping assays for the SNP, either by Pyrosequencing or by RFLP. In the cases of RFLP assay, the restriction enzymes were listed above, the enzyme cuts at the (N) allele.

10.2 Genotyping assays for Association Studies

dbSNP ID	Forward	Reverse	Size of Amplicon (bp)	Enzyme for RFLP assay
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 Haplotyping tagging SNPs for the *MAPT* association study (PSP, CBD (ch.4), PD (ch.5) and AD (ch.6))
Genotyping assays for the SNP by RFLP. The restriction enzymes were listed above; the enzyme cuts at the (N) allele.

10.3 Genotyping assays for the SNP in LD study of the Taiwanese Population

dbSNP ID	Forward	Reverse	Size of Amplicon (bp)
rs962885	GGTCCTGGAGATGAACATAA	TCTGAGAAATTCCTGTGTCC	158
rs2301689	GGAAGCCAGGTAGATTCTCT	GGGAACTGGGAATTAGAAAG	238
rs3744457	CGCGCTGTCACTTTAGTT	GAGCGCTGAGAAAGAAATC	247
rs2280004	TAACGAGGCAATGGTTTTAG	GTTCCTTTGCCCTACTTCA	124
rs1467967	CAACCTCACAGGGTACTTTC	CTCCAAACCCTGATAAAACA	365
rs3785880	CTGGTTTGAAAGGGAGGT	CTGAAAGAGAAACCTAGGAATG	235
rs1001945	CCACAGCAATGAGTGACATA	ACTTTGGCAACTCCATCTC	497
rs242557	ACAGAGAAAGCCCCTGTTGG	ATGCTGGGAAGCAAAAGAAA	384
rs242562	ACTGAAGGGATAAGGAGGAC	GGGAGCTGACGTTCATTA	316
rs2303867	GAGTTCGAAGTGATGGAAGA	ATTCTGGATGCAAACTGTTC	191
rs3785882	TCAGGAAGATTGCTGGAGT	TCGCTGATTCAACAGATAGA	161
rs3785883	GAATCTGCACTCAGAGTTGTG	CTAGCACTAGCATGACACAGA	293
rs3785885	TTACCTTTGTGTGTCCATGA	TTTGAACACCCCATCTAGTC	157
rs2258689	TTCCTCTGCTAAAACCTTGA	TTTCAAAGGTGGTTTCCTTA	208
rs2471738	ATCCGGGTTAAATTAAGGTC	AATGAGGACTGGAAAGTCTG	292
rs916896	AAGCTTCCAGAGACTGTGAG	GCAGTAAGCTTCTCTGCTGT	191
rs7521	ACCTCTGTGCCACCTCTCAC	AGGTGAGGCTCTAGGCCAGT	232
rs2074432	CACTTTGCTATGGACTCACC	GAGTTGAGGCAAACAAACAC	151
rs2277613	ACTGATACTCTGGGGGACTT	TCAAACCTCCCAAAAAGTTA	154
rs876944	AGGTTCAAATTACGGTCATC	TCTTCAGGCTTGTTCTTGAC	173
rs2301732	CCTGCTTCTGCTCTAGAATG	GAGATTCAGTCGTTGCTTCT	238

Table 10.4 PCR Primer pairs of linkage equilibrium structure study in the Taiwanese population
SNPs used to determine the linkage disequilibrium structure in the Taiwanese population

dbSNP ID	Pyrosequencing primer	Enzyme for RFLP assay
rs962885	GCGGGAGAGGGTCA	
rs2301689	GGCCTCCACTTCCTCT	
rs3744457	ACAGCCGCAGCCA	
rs2280004	CTGCTATTATTATCAGCATC	
rs1467967	---	***Dra I (A)***
rs3785880	GGTCTCCCCTGGAGTA	
rs1001945	GGAAGGCAGTGGAAA	
rs242557	---	***Apa L I (A)***
rs242562	GAGACCAGCCCGACT	
rs2303867	TCTGGGCCTGCTG	
rs3785882	CAACCAGTCCTGGAAC	
rs3785883	---	***Bsa H I (G)***
rs3785885	CCCCATCTAGTCCCA	
rs2258689	GGGAAGTGACAGAAGAGA	
rs2471738	---	***Bste II (C)***
rs916896	CAGCCTCGGGGCA	
rs7521	---	***Pst I (A)***
rs2074432	TTTTCTTGGGATGGTAA	
rs2277613	AGAGCACCCATGCC	
rs876944	CAAATTACGGTCATCCC	
rs2301732	AGTTCAACCTCTATTTGCT	

Table 10.5 Genotyping assays for the SNP in LD study of the Taiwanese Population
Genotyping assays for the SNP, either by Pyrosequencing or by RFLP. In the cases of RFLP assay, the restriction enzymes were listed above, the enzyme cuts at the (N) allele.

10.4 Selected reprints of publications arising from the work in this thesis

Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, Britton A, Gibbs JR, Langefeld C, Stiegert ML, Schymick J, Okun MS, Mandel RJ, Fernandez HH, Foote KD, Rodriguez RL, Peckham E, De Vrieze FW, Gwinn-Hardy K, Hardy JA, Singleton A. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol. 2006 Nov;5(11):911-6.

Fung HC, Xiromerisiou G, Gibbs JR, Wu YR, Eerola J, Gourbali V, Hellstrom O, Chen CM, Duckworth J, Papadimitriou A, Tienari PJ, Hadjigeorgiou GM, Hardy J, Singleton AB. Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts. Neurodegener Dis. 2006;3(6):327-33.

Fung HC, Evans J, Evans W, Duckworth J, Pittman A, de Silva R, Myers A, Hardy J. The architecture of the tau haplotype block in different ethnicities. Neurosci Lett. 2005 Mar 29;377(2):81-4.

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 Oct 21;369(3):183-5.

Pittman AM, Myers AJ, Abou-Sleiman P, **Fung HC**, Kaleem M, Marlowe L, Duckworth J, Leung D, Williams D, Kilford L, Thomas N, Morris CM, Dickson D, Wood NW, Hardy J, Lees AJ, de Silva R. Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. J Med Genet. 2005 Nov;42(11):837-46.