Parkinson's Disease Susceptibility: Genetic Mapping in an Isolated Population

Lisa Marie Skipper

Institute of Neurology, University College London, UK

Department of Neurogenetics, Mayo Clinic Jacksonville, Florida, USA

Submitted in requirement for the degree of Doctor of Philosophy

August 2004

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

Abstract

Parkinson's disease (PD) is an aetiologically complex, progressive and debilitating neurodegenerative disorder that primarily affects the elderly population. It is characterised clinically by the presence of motor symptoms including resting tremor, bradykinesia and rigidity; pathologically by neuronal loss within mid brain regions and intraneuronal inclusions comprising numerous protein aggregates. Disease risk factors are both environmental and genetic. To date, at least 10 genetic loci are implicated and specific mutations have been identified in *SNCA, PRKN, UCH-L1, DJ-1* and *PINK1*. Variability within the *MAPT* gene has also been associated with disease risk.

Population isolates are a powerful tool for the dissection of genetically complex disorders, due to expected genetic homogeneity and linkage disequilibrium (LD) levels. This project has focused upon a population isolate from Trondheim, central Norway, in which most individuals with PD show no family history of disease. In the first study of the genetic factors involved in PD within the Norwegian population, we chose to investigate recessive (*PARK2, PARK6* and *PARK7*) and susceptibility (*MAPT* and *PARK10*) loci, which may manifest as sporadic disease.

Analysis of *PRKN* (*PARK2*) suggested that this locus contributes to PD in the Norwegian population and that the high frequency of the A82E mutation in the Trondheim community is due to a founder effect. In addition, a novel proline insertion mutation was identified. Detailed examination of the *MAPT* H1 haplotype associated with parkinsonism, showed that 'H1' consists of a group of related but distinct haplotypes, one of which is preferentially associated with PD. The variability most associated with disease was shown to lie at the 5' end of MAPT, encompassing

exons 1 to 4. Candidate gene analysis and novel multipoint LD mapping methods at *PARK10* identified two genes, *EPS15* and *NRD1*, which may also contribute to PD risk. Further molecular genetic analysis will contribute to the understanding of pathogenic mechanisms through the use of cellular and animal models, and ultimately the development of both palliative and preventative therapies.

Table Of Contents

Acknowledgements

Firstly, gratitude to John Hardy (formerly Mayo Clinic, Jacksonville, Florida, USA) and Nick Wood (Institute of Neurology, UCL, UK) for the opportunity to undertake a PhD in such a wonderful location and special thanks to Matt Farrer (Mayo Clinic, Jacksonville, Florida, USA), my project supervisor and mentor.

I thank the Mayo Clinic for funding and also those colleagues there who gave technical assistance and advice during this project: Matt Baker, Gina Bisceglio, Julia Crook, Richard Crook, Mary Hulihan, Mike Hutton, Jennifer Kachergus, Sarah Lincoln, Paul Lockhart, Jonathon Manning, Minnie Schreiber, Mathias Toft, Andrew West and Kristen Wilkes.

Thankyou also to the fantastic friends who tried to keep me sane during the not so good times!: Kevin, Heather, Rich, Laurie, Amy, Simone, Mandy, Julie, Rosa and Dee.

Finally, huge thanks to my family for both emotional and financial support: Mum, Stephen, Nan, Clare, Lee, Philip and Julie; I dedicate this thesis to you.

Abbreviations

PD	Parkinson's disease
SNpc	substantia nigra pars compacta
MSA	multiple system atrophy
PSP	progressive supranuclear palsy
CBD	corticobasal degeneration
DLB	dementia with Lewy bodies
FTDP	fronto-temporal dementia with parkinsonism
LB	Lewy body
AD	Alzheimer's disease
ARJP	autosomal recessive juvenile parkinsonism
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MZ	monozygotic
DZ	dizygotic
PET	positron emission tomography
UPP	ubiquitin proteosome pathway
SNP	single nucleotide polymorphism
ROS	reactive oxygen species
DA	dopamine
GDNF	glial cell line-derived neurotrophic factor
LD	linkage disequilibrium
Kb	kilobase (1,000 bases)
Mb	megabase (1,000,000 bases)

Figures

Tables

Chapter 5

1 Introduction

1.1 Parkinson's Disease

It was almost 200 years ago that James Parkinson first described the disorder that today bears his name (Parkinson 1817). Parkinson's disease (PD) is an aetiologically complex, progressive and debilitating neurodegenerative disorder that primarily affects the elderly population. Current estimates indicate there may be four million people suffering from the disease worldwide and as life-expectancy increases, so the socioeconomic burden of PD will likely escalate (Guttman *et al.* 2003b) (http://www.roche.com/home/healthcare).

1.1.1 Clinical Features

Parkinson's disease can present with a range of clinical manifestations, primarily motor disturbances. A 4-6Hz tremor, most prominent in the resting state, is often the first sign. In addition, rigidity (resistance to passive movement, typically in 'cogwheel' fashion), bradykinesia (decreased capability to move rapidly, or rapidly alternate between different movements) and postural instability (particularly loss of balance) are other manifestations that comprise the 'cardinal features' of PD. Tremor and rigidity are typically worse in one side of the body in early stages of the disease (asymmetry at onset) (Lang and Lozano 1998). The term 'parkinsonism' refers to any combination of resting tremor, rigidity, bradykinesia, postural disturbances and the freezing phenomenon (feet transistantly 'glued to the ground') which may or may not be due to PD itself (Fahn 2003).

Other motor disturbances are often present: gait disturbances including shuffling and stooped posture with reduced arm swing; hypomimia (manifest as 'mask like' facial

expression); hypophonic, monotonal speech; drooling; micrographia and bradygraphia (Dauer and Przedborski 2003).

Non-motor disturbances can also present in varying degrees. Depression may be the most common of these, occurring in as many as half of all patients and may precede motor dysfunction (McDonald *et al.* 2003). Dementia is also an important feature in the elderly; a new diagnosis is made more than six times more often in PD patients than in elderly PD-free controls and occurs in more than half of patients over 85 years (Lang and Lozano 1998). Two-thirds of patients also report sleep disorders, including difficulties in initiating and maintaining sleep, parasomnia and excessive daytime sleepiness (Garcia-Borreguero *et al.* 2003). Other features can include emotional rigidity, diminished novelty seeking, increased apprehension and subordinate behaviour; characteristics associated with a 'parkinsonian' personality (Bodis-Wollner 2003). Mortality in PD is more than twice that expected compared with age-matched controls and life-expectancy is markedly reduced (Guttman *et al.* 2003a).

The spectrum of clinical features associated with PD contend that this is a complex disorder and there is no biological marker that unequivocally confirms a diagnosis (Lang and Lozano 1998). The age of onset is also variable (ranging from the 2nd to the 8th decade) with a mean around 55 years. A clinical diagnosis usually includes the presence of bradykinesia in addition to at least one other cardinal symptom (rigidity, tremor, postural instability) with asymmetry at onset, good response to *l*-DOPA and progressive disease course (Gelb *et al.* 1999; Litvan *et al.* 2003).

Other disorders can also present with parkinsonism ('parkinson-plus' syndromes), including multiple-system atrophy (MSA), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), dementia with Lewy bodies (DLB), fronto-temporal

dementia with parkinsonism (FTDP), Hallervorden-Spatz disease and several of the spinocerebellar ataxias (SCAs) (Lang and Lozano 1998; Gwinn-Hardy 2002). 'Secondary' parkinsonism can also be induced by drugs (dopamine receptor blockers and storage depletors), infection (encephalitis), brain tumour, head trauma or exposure to certain chemicals (e.g. MPTP (Section 1.1.6), cyanide) (Fahn 2003). A definitive diagnosis of PD is usually made *post-mortem*, through pathological examination of affected brain regions (Gelb *et al.* 1999). Pathological support for a clinical diagnosis is usually observed in ~80% of cases (Dekker *et al.* 2003).

1.1.2 Pathology

Pathologically, PD is characterized by the progressive loss of a selective but heterogeneous population of neurons. Cell loss primarily occurs in the *substantia nigra pars compacta* (SNpc) and the *locus coeruleus* (LC). These regions consist of neuromelanin-containing neurons that are deeply pigmented and as such their absence is readily noticeable at neuropathological examination (Aminoff 2001). Figure 1.1.2a shows PD-associated cell loss in the SNpc.

Nigral cells synthesize dopamine, providing dopaminergic input to the caudate nucleus and putamen (collectively termed 'striatum'). The striatum is one of the nuclei composing the basal ganglia, which is an important centre for the control of movement. Operation in the basal ganglia is mediated through the direct and indirect actions of the neurotransmitters dopamine, GABA, glutamate and acetylcholine. The actions of dopamine and acetylcholine are opposing, hence the reduction in striatal dopamine (particularly to the putamen) arising from SNpc neuronal degeneration means that the striatum is relatively over stimulated by cholinergic pathways. This results in an increase of inhibition of thalamo-cortical and midbrain tegmental

neurons and is believed to account for most of the motor symptoms of PD. Symptoms typically manifest when >60% of the SNpc cells are lost in addition to a striatal dopamine reduction of ~80%. Although the dopaminergic tract is primarily affected, the noradrenergic, serotonergic and cholinergic systems are also disturbed due to cell loss in the LC, raphe nuclei and the nucleus basalis of Meynert, generally in severe or late stages of disease (Lang and Lozano 1998; Dauer and Przedborski 2003). Figure 1.1.2b depicts the sites of neurodegeneration and neurochemical pathways disturbed in PD.

Another important pathological feature in PD is the Lewy body (LB). Lewy bodies are spherical intraneuronal cytoplasmic inclusions with a dense core and clear halo, usually greater than 15μM in diameter. They are found in surviving cells of affected brain stem and basal forebrain regions and are composed of numerous proteins including α-synuclein, ubiquitin and neurofilaments. Immunostaining with anti-α-synuclein reveals an intensely immunoreactive core, surrounded by a less reactive halo, whereas anti-ubiquitin reactivity is more diffuse (Figure 1.1.2c). Lewy-like pathology is also found in neuronal processes ('Lewy neurites') (Lang and Lozano 1998).

The specific role of LBs in PD is unclear, as they are found in other neurological disorders such as DLB and Alzheimer's disease (AD), Down's syndrome and also incidentally in old age (Gwinn-Hardy 2002; Dauer and Przedborski 2003). However, a correlation between LB burden and degree of dementia has been noted (Apaydin *et al.* 2002). The inclusions may provide a mechanism to detoxify potential cytotoxic species within the cell, but although dopaminergic cell loss reflects that of PD, LBs are not generally found in autosomal recessive juvenile parkinsonism (ARJP) (Feany and Pallanck 2003).

Figure 1.1.2a *Post Mortem* Examination of the SNpc

Pigmented neurons of the SNpc (arrowed) in normal individuals and their loss in PD is readily noticeable at *post mortem* examination.
From http://www.swmed.edu/stars/images/neurodisslides

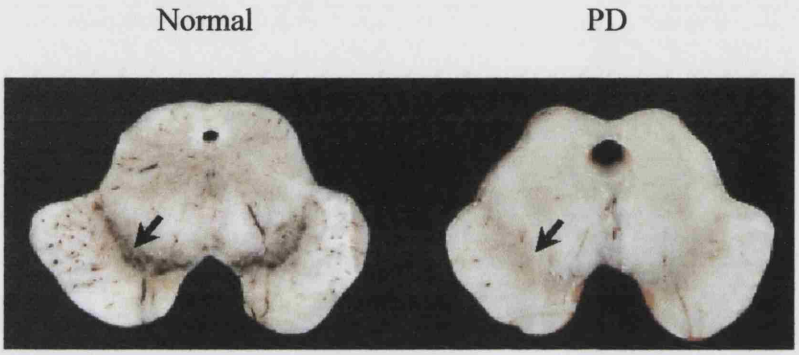


Figure 1.1.2c Lewy Bodies

Lewy bodies in dopaminergic cells of the SNpc immunostained with anti-α-synuclein (A) and anti-ubiquitin (B). From Dauer and Przedborski (2003).

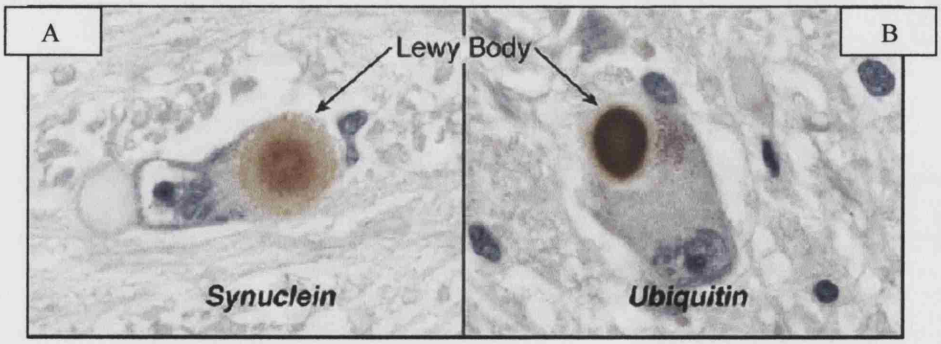
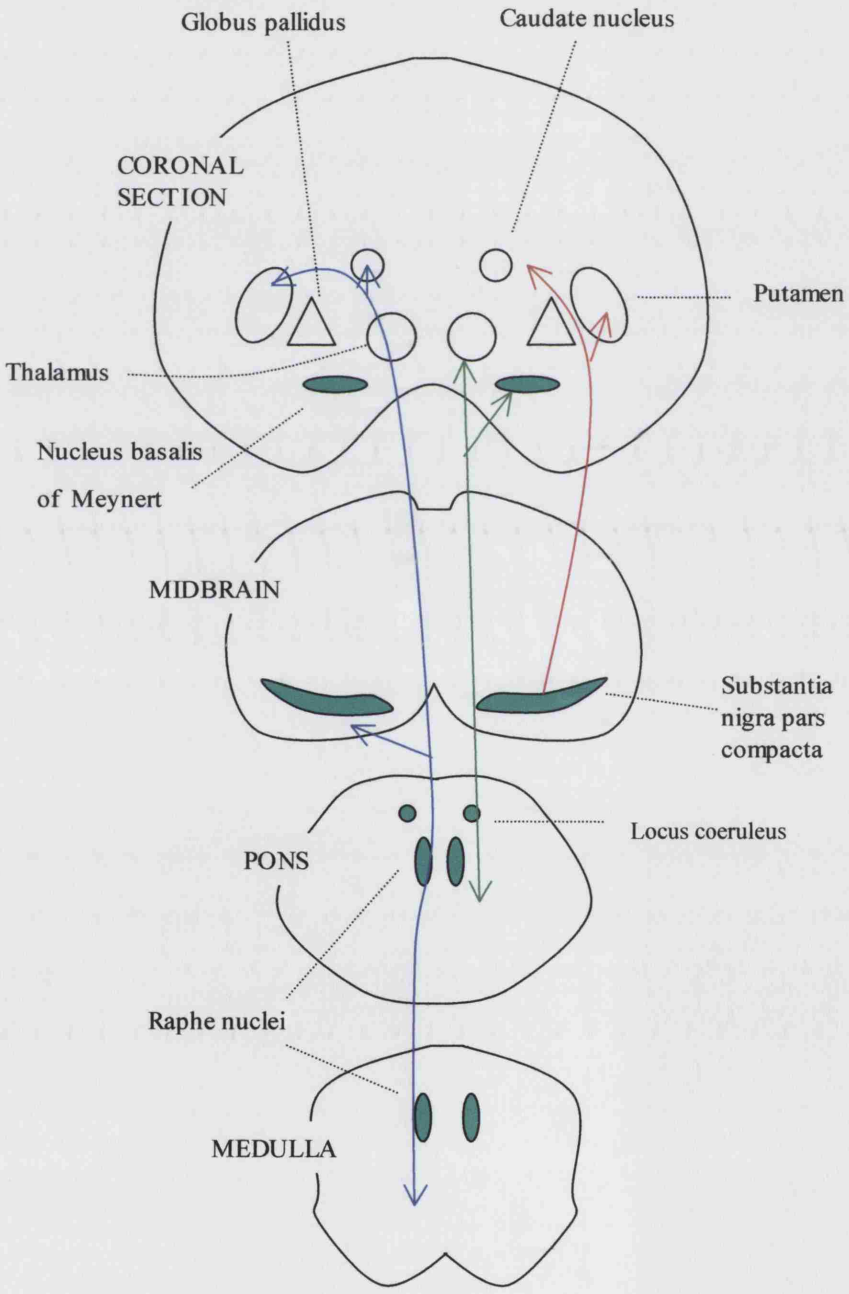


Figure 1.1.2b Sites of Neurodegeneration and Neurochemical Pathways Involved in PD

Sites of cell loss are shown in light blue. Neurochemical pathways are shown as coloured arrows (red, dopaminergic; green, noradrenergic; blue, serotonergic). Cholinergic pathways are also affected, as the nucleus basalis of Meynert is a wide-acting source of acetylcholine. Adapted from Lang and Lozano (1998).



1.1.3 Epidemiology

A number of epidemiological studies measuring the prevalence (proportion of the population who have disease at a point in time) and incidence (rate at which disease is diagnosed in a population during a specific time interval) of PD have been undertaken (Siderowf 2001). Estimations of the crude prevalence of PD in London, UK in 1997 were 128/100,000 individuals, amounting to 0.1% of the population at that time (Schrag *et al.* 2000a). Prevalence measures, however, are prone to underestimation due to disease-associated mortality, and thus incidence studies may more likely reflect the true pattern of PD in populations. Incidence estimates within different worldwide populations vary: the highest are reported for North America and Europe (11-20 and 5-26/100,000/year, respectively) and lowest for China (1.5/100,000/year) (Twelves *et al.* 2003).

Variation between different ethnic groups has also been noted. In a multi-ethnic population of California, USA, Van Den Eeden and colleagues report that incidence in Blacks is lowest (10.2/100,000/year) followed by Asians (11.3), non-Hispanic Whites (13.6) and Hispanic/Latinos (16.6) (Van Den Eeden *et al.* 2003). However, it is difficult to make meaningful geographical/ethnic comparisons as incidence studies rarely follow consistent methodologies (Twelves *et al.* 2003).

1.1.4 Risk Factors

One of the most consistent risk factors associated with PD is increasing age; incidence rapidly increases after the age of 60 years (Van Den Eeden *et al.* 2003). Incidence studies stratified for sex have also found up to a two-fold increase in PD amongst men compared to women, although these are not consistent findings (Twelves *et al.* 2003).

Family history of PD is also consistently associated with increased risk (Tanner and Aston 2000). Recent studies show patients report PD in family members almost three times more often than do controls (Kuopio *et al.* 2001). Risk was also increased almost eight-fold and three-fold in relatives of early-onset and late-onset PD, respectively (Payami *et al.* 2002). However, familial occurrence of PD does not necessitate a genetic influence but may reflect shared exposure to environmental risk factors. In addition, ~95% of PD cases are sporadic (show no family history) (Dauer and Przedborski 2003) although this could reflect recessive loss-of-function mutations.

A number of environmental risk factors have been reported. These include occupational exposure to heavy metals (e.g. iron, lead and copper), the herbicide paraquat, the pesticide rotenone and accidental exposure to a toxic compound named 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (reviewed by Di Monte 2003).

Increased risk has also been reported in association with increased dietary intake of iron and manganese, living in rural environments and the consumption of well water (Priyadarshi *et al.* 2001; Powers *et al.* 2003). Conversely, a decreased risk is associated with exposure to nicotine and caffeine (Ross and Petrovitch 2001). However, it is unclear whether these agents are protective or that 'addictive behaviours' are a phenotype of individuals less likely to develop PD for other reasons (Siderowf and Stern 2003). The consumption of a low-calorie diet is also inversely related to risk (Mattson 2003).

1.1.5 Genetic Defects

Most PD cases are sporadic and family history of disease could implicate a shared environmental aetiology (Section 1.1.4). However, twin studies can be useful in

of disease (Tanner *et al.* 1999). In such a study, Tanner and colleagues (1999) studied the concordance rates for PD in seventy-one pairs of monozygotic (MZ) and ninety pairs of dizygotic (DZ) twins. Where the first twin was diagnosed at >50 years, similar concordance rates were found (MZ=0.106; DZ=0.105). However where onset occurred before or at 50 years, complete concordance (1.0) was observed for the MZ group, albeit n=4 (DZ=0.167). This suggests that a genetic contribution to disease may be more important in earlier onset cases, although the authors concede that the cross-sectional nature of the study may have been a limitation.

Piccini and colleagues (1999) evaluated concordance amongst MZ and DZ twins, although this was longitudinal and used ^{18}Fdopa positron emission tomography (PET) to assess subclinical striatal dopamine dysfunction. Concordance based on imaging data was 55% and 18% in the MZ and DZ groups respectively (p=0.03). During a seven-year follow up period, asymptomatic MZ co-twins showed progressive loss of dopaminergic function and four developed clinical PD. Combined concordance for PD, based on dopaminergic dysfunction and clinical diagnosis, were 75% in MZ twins (n=12) and 22% in DZ twins (n=9) (p=0.02). Variability in rates of loss of dopaminergic function amongst MZ co-twins and variability in latency periods within MZ co-twins indicates that genetic causes are heterogeneous and may be modified by the environment (Piccini and Brooks 1999).

A large population based study undertaken in Iceland has also implicated the importance of genetic factors (Sveinbjornsdottir *et al.* 2000). Familial clustering of disease, beyond that of the nuclear family, with complex modes of inheritance, was demonstrated. Parkinson's disease patients were also found to be more related than unaffected individuals, on the basis of 'kinship coefficient' (defined as the probability that any particular allele in two individuals was inherited from a common ancestor) at

all ages of onset. Relative risks for siblings (6.3) and offspring (3.0) indicate that whilst genetics may play a role (in particular, recessive inheritance of susceptibility), shared environmental factors early in life may also contribute.

Debate on the relative contribution of genes and the environment in the aetiology of PD may continue. However, during the last seven years causal mutations in five genes have been identified in the pathogenesis of PD.

1.1.5.1 Dominant Genes – α-synuclein and UCH-L1

α-synuclein (SNCA, PARK1)

In 1990, an autosomal dominant form of PD was described in a large Southern Italian kindred (Golbe *et al.* 1990). Ensuing investigations led to the identification of a missense mutation in the *α-synuclein* gene (*SNCA*) on chromosome 4q22.1 (*PARK1*) (Polymeropoulos *et al.* 1997). The mutation, a G to A transition at base pair 209 causes the substitution of amino acid alanine with threonine at codon 53 (A53T). This mutation was also identified in a number of Greek PD families. Further investigation revealed a haplotype shared by affected members of both the Italian and Greek families, suggesting the mutation descended from a common founder (Athanassiadou *et al.* 1999). A second mutation, A30P, was identified in a German pedigree with early onset PD, albeit without pathological confirmation (Kruger *et al.* 1998). Screening in other sporadic and familial cases failed to reveal missense mutations (Farrer *et al.* 1998; Vaughan *et al.* 1998; Wang *et al.* 1998). A53T has been reported in 13 Greek/Italian families, whereas there have been no further reports of A30P (Mori *et al.* 2003). More recently, an E46K substitution was identified in a Spanish kindred (Zarranz *et al.* 2004). Genomic triplications of the *SNCA* locus have also been reported to cause a severe phenotype with early onset, cognitive decline and

also been reported to cause a severe phenotype with early onset, cognitive decline and dementia in two unrelated families (Singleton *et al.* 2003; Farrer *et al.* 2004). *SNCA* duplication has been reported in a further family with a phenotype more consistent with typical PD (Chartier-Harlin 2004).

Polymorphisms have been identified in the *SNCA* gene promoter (Xia *et al.* 1996) and have been associated with PD (Kruger *et al.* 1999; Tan *et al.* 2000; Farrer *et al.* 2001b; Pals 2004). Gene expression has also been correlated with promoter allelic variability (Chiba-Falek and Nussbaum 2001).

The α-synuclein protein is abundant in presynaptic terminals and is thought to exist in equilibrium as a soluble monomer and a membrane-bound oligomer. Its normal physiological function may have a role in the modulation of synaptic vesicle function and the regulation of dopamine release (reviewed by Lotharius and Brundin 2002; Orth and Tabrizi 2003).

UCH-L1 (PARK5)

Ubiquitin C-terminal hydrolase (UCH-L1) is abundantly expressed in brain, as well as being present in Lewy bodies (Lowe *et al.* 1990). Belonging to a family of deubiquitinating enzymes, it is thought to cleave polymeric ubiquitin to monomers playing a role in the the recycling of ubiquitin ligated to misfolded proteins after degredation by the ubiquitin proteasome pathway (UPP). In addition, it may hydrolyse bonds between ubiquitin and other small molecules (Larsen *et al.* 1998). A mutation in exon 4 of the *UCH-L1* gene at chromosome 4p14 (*PARK5*) was identified in an affected sib-pair of German descent, suggestive of autosomal dominant inheritence with clinical symptoms of typical PD. This mutation results in the

I93M is thought to be only a very rare cause of PD (Harhangi *et al.* 1999; Lincoln *et al.* 1999), however a polymorphism has been inversely associated with PD in a number of studies (Maraganore *et al.* 1999; Wintermeyer *et al.* 2000; Zhang *et al.* 2000a). This coding variant, resulting in the substitution of serine to threonine at codon 18 (S18Y) is also associated with a later age of onset within cases (Elbaz *et al.* 2003). *In vitro* studies have shown that 18Y has comparable hydrolase activity to the wild-type enzyme, but reduced ligase activity (Liu *et al.* 2002).

1.1.5.2 Recessive Genes – Parkin, DJ-1 and PINK1

Mutations have been identified in three genes which contribute to recessively inherited parkinsonism, these are *Parkin* (PARK2), *DJ-1* (PARK7) and *PINK1* (PARK6).

Parkin was initially implicated in a very early-onset form of parkinsonism in the Japanese (ARJP) (Kitada *et al.* 1998). However, *Parkin* mutations are now known to account for the most number of parkinsonism cases due to genetic factors. Approximately 50% of familial cases and 15% of sporadic cases with early onset (≤45 years) are accounted for by *Parkin*, in a variety of populations (Lucking *et al.* 2000; Periquet *et al.* 2003). Mutations in *DJ-1* can also cause early onset parkinsonism (Bonifati *et al.* 2003) but may be a somewhat rarer cause of disease (Ibanez *et al.* 2003). The most recent gene to be identified is *PINK1*, mutations in which have been reported in two families with early onset parkinsonism (Valente *et al.* 2004). There have been no reports, as yet, of screening in other early onset cases. *Parkin*, *DJ-1* and *PINK1* are discussed in Chapter 3.

1.1.5.3 Other Linked Loci

Linkage of autosomal dominant disease has also been reported for a further two loci, PARK3 (chr 2p13) and PARK8 (chr 12p11.2-q13.1). As well as these 'causal' loci (i.e. simple 'Mendelian'), susceptibility loci for late onset disease are thought to exist at PARK10 (Chr 1p32) and PARK11 (Chr 2q36-37). Known genes and other loci implicated in PD are summarised in Table 1.1.5a (Polymeropoulos *et al.* 1997; Gasser *et al.* 1998; Kitada *et al.* 1998; Leroy *et al.* 1998; Funayama *et al.* 2002; Hicks *et al.* 2002; Bonifati *et al.* 2003; Pankratz *et al.* 2003; Valente *et al.* 2004).

Genome scans have identified additional genomic regions that may harbour susceptibility genes for late-onset disease. These include chromosome 1q, 5q, 8p, 9q, 10q, 16q, 17q, and X (DeStefano *et al.* 2001; Scott *et al.* 2001; Hicks *et al.* 2002). Loci influencing the age of disease onset have been reported at chromosome 1p (overlapping PARK10), 2p13 (overlapping PARK3), 9q, 10q, 17p, 20 and 21 (DeStefano *et al.* 2002; Li *et al.* 2002).

Table 1.1.5a Loci Implicated in PD

Chromosome	Locus	Gene	Range of AOO in years (mean)	Phenotype	Reference
Autosomal dominant					
2p13	PARK3	Unknown	36-89 (58)	PD, LB	Gasser *et al*, 1998
4p14	PARK5	***UCH-L1***	49-51 (50)	PD, pathology unknown	Leroy *et al*, 1998
4q21.3	PARK1	***Alpha-synuclein***	20-85 (46)	PD, D, LB	Polymeropoulos *et al*, 1997
12p11.2-q13.1	PARK8	Unknown	38-68 (53)	PD, pathology unknown	Funayama *et al*, 2002
Autosomal recessive					
1p35-6	PARK6	***PINK1***	32-68	PD, pathology unknown	Valente *et al*, 2004
1p36	PARK7	***DJ-1***	27-40 (33)	PD, pathology unknown	Bonifati *et al*, 2003
6q25.2-27	PARK2	***Parkin***	6-71	PD	Kitada *et al*, 1998
Susceptibility					
1p32	PARK10	Unknown	(65)	PD, pathology unknown	Hicks *et al*, 2002
2q36-37	PARK11	Unknown	(58)	PD	Pankratz, *et al* 2003

AOO=age of onset; PD=Parkinson's disease; LB=Lewy bodies; D=dementia

1.1.5.4 Candidate Gene Analyses

Variability within a number of candidate genes has also been assessed, leading to the identification of further loci implicated in the pathogenesis of PD.

NR4A2

The nuclear receptor family member NR4A2 (also known as NURR1) is involved in the development and survival of dopaminergic cells (Bannon *et al.* 2002). Le and colleagues identified two heterozygous variations (exon 1 -291Tdel and -245T→G) in the *NR4A2* gene, in ~10% of their familial patients (n=107) with a phenotype of typical PD. The mutations result in a decrease in *NR4A2* mRNA levels and affect the transcription of the tyrosine hydrolase (TH) gene (the rate limiting factor in the synthesis of dopamine) and enhance transcription of the dopamine transporter (DAT) (Le *et al.* 2003). Mutations within *NR4A2* may be rare in PD (Wellenbrock *et al.* 2003; Zimprich *et al.* 2003), although allelic variability may be important. A one base-pair insertion in intron 6 (7048insG; NI6P) has been associated with PD both in the homozygous form (Xu *et al.* 2002) and in the heterozygous form (Zheng *et al.* 2003).

NF-M

Cytoskeletal proteins comprise a large proportion of the proteins found in LBs, including the heavy (NF-H), medium (NF-M) and light (NF-L) neurofilament subunits (Galvin *et al.* 1997). A heterozygous glycine to serine substitution was identified at amino acid 336 (G336S) of the gene encoding NF-M (*NF-M*) in a patient with an age of onset at 16 years. The substitution occurs within an evolutionary conserved region and is postulated to disrupt interactions with other neurofilament subunits. However, three siblings who were unaffected in the third and fourth decade

were also heterozygous carriers, indicating that either the mutation is not fully penetrant or is incidentally carried and not related to PD (Lavedan *et al.* 2002).

BDNF

Single nucleotide polymorphisms (SNPs) within 18 genes potentially involved in PD were analysed in an association study carried out in the Japanses population. These genes were chosen based on functions involving the metabolism, reception or transportation of dopamine, nerve growth factors and receptors and the metabolism or transportation of toxins. Remarkable results were found only for the brain derived neurotrphic factor gene (*BDNF*) in which a valine to methionine substitution at codon 66 (V66M) was found in the homozygous state more frequently in PD cases than in controls (p=0.019) (Momose *et al.* 2002).

MAPT

Mutations in the microtubule associated protein tau gene (*MAPT*) were initially implicated in FTDP (Hutton *et al.* 1998), but variation may play a role in susceptibility to a broader group of parkinsonian disorders. *MAPT* and its association with PD is discussed in Chapter 4.

1.1.6 Mechanisms of Neurodegeneration

Results of studies exploring both environmental and genetic causes of parkinsonism are beginning to shed light upon the mechanisms that result in neurodegeneration. These involve protein misfolding and impairments in the ubiquitin proteasome pathway (UPP), mitochondrial dysfunction and oxidative stress.

The normal process of cellular mitochondrial respiration generates reactive oxygen species (ROS) that can induce cellular damage by reacting with DNA, proteins and lipids, if not properly dealt with. The metabolism of dopamine (DA) generates the ROS hydrogen peroxide and superoxide radicals, and auto-oxidation of DA produces DA-quinone, which causes protein damage by reacting with cysteine residues. Therefore, the *substantia nigra* is particularly vulnerable to oxidative stress. Evidence of impairments in mitochondrial complex I activity are found in sporadic PD, which also lead to oxidative stress (Dauer and Przedborski 2003; Dawson and Dawson 2003).

The parkinsonism environmental risk factors MPTP, paraquat and rotenone (Section 1.1.4) are potent complex I inhibitors. The reactive metabolite of MPTP, MPP+, is selectively taken up by dopaminergic neurons through the dopamine transporter (DAT). The parkinsonian effects of MPTP were first noted in humans through accidental exposure to the compound (Langston *et al.* 1983). In non-human primates administration of MPTP causes tremor, rigidity, akinesia and postural instability (which are responsive to *l*-DOPA and dopamine agonists) and dopaminergic cell loss, similar to that seen in PD (Orth and Tabrizi 2003). Paraquat and rotenone administration to mice and rats, respectively, also result in dopaminergic cell loss

accompanied by α-synuclein containing inclusions (Betarbet *et al.* 2000; McCormack *et al.* 2002).

The α-synuclein protein itself may also play a part in promoting oxidative stress. Protofibrillar α-synuclein (the precursor of α-synuclein fibrils found within LBs) may have a permeabilising effect upon vesicles which is exacerbated by familial PD mutations. This could allow leakage of dopamine to the cytoplasm, where it may participate in reactions leading to oxidative stress (Kaplan *et al.* 2003). Additionally, familial PD mutations may inhibit the formation of vesicles from early endosomes, thereby reducing the availability of dopamine sequestering vesicles leading to an to accumulation of dopamine within the cytoplasm (Lotharius and Brundin 2002).

As well as the cellular energy failures resulting from oxidative stress, a defective UPP is implicated in neuronal death. In comparison to age matched controls, alpha subunits of the 26/20S proteasome are diminished and activity of 20S is impaired in the SNpc of sporadic PD patients (McNaught *et al.* 2003). Identification of familial mutations in *Parkin* (Section 3.1.1.1) and *UCH-L1* (Section 1.1.5.1), whose protein products are involved in the UPP, also support this notion. Aggregated protein products of *α-synuclein*, another gene implicated in familial PD, have also been shown to bind and inhibit function of the proteasome (Snyder *et al.* 2003).

Protein mishandling may also contribute to SNpc pathology. In a *drosophila* model, directed expression of the molecular chaperone HSP70 prevented α-synuclein induced dopaminergic neuronal loss and disruption of endogeneous chaperone function accelerated α-synuclein toxicity. The LBs in sporadic PD also immunostain for some chaperone proteins (Auluck *et al.* 2002). Loss of the putative chaperone function of

DJ-1, in the alleviation of misfolded proteins, may also contribute to proteolytic stress (Dawson and Dawson 2003).

1.1.7 Treatment

There currently are no preventative or curative treatments for PD although palliative therapies exist which may ameliorate symptoms of the disorder. *l*-DOPA (3,4-dihydroxy-L-phenyalanine, the metabolic precursor of dopamine) replacement therapy is particularly useful for easing bradykinesia and acts by restoring striatal dopamine levels (through the conversion of *l*-DOPA to dopamine in remaining neurons). Replenishment of dopamine, through the implantation of fetal dopaminergic cells has also been demonstrated to reduce motor deficits (Isacson *et al.* 2003). Although *l*-DOPA is the most effective drug to treat PD, response can begin to fluctuate ('wearing-off' effect) after ~5 years of treatment; ~60% of patients develop these complications (Fahn 2003). In addition, *l*-DOPA-induced dyskinesias, particularly chorea and dystonia, are a troublesome side-effect of this therapy and can effect up to 50% of patients on long-term *l*-DOPA therapy (Fahn 2000).

Other symptomatic therapies include dopamine agonists (which act by directly stimulating dopamine receptors) and catechol-methyl transferase (COMT) inhibitors (these indirectly increase available dopamine by reducing its conversion to a substrate competitor) (Aminoff 2001).

Surgical procedures may provide benefit. Stimulation of the sub-thalamic nucleus or globus pallidus have been associated with improvements in bradykinesis and rigidity (Guttman *et al.* 2003a). This can allow the reduction of *l*-DOPA dosage and therefore may also reduce associated dyskinesias (Fahn 2003).

Glial cell line-derived neurotrophic factor (GDNF) is a potent neurotrophic factor with restorative effects in a variety of rodent and primate models of PD. An initial trial involving infusion of GDNF to the putamen of human patients suggested this form of therapy may be potentially useful not only in restoring dopamine function but also reducing dyskinesias (Gill *et al.* 2003). The use of recombinant adeno-associated and lentiviral systems to deliver GDNF are underway in animal models of PD. These offer the prospect of longer term delivery (Kirik *et al.* 2004). Neuroprotective treatments that slow disease progression are also under investigation, and include administration of co-enzyme Q10 which has anti-oxidant properties (Muller *et al.* 2003).

1.2 Linkage Disequilibrium

Linkage disequilibrium (LD) is defined as the non-random association of alleles at adjacent loci. Loci and thus their allelic variants are expected to segregate independently in a randomly mating population. However, if a particular allele at one locus is found together on the same chromosome with a specific allele at a second locus, more often than expected by chance, then the loci are in linkage disequilibrium (Cardon and Abecasis 2003).

Linkage disequilibrium is initially created when a new allelic variant arises on the background of a particular haplotype; linkage disequilibrium between this and nearby marker loci will exist until it is broken down. Meiotic recombination is the most common destroyer of LD, through the physical shuffling of chromosomal segments, although there are many factors that can affect LD in the human genome. These include genetic drift, population growth, admixture, natural selection and gene conversion. Genetic drift refers to changes in haplotype frequencies during the production of a finite number of offspring; this will tend to increase LD in small, stable populations. Conversely, rapid population growth decreases LD through the reduction of genetic drift. Admixture, the mating of individuals from distinct populations, can increase LD if allele frequencies are sufficiently different between the populations. Natural selection can increase LD, through the 'hitchhiking' of haplotypes flanking favourable variants. The effect of gene conversion (the transfer of a short stretch of DNA from one chromosome to another during meiosis) is similar to that of two closely spaced recombination events and therefore can reduce LD (reviewed by Ardlie *et al.* 2002).

1.2.1 Measuring Linkage Disequilibrium

A simple quantification of LD is a measure known as D, which is given by P_{AB}-P_AP_B, where P_{AB} is the frequency of the observed haplotype containing alleles at loci A and B, and P_AP_B is the expected frequency of the haplotype dependent upon the relative frequencies of alleles at A and B (Weiss and Clark 2002).

D is dependent on the particular allele frequencies under observation and so is often scaled to D' (given as D/Dmax, where Dmax is the maximum value of D obtainable, depending upon allele frequencies). Values of D' can take the range –1 to 1. Values of |1| indicate 'complete LD' and occur when all copies of a particular allele are found exclusively with a particular allele at the other locus, such that three out of a possible four haplotypes are present (whether D' values are positive or negative depends on the arbitrary labelling of alleles). D' is also not without constraints, in particular small sample sizes can cause artificial inflation, especially when allele frequencies are low (Zondervan and Cardon 2004).

A third measure of LD is r^2 (sometimes denoted as Δ^2), calculated by $D^2/ P_AP_BP_aP_b$ (where P_a and P_b are the alternate allele frequencies at loci A and B). The measure ranges from 0 to 1, where 1 indicates 'perfect LD' (only two of a possible four haplotypes exist). Perfect LD between two loci means that information for one of these is redundant, but intermediate values are useful for mapping puposes. If one of the markers is in LD with a nearby unknown disease locus, the sample size must be increased by approximately $1/r^2$ in order to have the same power to detect an association with the disease locus itself (Pritchard and Przeworski 2001).

1.2.2 Variability Between Populations and Genomic Region

Early estimations of LD through simulation predicted that useful levels of LD (for mapping purposes – Section 1.3) would extend only 3 Kb in the general outbred human population (Kruglyak 1999). Empirical data from 19 randomly selected genomic regions showed that LD typically extends 60Kb in North American individuals of Northern European descent and <5 Kb in a Nigerian population (Reich *et al.* 2001). Lonjou and colleagues (2003) reported varying overall levels of LD in different populations, with Europeans and sub-Saharan Africans showing the most and least extensive LD, respectively (Lonjou *et al.* 2003).

Reich and colleagues (2001) also found intrapopulation LD variability between genomic regions. For example, $|D'|>0.5$ for at least 155 Kb around the WASL gene (chromosome 7) but for less than 6 Kb around the PCI gene (chromosome 14) (Reich *et al.* 2001). Other studies have indicated patterns of LD in the human genome exhibit a 'block-like' structure (Cardon and Abecasis 2003).

1.2.3 Haplotype Blocks and Haplotype Tagging Single Nucleotide Polymorphisms

Extensive regions showing strong LD interspersed with (usually smaller) regions in equilibrium are typically manifest as regions of limited haplotype diversity (hence the term 'haplotype block'). Block-like patterns of LD have been observed on chromosome 5q31 (Daly *et al.* 2001), the HLA locus on chromosome 6 (Jeffreys *et al.* 2001) and throughout chromosomes 21 and 22 (Patil *et al.* 2001; Dawson *et al.* 2002). A larger scale study over 51 genomic regions also suggests that haplotype blocks may be a general phenomenon in the human genome (Gabriel *et al.* 2002) and again, population variations (in terms of block-length) are apparent. The processes underlying the block structure of LD may be complex, although punctate

recombination may be a central factor (reviewed by Cardon and Abecasis 2003); this refers to the hypothetical distribution of 'recombination hot-spots' – localised chromosomal sites where recombination is recurrent.

The identification of haplotype blocks has led to the prospect of using 'haplotype tagging Single Nucleotide Polymorphisms'(htSNPs) in association studies, under the assumption that much of the variation can be identified with a smaller subset of SNPs, i.e. those that 'tag' a haplotype (Johnson *et al.* 2001). Their use has the potential to significantly decrease genotyping costs without loss of power. The HapMap project is attempting to resolve the haplotype block structure of the human genome with the view of making whole genomewide association studies (Section 1.3) feasible in all populations (Cardon and Abecasis 2003).

Haplotype blocks have been defined in terms of pairwise measures of LD (Daly *et al.* 2001; Gabriel *et al.* 2002) where a block is defined as a region where all pairwise statistics are above an arbitary level. Other studies define a block as a region where inferred haplotypic diversity is low and a small number of haplotypes account for the majority observed (Daly *et al.* 2001; Zhang *et al.* 2002). It has been proposed that it is inefficient to treat blocks as independent units when selecting htSNPs, since LD may extend over more than one block and that describing LD patterns in terms of common haplotypes is not useful for the study of rare variants. This has led to the concept of a 'metric LD map', with map distances akin to the centiMorgan (cM) scale of classical linkage maps and LD patterns expressed as linkage disequilibrium units (LDUs) at physical locations. One LDU corresponds to one 'swept radius', the average useful extent of LD, the distance in Kb at which LD has declined to ~1/0.37 of its starting value (Maniatis *et al.* 2002). Tapper and colleagues (2003) used this approach to construct a metric LD map of chromosome 22, providing a framework on

which to base optimal spacing of SNPs for genome-wide association mapping studies (Tapper *et al.* 2003). They also showed that LD patterns defined by LDUs shows a stronger association with recombination (measured by cMs) than that of D', on which the first generation LD map of chromosome 22 was constructed by Dawson and colleagues (2002).

1.3 Association Based Gene Mapping

Past successes in disease gene mapping have largely relied upon linkage analysis in mutligenerational families (pedigrees) which evaluates recombination events under a precise model of disease inheritance. Mutations that have been identifed in this way tend to be monogenic, highly penetrant, fairly uncommon in the general population and inherited consistent with Mendelian rules of segregation (Rannala 2001). In contrast, complex diseases are more frequent and likely controlled by multiple factors, each having only a modest effect on risk. These include both genetic (common, low penetrance variants) and environmental factors. In complex disorders, 'disease genes' are 'susceptibility genes' rather than true causal agents (Cardon and Abecasis 2003).

Linkage analysis has had some success in identifying genetic components of some non-Mendelian disorders, for example that of the human leucocyte antigen which plays a role in type 1 diabetes (Concannon *et al.* 1998). However, it is likely that susceptibility genes confering only modest to moderate genotypic relative risk (GRR – the ratio of disease risk associated with a susceptibility genotype compared to the risk associated with an alternative genotype) will not easily be identified through traditional linkage studies. Population-based association studies are also expected to be more powerful as they are not limited by the number of recombination events that have taken place during a few generations within families, but exploit the many recombination events in the history of a population. Most studies to date have focused upon localised regions based on biological function (candidate genes) or those highlighted by linkage analysis, although whole genome scans may soon become feasible (Zondervan and Cardon 2004).

1.3.1 Study Design

The phenomenon of LD is the key underlying principle in association based gene mapping and has led to the expectation that disease susceptibility can be identified through the assessment of neutral genetic variation. SNPs are convenient markers for use in association studies because of the relative ease of high-throughput genotyping and abundance throughout the genome (estimated at every 200-300 bp) (Salisbury *et al.* 2003). However, SNP frequencies are variable between populations (Consortium 2003).

At the population level, association studies are typically implemented in a case-control design. Genotype/allele frequencies at (usually) neutral loci are compared between unrelated individuals with disease (cases) and normal individuals (controls); a simple approach is to use a χ^2 test or Odds Ratio (Section 2.3) (Rannala 2001). Significantly higher frequencies or ORs>1 in the case group is taken as evidence that the allele/genotype is associated with increased risk of disease, conversely lower frequencies in cases or ORs<1 imply a protective effect. This may be because the locus itself affects disease risk. Alternatively and more often than not, it is in LD with and thereby acts as a marker for the disease susceptibility allele (indirect association) (Hirschhorn *et al.* 2002). The amount of LD that exists between the marker and susceptibility allele will therefore determine the power to detect an indirect association between the two (Zondervan and Cardon 2004). The pattern of LD across the human genome is likely to be highly variable so should be empirically assessed at the chromosomal location under study, before deciding upon the position and density of markers (Ardlie *et al.* 2002).

There are also other factors that can affect the power to detect an indirect association between a neutral variant and a susceptibility locus. The 'effect size' of (risk associated with) a susceptibility allele can influence the size of the sample required to detect association. It has been estimated that 500-1,000 cases and controls would be required for marker ORs>2-3; around 5,000 for ORs ~1.2-1.3 (assuming complete LD between marker and susceptibility variant). Frequency of marker and susceptibility alleles also influences the likelihood of detecting association, and in general detection power is greatest when the frequencies of both of these are equal (Zondervan and Cardon 2004).

Perhaps with the exception of a few studies, including the association between the apolipoprotein E (APOE) ε4 allele with Alzheimer's disease, case-control studies have typically shown inconsistent results (Hirschhorn *et al.* 2002). Common criticisms include small sample sizes, insufficiently stringent significance measures, biased publication of results and poorly matched control groups (Cardon and Bell 2001).

Population stratification bias can be introduced when case and control samples include individuals from genetically distinct populations; differences in allele frequencies may emerge at all loci that differentiate these groups, whether they are associated with disease or not, leading to false or 'spurious' association. Spurious associations may also arise in recently admixed populations (Pritchard and Rosenberg 1999). Methods have been developed to account for such confounding demographic factors. For example, population structure is expected to have a similar effect on all loci throughout the genome. Therefore, if a positive association is found at a candidate locus within a structured case-control sample, it is likely that there will be a high rate of significant association at other unlinked markers. Genomic control (GC)

methods use genomic information from a number of markers (~30 SNPs) to adjust the significance level for a standard χ^2 test of association (Devlin *et al.* 2001). Structured association (SA) methods use multilocus genotype data to identify subpopulations of genetically similar individuals; allowance for this is made when testing for association (Pritchard *et al.* 2000).

A number of family-based association tests have been developed to overcome the problem of population stratification – these use controls selected from the families of affected individuals. An early implementation was the transmission-disequilibrium test (TDT) (Spielman *et al.* 1993). This test classifies parental alleles into those transmitted and not transmitted to affected offspring. The untransmitted alleles serve as controls, and are therefore appropriately ethnically matched. This approach has the potential to assess parent-of-origin differences and possible genotype errors. However, some information is effectively thrown away; only heterozygous parental genotypes may be used, as the transmitted allele cannot be identified if homozygous in the parent. In addition, in its original form, the TDT requires genotypes from affected offspring and two parents, for the rough equivalent of one case and control; more time-consuming and also more expensive. Moreover, assembling family-based samples may be problematic for late-onset disease (Rannala 2001).

The use of discordant sibling pairs (pairs comprising one affected and one unaffected sibling) may be practical where parental DNA is unattainable and uses the unaffected sibling as the matched control (Spielman and Ewens 1998). Family-based approaches, however, may suffer a loss of power to detect association, in comparison to using unrelated population-based controls even when samples sizes are similar (Risch 2000).

Despite much concern over demographic factors that can potentially influence population case-control studies, there is little reported empirical evidence that could be the result of fundamental study design flaws. It has been suggested that related controls may not be required if investigators adhere to good study design principles (Cardon and Palmer 2003). Other bias can be introduced by inherent epidemiological differences between case and control groups due to cohort effects. These refer to the possibility that different birth cohorts may have different ethnic backgrounds; gene frequencies may be altered by demographic shifts in the population, related to historic migration patterns during a particular time period. Another concern is that cohorts may have different histories of exposure to environmental risk factors. Prevalent case bias can also be a concern, especially for chronic diseases of old-age. Genetic factors associated with survival may be over-represented in these cases (Edland *et al.* 2004).

1.3.2 Population Choice

Although admixture can be a potential confounder in case-control analyses at apecific loci, admixture can be advantageous for low resolution association mapping at the genome level (genome scan). When individuals from two genetically distinct populations mate, the next generation is likely to display increased amounts of LD at both linked and unlinked loci. Although the LD will decay over time, it will do so more rapidly for unlinked markers, leaving considerable local LD. This increased LD means that less markers are required for a first pass genome scan; optimal conditions for the identification of disease association will exist with minimal genetic variation (homogeneity) within the parental populations and maximum variation between them (McKeigue 1998; Terwilliger and Weiss 1998).

The reduced number of SNPs and disease mutations in isolated populations (also referred to as 'founder' populations) suggests that they have less genetic heterogeneity than the general population, making them attractive choices for association mapping studies (Shifman and Darvasi 2001). Population isolates start with a small group of founders and many experience 'bottlenecks' (an extreme reduction in population size followed by expansion of a random sample of the original population) alternated with population growth. War, famine and infectious diseases are among the factors contributing to population bottlenecks. As the population rebounds from the crash, it experiences more inbreeding and genetic drift. Immigration can counteract the isolation effects if it occurs before there is a marked increase in population size (Peltonen *et al.* 2000). Reduced heterogeneity also has the effect of increasing genotypic relative risk and hence the ability to identify disease alleles. Genetic similarities help to reduce background 'noise' which can obscure association signals (Zak *et al.* 2002). Population isolates can also be expected to display differing amounts of allelic diversity - the number of unrelated founders, timing of bottleneck and expansion events, ages and sizes of the population will all impact this variability (Laitinen 2002).

Small isolated populations may be expected to display increased levels of LD over physical distances >200kb, in comparison with large outbred populations. This is attributable to smaller generation number and therefore reduced recombination which is the most common destroyer of LD over longer distances (Shifman and Darvasi 2001). Long-range LD, such as is observed in some sub-populations of Finland (Varilo *et al.* 2003) and inhabitants of the Faroe Islands (Jorgensen *et al.* 2002) may be beneficial for low resolution whole genome scans, as the number of markers can be considerably reduced. Of course the trade off comes in the fine-scale mapping stages

- the extent of LD will determine the resolution that association mapping is able to localise the physical position of a susceptibility variant. It is in these stages of the mapping process that an outbred population, exhibiting shorter-range LD, may be expected to be more advantageous. Other advantages of outbred populations are that there may be more affected individuals, more opportunity for replication, and the susceptibility loci identified may be more applicable to the general population. Isolates, however, are more likely to have shared environmental exposures, which is desirable considering the common conception that many complex disorders are, to varying degrees, the result of an interaction between both genes and the environment (Peltonen *et al.* 2000; Laitinen 2002). Wright and colleagues suggest that genetically simplified isolates are more useful than diverse populations for mapping complex trait loci, under most disease model assumptions (Wright *et al.* 1999). However, the value of isolated populations has been contested, both with regard to the likely extent of LD (Eaves *et al.* 2000) and genetic homogeneity (Arnason *et al.* 2000) in comparison to outbred populations.

1.4 The Population of Norway

Up until ~13,000 years ago, much of the Scandinavian Peninsula in Europe was covered by icy vastness during the Last Glacial Maxim (LGM). However, climatic changes around that time exposed the region allowing for colonisation to take place. Archealogical and genetic data suggests this was undertaken by settlers from what is now central Europe (Germany/Poland and surrounding regions) (Passarino *et al.* 2002). The Kingdom of Norway, occupying the western and northern regions of the Scandinavian Peninsula, officially came into being in 900 AD, under Harald 'the Fairheaded' (http://www.worldinfozone.com/facts.php?country=Norway) (Figure 1.4a).

In 1349 the "Black Death" swept through the kingdom and killed over a third of its inhabitants. This dramatically reduced the population to 100,000-150,000 (http://www.worldinfozone.com/facts.php?country=Norway). Presently, Norway has a population of ~4.5 million (2002 estimate) and immigration to the country is low (http://www.wikipedia.org/wiki/Norway) (Gedde-Dahl 1973). Consequently, Norwegian ethnicity in the country is high; only ~2% of the population are of non-Norwegian descent (http://www.wikipedia.org/wiki/Norway). A Y chromosome haplotype that is common in Norwegians and rare within other Europeans also suggests that the population has been isolated within Europe (Passarino *et al.* 2002).

Population growth in Norway during the last two centuries has been relatively low and stable. Small populations that have remained stable in size over long periods have shown potential for mapping complex disease traits based on both experimental and empirical data. The advantages are that single founder mutations are not necessary and that extensive LD is likely to increase, not dissipate over time, around

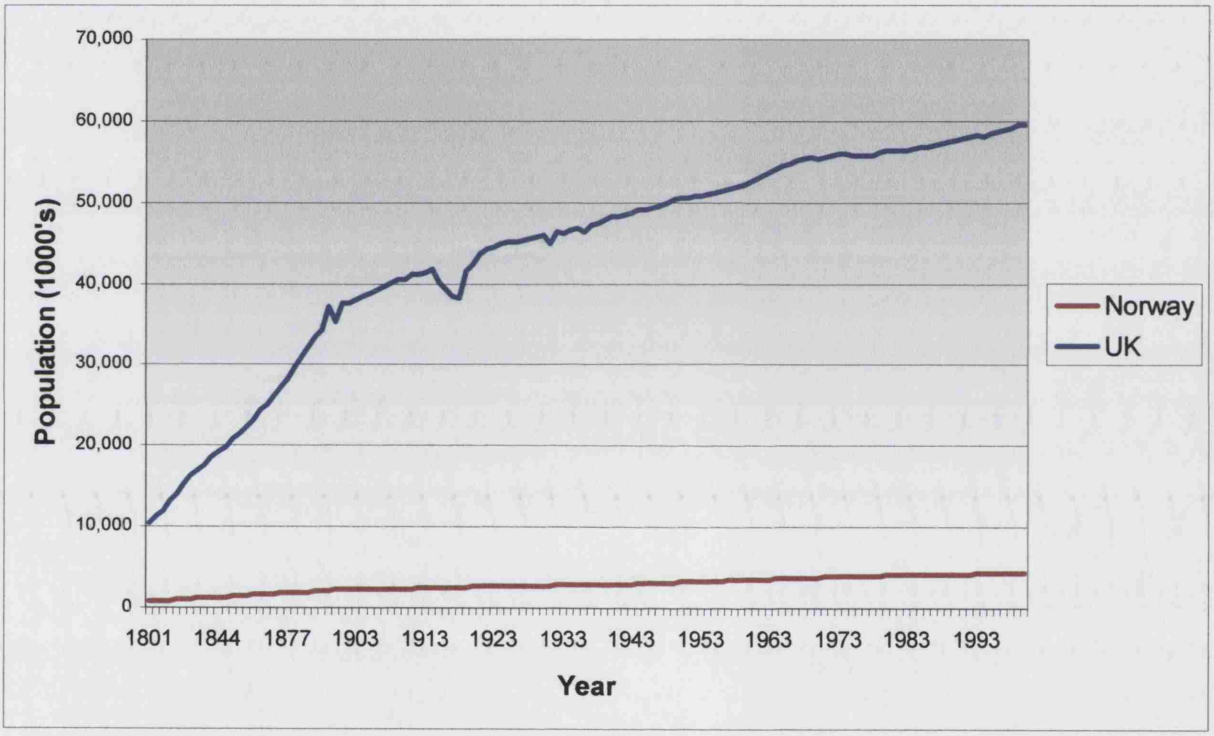
the disease locus due to the effects of genetic drift (Wright *et al.* 1999). Figure 1.4b shows Norwegian population growth, in comparison to that of the outbred population of the UK, during the last two centuries.

Figure 1.4a Map of Europe

The Kingdom of Norway occupies the western portion of the Scandinavian Peninsula (Adapted from www.abcn.com/images/map)

Figure 1.4b Population Growth in Norway and the UK During the Last Two Centuries

The Norwegian population is small and stable in comparison to that of the UK (Created using data from http://www.library.uu.nl/wesp/populstat/Europe/).



1.5 Project Aims

Although much progress has been recently made in the understanding of the processes involved in parkinsonism (Section 1.1.6), our knowledge is far from complete and current treatment strategies cannot halt or prevent the process of neurodegeneration (Section 1.1.7). Investigation into genetic causes of PD has and will likely continue the advancement of understanding (Dawson and Dawson 2003), such that the disease will no longer be the cause of a diminished quality of life to patients and family, and an economic burden to the rest of society (Schrag *et al.* 2000b; Guttman *et al.* 2003b).

Families with causal mutations account for only a small percentage of parkinsonism cases (Dekker *et al.* 2003) whereas susceptibility variants are likely to influence disease risk, prognosis and response to treatment in the larger population (Rannala 2001). The present study is fortunate in having access to a large case-control series from central Norway and aims to exploit the genetic heritage of this population isolate to investigate risk factors for parkinsonism. In the majority of Norwegian parkinsonism cases, parents were not affected so the focus will be upon recessively inherited and oligogenic (susceptibility) loci.

The first aim of this project is to examine monogenic recessive loci implicated in parkinsonism – *PARK2*, *PARK6* and *PARK7* (Chapter 3). The second aim is to investigate the *MAPT* H1 haplotype, variability in which is associated with parkinsonism susceptibility (Chapter 4). The third aim is to identify a novel parkinsonism susceptibility variant that has been linked to *PARK10* at chromosome 1p32 (Chapter 5).

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Agarose, High Melt/Medium Fragment (Mercury)

Boric Acid, Dimethyl sulfoxide [DMSO], EDTA disodium salt, NaOH, Tris base, Orange G, Glycerol (Sigma)

Hi-Di Formamide (Applied Biosystems)

2.1.2 Solutions

Conc NaOH solution (100 ml)

NaOH	200g
dH_2O	to 100 ml

0.5M EDTA pH 8.0 (500 ml)

EDTA disodium salt	93.05g
dH_2O	to 500 ml

Adjusted to pH 8.0 with conc NaOH solution

TBE Buffer 10X (1000 ml)

Tris base	108 g
Boric acid	55 g
0.5 M EDTA solution pH 8.0	20 ml
dH_2O	to 1000 ml

TBE Buffer 1X (1000 ml)

TBE Buffer 10X	100 ml
dH_2O	to 1000 ml

Agarose Gel Loading Buffer 6X (GLB) (10ml)

Orange G (2%)	1 ml
Glycerol	6 ml
dH_2O	3 ml

2.1.3 Enzymes

Restriction Enzymes for SNP genotyping (New England Biolabs or Roche)

Shrimp Alkaline Phosphatase [SAP] (Roche)

2.1.4 Size Markers

1Kb DNA Ladder (Invitrogen)

GeneScan® 400HD ROX Size Standard (PE Applied Biosystems)

GeneScan® 120LIZ™ Size Standard (PE Applied Biosystems)

2.1.5 Molecular Biology Kits

BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems)

Montage™ Seq96 Sequencing Reaction Cleanup Kit (Millipore)

Multiscreen™ PCR Cleanup Plates (Millipore)

QIAquick Gel Extraction Kit (QIAGEN)

SnaPshot™ Multiplex Kit (ABI PRISM)

Taq DNA Polymerase Kit (*Taq* DNA polymerase 1000U/ml, PCR Buffer 10X, Q Solution 5X, $MgCl_2$ 25mM) (QIAGEN)

2.1.6 DNA Samples

The samples used in this study originate from the Trondheim region of Norway. Trondheim is the ancient Norwegian capital dating back to the year 997AD. The population of the city today is around 150,000 individuals and immigration to the area is low (http://www.stud.ntnu.no/~ragnvald/trondheim/historie-eng.html).

DNA was available for 317 PD patients (cases). These were sequential new referrals to the Department of Neurology, University of Trondheim, Norway between May 1998 and January 2003. Eighty-one percent of these (n=258) resided within Trondheim or a 50 mile radius, 19% (n=61) resided in the surrounding district within a 200 mile radius. Median age of disease onset was 60 years (range 25-88 years), mean current age was 68 years (range 40-93 years) and 62% were male. Twenty-two percent had a reported family history of PD/parkinsonism.

All cases were examined using standardised protocols by a single neurologist specialised in movement disorders (Dr. Jan Aasly, Department of Neurology, University of Trondheim, Norway). Inclusion criteria were asymmetric rest tremor, rigidity and bradykinesia. Patients with three cardinal signs, together with a positive response to *l*-DOPA were grouped as probable PD, possible PD included two signs and at least moderate effect to *l*-DOPA and atypical PD showed no or minimal effects of *l*-DOPA. Follow-up visit attempts were made not to include patients with only transient effects of *l*-DOPA, early gait instability and restriction of ocular gaze who might represent PSP. One case (P038) was autopsy confirmed and thus classified as definite PD. Appropriate institutional review and informed consent was obtained. Appendix 6 details sex, age, age of disease onset, disease classification and family history of PD/parkinsonism information for all cases.

A total of 452 control DNA samples were available and came from volunteers recruited from the Department of Opthalmology, University of Trondheim or from the local blood bank. Median current age was 61 years (range 50-100 years). All control individuals resided within a 50 mile radius of Trondheim and 57% were male.

All case and control inviduals are Caucasian and have self-reported Norwegian ancestry dating back at least 4 generations. Sample collection was ongoing throughout the time course of this study, therefore analyses were performed on subsets of the complete series, depending upon availability at the time.

2.2 Laboratory Methods

2.2.1 Quantification of DNA

DNA which had previously been extracted and suspended in TE buffer was diluted 1:100 with dH20. Absorbance of this was measured at a wavelength of 260/280nm on a DU650 Spectrophotometer (Beckman Coulter). Concentration in ng/μl was calculated according to Beer-Lambert law, by the following equation:

$$(Abs260/280) * 50 * 100$$

50 = molar extinction coefficient of DNA, 100 = dilution factor

DNA samples were subsequently diluted to a concentration of 25ng/μl with dH_2O.

2.2.2 General PCR

PCR was carried out in a volume of 25µl (or doubled to 50 µl if product was to be subsequently used for sequencing reactions or DHPLC SNP genotyped), and contained:

	µl
a Forward primer (20µM)	1.0 (0.8µM)
b Reverse primer (20µM)	1.0 (0.8µM)
c dNTPs (25mM)	1.0
d 10X PCR buffer	2.5
e Template DNA (25µg)	1.0
f *Taq* DNA polymerase	0.1
g dH_20	variable
h Qiagen Q solution	*5.0*
i DMSO	*1.25*
j $MgCl_2$ (25 mM)	*1.0*
TOTAL	**25.0**

Ingredients a-g (blue) were standard to all PCR mixtures, h-j (italics) were incorporated as necessary depending on individual reaction requirements. Ingredients of the various PCR mixture are shown below:

PCR Mix 1 a-h
PCR Mix 2 a-g, i
PCR Mix 3 a-h, j
PCR Mix 4 a-g

Reactions were carried out on a Hybaid Multiblock system thermal cycler (Thermo-Hybaid) with a 'touchdown' method, whereby annealing temperature decreases through the cycle, according to the following:

	Temperature (°C)	Time (mm:ss)	No. of cycles
Initial denaturation	95	05:00	1
Denaturation	95	00:15	
Annealing	**A**	00:30	3
Extension	72	00:45	
Denaturation	95	00.15	
Annealing	**B**	00:30	18
Extension	72	00:45	
Denaturation	95	00:15	
Annealing	**C**	00:30	3
Extension	72	00:45	
Final extension	72	02:00	1

A= Starting annealing temperature

B=A-0.5, then decreasing by a further 0.5 every subsequent cycle until **C** is reached

C= Final annealing temperature

For most reactions A and B were 65°C and 55°C, or 60°C and 50°C respectively, depending upon specific melting temperatures of primers.

2.2.3 Restriction Enzyme Digest of PCR Product

Restriction enzymes digests were carried out in a total volume of 30µl, containing the following:

	µl
PCR product	25.0
10X enzyme buffer	3.0
Enzyme (2 units)	variable
dH_20	variable
10X BSA	variable
TOTAL	**30.0**

Reaction mixtures were incubated overnight at the temperature required for the specific activity of each enzyme, then stored at -20°C.

2.2.4 Agarose Gel Electrophoresis

Ethidium bromide agarose gels (agarose, 1X TBE, 0.005% ethidium bromide (v/v)) were cast at a percentage specific to the size and required separation of products. Typically a 1% (w/v) gel was used for checking PCR products, whereas the separation of restriction enzyme digest products required 2 or 3% gels. Gels were immersed in 1X TBE, then 6X GLB was added to samples before loading. A voltage of 15V/cm gel length was applied until the required resolution of products was achieved. DNA fragments were visualised using a TFM-30 transilluminator (UVP, Inc.) and digital images captured by the LAS-1000plus imaging system (Fujifilm).

2.2.5 Purification of DNA Fragments From Agarose Gel

Products to be purified were electrophoresed until resolved then excised with clean blades. QIAquick Gel Extraction Kits were used to extract DNA from the agarose

gel. Products were typically eluted into 30 µl of elution buffer, quantified and diluted as Section 2.2.1.

2.2.6 Single Nucleotide Polymorphism Genotyping

2.2.6.1 Restriction Enzyme Digest

Some SNPs create or destroy a restriction enzyme recognition site and so can be genotyped by digesting a suitably sized PCR product with an appropriate enzyme. Primers were designed to amplify the region around the SNP by PCR (typically between 300 and 500 bp in length). PCR products ideally include a second site for the enzyme, which is not polymorphic, as an internal check that the enzyme is functioning as expected, although this was not possible for all SNPs typed in this manner. The PCR products were then digested and resulting fragments agarose gel electrophoresed. An example is shown in Figure 2.2.6.1a. The SNP variants are G and T. The G variant creates a recognition/digest site for the *Dde I* enzyme. The PCR product also contains another *Dde I* site, hence the 128 bp band is the internal control; other combinations of bands are seen according to genotypes of either GG, GT or TT.

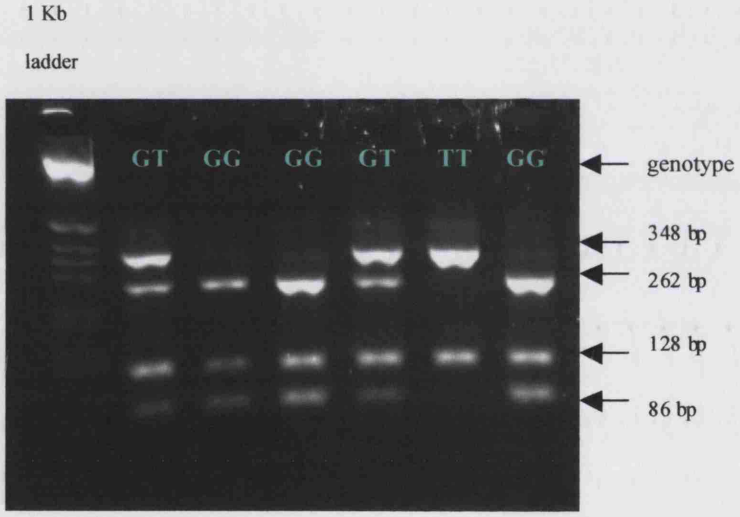
Figure 2.2.6.1a SNP Genotyping by Restriction Enzyme Digest

A) SNP assay design, with resulting fragment (band) sizes according to possible genotypes.
B) Individuals are genotyped by agarose gel electrophoresis of digest fragments.

A

SNP variants	Enzyme	Recognition Sequence (SNP)	Band sizes (bp)		
			GG	GT	TT
G/T	Dde I	C^TNAG	262, 128, 86	348, 262, 128, 86	348, 128

B



2.2.6.2 Primer Extension

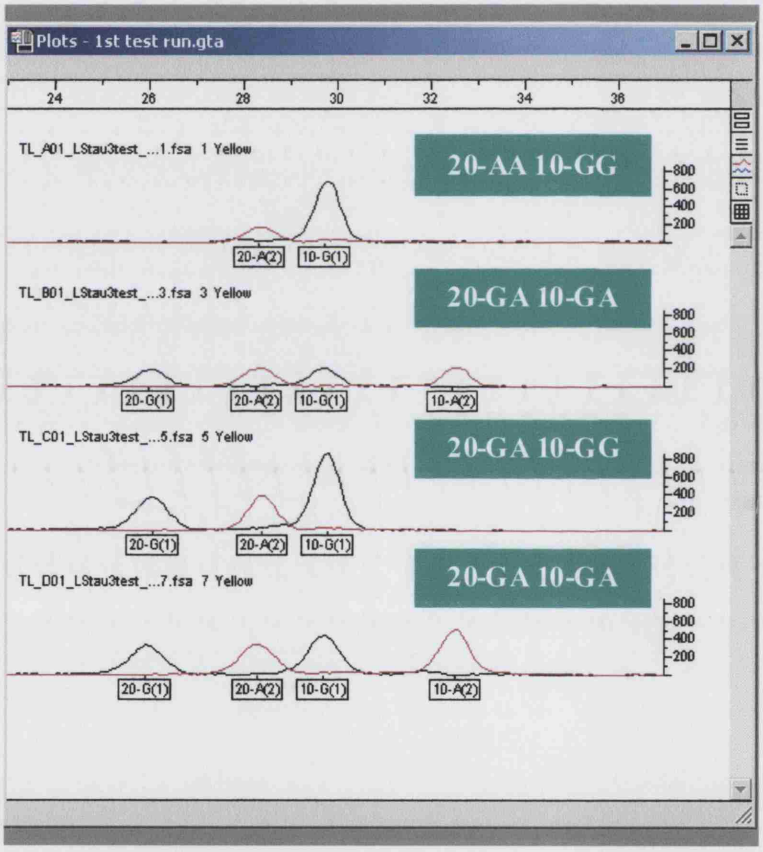
SNP genotyping using a primer extension method involves the single base pair extension of an oligonucleotide primer. PCR amplification of the region surrounding the SNP is firstly performed. Secondly, an oligonucleotide (sequencing primer, typlically 15-35 bp in length) is designed to anneal to the sequence one base 5' of the SNP. A sequencing reaction is then carried out to extend one base pair 3' from this primer. As flourescently labelled deoxyribonucleoside triphosphates (adenosine, cytidine, guanosine, thymidine; dNTPs) are used, they can be detected and will determine the SNP variant, i.e. if the extension nucleoside is G, then the SNP variant is C. Multiple SNPs can be typed in one reaction, provided the sequencing primers differ in size (sequence which will not anneal to human DNA is added to the 5' end of the sequencing oligonucleotide to create size differences).

SNP genotyping was carried out follows: PCR amplification of region surrounding SNP (200-300 bp in length) (Section 2.2.2). 5µl of PCR product was cleaned with Multiscreen™ PCR Cleanup Plates and eluted into 80µl of dH_2O. If multiple SNPs were to be genotyped, 5µl of PCR product for each SNP was pooled before cleaning. 1µl of cleaned product was added to 3µl of SNaPshot™ Multiplex Ready Reaction Mix and 1µl of sequencing primer (0.2µM). For multiple typings, primers were premixed to give a final concentration of 0.2µM. Thermal cycling was then carried out on a Hybaid Multiblock system thermal cycler, with the following program: 96°C for 10 seconds, 50°C for 5 seconds, 60°C for 30 seconds, over 25 cycles. 1µl of shrimp alkalaine phosphatase (SAP) was added (to remove unincorporated nucleotides) and the mixture incubated at 37°C for 1 hour. SNaPshot products were then electrophoresed on an ABI Prism® 3100 Genetic Analyzer (Applied Biosystems)

and analyzed using Genotyper Ver 3.7 (Applied Biosystems). 12.5µl of GeneScan® 120LIZ™ Size Standard was mixed with 1ml of Hi-Di Formamide. 9µl of this was mixed with 1µl of SNaPshot product, denatured at 96°C for 3 minutes and electrophoresed. An example of Genotyper analysis is shown in Figure 2.2.6.2a. Two markers have been genotyped in this example ('20' and '10'). Each lane is an individual sample, and based on the known length of the sequencing primers for each SNP it is genotyped according to the peak colour. The dye labels assigned to each dNTP (and the colour of analyzed data) are as follows: adenine (A)=dR6G (green), cytosine (C)=dTAMRA™ (black), guanine (G)=dR110 (blue), thymine (T)=dROX™ (red). As the flourescent nucleotides are complementary to the SNP being typed, a green peak indicates the presence of the nucleotide T, black peak for G, blue peak for T and red peak for A.

Figure 2.2.6.2a SNP Genotyping by Primer Extension Using SNaPshot™

SNP genotype assays can be multiplexed with SNaPshot™. In this example, two SNPs, '20' and '10' have been genotyped. Each lane is an individual sample. The presence of only one peak for a particular SNP indicates homozygosity. Genotytpes for the four samples are shown in blue boxes.



2.2.6.3 Denaturing High Performance Liquid Chromatography

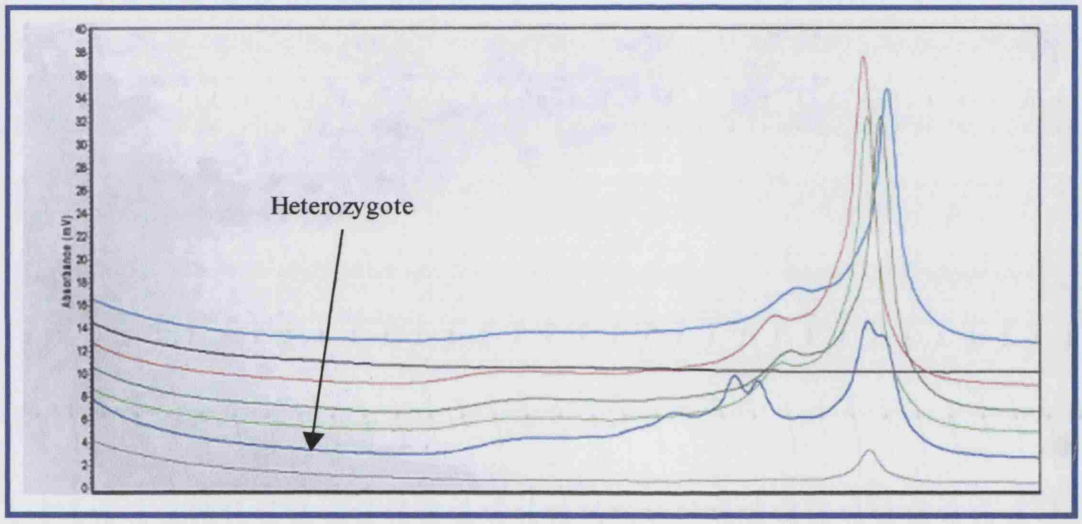
Denaturing High Performance Liquid Chromatography (DHPLC) can detect sequence variation through a change in the melting profiles of heterogeous strands of DNA. PCR amplification of the region surrounding the SNP is performed, then the product is denatured and allowed to reanneal, forming heteroduplexes if the original template contains sequence variation (i.e. a heterozygous SNP). The heteroduplexes are passed through a column that has a linear gradient of DNA binding efficiency. Binding affinity is related to the nucleic acid content of each fragment, hence differences in sequence determine at which point a fragment will separate from the column and variants can be identified (http://www.cagt.uga.edu/dhplc.html).

SNP genotyping in this study was carried out using a WAVE DHPLC system (Transgenomics). Primers were designed to amplify a region surrounding the SNP, with a melting temperature between 50 and 60 °C. 10μl of product from a 50μl PCR reaction was used for analysis. This was denatured at 96°C for 3 minutes and allowed to reanneal at room temperature for 30 minutes before loading. Chromatograms from initial output were used to identify the peak profile of the homozygous wild-type (major allele) variants (assuming that these would account for the majority of variants) and were genotyped on this basis. These usually show one peak as the duplexes formed by reannealing are identical. Heterozygotes typically show two distinct peaks and were also genotyped on this basis (although four heteroduplexes are actually formed). In order to identify mutant (minor allele) homozygotes (which after the first round of genotyping appear as homozygous wild-type), 5μl of a wild-type homozygous (the status of which was confirmed through sequence analysis) sample was added to the samples already processed, denatured and reannealed as before, then

then loaded. Any samples which then appeared heterozygous were genotyped as mutant homozygotes, as the addition of DNA carrying the alternative allele induces heteroduplex formation. An example of chromotogram output is shown in Figure 2.2.6.3a.

Figure 2.2.6.3a SNP Genotyping Using the WAVE DHPLC System

Each chromatogram is an individual sample. The double-peaked profile of the heterozygote is distinguishable from that of homozygotes.



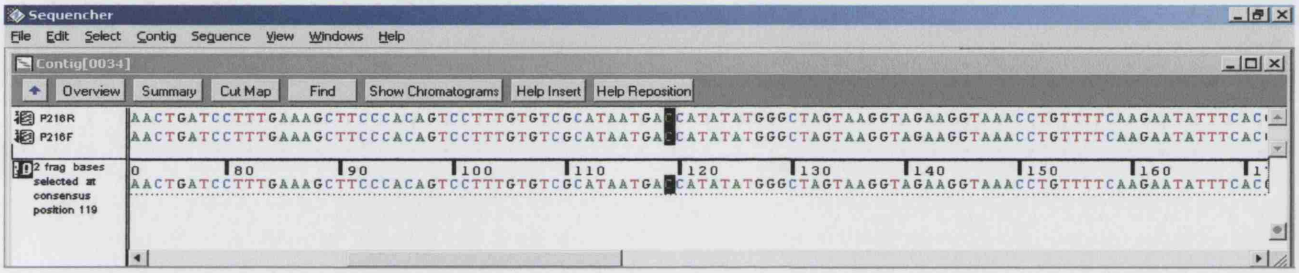
2.2.7 DNA Sequencing

A 50µl PCR amplification of the region to be sequenced was performed. This was cleaned using Multiscreen™ PCR Cleanup Plates and eluted into 40µl of dH_2O. 3.5µl of this was agarose gel electrophoresed to check the amount of the product (as shown by band intensity). Between 3 and 16µl of product were added to 3µl of BigDye® Terminator v3.1 Cycle Sequencing reaction mix and 1µl (3.2 pM) of forward or reverse primer (as for original PCR) in a total of 20µl (with dH_2O as necessary). This was thermal cycled on a Hybaid Multiblock system with the following program: 96°C, 10 seconds; 50°C, 5 seconds; 60°C, 4 minutes for 25 cycles. The total reaction volume was cleaned using Montage™ Seq96 Sequencing Reaction Cleanup Kits and eluted into 20µl. 5µl of this was added to 10µl of Hi-Di Formamide, denatured at 96°C for 3 minutes and electrophoresed on an ABI Prism® 3100 Genetic Analyzer. Resulting chromatograms were visualised using Sequencher™ v 4.1.4 (Gene Codes Corporation). An example is shown in Figure 2.2.7a. Sequencing reactions were carried out in both directions for accuracy. The forward original PCR primer was used to sequence the sense template strand and the reverse original primer to sequence the antisense strand.

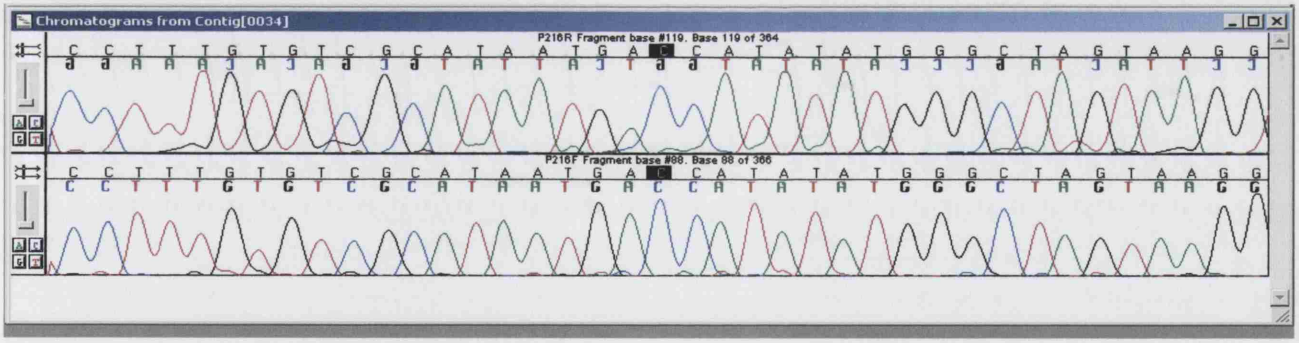
Figure 2.2.7a DNA Sequencing

Example of sequence output, visualised using Sequencher™ v 4.1.4. Panel A shows sequence from both sense and antisense template strands and a consensus, based on chromatograms shown in panel B.

A



B



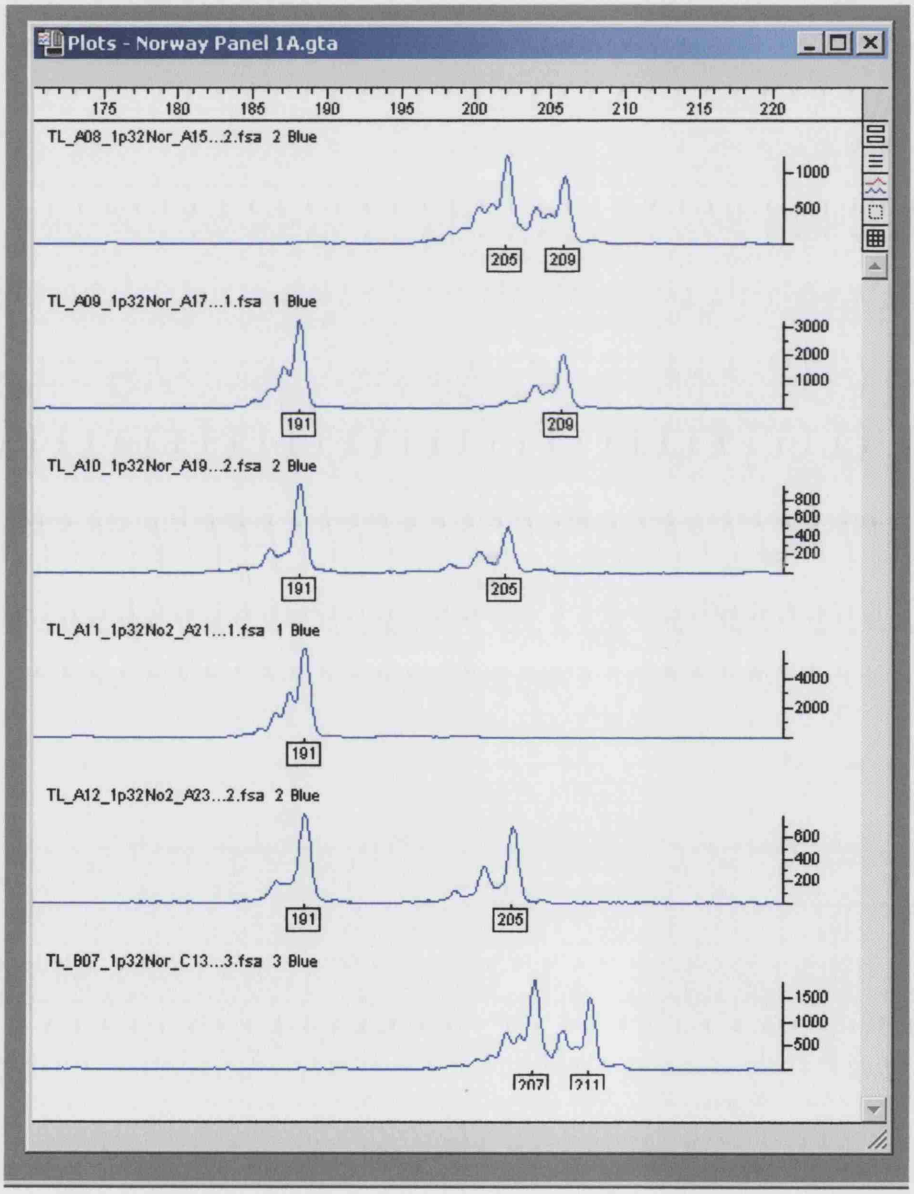
2.2.8 Short Tandem Repeat Polymorphism Genotyping

Genotyping STRPs is carried out by sizing the DNA fragments surrounding the polymorphic site, which vary in the number of repeats they contain. A PCR reaction is typically carried out with the reverse primer fluorescently labelled to allow visualisation of the product.

For genotyping in this study, a 25µl PCR amplification of the polymorphic region was performed in samples and CEPH individuals 1331-01 and 1331-02 (Section 2.2.2), with the reverse primer labelled with [5'-HEX] or [5'-FAM]. 2µl of this was added to 9µl of a size standard/formamide mixture (35µl GeneScan® 400HD [ROX] Size Standard with 1ml Hi-Di Formamide), denatured at 96°C for 3 minutes and electrophoresed on an ABI Prism® 3100 Genetic Analyzer. Fragment analysis was carried out using Genotyper Ver 3.7. Published genotypes of the CEPH samples (http://www.cephb.fr/cephdb/) allowed accurate sizing of repeat numbers and genotypes to be characterised. Reactions were multiplexed into 'panels' such that more than one STRP could be genotyped in a single reaction, by using pooled PCR products with different sizes/dye labels. An example of the chromatograms used for genotyping is shown in Figure 2.2.8a.

Figure 2.2.8a STRP Genotyping by Fluorescently Labelled Primer PCR Fragment Analysis

Each lane contains an individual sample. The peaks identify a polymorphic (dinucleotide repeat) fragment (allele) which is sized as shown in boxes under each peak. Homozygosity was assumed where only one peak was present.



2.2.9 Semi-Quantitative PCR for Exon Dosage Analysis

Exon dosage (to identify heterozygous deletions of exons) can be assessed through semi-quantitative PCR using flourescently labelled primers (as used for STRP genotyping). Products amplified from a single exon will be typically half the quantity of those amplified from a normal diploid template. Exon dosage analyses in this study was carried out as follows: Primers were designed to amplify the exon of interest, the forward primer labelled with [5'-HEX]. A 25µl PCR was carried out, with the inclusion of a control set of primers to amplify a different genomic region (of sufficiently different size to that of the exon product to allow resolution). The number of cycles for thermal cycling was dependent upon the particular reaction and had been empirically determined using known heterozygous deletion carriers, ensuring that the PCR did not not exceed the log-linear range, allowing product dosage to be resolved (S. Lincoln, personal communication).

2-4 µl of PCR product was mixed with 9µl of a size standard/formamide mixture, denatured at 96°C for 3 minutes, then electrophoresed and analyzed (as in Section 2.2.8). Assays were designed in a multiplex fashion to allow dosage of more than one exon to be quantified in a single reaction.

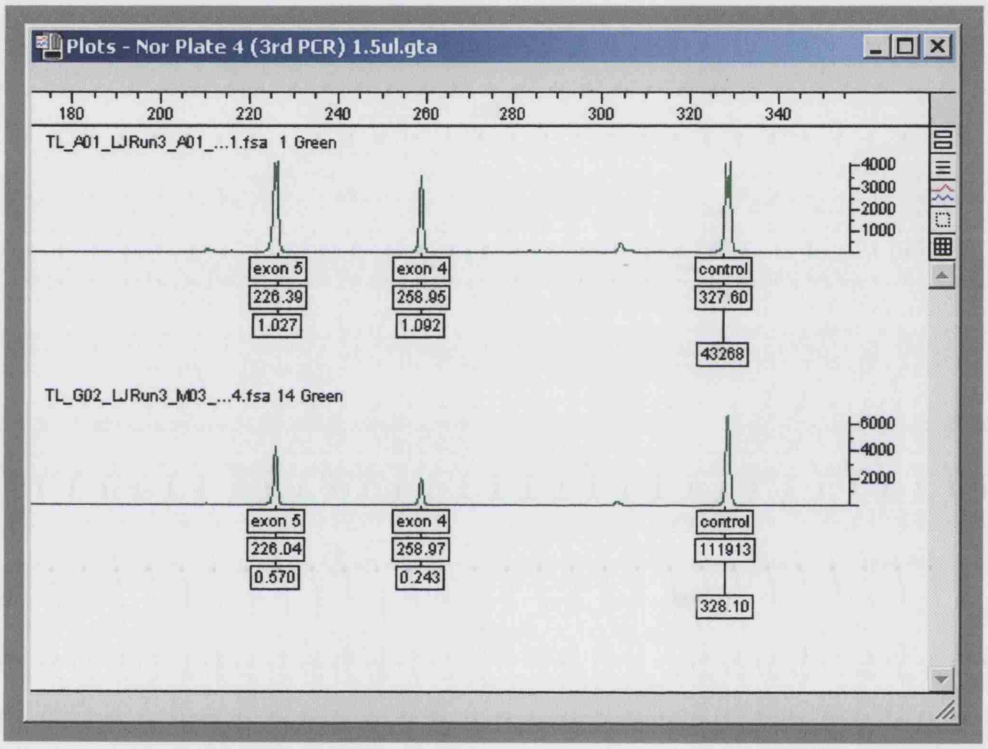
An example of the chromatograms used for genotyping is shown in Figure 2.2.9a. The area of the control product peak is used to calculate a ratio value for each exon product (shown in the bottom box under exon 4 and 5 peaks). This value is a measure of how much product has been amplified for that exon, in comparison to the control product, the template of which is assumed to be normal (diploid). The top lane of panel A shows normal diploid dosage of both exons as the ratios are ~1. Ratios <=0.5 indicate lowered dosage and heterozygous deletions of both exons are assumed

for the sample in the bottom lane. However, ratios are not always close to 0.5 for heterozygous deletions, which appears to be an artefact of the particular PCR. Therefore, wherever normal diploid dosage was suspected, PCRs were repeated and analysed three times. This method also detects homozygous exon deletions (as shown in panel B of Figure 2.2.9a) and exon multiplications.

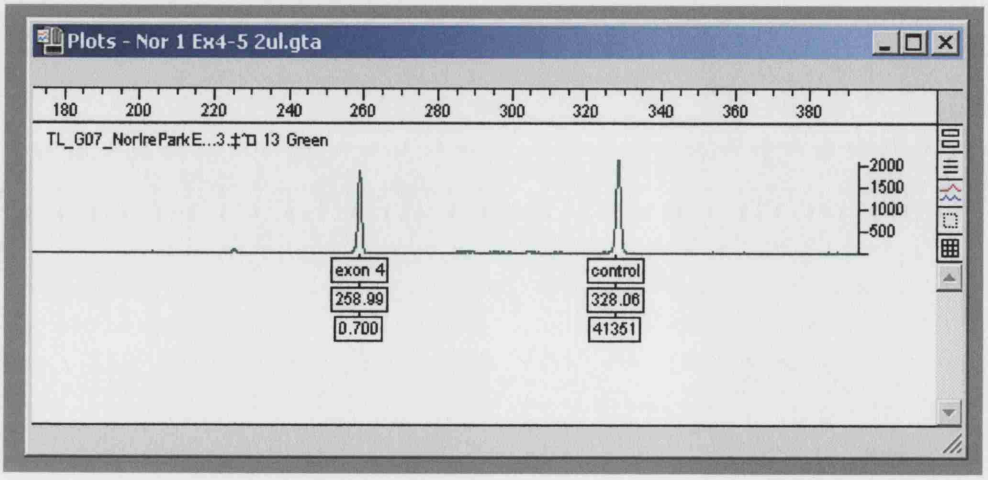
Figure 2.2.9a Semi-Quantitative PCR for Exon Dosage Analysis

The size and area is shown under the control peak. The sizes of exon 4 and 5 products are shown above the ratio of peak area compared with that of the control.
A) Top lane shows normal diploid dosage fror exons 4 and 5. Bottom lane shows heterozygous deletions of both exons (ratios are <= half normal).
B) Lack of exon 5 product indicates a homozygous deletion of exon 5.

A



B



2.3 Computational and Statistical Methods

2.3.1 Genome Sequence Analysis

Human genome sequence assemblies used for determining the gene content of genomic regions, gene and protein sequences, SNPs and STRPs were:

National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) Human Genome Build 33 (December 2003).

University of California at Santa Cruz (UCSC) (http://genome.ucsc.edu) July 2003 Assembly.

Celera Discovery System ™ (CDS) (http://celeradiscoverysystem.com) Human Genome Assembly as at Dec 2003.

These sequence assemblies vary in the exact base pair position of loci. All reference positions and sequence data for loci analysed in this study were based on those reported by CDS, unless otherwise stated. SNP identification numbers preceded by 'rs' and 'hCV' are deposited on NCBI and CDS databases, respectively.

2.3.2 Identification of Human/Mouse Genetic Sequence Conservation

Comparisons of human and mouse sequence data, and identification of conserved regions between the two were made using the online mVISTA program (http://www-gsd.lbl.gov/vista/). All program parameters were default, except 'window size' was changed to 50 bp, to give a higher resolution of sequence comparison.

2.3.3 Primer Design

Primers for PCR, sequencing and SNaPshot reactions were designed using Gene Runner Version 3.05 (Hastings Software, Inc.). This program was also used to calculate theoretical melting temperatures of PCR products.

2.3.4 Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium (HWE) refers to the balance of allele and genotype frequencies in a randomly mating population (not subject to other pressures which might alter genotype frequencies). The expected genotype frequencies for a diallelic locus with allele frequencies of p (major) and q (minor) is given by p^2 (homozygote major allele), 2pq (heterozygote) and q^2 (homozygote minor allele) in a population that does not significantly differ from a population in HWE. Departures from HWE are typically caused by non-random breeding (inbreeding, assortative mating) but often reflect genotyping errors and sampling bias, as in most normal populations HWE is normally maintained (Mueller and Young 1996).

Tests of departure from HWE were performed for all loci analysed in this study, primarily to identify possible genotyping inaccuracies. For diallelic loci, this was carried out by a chi-squared test (2 df), comparing expected genotype frequencies with those observed.

For multi-allelic loci (STRPs), a modified version of Guo and Thompson's exact test of Hardy-Weinberg proportion for multiple alleles was used, as implemented in the Arlequin Population Genetics and Data Analysis software, version 2.0 (http://anthro.unige.ch/arlequin).

2.3.5 Haplotype Estimation

Bilocus and multilocus haplotypes and frequencies were estimated from gametic phase unknown genotype data using the Arlequin Population Genetics and Data Analysis software version 2.0 (http://anthro.unige.ch/arlequin). This program uses an Expectation Maximization (EM) algorithm to estimate haplotypes. The algorithm starts by randomly assigning haplotype frequencies. It uses these to estimate genotype frequencies assuming HWE (the Estimation (E) step). Genotype frequencies are then used as weights for their relative haplotype frequencies, by counting alleles, leading to new estimates of haplotype frequencies (the Maximization (M) step). The E and M steps are repeated until the haplotype frequencies do not change more than a predefined value (reach equilibrium) which was 1.0^{e-7} for all analyses in this study.

2.3.6 htSNP Identification

Multilocus haplotypes and frequencies were estimated as in Section 2.3.5. These were input to the SNPtagger program, a web tool for choosing minimal sets of non-redundant markers to capture information in haplotypes (Ke and Cardon 2003)(http://www.well.ox.ac.uk/~xiayi/haplotype/) (Section 1.2.3). This program allows the user to specify how much haplotypic variation is to be captured by htSNPs. Initially, 100% of variation was input for this parameter. If the resulting htSNP set was too large (defeating the purpose of using htSNPs), frequencies of 95% and 90% were used.

2.3.7 Cladistics

Cladistic analysis, traditionally applied to the study of evolution and taxonomy, has been suggested as a tool for association studies in complex disease. These analyses typically link haplotype data to form 'cladograms' (trees) based upon evolutionary relationships. The assumption is that susceptibility variants arising at some point in the evolutionary history of a haplotype will be fixed within the framework of the cladogram. This would then facilitate identification of haplotypes associated with disease (Heng and Low 2000).

A 'maximum parsimony' tree is a cladogram which 'best fits' the haplotypic data. Maximum parsimony trees for haplotypes were generated using the PENNY and DRAWTREE algorithms of the PHYLIP suite of programs for inferring phylogenies (evolutionary trees) (http://evolution.genetics.washington.edu/phylip/general.html).

The search strategy used by PENNY is to initially make a tree consisting of the first three haplotypes for an unrooted tree. An unrooted tree was assumed as frequencies of some SNPs were close to 0.5, so identification of the ancestral state was not possible. The algorithm then tries to add the next haplotype in all possible places. For each of the resulting trees it evaluates the number of mutational steps between haplotypes. A maximum parsimony tree is created assuming the minimum number of mutational steps between haplotypes. Output from PENNY was used to produce graphical output using the DRAWTREE algorithm.

2.3.8 Linkage Disequilibrium Statistics

Pairwise measures of LD (D, D' and r^2) were calculated using the Excel macro Genotype Transposer (Cox and Canzian 2001). Graphical overview of LD (GOLD) was generated using the GOLD program (Abecasis and Cookson 2000).

2.3.9 Metric LDU Maps

Metric LDU maps were constructed using the LDMAP program running under the UNIX operating system (Maniatis *et al.* 2002).

(http://cedar.genetics.soton.ac.uk/public_html).

2.3.10 χ^2 Analysis using CLUMP

CLUMP is a program designed to assess the significance of the departure of observed values in a contingency table from the expected values conditional on the marginal totals and was designed specifically for use in genetic case-control association studies (Sham and Curtis 1995). The significance is assessed using a Monte Carlo approach, by performing repeated simulations to generate 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 real table is achieved by the randomly simulated data. This means that the significance levels assigned should be unbiased and that no special account needs to be taken of small expected cell counts and multiple testing. All p values reported in this study are the result of 10,000 simulations.

2.3.11 Odds Ratios

The Odds Ratio (OR) is a statistic used to assess the risk of a particular outcome with the presence of a certain factor. In this study, ORs are used to assess the risk of disease associated with allelic variants (SNPs). The OR is given by $(ad)/(bc)$ where a is the number of normal controls without the variant, b is the number of cases without the variant, c is the number of controls with the variant and d is the number of cases with the variant. 95 % confidence intervals (CIs) were also calculated, such that if the interval is above and does not include an OR of 1.0, a significant risk of disease is

associated with the variant. ORs and their 95% CIs were calculated using SPSS for Windows, release 10.0.0 (SPSS, Inc.).

2.3.12 Multipoint Linkage Disequilibrium Mapping

Multipoint disequilibrium mapping was performed using the COLDMAP program (Morris *et al.* 2003), run on a 933 MHz dual Pentium III processor (2Gb RAM) under the Linux 7.2 operating system. Typical run times analysing 49 markers in 369 samples with 20,000 iterations were ~30 days.

3 Monogenic Recessive Parkinsonism

3.1 Introduction

3.1.1 Parkin

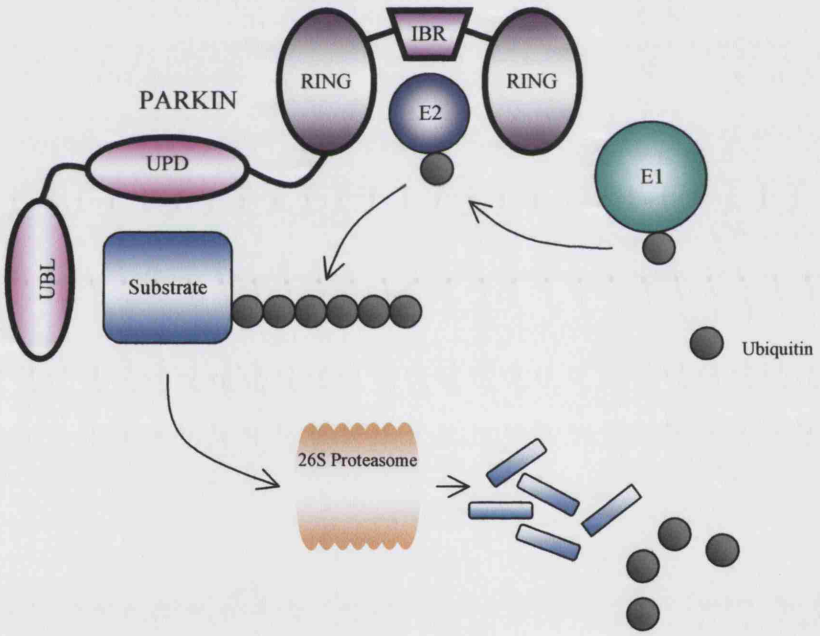
Parkin is a 52kDa protein found in a variety of different cell types, including heart, testes, skeletal muscle, kidney and brain. Within the brain, parkin is abundant in the neocortex, hippocampus and *substantia nigra*. It is highly conserved across species not only amongst veterbrates but invertebrates such as *C. elegans* and *D. melanogaster*. At the N terminus, Parkin is 62% homologous to the protein ubiquitin (the UBL domain), the C-terminal half contains 2 RING finger motifs flanking a cysteine-rich domain, termed the IBR domain. The region between the UBL and RING fingers appears unique to parkin and is termed the UPD (reviewed by Tanaka *et al.* 2001).

Parkin plays an important role in the Ubiquitin Proteasome Pathway (UPP). The UPP is one of the principle mechanisms through which damaged, misfolded or short-lived regulatory proteins are recognised and destroyed in the cell. The pathway employs an enzymatic cascade which results in the covalent attachment of 4 or more ubiquitin (Ub) molecules which acts to direct the substrate to the 26S proteosome complex for degradation. The process begins with the activation of Ub by an E1 (Ub activating) enzyme, the activated Ub is then transferred to an E2 (Ub conjugating) enzyme. In some cases, the E2 enzyme directly transfers Ub to the target protein, but the process often requires an E3 (Ub ligating) enzyme (reviewed by Miller and Wilson 2003).

Parkin is an E3 ubiquitin ligase and works in conjuction with the E2 enzymes UbcH7 and UbcH8 (Shimura *et al.* 2000). Figure 3.1.1a shows a model of parkin-mediated protein degradation by the UPP.

Figure 3.1.1a Parkin in the Ubiquitin Proteasome Pathway

Proposed mechanism of parkin as a RING-type E3 ubiquitin ligase that mediates protein degradation through the 26S proteasome, in partnership with the E2 enzymes UbcH7 and UbcH8. Substrates include α-Sp22, synphilin-1, CDCrel-1, Pael-R, cyclin E, synaptogamin XI, PACRG and SEPT5_v2.
(Adapted from Tanaka *et al.*, 2001).



A number of substrates for Parkin mediated ubiquitination have been identified, including an O-glycosylated form of α-synuclein (α-Sp22) (Shimura *et al.* 2001), the α-synuclein-binding protein synphilin-1 (Chung *et al.* 2001b), the synaptic vesicle-associated CDCrel-1 (Zhang *et al.* 2000b), the endothelin receptor-like receptor Pael-R (Imai *et al.* 2001), cyclin E (Staropoli *et al.* 2003), synaptogamin XI (Huynh *et al.* 2003), PACRG (Imai *et al.* 2003) and SEPT5_v2 (Choi *et al.* 2003).

In general, disease involving loss of parkin function lacks the LBs found in idiopathic PD, suggesting that either parkin-mediated neurodegeneration proceeds through mechanisms that are distinct from those in idopathic PD or that parkin may promote the formation of LBs. Indeed, *in vitro* studies suggest that LB formation is dependent upon parkin expression; the formation of these inclusions may be a mechanism to sequester potentially toxic proteins (reviewed by Chung *et al.* 2001a). Recent studies have also demonstrated that parkin may provide neuronal protection against a variety of insults. These include attenuation of Pael-R mediated dopaminergic cell loss in *drosophila* (Yang *et al.* 2003) and rescue of toxicity associated with proteasome inhibition and mutant α-synuclein expression, *in vitro* (Petrucelli *et al.* 2002). The identification of parkin substrates, and in some cases their specificity for dopaminergic cells may explain the select vulnerability of dopaminergic neurons, considering parkin is also expressed in other brain and non-brain regions (Feany and Pallanck 2003).

The gene encoding the parkin protein (*PRKN*) is one of the largest in the human genome at ~1.38 Mb, although the coding region is small by comparison: 12 exons give rise to a peptide comprised of 465 amino acids. *PRKN* is located at the telomeric end of chromosome 6q, within the third most common fragile site within tumour tissue (*FRAE6*) (Cesari *et al.* 2003) and shares a bi-directional promoter with the gene

encoding one of its substrates, *PACRG* (West *et al.* 2003). Mouse models may potentially provide insight into the mechanisms of parkin-linked neurodegeneration, as behavioural deficits and alterations in dopamine metabolism were observed in transgenic *PRKN* knockouts (Goldberg *et al.* 2003; Itier *et al.* 2003) and the *PRKN* null, spontaneous mutant 'Quaker' (Lockhart *et al.* 2004b), although dopaminergic cell loss was not observed. However, understanding why these cells remain resistant may be of therapeutic importance (West and Maidment 2004).

PRKN was first implicated in parkinsonism in consanguineous Japanese families with ARJP, in which homozygous deletions of whole and multiple exons were found (Kitada *et al.* 1998). Subsequent screening has identified a myriad of mutation within the gene: at least 55 point mutations and 34 exon rearrangements have been reported (reviewed by Periquet *et al.* 2003; Mata *et al.* 2004) and these are not restricted to the Japanese population, or cases with juvenile onset (≤25 years). *PRKN* is thought to account for up to 50% of recessively inherited familial early onset parkinsonism, 15% of sporadic early onset (≤45 years) parkinsonism in Europe (Lucking *et al.* 2000; Periquet *et al.* 2003). North American studies suggest that 2-11% of late-onset disease is attributable to *PRKN* (Foroud *et al.* 2003; Oliveira *et al.* 2003). However, a community-based study emphasized the value of comparison subjects when interpreting the causal relationship of *PRKN* variants, as overall mutation rates were similar in cases and controls (Lincoln *et al.* 2003).

Most *PRKN* mutations are homozygous or compound heterozygous, indicating that loss of function of the parkin protein is important in disease pathogenesis. However, other pathogenic mechanisms may exist (haploinsufficiency, dominant-negative effect), as in some cases only a single *PRKN* allele is affected (West *et al.* 2002a). Incidental carrier status of some heterozygous cases may account for an apparent

dominant mode of action, as pathogenic alterations found in the homozygous and compound heterozygous state are carried by normal individuals as well (Lincoln *et al.* 2003; Oliveira *et al.* 2003). However, it has been suggested that heterozygous alterations that attenuate, not ablate, parkin activity may result in a later age of onset (Foroud *et al.* 2003) and the presence of a second mutation is not thought likely to account for all heterozygous cases (30% of all *PRKN* cases) (Oliveira *et al.* 2003).

Due to the enormity of genomic *PRKN,* current screening strategies may have overlooked further mutations within heterozygous cases, although it is possible that different mutations alter parkin function to differing degrees, perhaps jointly with additional susceptibility genes/alleles (West *et al.* 2002a). This is not inconsistent with the complex nature of PD. *In vitro* studies showed that the RING finger-1 point mutations R256C and R275W induce altered protein localisation and aggresome-like aggregation consistent with wild-type protein overexpression (Cookson *et al.* 2003). This suggests that mutant *PRKN,* as well as causing recessive loss-of-function, may confer dominant gain of function, depending upon the protein domain affected. Interestingly, the presence of LBs has been demonstrated in a compound heterozygous Ex3Δ40/R275W individual (Farrer *et al.* 2001a) and LBs have been associated with aggresome formation (McNaught *et al.* 2002). Subtle variation in parkin protein levels arising from promoter polymorphisms may also be a risk factor for common, idiopathic PD (West *et al.* 2002b).

3.1.2 PARK7/DJ-1

Van Duijn and colleagues (2001) utilised homozygosity mapping in a large, multiply consanguineous Dutch kindred to identify the novel PARK7 locus at chromosome 1p36. The phenotype in this family was early-onset parkinsonism (≤40 years) with

slow disease progression and good response to *l*-DOPA. Psychiatric disturbances (severe anxiety, psychotic episodes) and dystonic features were also present in some individuals. Pathological data was unavailable, although PET evaluation was consistent with nigrostriatal dopaminergic dysfunction (van Duijn *et al.* 2001).

Preliminary analysis of STRP marker genotypes in affected individuals showed homozygosity at loci on chromosomes 1, 5 and 17. The chromosome 5 and 17 homozygous alleles were frequent (0.34 and 0.42 respectively) and nearby markers showed heterozygosity, indicating that the observed alleles were probably identical by state. However, the homozygous allele at the chromosome 1 marker (D1S468) had a population frequency of 0.09, and genotypes of the 3' adjacent marker (D1S214) were also homozygous, indicating that these alleles were probably identical by descent. The region around D1S468 and D1S214 was further genotyped, resulting in a maximum LOD score of 4.3 and a disease haplotype spanning 16cM (van Duijn *et al.* 2001).

Subsequently, linkage to PARK7 was reported in a further two consanguineous families from Italy (Bonifati *et al.* 2002). Further genotype analyses in the Dutch kindred allowed the critical region to be reduced to 5.6 Mb and ~90 genes. Screening of transcripts in the region identified a 14 Kb homozygous deletion containing the first 5 exons of the *DJ-1* gene (Δ1-5). A homozygous point mutation was also identified in an Italian PARK7-linked patient, resulting in the substitution of a conserved Leucine with Proline at amino acid 166 (L166P). Both mutations showed complete segregation with disease in the families and were not found in the homozygous state in other early-onset cases. Heterozygosity was identifed for Δ1-5, but carriers were asymptomatic (Bonifati *et al.* 2003).

Further screening in early onset cases has revealed additional exonic deletions (Δ5, Δ5-7), point mutations (M26I, E64D, A104T, D149A) splice site mutations (IVS6-1G→C, IVS5+2-12del) and a small deletion+substitution (c.56delC c.57G→A) (reviewed by Bonifati *et al.* 2004). However, in contrast to *PRKN*, *DJ-1*-linked disease is likely to account for only a small percentage of early-onset parkinsonism (~1%) (Abou-Sleiman *et al.* 2003; Bonifati *et al.* 2003; Hague *et al.* 2003; Ibanez *et al.* 2003; Healy *et al.* 2004; Hedrich *et al.* 2004; Lockhart *et al.* 2004a).

Most *DJ-1* mutations are found in the homozygous or compound heterozygous state, although heterozygous mutations have been reported as pathogenic. These may act in a dominant manner, or increase susceptibility in concert with unidentified variation (e.g. intronic or promoter mutations) within *DJ-1* or other susceptibility genes (Moore *et al.* 2003). Heterozygosity may also be associated with a later age of onset, as shown for *PRKN* carriers. However, neuroimaging studies in patients and asymptomatic carriers will be needed to address the role of heterozygous *DJ-1* mutations (Djarmati *et al.* 2004).

The *DJ-1* gene is evolutionarily conserved and contains 8 exons, the first two of which are non-coding and alternatively spliced in mRNA. The major transcript encodes a 189 amino acid, 20kDa protein and is expressed in various human tissue, including heart, brain, liver, skeletal muscle, liver and pancreas (Bonifati *et al.* 2003). In human brain, sub-cortical expression is higher than in cortical regions (Bonifati *et al.* 2003) and immunoreactivity has been reported in both neuronal and glial cells (Bandopadhyay *et al.* 2004; Rizzu *et al.* 2004).

Structural studies of the protein reveal that the crystalline architecture is an asymmetric, 2-fold axis, face-to-face dimer (Honbou *et al.* 2003). The L166P

mutation is located within an α-helix that forms part of a hydrophobic core. It has been shown to disrupt folding of the monomer, inhibiting the formation of both mutant/mutant and mutant/wildtype dimers (Olzmann *et al.* 2003). Met26 is also positioned within the hydrophobic core, spatially close to Leu166 (Tao and Tong 2003).

DJ-1 cellular functions are likely to be multiple. Human DJ-1 cDNA was first cloned as a result of a yeast two-hybrid screen to identify *c-myc* interacting proteins and was speculated to be involved with *ras*-mediated signalling (Nagakubo *et al.* 1997). Roles in RNA stabilisation (Hod *et al.* 1999), tanscription (Takahashi *et al.* 2001) and the fertilization process (Klinefelter *et al.* 2002) have also been reported. DJ-1 may also exhibit protease (Olzmann *et al.* 2003) and chaperone activity (Lee *et al.* 2003). The rat DJ-1 orthologue (SP22) has been implicated in spermatogenesis and the fertilisation process (Welch *et al.* 1998) although a similar function in humans has not been explored.

It is not clear exactly how mutant DJ-1 manifests as parkinsonism, although *in vitro* studies have shown the protein plays a role in the antioxidative stress reaction. Mutations, including L166P can also induce cell death (Taira *et al.* 2004). DJ-1 may also be implicated in tauopathies, as immunoreactivity co-localised within a subset of pathological tau inclusions in PSP, FTDP-17, PiD, DLB and AD, although the same antibodies failed to label LBs or LNs in PD or DLB (Rizzu, 2004). However, Neumann and colleagues (2004) report immunoreactivity in reactive brainstem astrocytes in PD and DLB (Neumann *et al.* 2004).

3.1.3 PARK6/PINK1

In addition to PARK7, a further region of chromosome 1p35-6 is linked to early onset recessive parkinsonism. This locus, PARK6, is ~25 cM centromeric to PARK7/*DJ-1* and was initially mapped in a large consanguineous Italian family. The ages of onset of affected individuals were 32-48 years, with a typical parkinsonian phenotype including slow progression and sustained response to *l*-DOPA, although no pathological data were available. Valente and colleagues (2001) employed homozygosity mapping to identify a 12.5 cM disease-linked interval which gave a maximum LOD score of 4.01 (Valente *et al.* 2001).

Linkage to PARK6 was subsequently confirmed in a further 8 families from Italy, the Netherlands, Germany and the UK. Haplotype analysis suggested the occurrence of independent mutational events and allowed the critical region to be narrowed to 9cM, flanked by STRPs D1S483 (telomeric) and D1S2674 (centromeric). Clinical features of PARK6-linked cases were similar to *PRKN*-positive ARJP of non-Japanese origin, showing a good response to *l*-DOPA and slow disease progression, without dystonia at onset and sleep benefit. Although mean age of onset was <45 years, in a quarter of cases onset was later (latest=68 years). In the later onset cases, the clinical features were indistinguishable from those of typical idiopathic PD (Valente *et al.* 2002).

Very recently, the PARK6 gene was identified as *PINK1*, encoding the PTEN-induced kinase 1 (PINK1). Two homozygous mutations, a G309D substitution in exon 4 and a W437OPA substitution in exon 7 were found in one Spanish and two Italian families, respectively. The exon 4 mutation is at a highly conserved amino acid within the putative kinase domain and the exon 7 mutation causes premature product truncation. Preliminary *in vitro* studies suggested that PINK1 may protect

product truncation. Preliminary *in vitro* studies suggested that PINK1 may protect neurons from stress-induced mitochondrial dysfunction and apoptosis, and that G309D negatively affects these processes(Valente *et al.* 2004).

3.1.4 Aims of the Work Described in Chapter 3

Most PD cases (79%) in Norway are sporadic in nature (there is no family history of disease.) This could reflect the presence of recessive loss of function mutations. Therefore, this chapter aimed to investigate the contribution of recessive loci (Section 1.1.5.2) to PD in the Norwegian population.

Referral studies suggest a significant proportion of PD is due to *PRKN* mutation, however community based studies may be more useful in interpreting the causal relationship of *PRKN* variants (Section 3.1.1). The first aim of this chapter was to determine the contribution of *PRKN* to PD in Norway through a case/control asssessment of *PRKN* mutation prevalence.

At the time this study was initiated, two further loci had been linked to recessively inherited early onset parkinsonsim. These were *PARK7* at chromosome 1p36 and *PARK6* at chromosome 1p35-36. Both loci had been identified in consanguineous families of European origin. It was hypothesised that *PARK6* and *PARK7* may also contribute to parkinsonism in the Norwegian population. The second and third aims of this chapter were to identify individuals that might be linked to either loci through assessing homozygosity. In particular regions that were likely identical by descent rather than by state were sought, as these are more likely to harbour recessive mutations. At this time, neither *DJ-1* (*PARK7*) nor *PINK1* (*PARK6*) had been identified as the causative genes.

3.2 Results

3.2.1 PRKN Mutation Screening

PRKN mutation screening was carried out in 193 cases (probable and possible PD) and 81 controls. Median age of cases was 71 years (range 46-94 years) with a median age of disease onset of 60 years (range 25-85 years); 59% were male (n=114). Twenty-one cases had an AOO ≤45 years (11%). Median age of controls was 81 years (range 69-100); 47% were male (n=38).

Sample ascertainment was ongoing at the time of this study, and only common dosage alterations of exons 3, 4 and 5 and common point mutations in exons 3, 7, 9 and 11 were assessed. Exon dosage was assessed by semi-quantitative PCR (Section 2.2.9). Point mutation screening was carried out by DHPLC (Section 2.2.6.3). Samples that demonstrated unusual melting profiles (in comparison to known normal controls) were fully sequenced (Section 2.2.7). Primers and PCR conditions are detailed in Appendix 1. Table 3.1.2a summarises the mutations identified.

Overall, 16 cases were identified with at least one mutation (8%). Four controls were also carriers (5%). Five individuals (4 cases and 1 control) also had mutations affecting more than one exon. However, phase was not determined so with the exception of P029 (homozygous deletion of exon 5), it was unclear whether these mutations affect both alleles. A family history of PD was noted in 5 cases with a *PRKN* mutation in none of the controls. *PRKN* mutation carrier frequency was 0.10 in cases and 0.07 in controls.

Table 3.2.1a *PRKN* Mutations Identified in PD Cases and Controls

Sample	Diagnosis (Family History)	Gender	Age of Onset/Age (Years)	Mutation
P029	Possible PD (Cousin PD)	M	35	Homozygous Deletion Exon 5
P171	Probable PD	F	38	A82E
P221	Probable PD (Father PD)	F	40	A82E
P176	Probable PD	M	40	R275W
P234	Probable PD	M	44	Deletion Exon 3
P184	Probable PD	M	46	A82E
P033	Probable PD	F	47	R256C
P214	Probable PD	M	52	A82E
P011	Possible PD (Cousin PD)	M	58	Deletion Exon 3 and 4*
P206	Probable PD	M	63	R275W
P091	Possible PD (Cousin PD)	M	64	A82E
P085	Probable PD	M	64	R275W
P154	Probable PD (Aunt PD)	F	66	3 bp Insertion Exon 3 / A82E*
P126	Probable PD	M	67	A82E
P032	Probable PD	M	77	Deletion Exon 4 and 5*
P164	Probable PD	F	88	R275W
K043	Normal	M	72	A82E
K035	Normal	M	72	Deletion Exon 3 and 4*
K003	Normal	F	86	A82E
K058	Normal	F	98	A82E

* Phase undetermined

Table 3.2.1b *PRKN* Exon 3 Region Genotype Analysis in A82E Carriers

	K003		K043		K058	
D6S1599	131	**133**	131	**133**	153	**133**
IVS2-P2	178	**180**	178	**180**	200	**180**
IVS2-P3	252	**264**	252	**264**	254	**264**
A82E	1	**2**	1	**2**	1	**2**
D6S980	279	297	301	**273**	289	**273**

	P091		P126		P154		P184		P214		P221		P171	
D6S1599	133	**133**	131	**133**	133	**133**	131	**133**	131	**133**	131	**133**	133	133
IVS2-P2	182	**180**	178	**180**	180	**180**	178	**180**	178	**180**	178	**180**	180	180
IVS2-P3	264	**264**	254	**264**	264	**264**	252	**264**	252	**264**	264	**264**	252	266
A82E	1	**2**	1	**2**	1	**2**	1	**2**	1	**2**	1	**2**	1	2
D6S980	293	**273**	281	**273**	NA	NA	265	**273**	293	**273**	297	**273**	287	291

Phase could not be determined, hence genotypes are presented according to the most parsimonious haplotypes. The inferred ancestral A82E haplotype is highlighted in blue. NA=not assessed.

The A82E mutation (exon 3) was found to be more common than has previously been noted, with an overall allele frequency of 0.02. This was postulated to be the result of a common founder in this relatively isolated population. Therefore, genotypes surrounding the mutation were investigated with additional STRPs, spanning a region of ~200 Kb. D6S1599, IVS2-P2 and IVS2-P3 (intron 2) and D6S980 (intron 3) were genotyped in A82E carriers (Section 2.2.8). Primers and PCR conditions are detailed in Appendix 1. Phase could not be determined, hence the most parsimonious haplotypes are presented in Table 3.1.2b. Common haplotypes could be reconstructed from allelic information in 9 out of 10 carriers.

3.2.2 PARK7/DJ-1

3.2.2.1 STRP analysis at PARK7

The *PARK7* critical region was initially reported to lie between STRP markers D1S243 (telomeric) and D1S244 (centromeric) (van Duijn *et al.* 2001) and *PARK7*-linked cases had AOOs ≤40 years (Bonifati *et al.* 2001). Seven cases with onset ≤40 years were available at this time. Only one of these cases had a family history of parkinsonism (P123), with possible recessive transmission, as both paternal and maternal aunts were reportedly affected. Lack of family history in remaining cases was also possibly indicative of recessive inheritence. Cases were genotyped for STRP markers (telomeric to centromeric) D1S243, D1S468, D1S2870, D1S1646, D1S1612 and D1S244 (Section 2.2.8) to identify regions of extended homozygosity. Primers and PCR conditions are detailed in Appendix 1.

Two cases, P016 and P113, were homozygous for 2 adjacent markers (Table 3.2.2.1a; highlighted in yellow). To identify whether these individuals likely carried alleles that were identical by descent (rather than by state), Norwegian population allele frequencies were assessed in control individuals (Table 3.2.2.1b).

Between 42 and 77 control individuals were genotyped for the 6 STRPs, revealing between 7 and 15 different alleles at each locus. Alleles carried by P016 at D1S468 and D1S2870 were found to be the most common in Norwegians (frequencies of 0.40 and 0.35 respectively) indicating they were likely to be identical by state. Allele 5 at D1S1646 in P113 was also found to be fairly common (frequency = 0.28, third most common) although allele 5 at D1S1612 had a frequency of 0.14, suggestive of identity by descent. No further clinical information was available for this individual. Since psychiatric disturbance was a feature of *PARK7*-linked cases, clinical information

would have been useful in determining the likelihood that P113 represented a *PARK7*-linked case. Whilst plans for follow-up work were underway (including procurement of extended family samples and clinical information, and haplotype analysis), the causative gene at *PARK7* was identified as *DJ-1*, flanked by markers D1S1646 and D1S1612 (Bonifati *et al.* 2003).

Table 3.2.2.1a *PARK7* STRP Marker Case Genotypes

	Sample	P095	P123	P112	P016	P014	P102	P113
	AOO	25	33	36	39	40	40	40
	Marker							
	D1S243	5 9	7 9	1 8	1 3	1 8	7 8	4 6
4.2 cM (1.4 Mb)								
	D1S468	4 6	3 6	1 3	6 6	6 6	6 6	3 3
9.8 cM (2.6 Mb)								
	D1S2870	3 13	6 12	11 12	6 6	4 8	5 12	5 11
0.6 cM (0.9 Mb)								
	D1S1646	NA	NA	3 4	4 5	4 5	3 4	5 5
1.6 cM (1.0 Mb)								
	D1S1612	5 6	2 4	1 2	1 8	7 8	1 5	5 5
4.4 cM (2.5 Mb)								
	D1S244	NA	NA	7 7	5 10	11 11	5 9	9 10

STRP alleles are recoded (1=shortest repeat length). Intermarker genetic (and physical) distances are shown on the left. Individuals with 2 adjacent homozygous genotypes are highlighted in yellow. NA=not assessed.

Table 3.2.2.1b *PARK7* STRP Norwegian Population Allele Frequencies

D1S243	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
(Product size)	(142)	(150)	(154)	(156)	(158)	(162)	(164)	(166)	(168)	(170)	(172)	
Control n	17	1	11	24	5	8	5	13	7	4	1	96
Control Frequency	0.18	0.01	0.11	0.25	0.05	0.08	0.05	0.14	0.07	0.04	0.01	

D1S468	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
(Product size)	(173)	(181)	(183)	(185)	(187)	(189)	(191)	
Control n	10	2	46	15	3	62	16	154
Control Frequency	0.06	0.01	0.30	0.10	0.02	0.40	0.10	

D1S2870	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**Total**
(Product size)	(194)	(200)	(202)	(204)	(206)	(208)	(210)	(212)	(214)	(216)	(218)	(220)	(222)	(224)	(228)	
Control n	0	1	17	4	4	31	5	0	1	6	6	7	1	4	1	88
Control Frequency	0.00	0.01	0.19	0.05	0.05	0.35	0.06	0.00	0.01	0.07	0.07	0.08	0.01	0.05	0.01	

D1S1646	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
(Product size)	(130)	(134)	(138)	(142)	(146)	(150)	(154)	
Control n	1	1	29	29	25	2	1	88
Control Frequency	0.01	0.01	0.33	0.33	0.28	0.02	0.01	

D1S1612	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
(Product size)	(94)	(98)	(102)	(106)	(110)	(114)	(118)	(122)	(126)	(130)	(134)	
Control n	6	28	3	0	14	15	16	5	10	0	1	98
Control Frequency	0.06	0.29	0.03	0.00	0.14	0.15	0.16	0.05	0.10	0.00	0.01	

D1S244	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**Total**
(Product size)	(281)	(284)	(286)	(287)	(288)	(289)	(290)	(291)	(292)	(293)	(294)	(295)	(296)	(297)	(298)	
Control n	1	1	1	14	7	2	15	9	17	7	8	0	1	1	0	84
Control Frequency	0.01	0.01	0.01	0.17	0.08	0.02	0.18	0.11	0.20	0.08	0.10	0.00	0.01	0.01	0.00	

3.2.2.2 DJ-1 Mutation Screening

The *DJ-1* mutations identified by Bonifati and colleagues (2003) were a 14kb homozygous deletion that encompassed the promoter and exons 1 to 5 (Δ1-5) and a homozygous point mutation, L166P. P113 was screened for homozygous Δ1-5 by PCR (Section 2.2.2), heterozygous Δ1-5 by semi-quantitative PCR (Section 2.2.9) and L166P through direct sequencing (Section 2.2.7) (primers and PCR conditions are detailed in Appendix 1).

Exons 2 and 4 of *DJ-1* were successfully amplified from genomic DNA by PCR, indicating that no homozygous deletions of these exons were present. To test for carrier status (heterozygous exonic deletions) semi-quantative PCR was performed for exons 2 and 4. A known heterozygous Δ1-5 idividual was used as a control. Results are shown in Figure 3.2.2.2a. Normalised (according to the peak area for control PCR product) peak areas for exon 2 and 4 in the Δ1-5 idividual are similar to those for P113, and ~half that of normal individuals. This suggested P113 may carry a heterozygous deletion that includes exon 2 and 4. Ideal peak heights (scale is shown to the right of each track in Figure 3.2.2.2a) are within the 1000-4000 range and those for this particular assay were low (~100). Therefore, these results needed to be confirmed by repeating the initial PCR. Unfortunately, technical issues prevented this and carrier status of an exon 2-4 deletion in P113 could not be confirmed. Sequence data from exon 7 was also of poor quality and the presence of an L166P mutation could not be ruled out. Fresh blood samples are currently being collected for DNA extraction and immortalisation for cDNA analysis.

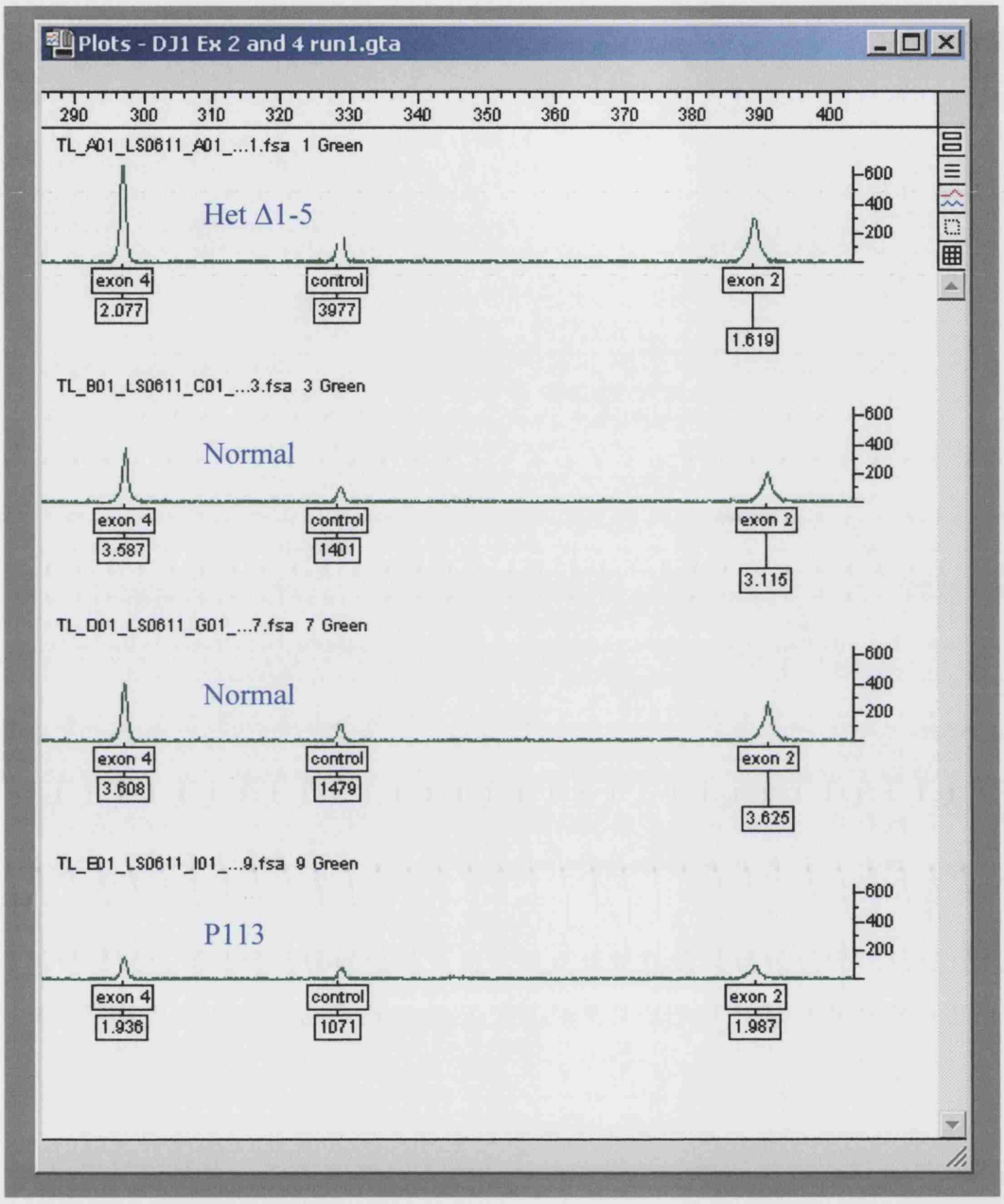


Figure 3.2.2.2a *DJ-1* Exons 2 and 4 Dosage Analysis by Semi-Quantitative PCR

Peak areas (normalised, according to the control PCR product peak) for exons 2 and 4 are shown beneath the peaks. Peak areas for P113 are similar to Het Δ1-5.

3.2.3 PARK6

At the time this study was initiated, *PINK1* had not yet been identified and the *PARK6* critical region spanned 9cM, flanked by STRPs D1S483 (telomeric) and D1S2674 (centromeric). Homozygosity across this region was assessed in cases without family history, with onset up to 68 years (consistent with the latest age of onset reported in a *PARK6*-linked family (Valente *et al.* 2002)). STRP markers genotyped were D1S552, D1S199, D1S478, D1S2828 and D1S2885 (Section 2.2.8). Primers and PCR conditions are detailed in Appendix 1.

Five cases were homozygous for two or more adjacent markers (Table 3.2.3a highlighted in yellow). Normal Norwegian control allele frequencies were assessed to determine the likelihood that homozygosity in cases was by descent rather than state. At least forty-six control individuals were genotyped and between 5 and 15 alleles were identified at each locus (Table 3.2.3b).

P052, P130 and P059 were homozygous for two adjacent markers, but at least one of these involved one of the most common alleles found in Norwegians (frequency >0.33). Of interest is P023 who was homozygous at D1S478, D1S2828 and D1S2885 with alleles that are rare in the normal Norwegian population (allele frequency <0.10). Further, allele 3 at D1S478 was extremely rare, identified only once in 98 control chromosomes. P095 was also homozygous for rare alleles at D1S2828 and D1S2885. Both cases have been clinically diagnosed with probable PD and neither has a reported family history of parkinsonism. Age of onset in P095 (25 years) was the youngest of the Norwegian cases, whereas it was considerably later in P023 (59 years). It was postulated that both cases potentially carried recessive mutations in the PARK6 gene. It was also assumed that the mutations would differ between the two

individuals as they carried different haplotypes across D1S479-D1S2828-D1S2885 (P095=4/13/9; P023=3/12/11); disparity in ages of onset might also suggest this. Follow-up work was initiated and included the procurement of clinical details and family DNA for further haplotype analysis. Prior to subsequent analyses, Bonifati and colleagues (2004) identified the *PARK6* gene, *PINK1*, which is located ~1Mb 3' of D1S199 and ~0.7 Mb 5' of D1S478 (Valente *et al.* 2004). However, time constraints did not allow sequence analysis to be undertaken in Norwegian samples.

Table 3.2.3a *PARK6* STRP Marker Case Genotypes

Sample	P095	P123	P112	P016	P014	P102	P113	P103	P099	P033	P052	P028	P012	P100	P120
AOO	25	33	36	39	40	40	40	44	45	47	47	48	49	49	49
Marker															
D1S552	4 4	2 3	3 4	1 3	3 4	1 3	3 3	3 3	2 3	3 3	3 5	3 4	1 3	1 2	3 4
D1S199	4 10	3 4	2 10	2 4	9 11	4 5	4 6	2 4	3 11	2 9	4 4	2 2	2 4	2 2	2 9
D1S478	4 4	4 6	5 7	5 7	4 7	6 7	5 7	2 7	4 7	7 7	7 7	5 7	5 7	1 4	7 7
D1S2828	13 13	3 7	3 14	3 7	3 12	3 14	4 11	5 5	5 8	3 12	9 14	3 12	3 7	3 7	6 13
D1S2885	9 9	8 10	3 8	3 7	8 9	3 10	7 15	6 12	5 12	6 7	6 8	3 9	3 7	6 6	3 6

Sample	P026	P098	P105	P073	P002	P046	P063	P071	P060	P042	P051	P055	P132	P058	P101
AOO	50	50	50	51	52	52	53	53	54	55	55	55	55	56	57
Marker															
D1S552	4 4	3 4	2 3	4 4	1 4	3 4	3 4	1 4	1 4	4 4	1 3	3 4	2 3	2 4	2 3
D1S199	3 9	2 8	2 2	1 2	8 11	1 12	3 4	4 6	2 2	3 7	4 4	2 6	2 3	3 6	2 4
D1S478	5 7	4 7	4 5	5 6	4 6	4 6	6 7	2 8	7 8	7 7	6 7	6 7	4 6	6 7	6 8
D1S2828	7 12	3 5	6 12	3 9	5 7	13 13	7 13	3 3	4 8	5 7	3 15	3 7	3 4	3 13	5 7
D1S2885	9 9	6 10	14 14	3 8	6 14	10 12	3 3	6 9	8 14	3 3	3 8	1 8	6 9	5 8	3 9

Sample	P019	P023	P025	P040	P056	P065	P072	P104	P003	P077	P130	P059	P096	P097	P088
AOO	59	59	60	60	60	60	60	60	61	61	61	62	62	62	63
Marker															
D1S552	4 4	3 4	1 5	1 3	3 5	3 4	3 4	1 4	1 3	3 4	2 3	1 1	2 4	2 3	4 4
D1S199	2 9	2 4	3 7	2 3	2 4	9 10	2 2	2 12	2 9	3 9	2 2	2 2	8 9	10 10	8 9
D1S478	4 7	3 3	2 7	6 7	5 7	1 6	6 7	4 7	4 7	7 7	6 6	1 7	4 7	4 7	4 7
D1S2828	3 13	12 12	9 13	5 7	3 12	4 12	3 12	3 5	3 4	6 13	3 9	3 13	8 8	3 4	3 6
D1S2885	2 11	11 11	6 11	3 10	6 13	8 14	8 9	3 3	6 6	2 9	2 2	7 8	4 9	3 6	5 10

STRP alleles are recoded (1=shortest). Individuals with 2 or more homozygous genotypes are highlighted in yellow.
Intermarker genetic (cM) and physical (Mb) distances are:
D1S552-(0, 0.7)-D1S199-(3.2, 1.6)-D1S478-(0, 0.3)-D1S2828-(6.6, 4.3)-D1S2885

Table 3.2.3b *PARK6* STRP Norwegian Population Allele Frequencies

D1S552	1	2	3	4	5	Total
(Product size)	(244)	(248)	(252)	(256)	(260)	
Control n	10	18	42	22	4	96
Control Frequency	0.10	0.19	0.44	0.23	0.04	

D1S199	1	2	3	4	5	6	7	8	9	10	11	12	Total
(Product size)	(94)	(96)	(98)	(100)	(102)	(104)	(106)	(108)	(110)	(112)	(114)	(116)	
Control n	2	32	6	13	2	2	5	4	8	10	9	5	98
Control Frequency	0.02	0.33	0.06	0.13	0.02	0.02	0.05	0.04	0.08	0.10	0.09	0.05	

D1S478	1	2	3	4	5	6	7	8	9	Total
(Product size)	(153)	(163)	(165)	(167)	(169)	(171)	(173)	(175)	(177)	
Control n	9	3	1	11	11	15	38	7	3	98
Control Frequency	0.09	0.03	0.01	0.11	0.11	0.15	0.39	0.07	0.03	

D1S2828	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
(Product size)	(247)	(249)	(251)	(253)	(255)	(263)	(265)	(267)	(269)	(271)	(273)	(275)	(277)	(279)	(281)	
Control n	5	1	31	1	7	13	8	5	2	2	2	6	5	2	2	92
Control Frequency	0.05	0.01	0.34	0.01	0.08	0.14	0.09	0.05	0.02	0.02	0.02	0.07	0.05	0.02	0.02	

D1S2885	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
(Product size)	(235)	(239)	(241)	(243)	(245)	(247)	(249)	(251)	(253)	(255)	(257)	(259)	(261)	(263)	(265)	
Control n	0	3	19	4	3	20	3	3	8	16	5	6	3	1	0	94
Control Frequency	0.00	0.03	0.20	0.04	0.03	0.21	0.03	0.03	0.09	0.17	0.05	0.06	0.03	0.01	0.00	

3.3 Discussion

3.3.1 PRKN

This study is the first to assess the frequency of *PRKN* mutations in Norway. As sample collection was ongoing at the time, only common mutations were screened for. However, from this initial assessment, it is likely that *PRKN* contributes towards parkinsonism in the Norwegian population, as 8% of cases (n=16) had at least one *PRKN* mutation; 5 had mutations affecting more than one exon. For one case (P029), it was certain that both alleles were affected as the individual was homozygous for a deletion of exon 5. Exon 5 deletion results in expression of an intact protein without amino acids 179-206 within the UPD (Kahle *et al.* 2000). Although it is not known exactly how this genotype manifests as parkinsonism, the age of disease onset was the youngest of all *PRKN* mutation carriers in this series. With onset of 35 years and a cousin who is also reportedly affected with PD, this is consistent with a recessive loss of parkin function. It would be useful to include the extended kindred in future *PRKN* screening, to further investigate pathogenesis in this family.

Three other cases and one control had mutations affecting more than one allele, although it is uncertain whether these occur *in trans* (on alternative alleles) or *in cis* (on the same allele). Samples are currently being recollected for immortalisation, therefore future cDNA analysis will determine the configuration of exonic deletions in individuals P011, P032 and K035. In individual P154, one of the mutations identified (in addition to A82E) was a novel, in-frame 3 bp insertion (CCA) in exon 3, theoretically resulting in the addition of a proline amino acid at position 167 within the UPD. The functional significance of this variant remains to be assessed.

The overall frequency of *PRKN* mutations identified in this community (cases=0.10, contros=0.07) is higher than that previously reported in a community-based study, although assessment of the contribution of *PRKN* to PD in the Norwegian population awaits completion of the screening process. Lincoln and colleagues (2003) evaluated the frequency of *PRKN* coding variants in a large clinical PD case control series from North America. Mutation frequencies ranged from 0.03 (controls) to 0.04 (cases) (Lincoln *et al.* 2003). The frequencies observed in the present study are also likely conservative estimates, as 6 exons remain to be analysed. Higher frequencies in Norwegians may in part be explained by the isolated nature of the population, in comparison to that of North America assuming that *PRKN* mutations do not affect biological fitness. Indeed, the frequency of A82E at 0.02 in the series as a whole, is the highest reported to date. Frequencies were not significantly different between cases and controls (χ^2=0.14, p=0.71), although it is not possible to speculate on the pathogenesis of this variant, as additional undetected mutations may be present. A82E has been reported as pathogenic in concert with a deletion of exon 7 (compound heterozygous) (Hedrich, 2001), as pseudo-dominant (the only mutation present) (West, 2002) and as a polymorphism (Oliveira *et al.* 2003). However, the results of the present study suggest the variant may represent a susceptibility allele, as the three control carriers are aged between 72 and 98 years. Unlike the point mutations R275W and R256C, A82E does not cause mislocalisation or aggreagation of the protein *in vitro*, although associated functional variations remain to be investigated (Cookson *et al.* 2003).

Periquet and colleagues suggest that *PRKN* exonic anomalies result from recurrent *de novo* events, whereas the prevalence of (some) point mutations may result from a small number of common founders, although A82E was not investigated in this study

(Periquet *et al.* 2001). Analysis of haplotypes around exon 3 suggests that high frequencies of A82E in Norwegians also result from a common founder (Table 3.1.2b) and further supports that Norwegians represent an isolated founder population (Section 1.4). Analysis of *PRKN*, a gene originally implicated in rare Mendelian parkinsonism, has therefore highlighted the potential utility of the Trondheim case control series in the study of the genetic aetiology of more common forms of PD (Section 1.3.2).

3.3.2 PARK7/DJ-1

At the time this study was initiated, homozygosity mapping in recessive families from a genetically isolated community in the Netherlands had identified PARK7 as a locus for early-onset parkinsonism (van Duijn *et al.* 2001). The assignment was confirmed in further Dutch and Italian families (Bonifati *et al.* 2002) and it was postulated that early-onset Norwegian cases with possible recessive modes of disease transmission may also be linked to this locus. Two cases (P016 and P113) were initially identified for further analyses. However, the causative gene (*DJ-1*) was subsequently identified by Bonifati and colleagues (2003) to lie between STRPs D1S1646 and D1S1612, and harboured a homozygous deletion of exons 1-5 (Δ1-5) and a homozygous point mutation (L166P) (Bonifati *et al.* 2003). Exon dosage analyses in P113 (homozygous for STRPs D1S1646 and D1S1612) suggested the presence of a heterozygous deletion of exons 2 and 4. Unfortunately, technical issues and time contstraints prevented confirmation of this in addition to screening for L166P.

It has been reported that heterozygous Δ1-5 individuals are not at increased risk for parkinsonism, and the mutation is not likely to exist outside the founder population where it was originally identified (Bonifati *et al.* 2003). However, it is not known

whether carrier status could influence risk in conjuction with other susceptibility alleles as has been proposed for heterozygous Parkin mutations (West *et al.* 2002a) (Foroud *et al.* 2003).

After the initial identification of Δ1-5 and L166P, further homozygous and heterozygous exonic deletions, homozygous and heterozygous missense and compound heterozygous mutations were reported in other populations (reviewed by Bonifati *et al.* 2004). It is interesting that *PRKN* was also originally identified in a population isolate, manifesting as homozygous deletions in recessive families, but since has been found to harbour multiple types of mutations and may be responsible for seemingly sporadic cases of disease. *PRKN* and *DJ-1* also reside within fragile regions of the genome which are prone to genetic aberrations (Schramayr *et al.* 1990; Smith *et al.* 1998). Although it has been reported that DJ-1-linked parkinsonism is likely to be rare (Abou-Sleiman *et al.* 2003; Bonifati *et al.* 2003; Hague *et al.* 2003; Ibanez *et al.* 2003; Healy *et al.* 2004), in some studies dosage analysis was either not carried out, or only the Δ1-5 mutation was screened for. In light of the recent discovery of further *DJ-1* mutations and the fact that they are not always recessive indicates that complete exon dosage and sequence analysis may be worthwhile in those early-onset Norwegian cases. Even if *DJ*-1 does transpire to cause only relatively few cases of parkinsonism overall, it is at present the second most common genetic cause of early-onset parkinsonism (Abou-Sleiman *et al.* 2003) and the identification of additional mutations will be useful for the development of cellular and animal models to help elucidate the pathogenesis of mutant *DJ-1*. In addition, later ages of onset may be considered for screening as recently a heterozygous Δ5 mutation was found in an individual with onset at 45 years (Djarmati 2004). Interestingly disease course was rapid in contrast to that in recessive *DJ-1*-linked or

PRKN-linked disease, hinting that dominant gain of function mutations may be more toxic than complete loss of protein, as suggested for parkin. However, additional promoter or intronic mutations may have been overlooked by screening methods in this individual.

Although no pathological data is available for *DJ-1*-linked disease, there has been no evidence of DJ-1 immunoreactivity within LBs of typical PD. In contrast, co-localisation of immunoreactivity with some tau inclusions and aggregation of wild-type protein has been reported in tauopathies (Rizzu, 2004; Neumann, 2004). Tau pathology has been reported in *PRKN*-linked disease which is not usually associated with LBs (van de Warrenburg *et al.* 2001); it will remain to be seen whether DJ-1-linked disease (as it does at the clinical level) mimics parkin disease at the pathological level and whether DJ-1 pathology is present in parkin brain. It would also be of value to assess the contribution of *DJ-1* in tauopathies and mutation screening is warranted in CBD, PiD, PSP, FTDP cases.

3.3.3 PARK6/PINK1

Initial studies suggested that *PARK6* may be a common cause of recessively transmitted familial parkisonsim in Europe, as linkage has been reported in families from Italy, Britain, the Netherlands and Germany. Clinical features of disease in these families overlapped somewhat with typical late-onset PD, with an age of onset of up to 68 years (Valente *et al.* 2002). Homozygosity screening in Norwegian cases revealed two individuals (P095 and P023) that may have been useful in follow-up studies to narrow the *PARK6* candidate region, which was 9cM and prohibitively large for candidate gene screening, at the time this study was initiated.

PINK1 was very recently named as the causal gene at *PARK6*; two different homozygous mutations were identified in consanguineous families from Spain and Italy (Valente 2004). Haplotype analysis in earlier PARK6-linked families suggested the occurrence of independent mutational events (Valente *et al.* 2002), and indeed the *PINK1* mutations were found to be dissimilar in Spain and Italy. Both mutations were found only in the homozygous state and were not carried on any of 400 control chromosomes analysed (Valente *et al.* 2004). Results of *PINK1* mutation screening in other *PARK6*-linked families, or other early-onset cases, are not available as yet. For cases P095 and P023, it is not known as yet whether they are part of a consanguineous kindred, although STRP allele frequencies in the vicinity of *PINK1* suggest that some degree of inbreeding has occurred. This is not unexpected in this isolated community. It will now be important to undertake *PINK1* mutation screening, including exon dosage analyses, in these individuals. The two other genes that contribute to recessively inherited early-onset parkinsonism, *PRKN* and *DJ-1* also reside within fragile chromosomal regions and exon copy number alterations are common. It is therefore hypothesised that *PINK1* may also harbour these types of mutation.

3.4 Summary and Conclusions

At the time the present study was initiated, two distinct regions of chromosome 1 (*PARK7* and *PARK6*) had been linked to early-onset recessive parkinsonism in European populations. *PARK7* was a 16 cM region and *PARK6* was 12.5 cM, both prohibitively larage for candidate gene searching. We therefore attempted to identify Norwegian individuals that might be linked to either loci, by seeking regions of homozygosity that were likely ro be identical by descent. The *PARK7* gene, *DJ-1,* was identified within the time-frame of the present study, therefore the two known mutations at that time, Δ1-5 and L166P, were screened for in individual P113 who showed extended rergions of homozygosity for rare alleles in the region. Results were suggestive of a heterozygous exon 2 and 4 deletion, although technical issues prevented confirmation of this.

The *PARK6* gene *PINK1* was only very recently identified, after the time-frame of this study and therefore could not be screened. However, haplotype analysis in Norwegians identified two cases (P095 and P023) that are homozygous for very rare STRP alleles; *PINK1* mutation screening should therefore be prioritised in these individuals.

Mutations in the *PRKN* gene were initially identified in consanguineous Japanese families and found to be causative for a juvenile onset form of parkinsonism, ARJP. These mutations included deletions of whole/multiple exons. *PRKN* screening in a variety of populations has identified an array of variation that has been associated with not only familial early-onset but cases of sporadic late-onset disease also. Studies have indicated that *PRKN* mutation accounts for a considerable proportion of parkinsonism, however community-based studies, in which variation has been

assessed in normal individuals, hint that previous referral-based studies may have overestimated the pathogenesis associated with this gene. Results of the present study indicate that mutation carrier frequency is comparable between cases and normal controls, although assessment of the contribution of *PRKN* to PD in Norway awaits further screening. These preliminary results, however, indicated that carrier frequency, in particular for A82E, was high compared with an outbred, US case-control series. This was hypothesised to result from a common founder in the relatively isolated population and haplotype analysis about A82E indicated that this was likely.

We can therefore conclude that 1) *PRKN* variation is relatively frequent in Norwegians 2) the population is likely to represent an isolated founder population, useful for the further dissection of genetic causes of parkinsonism and 3) *DJ-1* variation may be present in this population.

4 Investigation of the MAPT H1 Haplotype

4.1 Introduction

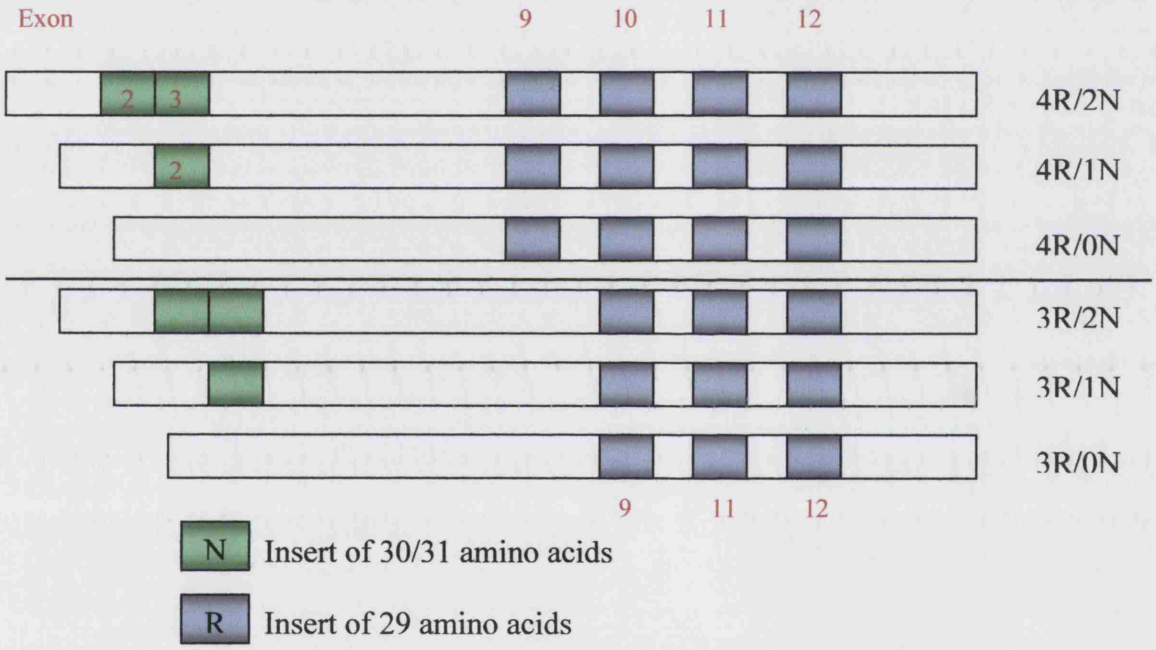
4.1.1 Microtubule Associated Protein Tau

Tau proteins are microtubule-associated proteins that are abundant in the central nervous system (CNS), predominantly expressed in neurons and particularly enriched in axons. They are are encoded by a single gene (*MAPT*) at chromosome 17q21 that is ~130 Kb, containing 16 exons (Poorkaj *et al.* 2001). Alternate splicing of exons 2, 3 and 10 results in the expression of six different polypeptides in the human CNS. Exclusion of exons 6 and 8 and inclusion of exon 4A also results in a higher molecular weight species which is expressed in the peripheral nervous system (reviewed by Schraen-Maschke *et al.* 2004).

CNS tau isoforms differ by the presence of either 3 (3R) or 4 (4R) carboxy-terminal tandem repeats of 31 or 32 amino acids, encoded by exons 9-12. Alternative splicing of exon 10 produces the 3R (exon 10-) and 4R (exon 10+) isoforms. The ratio of 3R and 4R in the adult human brain is ~1:1 (reviewed by Shahani and Brandt 2002). These isoforms also differ in that they contain either 0 (0N), 29 (1N) or 58 (2N) amino acid inserts in the amino terminal half mediated by the inclusion or exclusion of exons 2 and 3. 0N, 1N and 2N isoforms comprise ~50%, ~40% and ~10% of total tau, respectively. Developmental regulation is also apparent, as only the shortest isoform (3R/0N) is expressed in fetal brain (reviewed by Lee *et al.* 2001). A schematic of the tau isoforms and their nomenclature in the adult CNS is shown in Figure 4.1.1a.

Figure 4.1.1a Tau Isoforms and Nomenclature in the Adult CNS

The six isoforms differ with respect to the number of carboxy-terminal repeats of 30/31 amino acids (3 or 4). These constitute the microtubule binding domains and are encoded by exons 9-12. The number of inserts at the amino-terminal (0=0, 1=29 or 2=58 amino acids) also differs. These are generated by exclusion of exon 2 and 3 (0N), inclusion of exon 2 only (1N) or inclusion of exons 2 and 3 (2N). Exons corresponding to domains are numbered in red.



Tau is thought to have multiple functions. Most notably it binds, stabilises and promotes microtubule (MT) polymerisation. The MT interaction is mediated through the carboxy-terminal repeats encoded by exons 9-12 (Figure 4.1.1a). 4R isoforms have greater MT binding affinity and more efficiently promote MT assembly than do the 3R isoforms. The region in between the amino and carboxy termini, which is rich in Proline ('Proline-rich region'), may also mediate MT interactions. Other studies have suggested roles for tau as a linker protein between MTs and the actin cytoskeleton, in signalling mechanisms involved in neuronal polarity and in the trafficking of organelles including vesicles and mitochondria (Shahani and Brandt 2002) (Lee *et al.* 2001).

Phosphorylation plays an important role in regulating tau function. This occurs at numerous sites, particularly clustered around the MT binding repeats and the Proline-rich region. Phosphorylation events can change the conformation of tau, decrease MT binding and increase the dynamic instability of MTs. This post-translational modification is also developmentally regulated; phosphorylation is higher in fetal neurons and decreases with age (reviewed by Avila *et al.* 2004). Abnormally and hyperphosphorylated tau is a major component of the pathological neuronal neurofibrillary tangles (NFTs) found in multiple neurodegenerative disorders, including Alzheimer's disease (AD), progressive supranuclear palsy (PSP), corticobasal ganglionic degeneration (CBD), argyrophilic grains disease (AgD) and frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) (Shahani and Brandt 2002).

A direct role for the *MAPT* gene as a causative factor in neurodegeneration was established with the discovery of mutations in familial FTDP-17 cases (Hutton *et al.* 1998; Poorkaj *et al.* 1998). Since then, more than 30 mutations have been reported in

over 80 FTDP-17 families. These consist of coding missense, deletion or silent mutations, or exonic and intronic mutations around the splice donor site of the intron following exon 10 (reviewed by Goedert 2003). These splice site mutations may disrupt a stem–loop structure that is thought to negatively regulate exon 10 splicing, affect splicing enhancer or silencer sequences in exon 10, or affect a novel *trans*-acting splicing factor. Irrespective of mechanism, the consequence is an increased recognition of exon 10, augmenting the ratio of exon 10+ to exon 10- *MAPT* mRNA and increased expression of 4R isoforms. Missense mutations can disrupt MT binding, promote filament formation and decrease tau degradation (Shahani and Brandt 2002).

4.1.2 MAPT Variability in Neurodegeneration

Mutations in *MAPT* are rare in tauopathies other than FTDP-17 (Schraen-Maschke *et al.* 2004). However, the 4R tauopathies PSP and CBD are associated with common polymorphic variability within *MAPT*. Conrad and colleagues (1997) were the first to report overrepresentation of specific *MAPT* alleles in Caucasian PSP cases compared to controls. They identified a dinucleotide repeat ($TG_{11\text{-}15}$) within the intron following exon 9 (intron 9) and found homozygosity for the shortest allele (TG_{11}, designated A0) was 96% in PSP cases, compared to 57% in controls ($p=0.001$) (Conrad *et al.* 1997). Analysis of SNPs within *MAPT* has revealed the presence of two major haplotypes, termed H1 and H2. The A0 allele is found only the H1 haplotype (Baker *et al.* 1999). Homozygosity for A0 or H1 has been further associated with PSP (Baker *et al.* 1999; Hoenicka *et al.* 1999; Morris *et al.* 1999; Pastor *et al.* 2000; de Silva *et al.* 2001) and CBD in Caucasian populations (Di Maria *et al.* 2000; Houlden *et al.* 2001).

Although primarily regarded as an α-synucleinopathy and not usually associated with tau pathology, the clinical and pathological characteristics of PD overlap with those of the tauopathies. PSP clinically presents with parkinsonism (Litvan and Lees 1999) and LBs have been identified in some cases (Mori *et al.* 2002). Tau aggregates have been demonstrated in ARJP caused by *parkin* mutation (van de Warrenburg *et al.* 2001), *α-synuclein*-linked PD (Duda *et al.* 2002) and tau co-localises with α-synuclein in LBs of sporadic PD (Ishizawa *et al.* 2003). α-synuclein has been shown to bind to and modulate tau phosphorylation (Jensen *et al.* 1999) and parkin may ubiquitinate tau (Shimura *et al.* 1999). In addition, overexpression of tau in the fly is sufficient to cause neuronal death without NFT formation (Wittmann *et al.* 2001). Evidence for linkage to chromosome 17q, close to *MAPT*, was also identified in a large genome scan of idiopathic PD (Scott *et al.* 2001). It is therefore possible that variability within the *MAPT* gene may contribute towards the clinical outcome of a spectrum of neurodegenerative disorders including PD.

Lazzarini and colleagues initially showed an association between the A0 alelle and PD (Lazzarini 1997). An expansion of this original series and meta-analysis of other independent, but largely equivocal studies, confirmed these findings. The overall frequency of the A0A0 genotype in PD was 61% (after correction for misdiagnosis of PSP as PD) compared to 52% in controls (p=0.002) (Golbe *et al.* 2001). An increase in H1/H1 genotypes over controls was also observed in three US (Maraganore *et al.* 2001; Martin *et al.* 2001; Clark *et al.* 2003) and a UK PD case control series (de Silva *et al.* 2002). However, these findings did not always reach statistical significance, hence the association between *MAPT* H1 and PD remains controversial.

4.1.3 The MAPT Extended Haplotype

A number of polymorphisms have been identified which are in complete disequilibrium with, and therefore define the H1 and H2 haplotypes (Baker *et al.* 1999; de Silva *et al.* 2001; Martin *et al.* 2001; Verpillat *et al.* 2002) with no evidence of meiotic recombination between the two. Baker and colleagues originally showed that the non-recombining region spans the entire *MAPT* gene (Baker *et al.* 1999). More recently, Pastor and colleagues further examined the region and both STRPs and SNPs were used to show the region extends some 60Kb north (5' and centromeric) and ~600 Kb south (3' and telomeric) of *MAPT* (Pastor *et al.* 2002).

The 3' end of the non-recombining region extends to the STRP D17S810 (Pastor *et al.* 2002). Although human genome sequence assemblies are largely complete, there are still gaps. In the chromsome 17q21 region, the orders and positions of some of the smaller sequenced units (contigs) dramatically changed during the year 2002. NCBI, UCSC and CDS human genome sequence data currently place D17S810 5' of *MAPT*. The 'relocation' of D17S810 since the report of Pastor and colleagues (2002) meant that the estimated extent of the haplotype is unreliable.

4.1.4 Aims of the Work Described in Chapter 4

Since previous *MAPT* H1 (or A0)/PD association studies have been carried out in outbred populations, it was reasoned that equivocal results and failure to reach statistical significance may be due to underlying heterogeneity in these samples and/or the use of convenience control samples. Hence, the first aim was to test the association of H1 with PD in the genetically homogenous Trondheim case-control series. The initial study that was carried out demonstrated a significant association between H1 and PD in Norwegians (Farrer *et al.* 2002). The first aim was to replicate

this finding in a larger, distinct set of samples. As the H1 haplotype probably includes neighbouring genes (Pastor *et al.* 2002), it is conceivable that a susceptibility locus resides outside of *MAPT*. Therefore, the second aim was to map the boundaries of the non-recombining region to determine the extent of H1 and hence the PD susceptibility candidate region.

Microsatellite variability within *MAPT* suggests the H1 haplotype may be partitioned into H1-specific sub-haplotypes (Golbe *et al.* 2001). We hypothesised that elevated copies of H1 in PD cases could be accounted for by the presence of a particular 'sub-H1' haplotype carrying disease risk, that is absent or present only at low frequencies in normal controls. Hence, the third aim was to dissect the genetic architecture of H1 sub-haplotypes and explore their association with PD.

4.2 *MAPT* H1 and Parkinson's Disease

4.2.1 Initial Studies

The association of PD with the *MAPT* H1 haplotype was investigated in the Trondheim case-control series. The ensuing publication can be found in Chapter 8 (Farrer *et al.* 2002). To summarise, 96 unrelated PD cases (clinically defined probable and possible PD) and 68 controls were genotyped for the *MAPT* H1 and H2 haplotypes, as defined by the A>G SNP rs1800547, by *BanII* restriction enzyme digest (A carried on H1, G carried on H2, (Baker *et al.* 1999)). The H1/H1 genotype was significantly associated with PD, even after adjustment for misdiagnosis of PD as PSP (OR=5.6, 95% CI 2.6-11.9, $p<3.1^{e-6}$). As the Trondheim series expanded, further samples became available for analysis.

4.2.2 MAPT H1 Haplotype Case-Control Analysis

The sample set used in this analysis was distinct from that previously reported (Farrer *et al.* 2002). It included 200 cases with a median age of 70 years (range 40-94 years), male frequency 0.59 and 373 controls with a median age of 58 years (range 50-93 years), male frequency 0.61.

MAPT H1 and H2 genotypes were identified by PCR fragment length polymorphism genotyping. The H2 variant contains a 238 bp deletion located in intron 9 of the gene (Baker *et al.* 1999) and is denoted 'int9 indel'. PCR amplification (Section 2.2.2) of the surrounding region was carried out (primers and PCR conditions are detailed in Appendix 1) and products were electrophoresed on a 1% agarose gel (Section 2.2.4). An example of the banding pattern for possible genotypes is shown in Figure 4.2.2a.

Odds ratios were calculated using SPSS (Section 2.3.11). The association between H1/H1 and PD remained highly significant in this replication series (OR=2.3; 95% CI=1.6-3.4, $p<1.5^{e-5}$) and also when all samples (those previously reported by Farrer *et al.* 2002 and in this section) were considered (OR=1.9, 95% CI=1.3-2.6, $p<1.5^{e-4}$; Table 4.2.2b).

Figure 4.2.2a *MAPT* H1/H2 Genotyping

PCR amplification of the region surrounding a 238bp deletion within intron 9 (int9 indel) gives a 498bp product for the H1 allele and a 260bp product for the H2 allele. An example of the banding pattern for H1/H2, H1/H1 and H2/H2 genotypes is shown.

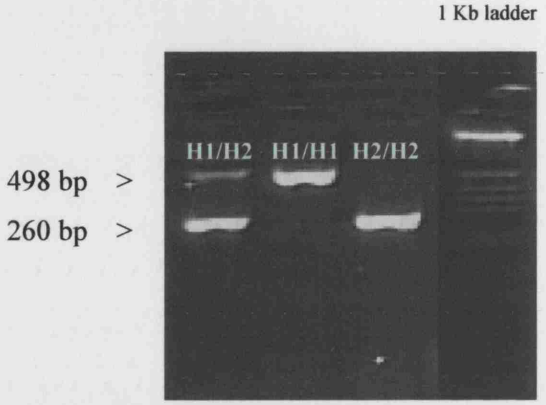


Table 4.2.2b Association Between PD and *MAPT* Haplotypes

	n	No. of genotypes (%) H1/H1	H1/H2	H2/H2	H1/H1 vs H1/H2 & H2/H2 OR	95% CI	*p*	H1/H1 & H1/H2 vs H2/H2 OR	95% CI	*p*
Controls	441	282 (63.9)	143 (32.4)	16 (3.6)	1	(ref)		1	(ref)	
Cases										
All (Prob, Poss, Atyp)	296	227 (76.7)	62 (20.9)	7 (2.4)	1.86	1.33-2.59	0.0002	0.64	0.26-1.58	0.33
Probable & Possible	280	213 (76.1)	60 (21.4)	7 (2.5)	1.79	1.28 2.51	0.001	0.68	0.28-1.68	0.40
Probable	223	169 (75.8)	50 (22.4)	4 (1.8)	1.77	1.23-2.54	0.002	0.49	0.16-1.47	0.19
Adjustment for possible PSP cases*	278	213 (76.6)	58 (20.9)	7 (2.5)	1.85	1.32-2.59	0.0004	0.69	0.28-1.69	0.41

* Genotype frequencies were adjusted for possible PSP cases misdiagnosed as PD by randomly removing 6% of cases (Golbe *et al.* 2001) (6%*76.7%=14 H/1H1s, 6%*20.9%=4 H1/H2s, 6%*2.4%=0 H2/H2s)

4.3 The Extended *MAPT* Haplotype

Pastor and colleagues estimated the minimum *MAPT* extended haplotype to be flanked by markers rs937 (~60Kb 5' of *MAPT*) and D17S810 (~600Kb 3' of *MAPT*) (Pastor *et al.* 2002). Subsequent to this publication, STRP D17S810 was relocated ~0.5 Mb 5' of *MAPT*, therefore the extent of the haplotype became unreliable. We sought to map the boundaries of the non-recombining region, and hence the maximum extent of H1, using the most recent genome sequence assembly data.

Fifteen individuals with H1/H1, H1/H2 or H2/H2 genotypes (as defined by the *MAPT* intron 9 indel; Section 4.2.2) were genotyped for SNPs extending 5' and 3' of rs937 and rs1816 respectively, which marked the minimum extent of the *MAPT* non-recombining region (rs937 was substituted by hCV7450783 which is ~200 bp 3') (Pastor *et al.* 2002). SNPs were chosen from the CDS SNP database, although it was not known *a priori* whether these polymorphisms were present in the Norwegian population.

SNP genotyping was carried out by restriction enzyme digest (Section 2.2.3) (except for ss24821057 and hCV2265263, which were detected through direct sequencing of a CDS putative gene hCG1640620; Section 2.2.7). Assays are detailed in Appendix 1. SNPs and their physical positions are shown in Table 4.3a. Some SNPs were not identified in these samples. An association of a SNP with the *MAPT* haplotype was identified by comparing *MAPT* int9 indel genotypes with SNP genotypes: a SNP associated (in linkage disequilibrium) with the *MAPT* haplotype gives the same set of genotypes in all samples; all H1/H1 individuals have a SNP genotype of 11, all H1/H2 individuals have a SNP genotype of 12 and all H2/H2 individuals have a SNP genotype of 22. Disparity between SNP genotype and *MAPT* haplotypes was

assumed to be an indication of recombination between the two. When genotyping was initially carried out, hCV7540752 was located ~60Kb 5' of SNP and analyses of *MAPT* haplotype/SNP association allowed the maximum and minimum extent of the haplotype to be established. At maximum, it extended over a ~700Kb region with the 5' boundary flanked by hCV7540752 and ss24821057 and the 3' end flanked by rs1816 and hCV11411459. However, hCV7540752 was later repositioned 3' of rs1816. This redefined the 3' boundary and also meant that the 5' boundary was no longer known. Based on the present physical locations of the SNPs genotyped in this study and the most recent position of D17S810, we can surmise that at minimum, the non-recombining region is ~800Kb in length and that the 3' boundary lies within a ~64Kb interval between hCV11936211 and hCV7450752), ~200 Kb 3' of *MAPT* (Figure 4.3b). There are therefore at least nine putative/known genes (including *MAPT*) in LD with the H1 haplotype; all are candidates for harbouring a risk variant associated with PD and other disorders in which H1 is over-represented.

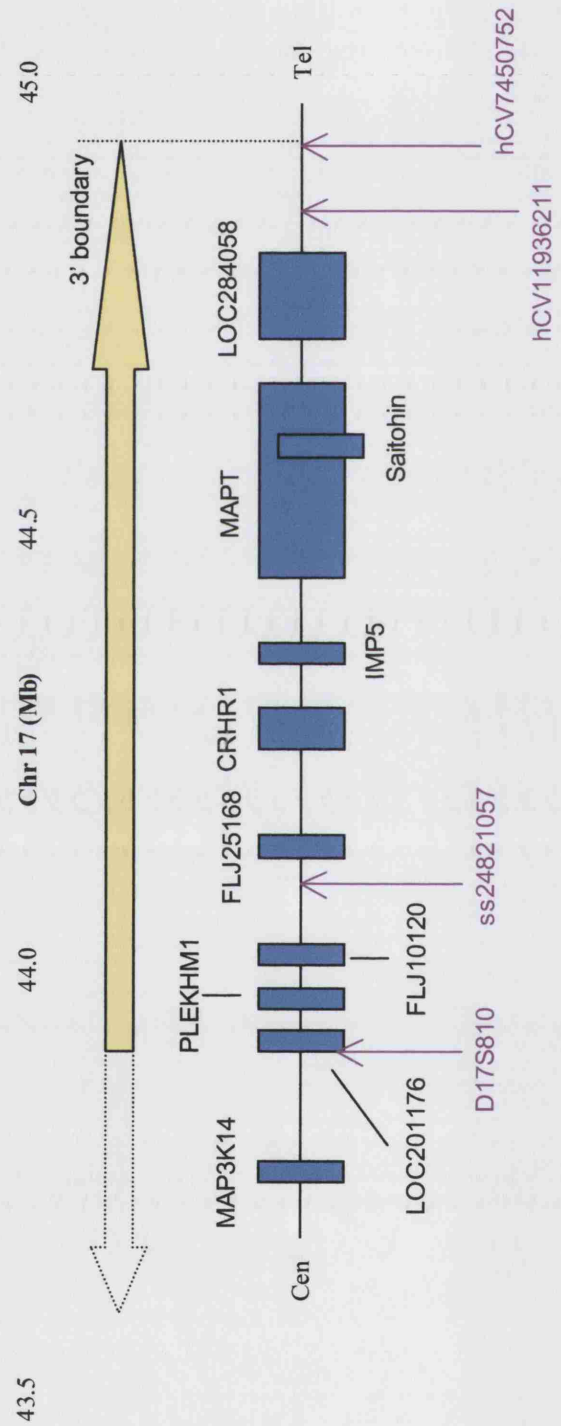
Table 4.3a SNPs Genotyped to Determine Boundaries of the Extended *MAPT* Haplotype

SNPs were genotyped in known H1/H1, H1/H2 and H2/H2 individuals. Not all SNPs were polymorphic in these samples. An association with the extended *MAPT* haplotype was identified when SNP genotypes were identical to *MAPT* haplotypes in all individuals. The 3' boundary of the non-recombining region lies between hCV11936211 and hCV7450752.

SNP	Celera ID	NCBI ID	Change	Chr 17 Position (bp)	Polymorphism Identified?	Association With int9 indel?
1		ss24821057	G > A	44,185,793	yes	yes
2	hCV2265270	rs413778	G > A	44,192,307	no	
3	hCV2265267	rs439558	T > C	44,193,225	no	
4	hCV2265263	rs3418	C > T	44,198,884	yes	yes
5	hCV7450783	rs878886	C > G	44,387,910	yes	yes
	MAPT			44,447,262 - 44,578,250		
6	hCV11936211	rs1816	A > G	44,752,034	yes	yes
7	hCV7450752		C > G	44,817,359	yes	no
8	hCV11416920		T > C	45,611,015	no	
9	hCV11417996		A > C	63,300,284	no	
10	hCV11411459	rs8866	C > G	65,924,080	yes	no
11	hCV2077569		G > C	66,022,999	no	
12	hCV11618556		C > T	66,821,253	yes	no

Figure 4.3b *MAPT* Extended Haplotype Region

The 3' boundary of (*MAPT* extended haplotype (yellow arrow) is flanked by markers hCV11936211 and hCV7450752. 5' of *MAPT*, the haplotype extends at least as far as gene LOC201176 (D17S810; Pastor et al. 2002).



4.4 Investigation of the *MAPT* H1 Haplotype

To further explore the association of *MAPT* H1 and PD, an epidemiologically-matched subset of 81 cases and 81 controls from the Trondheim series, homozygous for H1 (defined by int9 indel, Section 4.2.2) was selected for analyses. Median age in the case group was 76 ± 3.9 years (range 71-86) and 81 ± 5.1 years (range 72-93) in the control group and 58% were male. Microsatellite variability within *MAPT* suggests that H1 may be partitioned into H1-specific sub-haplotypes (Golbe *et al.* 2001). We chose to investigate H1 variability by the analysis of SNPs that are H1-specific as SNPs are preferable to microsatellite polymorphisms for the assessment of LD (Ardlie *et al.* 2002). Initially, SNPs were identified from NCBI and CDS databases. Genotyping of these putative variants in H1/H1 (n=11) and H2/H2 (n=11) individuals indicated whether they were polymorphic in Norwegians and H1-specific. SNPs that were chosen for further analyses were polymorphic within the population of H1 chromosomes and not present on H2 chromosomes (H1-SNPs).

4.4.1 Linkage Disequilibrium Within MAPT H1

Initially, 5 H1-SNPs within *MAPT* were genotyped in cases and controls. Pairwise LD statistics and GOLD figures were generated for each sample group using Genotype Transposer and GOLD (Section 2.3.8), shown in Figure 4.4.1a.

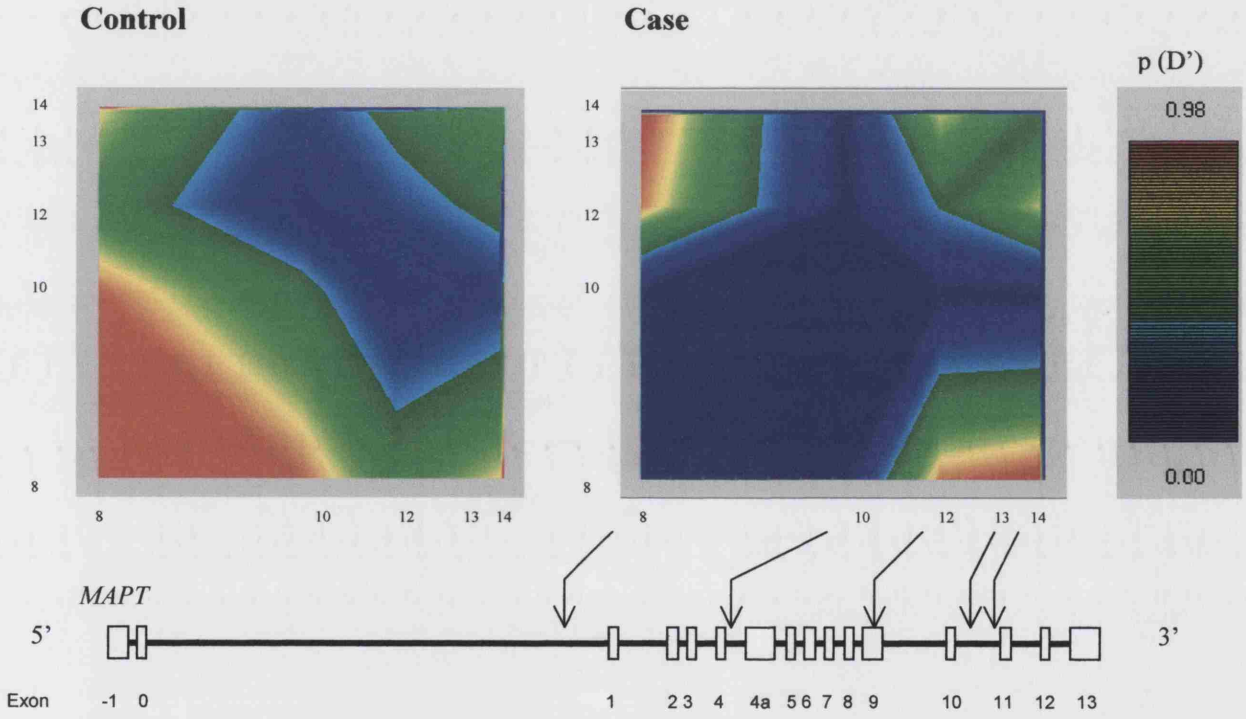
Figure 4.4.1a Graphical Overview of LD for *MAPT* H1-SNPs

a) H1-SNPs within *MAPT* for which LD was assessed. b) GOLD figures for H1-SNPs in controls and cases. Relative SNP positions are shown within an ideogram of genomic *MAPT* below (exons are denoted by boxes). c) Raw LD data used to construct GOLD figures.

a)

SNP	CDS ID	NCBI ID	Change
8	hCV3202957	rs242562	G > A
10	hCV16089259	rs2435207	G > A
12		rs11568305	G > A
13		rs11079728	T > C
14		rs11568306	CT > AA

b)



c)

CONTROL					CASE				
Marker1	Marker2	D	D'	p	Marker1	Marker2	D	D'	p
8	10	0.00	-0.01	0.91	8	10	0.09	0.39	0.00
8	12	0.00	0.19	0.57	8	12	0.00	-0.08	0.87
10	12	0.01	0.55	0.07	10	12	0.02	1.00	0.02
8	13	0.00	-0.09	0.79	8	13	0.00	0.01	0.97
10	13	0.01	0.27	0.21	10	13	0.06	1.00	0.00
12	13	0.00	-1.00	0.48	12	13	0.00	-0.36	0.77
8	14	0.00	0.00	0.98	8	14	0.04	0.28	0.02
10	14	-0.01	-0.11	0.50	10	14	-0.05	-0.41	0.00
12	14	0.03	1.00	0.00	12	14	0.02	1.00	0.00
13	14	0.00	-0.23	0.61	13	14	-0.03	-1.00	0.01

LD between rs242562 and rs2345207 was considerably greater in cases than in controls, and in the former it appears to extend further 5' of rs242562, which is located in the intron following *MAPT* exon 1 (intron 1). The non-recombining region extends further 5' of *MAPT*, including the *CRHR1* locus (Section 4.3). This gene encodes one of the two receptors (CRHR1 and CRHR2) that mediate the action of corticotropin-releasing hormone (CRH). This peptide has a role in endocrine, autonomic, behavioural, and immune responses to stress and has been implicated in a variety of neuroendocrine, neurological and psychiatric disorders (Grammatopoulos and Chrousos 2002). CRH levels in cortex of AD, PSP and PD brains were reduced when compared with normal controls (Rehman 2002) and CRHR1 has been shown to mediate the neuroprotective properties afforded by the CRH family-member urocortin, in rats (Pedersen *et al.* 2002).

4.4.2 H1-SNP Analyses within the CRHR1-MAPT Interval

To test the hypothesis that variation within *CRHR1* was responsible for the association of H1 and PD, additional H1-SNPs within the *CRHR1-MAPT* interval were sought. A further nine H1-SNPs were genotyped in cases and controls; the total fourteen H1-SNPs are detailed in Table 4.4.2a. They cover ~210Kb of the non-recombining region and average marker spacing is ~15Kb. Figure 4.4.2b illustrates relative positions within gene loci.

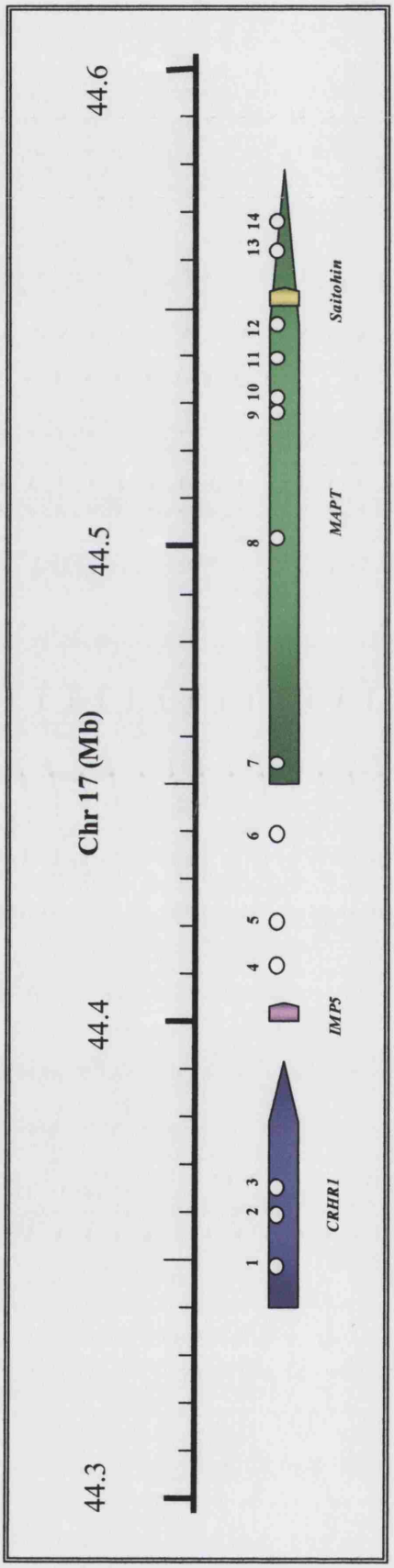
Table 4.4.2a H1-SNPs

14 H1-SNPs in the *MAPT-CRHR1* (non-recombining) interval, genotyped in case and control H1 homozygotes.

SNP	CDS ID	NCBI ID	Change
1		rs110402	C>T
2	hCV2544834	rs171440	G>A
3	hCV2544832	rs242937	G>A
4	hCV2544799	rs242935	T>C
5	hCV2544792	rs242928	A>G
6	hCV2257661	rs2019820	C>T
7	hCV3202942	rs11079727	C>A
8	hCV3202957	rs242562	G>A
9		rs3785883	G>A
10	hCV16089259	rs2435207	G>A
11	hCV16017251	rs2258689	C>T
12		rs11568305	G>A
13		rs11079728	T>C
14		rs11568306	CT>AA

Figure 4.4.2b H1-SNPs Within the *CRHR1-MAPT* Interval

H1-SNPs are shown as white circles and numbered 1-14 as shown in Table 4.4.2a.

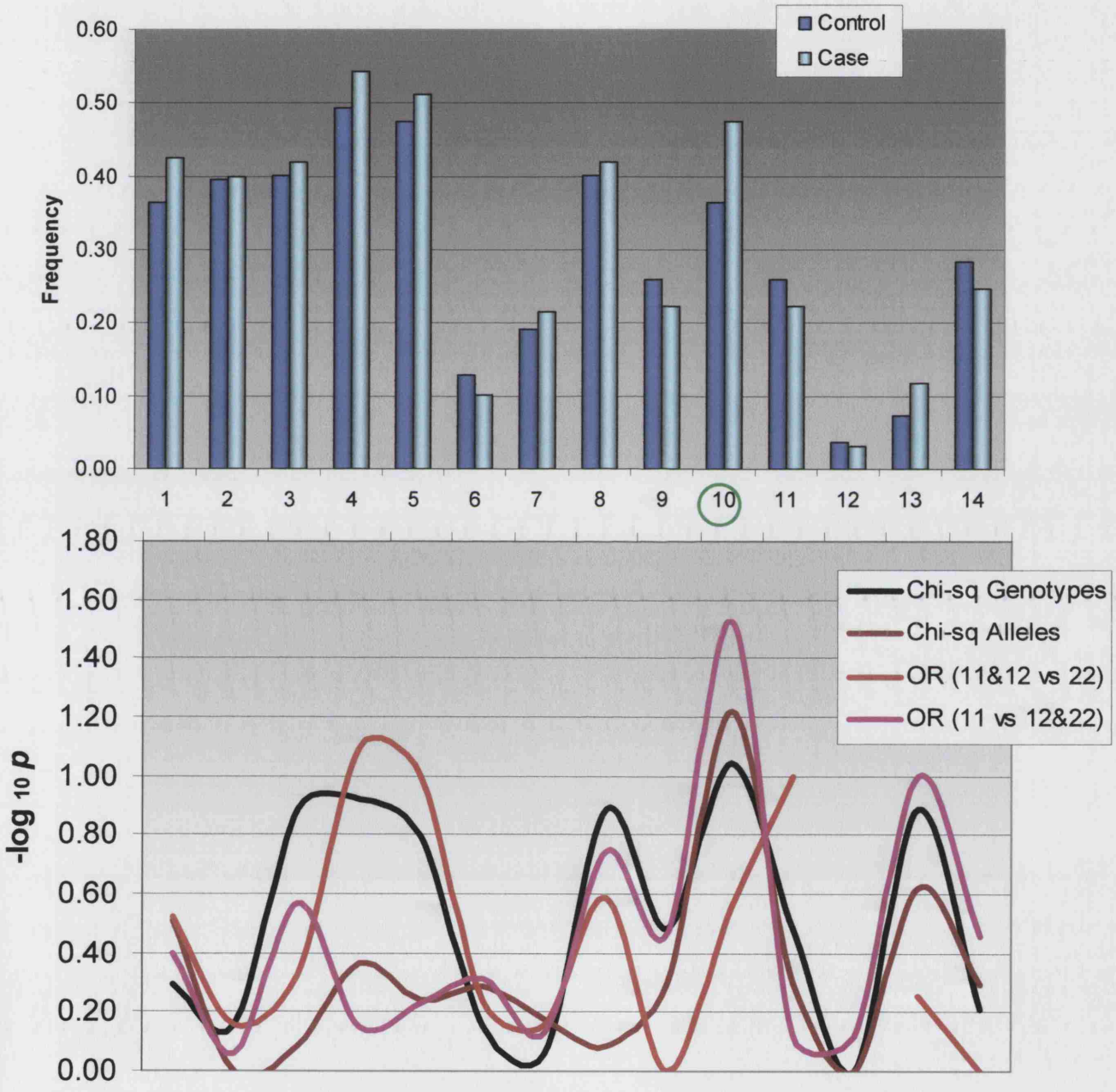


Single marker tests for association were carried out using CLUMP (Section 2.3.10) for allele and genotype frequencies. Odds ratios for each genotypic variant (11&12 vs 22 and 11 vs 12&22) were calculated using SPSS (Section 2.3.11). Allele frequencies and p values for all tests are shown in Figure 4.4.2c.

Most of the 14 H1-SNPs exist at appreciable frequencies (>0.10) and half have frequencies >=0.40, suggesting these are ancient polymorphisms that are likely to be informative. Notably, the minor allele (2=A) at SNP 10 (rs2435207) was over represented in cases (OR [11 vs 12&22] =2.1; 95% CI=1.1-4.2; p=0.03).

Figure 4.4.2c H1-SNP Association Tests

Top panel shows H1-SNP (1-14) allele frequencies in cases and controls. Bottom panel shows $-\text{Log}_{10}$ p values for χ^2 tests comparing genotype and allele frequencies and for Odds Ratios (11&12 vs 22 and 11 vs 12&22) between cases and controls. Notably, the minor allele (2=A) for SNP 10 (rs2435207) is significantly over-represented in cases (OR [11 vs 12&22] =2.1; 95% CI=1.1-4.2; p=0.03).



4.4.3 H1-SNP Multilocus Haplotype Analysis

To maximise information provided by H1-SNP genotypes, a multilocus approach was taken next. SNPtagger (Section 2.3.6) was used to identify a minimum set of H1-SNPs required to capture >99% of the haplotypic diversity within the *CRHR1-MAPT* interval, thereby removing redundant SNP data and allowing a smaller dataset to be analysed without loss of power. htSNPs were sought using data from cases, controls, and cases & control combined. The htSNP set chosen included htSNPs identifed from all tests (Table 4.4.3a).

For these H1 htSNPs, multilocus haplotype frequencies were estimated using Arlequin (Section 2.3.5) and compared between cases and controls using CLUMP (Section 2.3.10). Seventeen different 'H1 sub-haplotypes' have frequencies >0.02 and account for 74% of the haplotypes present in both groups combined (Table 4.4.3b). Of note, there are six H1 sub-haplotypes in the PD case group that are not represented in the control group (XII-XVII), and seven in the control group not present in cases (V-XI). Comparison of multilocus haplotype counts between cases and controls was highly significant, $\chi^2=78.8$, $p<1.0e^{-6}$ (16 d.f.). Thus, the *MAPT* H1 haplotype term is a misnomer; H1 represents a clade of haplotypes on the same backbone (i.e. $H1^{I}$, $H1^{II}$, *etc.*), that are not H2.

Table 4.4.3a H1 htSNPs

htSNPs sufficient to capture >99% of the H1 haplotypic diversity (based on H1 SNP genotypes) within the *CRHR1-MAPT* interval are shown in blue.

Group	htSNPs
Case	1 2 3 4 5 6 7 8 9 10 11 12 13 14
Control	1 2 3 4 5 6 7 8 9 10 11 12 13 14
Case & Control	1 2 3 4 5 6 7 8 9 10 11 12 13 14
htSNP set used	**1 2 3 4** 5 6 **7 8 9 10 11** 12 13 **14**

Table 4.4.3b H1 Sub-Haplotype Frequencies in Cases and Controls

Haplotypes with frequencies >0.02 are shown.

H1 Sub-haplotype (Allelic conformation)	Control Frequency (n)	Case Frequency (n)
I (GGTACGCGT)	0.16 (26)	0.22 (35)
II (GGCGTGTGT)	0.09 (14)	0.08 (13)
III (GGTACGTGT)	0.06 (9)	0.04 (7)
IV (AGTACGCGT)	0.05 (8)	0.06 (10)
V (AACACATGC)	0.05 (8)	-
VI (AACGCTGCAT)	0.04 (7)	-
VII (GGCACATGC)	0.04 (6)	-
VIII (GGTATGCAC)	0.03 (5)	-
IX (GGCACGCGT)	0.03 (5)	-
X (GGTATGCAT)	0.03 (5)	-
XI (GGCGTGCGT)	0.02 (3)	-
XII (AACGTGTGT)	-	0.08 (12)
XIII (GGCACACGC)	-	0.04 (6)
XIV (AACACATGT)	-	0.03 (5)
XV (GGCATGCGT)	-	0.03 (4)
XVI (GGCACGTGT)	-	0.02 (4)
XVII (GGCATATAC)	-	0.02 (3)

4.4.4 Modelling H1 Evolutionary History

One or more H1 sub-haplotypes are associated with PD (Section 4.4.3), although these are >200 Kb in length and there is little indication of the possible location of risk variation. To narrow the candidate region, we attempted to perform cladistic analysis as proposed by Templeton and colleagues (Templeton *et al.* 1988). Cladistic analyses link haplotype data by evolutionary relationships, under the assumption that disease associated variation is embedded, and can be localised, within the framework of haplotype clades (Heng and Low 2000). Cladograms were constructed for H1 sub-haplotypes using the PHYLIP software suite (Section 2.3.7). More than 100 cladograms provided an equally good fit to the data, one of which is shown in Figure 4.4.4a. This is indicative of ancestral recombination between H1 sub-haplotypes (Templeton *et al.* 1988). Although this is an interesting finding and contrasts that observed for H1 and H2 haplotypes (Pastor *et al.* 2002), cladistic analyses may not be appropriate for regions with high levels of recombination (Templeton *et al.* 1988; Heng and Low 2000). A metric LD map was constructed using H1-SNP data for both cases and controls using LDMAP (Section 2.3.9) (Figure 4.4.4b). This is also indicative of recombination within the H1 *CRHR1-MAPT* interval.

Figure 4.4.4a Cladogram of H1 Sub-Haplotypes

One of a possible ~100 models of the evolutionary history of H1 sub-haplotypes. Haplotype names correspond to those in Table 4.4.3b (i.e. Hap1=I, Hap2=II, *etc.*)

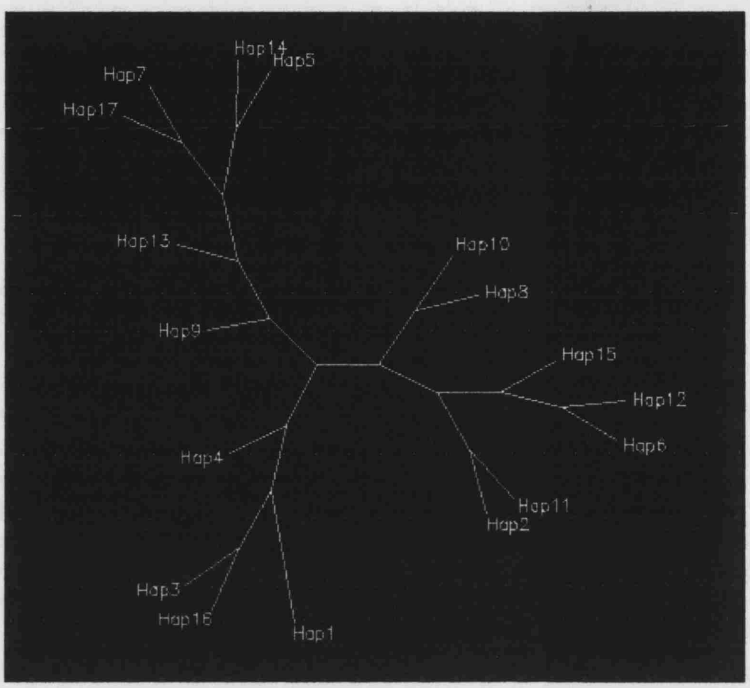
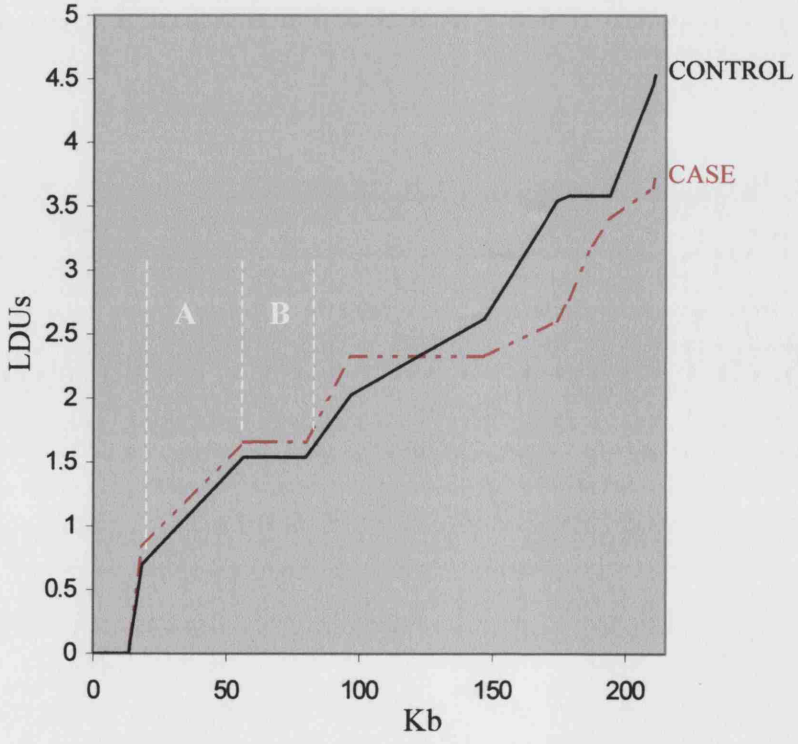


Figure 4.4.4b Metric LD Map of the H1 *CRHR1-MAPT* Interval

Regions of low LD (recombination) appear as sloping 'steps' (A). Regions of high LD appear as horizontal 'plateaus' (B).



4.4.5 Sliding 3-Marker Haplotypes

To refine the genomic region within *CRHR1-MAPT* interval contributing most to disease association, 3-marker haplotypes consisting of sliding trios of markers (1-2-3; 2-3-4; *etc.*) for all H1-SNPs were estimated using Arlequin (Section 2.3.5). As haplotype frequencies are dependent upon the frequencies of SNPs that comprise them, we calculated the theoretical minimum and maximum possible frequencies of each of the 3-marker haplotypes in the control group. The expected maximum frequency of a haplotype is P_{ABC} where ABC is the frequency of major alleles at loci A, B and C. The mimimum expected frequency is P_{abc}, where abc is the frequency of the minor alleles at each loci (Figure 4.4.5a). The results suggested that H1 3-marker haplotypes would be potentially informative and most are not appreciably restricted by SNP frequencies.

χ^2 analysis of 3-marker haplotype counts in cases and controls was performed using CLUMP (Section 2.3.10). The number of haplotypes containing markers 8-9-10 were strikingly different between cases and controls ($-\text{Log}_{10}$ $p<3.23$, $p<0.0006$; Figure 4.4.5b).

Figure 4.4.5a Theoretical Minimum and Maximum 3-Marker Haplotype Frequencies

The expected maximum frequency of a haplotype is P_{ABC} where ABC is the frequency of major alleles at loci A, B and C. The mimimum expected frequency is P_{abc}, where abc is the frequency of the minor alleles at each locus.

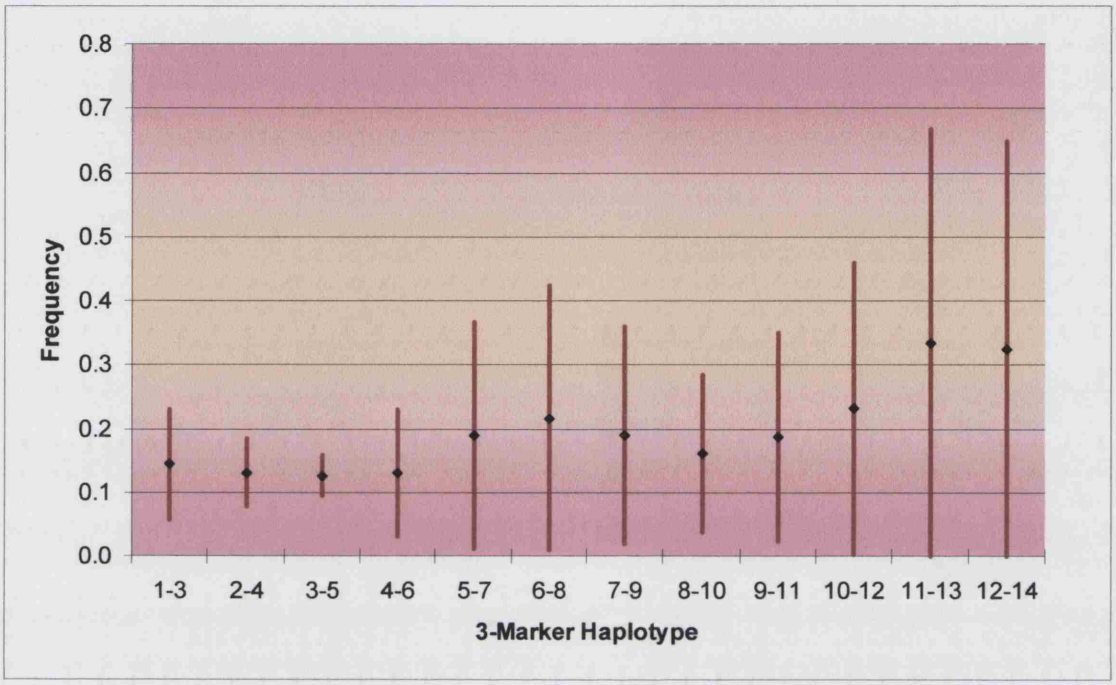
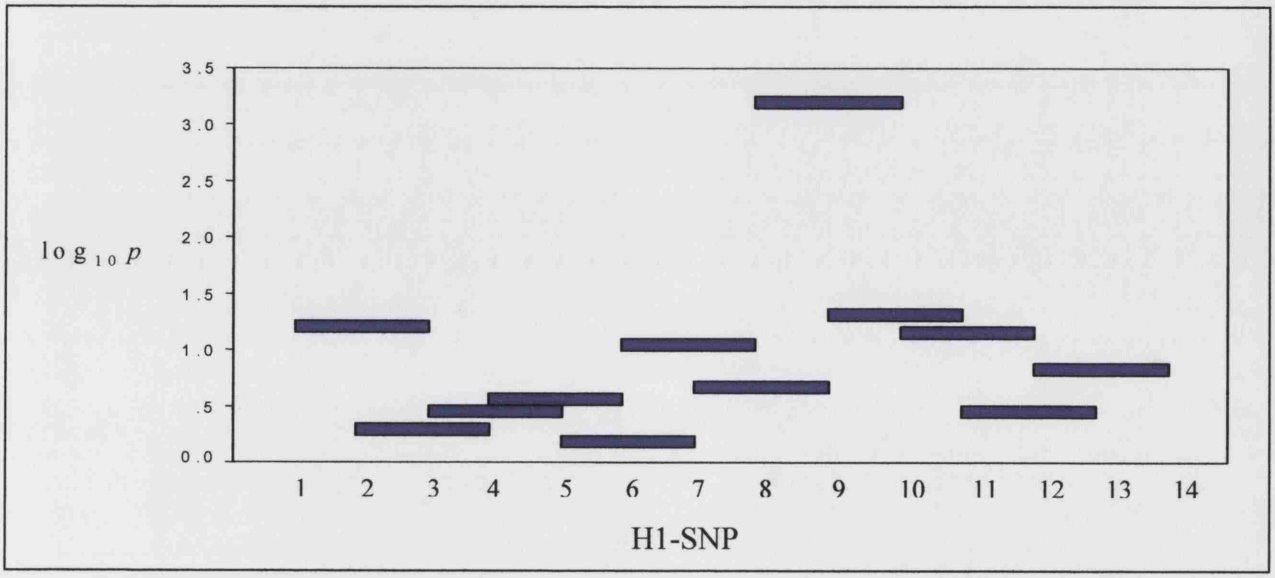


Figure 4.4.5b H1-SNP Haplotype Trios in Cases and Controls

$-\text{Log}_{10}$ p values (from χ^2 test) are shown for haplotype trio counts compared between cases and controls. The number of haplotypes containing markers 8-9-10 were strikingly different ($-\text{Log}_{10}$ p<3.23, p<0.0006).



Closer inspection of all pairwise combinations showed that the 'A-A' haplotype for markers 8 (rs242562) and 10 (rs2435207) is most significantly associated with disease ($p<0.003$) and cases carry twice as many as controls (Table 4.4.5c). The magnitude of this difference (15%) accounts for the increase in H1/H1 genotypes in cases compared with controls (13%; Table 4.2.2b). This H1 sub-haplotype is denoted '$H1^{PD}$'.

In view of the strong association of H1 with the 4R tauopathy, PSP (Section 4.1.2), frequencies of $H1^{PD}$ were also assessed in this disorder. Eighty-six case samples (median age=70 years, range 42-85) and 54 controls (median age=75 years, range 46-91) of North American extraction were available for analysis. SNP 8-10 haplotype frequencies differed between cases and controls (χ^2 $p<0.03$), although in contrast to PD, the 'A-G' haplotype was over-represented in cases (13%; Table 4.4.5d). This suggests that H1 PSP susceptibility may be dissimilar to that associated with PD risk.

Table 4.4.5c SNP 8-10 Haplotype Frequencies

The 'A-A' haplotype formed by H1-SNPs 8 (rs242562) and 10 (rs2435207) is over-represented in cases by 15%.

CONTROL				CASE			
Haplotype	Frequency	Counts	Allelic Conformation (SNP 8 - SNP 10)	Haplotype	Frequency	Counts	Allelic Conformation (SNP 8 - SNP 10)
1	0.38	61	G-G	1	0.39	63	G-G
2	0.26	42	A-G	2	0.13	22	A-G
3	0.22	36	G-A	3	0.19	31	G-A
4	0.14	23	A-A	4	0.29	46	A-A
	Total	162			Total	162	

Table 4.4.5d SNP 8-10 Haplotype Frequencies in PSP

The 'A-G' haplotype is over-represented in cases by 13%.

CONTROL				CASE			
Haplotype	Frequency	Counts	Allelic Conformation (SNP 8 - SNP 10)	Haplotype	Frequency	Counts	Allelic Conformation (SNP 8 - SNP 10)
1	0.47	50	G-G	1	0.30	51	G-G
2	0.17	19	A-G	2	0.30	52	A-G
3	0.12	13	G-A	3	0.15	26	G-A
4	0.24	26	A-A	4	0.25	43	A-A
	Total	108			Total	172	

4.4.6 Investigation of $H1^{PD}$

The $H1^{PD}$ haplotype is formed by H1-SNPs 8 (rs242562) and 10 (rs2435207). These span a ~32 Kb interval that includes the alternatively spliced *MAPT* exons 2 and 3. To further investigate $H1^{PD}$, individuals unambiguously carrying this haplotype (homozygous for the 'A' allele at SNPs 8 and 10) were identified for further analyses. They included one normal control, 4 possible PD cases and 1 probable PD case. We hypothesised that this haplotype may harbour sequence risk variation in the vicinity of SNPs 8 and 10. However, since LD mapping cannot be used to bound an interval associated with disease (in contrast to low resolution linkage mapping in families, where obligate recombinants define a candidate region), we also considered the region extending to SNPs 7 and 11 for analysis.

To identify regions potentially containing functionally important variation, we firstly used mVISTA (Section 2.3.2) to identify sequence conservation between human (*MAPT*) and mouse (*mapt*) genes (Figure 4.4.6a). High conservation (>75%) was apparent around exons and also in some intronic regions, including intron –1 (position ~1Kb), intron 0 (46.5-49.5Kb and ~59Kb) and intron 2 (78-79kb). These regions, exons 2 and 3 (and flanking intronic sequences) and ~1Kb 5' of exon –1 were sequenced in $H1^{PD}$ homozygotes (Table 4.4.6b). Sequencing primers/assays are detailed in Appendix 1.

No gross sequence anomalies were identified, although a further 5 SNPs were identified (Figure 4.4.6c).

Figure 4.4.6a Sequence Conservation between Human and Mouse Tau Genes

Percentage of sequence conservation (peaks) over a sliding 50 base pair window is indicated on the Y-axis, calculated using mVISTA. Positions of human *MAPT* exons (blue bars), regions sequenced in $H1^{PD}$ homozygous individuals (grey bars) and SNPs 7-11 (arrows) are shown.

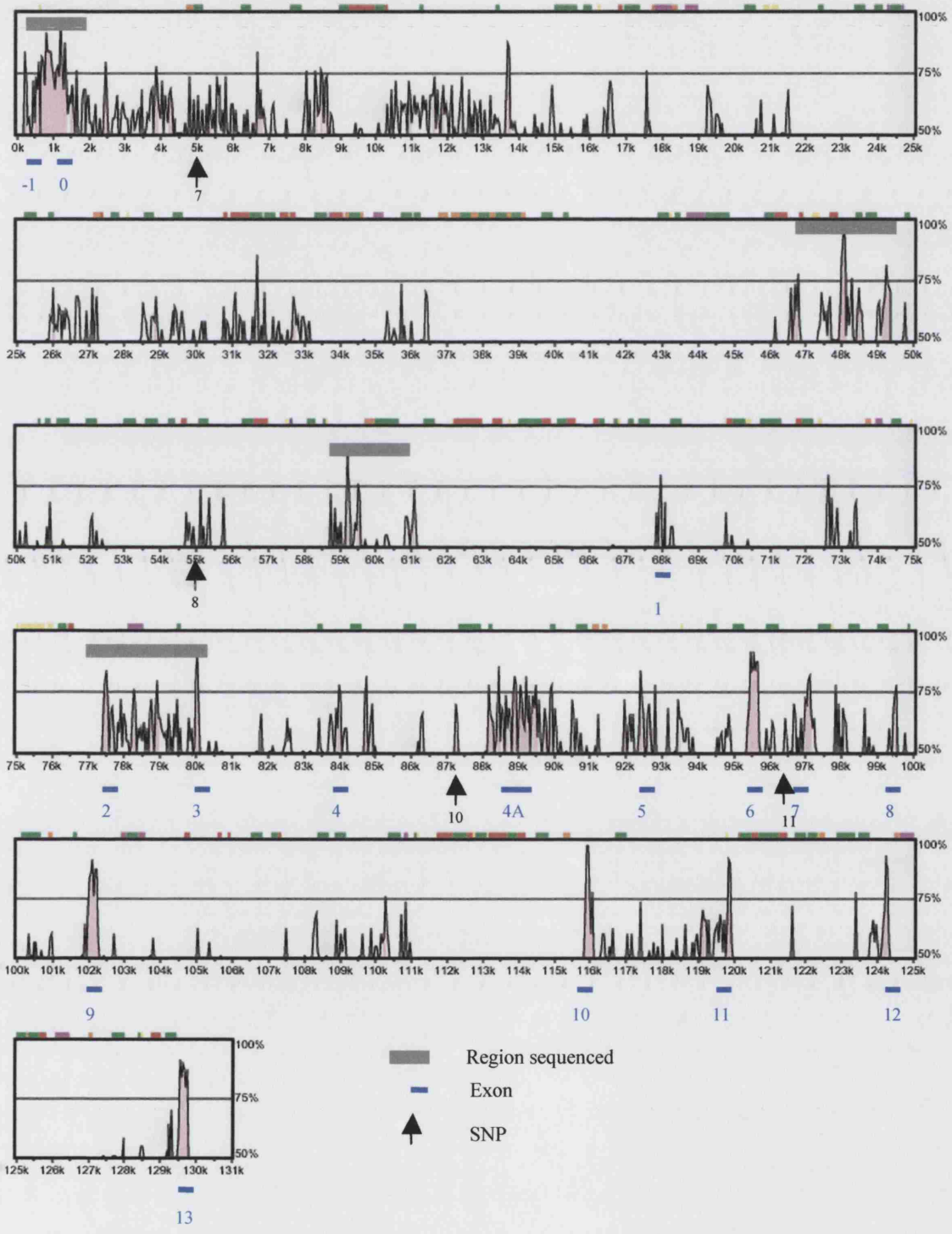


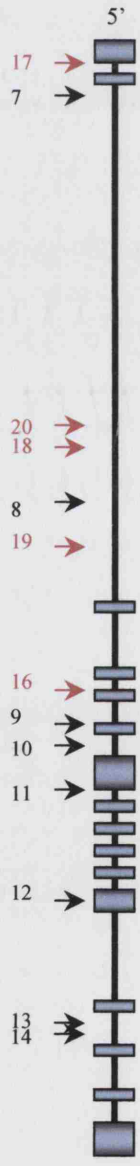
Table 4.4.6b $H1^{PD}$ Homozygotes

Sample	Age of Onset	Age	Diagnosis
K118		84	Normal Control
P263	40	46	Possible PD
P250	47	49	Possible PD
P204	70	74	Possible PD
P196	73	75	Probable PD
P057	75	88	Possible PD

Figure 4.4.6c SNPs Identified in $H1^{PD}$ Homozygotes

SNP	CDS ID	NCBI ID	Change
17		rs3744456	G>C
20		rs2435205	G>A
18	hCV1016016	rs242557	G>A
19		rs878917	C>T
16		rs10514889	A>G

Relative positions of SNPs within the *MAPT* gene (in addition to SNPs 1-14) are shown in red. *MAPT* exons are shown as blue boxes.



4.4.7 SNPs 16-20

To assess association between variants found on $H1^{PD}$ haplotypes and disease, SNPs 16-20 were genotyped in the epidemiologically matched case-control H1/H1 series (n=81+81). Assay details are shown in Appendix 1. Genotype and allele frequencies were compared between cases and controls (χ^2 test) using CLUMP (Section 2.3.10) and ORs were calculated using SPSS (Section 2.3.11). Genotype frequencies and p values for tests are shown in Table 4.4.7a. A marginally significant difference was observed for SNP 20 when comparing genotype frequencies (p=0.03). Comparisons of sliding haplotype frequencies for trios of markers (as Section 4.4.5), were also compared between cases and controls (Table 4.4.7b); none were preferentially associated with disease. SNP 20 was also genotyped in aditional H1/H1 homozygotes (case n=146, control n=201) and allele/genotype frequencies were compared between groups (case n=227, control n=282); no significant differences were found (p=0.98; p=0.56). In the absence of any remarkable sequence changes or SNPs that were associated with disease risk, we further examined the genotypic make-up of $H1^{PD}$.

Table 4.4.7a Case Control Analysis of SNPs 16 to 20

Alleles are recoded: 1=major allele, 2=minor allele (alleles are detailed in Figure 4.4.6c).

		Genotype frequency (n)						Chi-sq p		11&12 vs 22			11 vs 12&22		
SNP	Group	11		12		22		Genotypes	Alleles	OR	95% CI	p	OR	95% CI	p
17	Control	0.61	(49)	0.35	(28)	0.04	(3)	0.84	0.82	1.5	0.3-9.5	0.64	1.1	0.6-2.0	0.80
	Case	0.59	(48)	0.38	(31)	0.02	(2)								
20	Control	0.17	(13)	0.64	(49)	0.18	(14)	0.03	0.13	0.4	0.19-0.83	0.01	1.0	0.4-2.4	0.97
	Case	0.17	(13)	0.47	(36)	0.36	(28)								
18	Control	0.32	(6)	0.53	(43)	0.15	(12)	0.44	0.82	1.4	0.6-3.5	0.48	1.4	0.7-2.7	0.39
	Case	0.26	(21)	0.63	(51)	0.11	(9)								
19	Control	0.86	(70)	0.14	(11)	0.00	(0)	0.97	0.82		NA		0.9	0.4-2.2	0.82
	Case	0.88	(71)	0.12	(10)	0.00	(0)								
16	Control	0.26	(20)	0.50	(39)	0.24	(19)	0.48	0.24	0.7	0.4-1.5	0.41	1.5	0.7-3.3	0.28
	Case	0.18	(14)	0.51	(39)	0.30	(23)								

NA: not applicable (no 22 genotypes)

Table 4.4.7b SNP 17 to 10 Haplotype Trio Frequency Comparisons

Frequencies of haplotypes formed by trios of SNPs sliding across markers 17 to 10 were compared (χ^2 test) between cases and controls (n=81+81). See Figure 4.4.6c for SNP positions within *MAPT*.

Haplotype	p
17-7-20	0.56
7-20-18	0.13
20-18-8	0.47
18-8-16	0.27
8-16-9	0.22
16-9-10	0.08

4.4.8 H1-SNP Genotypes

H1-SNP genotypes within *MAPT* were further examined in unambiguous carriers of $H1^{PD}$. P024 and P250 were homozygous for the same alleles at all markers within *MAPT* (~210Kb). Other case individuals were homozygous for shorter regions, however all shared genotypes for SNPs 20 to 16, spanning a distance of ~33 Kb from the 3' half of intron 0 to intron 2, including exons 1 and 2 (Table 4.4.8a). Genotyping assays that were used in this study detected specific variants, *i.e.* if only one allele was identified then homozygosity for that allele was assumed and heterozygous deletions would have been overlooked. Interestingly, the region covered by SNPs 20 to 16 overlaps the region of high LD in cases as measured by LDUs within H1 *CRHR1-MAPT* (100-150Kb, Figure 4.4.4b). In addition, this is the only region where the case LDU profile is distinctly different to controls. We hypothesised that some cases may carry a heterozygous genomic deletion spanning exons 1 and 2 of *MAPT*; this would be consistent with apparent homozygosity for markers and the paucity of recombination (high LD) in the region, although this remains to be formally tested. Two other controls who were asymptomatic at 92 and 89 years (K047 and K009) also share genotypes at SNP 20 to 16, although these individuals carry only one copy of $H1^{PD}$ (they are heterozygous at SNP 10).

Table 4.4.8a Shared Homozygosity for *MAPT* H1-SNPs

Unambiguous carriers of $H1^{PD}$ (blue) are homozygous at numerous H1-SNPs across *MAPT* (yellow). All share identical homozygous genotypes for markers 20 to 16 (boxed). Two other normal control individuals (K047 and K009) also share these genotypes, although they carry only 1 copy of $H1^{PD}$.

Sample	Diagnosis	SNP 17	7	20	18	8	19	16	9	10	11	12	13	14
K118	Con	11	22	12	22	22	11	22	11	22	11	11	11	11
P057	Pos	12	12	11	22	22	11	22	12	22	11	11	12	11
P196	Pro	11	12	11	22	22	11	22	11	22	11	11	11	11
P204	Pos	11	22	11	22	22	11	22	11	22	11	11	11	11
P250	Pos	11	22	11	22	22	11	22	11	22	11	11	11	11
P263	Pos	11	12	11	22	22	11	22	12	22	11	12	11	12
K047	Con	11	12	11	22	22	11	22	11	12	12	11	11	12
K009	Con	11	12	11	22	22	11	22	12	12	11	11	11	12

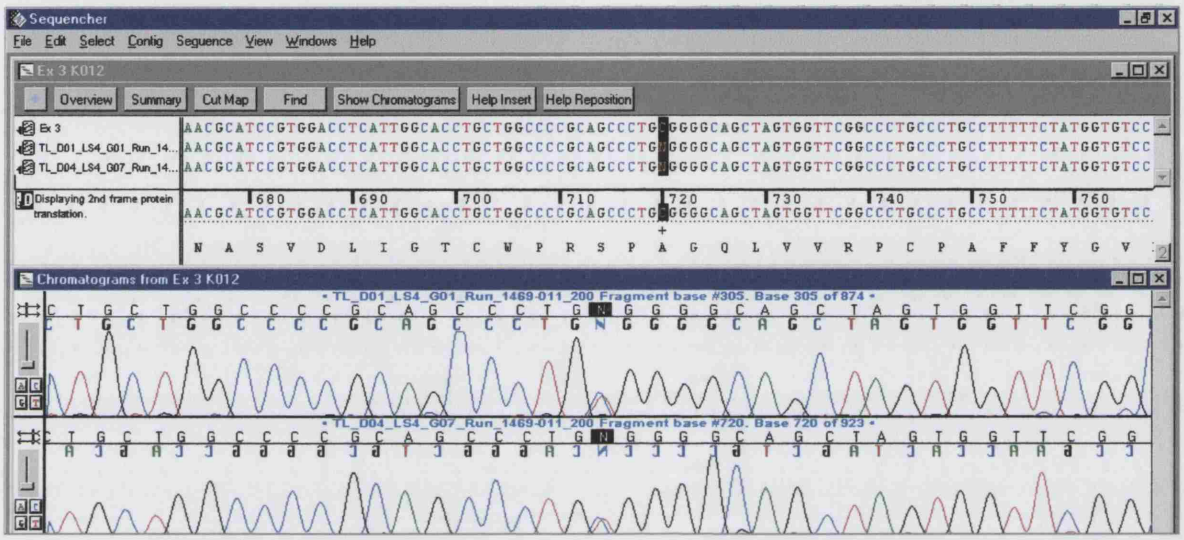
4.5 *CRHR1* Analysis

4.5.1 Sequencing *CRHR1*

Genomic sequencing of *CRHR1* was undertaken in 6 normal controls to identify SNPs useful for i) mapping the extent of the *MAPT* extended haplotype (Section 4.3) and ii) investigating the H1 haplotype (Section 4.4). Eleven out of the 14 exons (exons 2-4, 6-10, 12-14), including those alternatively spliced (exons 3, 6 and 13) and their flanking intronic sequences were successfully sequenced (Section 2.2.7). Primers and assay details are shown in Appendix 1. A total of 33 SNPs (intronic or coding synonomous) and a 5 bp indel polymorphism were identified (intron 2). Thirty-two of these were in LD with the extended *MAPT* haplotype, including the 5 bp indel (H2 haplotype carries the deletion). Of interest was an apparent H1-specific, previously unreported coding non-synonomous single base-pair change, an exon 3 C>T transition resulting in the putative substitution of alanine with valine at residue 60 of the CRHR1 protein (individual K012), termed 'A60V' (submitted to NCBI SNP database; ss26161961). The sequence chromatogram is shown in Figure 4.5.1a.

Figure 4.5.1a Sequence Chromatogram of *CRHR1* A60V

The upper panel shows the sequence based on the chromatograms in the lower panel (sense and antisense strand sequencing). The heterozygous C>T change (resulting in a putative A60V substitution) in K012 is highlighted.



4.5.2 Case-Control Analysis of *CRHR1* A60V

A60V was assumed to be H1-specific, since it was found in the heterozygous state in an H1/H1 homozygous individual and *CRHR1* was also known to lie within the *MAPT* non-recombining region (Section 4.3). We therefore tested the hypothesis that this variant may be pathogenic and responsible for the association of H1 and PD.

The epidemiologically-matched H1/H1 homozygote case control series (n=81+81) was genotyped for A60V (assay details are shown in Appendix 1); frequencies are shown in Table 4.5.2a. The frequency of A60V was low in both cases and controls and is likely to represent a rare, inocuous polymorphism within the Norwegian population.

Table 4.5.2a Case Control Analysis of *CRHR1* A60V

A60V is rare in both cases and controls.

	Genotype frequency (n)						Allele frequency (n)			
Group	CC		CT		TT		C		T	
Control	0.95	(77)	0.05	(4)	0	(0)	0.98	(158)	0.02	(4)
Case	0.96	(78)	0.04	(3)	0	(0)	0.98	(159)	0.04	(3)

4.5.3 Linkage Disequilibrium within *CRHR1*

Sequence analysis of the Saitohin gene in primates has suggested that H2 is the ancestral variant, and that the higher frequency of H1 (compared to H2) in humans is due to positive selection for that allele (Conrad *et al.* 2004). An important characteristic of positive selection is that it causes an unusually rapid increase in allele frequencies during a short enough time interval that recombination does not have a chance to destroy LD (Sabeti *et al.* 2002). The *CRHR1* gene is a candidate for positive selection due to its role in immunity (Grammatopoulos and Chrousos 2002). We tested for evidence of selection at *CRHR1* by calculating LD between high frequency H1-SNPs 1, 2 and 3 (Figure 4.4.2b), within the epidemiologically-matched H1/H1 case-control series (n=81+81). Pairwise measures of LD were calculated using GOLD (Section 2.3.8) and are shown in Table 4.5.3a. All SNPs exist at high frequencies in both cases and controls (0.36-0.42) and pairwise LD is high for all SNP combinations. This is in contrast to the LD between similarly frequent H1-SNPs 8 and 10 within *MAPT* in controls (D'=-0.01, p=0.91). These data firstly suggest that the H1 *CRHR1* locus may be subject to selective forces and secondly, since there is no appreciable diffference between cases and controls, is further support that this locus is not responsible for the association of H1 and PD.

Table 4.5.3a Pairwise LD between High Frequency H1-SNPs within *CRHR1*

CONTROL SNPs		D	D'	p	r	SNP	freq
1	2	0.23	1.00	1.1E-32	0.97	**1**	0.36
1	3	-0.14	-0.75	1.2E-14	0.63	**2**	0.40
2	3	-0.14	-0.75	1.2E-14	0.63	**3**	0.40
CASE SNPs		**D**	**D'**	**p**	**r**	**SNP**	**freq**
1	2	0.22	0.97	2.2E-28	0.93	**1**	0.43
1	3	-0.10	-0.70	9.3E-09	0.48	**2**	0.40
2	3	-0.12	-0.77	3.3E-11	0.56	**3**	0.42

4.6 Discussion

This study has undertaken a detailed analysis of the *MAPT* H1 haplotype that is associated with a variety of neurodegenerative disorders, including PD. Firstly, we have replicated and extended the association with clinically defined PD in the Norwegian population as previously reported (Farrer *et al.* 2002). The present study employed a large case-control series (case n=296, control n=441) from a genetically homogeneous population within a restricted geographical region. The association between H1/H1 homozygotes and disease was highly significant, even after adjustment for possible misdiagnosis of PSP as PD (OR=1.9, 95% CI=1.3-2.6, p=0.0004). To do this, we randomly removed 6% of cases, as has been previously suggested (Golbe *et al.* 2001) as there is no evidence to infer that the prevalence of PSP in Norway is appreciably different to that in other Caucasian populations. In addition, when considering only cases fitting the strictest diagnostic criteria ('probable' PD), the association remains significant (OR=1.8, 95% CI=1.2-2.5, p=0.002; Section 4.2) and is the strongest reported association to date. This is evidence of the utility of a population isolate in identifying susceptibility in complex diseases.

We showed that the *MAPT* extended haplotype extends at least as far as, and includes the *CRHR1* locus (Section 4.3). By genotyping H1-SNPs within the *CRHR1-MAPT* interval, we explored the possibility that variation in a neighbouring gene, within the extended haplotype, may be responsible for disease association. However, the H1 sub-haplotype associated with disease ($H1^{PD}$) mapped to a genomic interval of <90 Kb (SNP 7 [rs11079727] through SNPs 11 [rs2258689]), containing *MAPT* exons 1-4 (Section 4.4.4). Aberrant splicing of *MAPT* exon 10 and its subsequent

inclusion/translation is responsible for the over-production of 4R tau (and associated NFTs) in many familial cases of FTDP-17 (Hutton 2001). Conversely, tau isoforms without exon 2 and 3 have less propensity for aggregation *in vitro* (King *et al.* 2000). We hypothesised that $H1^{PD}$ harboured variability which influences exon 2-3 splicing. Although variation was identified through sequencing evolutionarily conserved regions in and around exon 2 and 3, intron 0 and the promoter, in a limited number of samples, none were associated with disease, either in isolation (Table 4.4.7a) or when considered in 3-marker haplotypes (Table 4.4.7b).

There still remain, however, further $H1^{PD}$ haplotypes that have yet to be analysed. Haplotype frequencies were estimated by a statistical algorithm (Section 2.3.5) and therefore not determined directly for individuals. Hence, individuals carrying $H1^{PD}$ were identified on the basis of homozygosity for allele '2' at both SNPs 8 and 10. In theory, individuals that are heterozygous for SNPs 8 and 10, or heterozygous at one marker and homozygous at the other, could carry $H1^{PD}$. There are 41 cases that fit this criteria. Empirical determine of haplotypes, using long range PCR and intramolecular ligation (McDonald *et al.* 2002) would identify further $H1^{PD}$ chromosomes for sequence analysis.

The finding that all unambiguous patient carriers of $H1^{PD}$ share homozygosity for SNPs 20 to 16 (Section 4.4.8) may be indicative of a genomic deletion spanning exons 1 and 2. This particular allelic conformation does not occur in any of the remaining 81 H1/H1 cases and in only two controls. However, in these control individuals genotypes at SNP 10 are heterozygous, suggesting they carry only one copy of $H1^{PD}$. This would be consistent with a sub-population of $H1^{PD}$ haplotypes carrying a deletion: homozygotes could carry both a normal and a deleted copy and the controls would be expected to carry the normal copy. The hypothesis could be

tested using *MAPT* exon dosage analysis, similar to that used to detect dosage alterations within *PRKN* (Section 2.2.9). Deletion of exons 1 and 2 could theoretically result in the production of mRNA lacking exons 1, 2 and 3, since inclusion of exon 3 is dependent upon the presence of exon 2 (Andreadis *et al.* 1995). Translation may then be initiated at the next methionine site (exon 5 +6), resulting in a mutant peptide lacking the projection domain, which could be subject to cellular degradation mechanisms. This would result in tau haploisufficiency, although this may not be expected to have appreciable consequences as *MAPT* knockout mice appear phenotypically normal, perhaps because of upregulation and compensation by other MAPs. However, microtubule organisation is altered in small calibre axons in these animals (Harada *et al.* 1994). Alternatively, a heterozygous deletion of exon 2 only would be predicted to result in an increase of 0N isoforms and a reduction in 1N and 2N isoforms. The projection domain, composed of the amino terminal repeats, is thought to play roles in microtubule spacing (Harada *et al.* 1994) and interact with motor proteins (Seitz *et al.* 2002). An attractive mechanism to account for the selective cell loss in PD is that distorted projection domain size ratios can directly (through interaction with motor proteins) or indirectly (through alteration of microtubule organisation) impair dopamine vesicle trafficking, leading to a potentially toxic build-up within the cell. During the normal aging process, neurons shrink and hence cellular spatial dymanics are altered (Terry 1995). A change in microtubule organisation due to dysfunctional tau may be envisaged to have similar consequences as aging in the brain, which includes both cell loss (Turlejski and Djavadian 2002) and inclusion formation (Peters 2002).

Although population genetics exploits ancestral recombination events and may facilitate high resolution mapping of disease loci, LD cannot be used to bound an

interval associated with disease (this is in contrast to lower resolution linkage mapping, within families, in which obligate recombinants are used to define a candidate region). Therefore, variability within H1 in regions outside the SNP 8-10 interval may be important. Overexpression of *MAPT* is sufficient to cause neurodegeneration without NFT formation in the fly (Wittmann *et al.* 2001), hence both *MAPT* gene splicing and expression may be important. A useful analogy are the genetics of amyloid precursor protein (*APP*); mutations in the gene may lead to over-expression (Prasher *et al.* 1998) or directly affect protease cleavage and the production of amyloidogenic Aβ40/42 peptides, either mechanism may predispose to Alzheimer's disease (Sambamurti *et al.* 2002).

Studies have suggested (de Silva *et al.* 2001) and demonstrated (Kwok *et al.* 2004) that the H1 haplotype may effect tau expression levels. However this does not account for the increased frequency of H1 over H2 in normal individuals, although the presence of other susceptibility factors in patients may influence the extent of the effect. For example, subtle variation in the *PRKN* (West *et al.* 2002b) and *SNCA* promoters, whose proteins are both linked to tau (Jensen *et al.* 1999; Shimura *et al.* 1999) may have additive or multiplicative effects in conjuction with the high expressing *MAPT* H1. This is not inconsistent with that envisaged for a complex disorder. It would therefore be useful to compare 'multiple risk factor genotypes' with tau expression and pathology in PD and control brain.

It is interesting that a haplotype composed of SNPs 8 and 10 is also over-represented in PSP cases, compared with controls albeit with different allelic conformation (Section 4.4.5). This may indicate that the variability associated with PD is different in PSP and may be expected, as tau aggregates are not typically a feature of PD, in contrast to PSP. However, overlapping pathologies in some cases has led to

speculation that H1 may be associated with a pathological sub-phenotype of PD (Kwok *et al.* 2004). 4 out of the 5 case $H1^{PD}$ homozygotes were diagnosed as 'possible' PD cases, and therefore showed no or minimal response to *l*-DOPA. This is also a feature of the parkinsonism seen in both FTDP-17 and PSP, therefore it is possible that common H1 variability may underlie susceptibility in a subset of tauopathies and α-synucleinopathies. It also could indicate, however, that these possible cases may be actually misdiagnosed PSP cases. Therefore, complete H1 SNP analysis is warranted in PSP cases with suitably matched controls. It would also be useful to include PSP brain in 0N/1N/2N ratio comparison studies.

It is recognised that only half the genomic region within H1 was investigated in this study. Attempts were made to precisely map the extent of H1, allowing a more complete analysis, but this was hindered by inconsistencies in the chromosome 17q21 physical map. During the timecourse of this study, both NCBI and CDS physical maps have been revised on more than one occasion. In addition to SNP, STRP and gene relocations (e.g. at the time of Pastor and colleagues' publication, *MAP3K1A* was located 3'of *MAPT,* but is now located 5'), duplicate locations had been reported for some markers, with a placement both 5' and 3' of *MAPT.*

The *saihtohin* gene, nested within intron 9 of the MAPT gene is an interesting candidate, although mutations have not been identified in some PD cases (Levecque *et al.* 2004). Sequence analysis of *saihtohin* orthologues in non-human primates show that polymorphisms are consistent with the human H2 haplotype, suggesting that H2 is the ancestral allele (Conrad *et al.* 2004). This raises the question of why H1 is now the predominant allele in human populations. Selection for a beneficial variant may account for this (Schlotterer 2003). Indeed, we found strong LD between high frequency SNPs in both case and control H1 chromosomes, suggesting the region has

undergone recent selective mechanisms (Table 4.5.3a). Analysis of extended haplotype homozygosity (Sabeti *et al.* 2002) would be useful to determine this.

The apparent lack of recombination between H1 and H2 is also an interesting phenomenon; it lends the hypothesis that recombination is suppressed between the two variants, or that recombinant haplotypes are selected against (Baker *et al.* 1999). Chromosome 17q21 is an ancestral breakpoint marking a pericentric inversion by which human chromosome 17 and the chimp orthologue (chromosome 19) differ (Kehrer-Sawatzki *et al.* 2002). It is therefore possible that H1/H2 is an inversion polymorphism and samples carrying both H1 and H2 haplotypes could account for inconsistencies in the chr 17q21 physical map.

Although the inversion hypothesis remains to be formally proven, physical organisation of H1 and H2 haplotypes could be examined by pulse-field macrorestriction mapping or flourescence *in situ* hybridisation. Chromosomal inversions are known to exert a strong suppressing effect upon recombination (Hey 2003). Interestingly, Lui and colleagues also reported evidence of effective suppression of recombination at chromosome 17q21 in a region around the breast cancer susceptibility gene *BRCA1* (Liu and Barker 1999). In this study, a series of biallelic polymorphisms were shown to define two common haplotypes with frequencies of 0.66 and 0.34, not dissimilar to those observed for *MAPT* H1 and H2. *BRCA1* is located ~2.7 Mb 5' of *MAPT.* It would be tempting to speculate that the extended haplotype extends to and includes the *BRCA1* locus. However, Pittman and colleagues (2004) have recently reported that the 5' boundary does not exceed *MAP3K14* (Pittman *et al.* 2004).

Despite the paucity of recombination between H1 and H2, we showed that the *MAPT* H1 haplotype term is a misnomer; H1 truly represents a clade of haplotypes on the same H1 backbone (e.g. $H1^{I}$, $H1^{II}$, etc.), *that are not H2* (Section 4.4) and it was this finding that allowed us to refine the region associated with PD to a ~90 kb interval at the 5' end of *MAPT*.

4.7 Summary and Conclusions

The *MAPT* gene encodes a variety of microtubule associated tau protein isorforms in the CNS. The common *MAPT* H1 variant is associated with a number of 4R tauopthies and PD. This variant also appears not to recombine with the other major variant in human populations, H2. Using H1-specific single nucleotide polymorphisms (H1-SNPs) we demonstrated *MAPT* H1 is a misnomer, and actually consists of a family of recombining H1 alleles. Population genetics, linkage disequilibrium and association analyses have shown that one *MAPT* H1 sub-haplotype is preferentially associated with Parkinson's disease. Using a sliding scale of *MAPT* H1-specific haplotypes in age/gender matched PD cases and controls, we refined the disease association to within a ~90kb interval in 5' *MAPT*, encompassing exons 1-4 and speculate that aberrent exon 2-3 splicing may account for the association of H1 in PD. We also suggest that the *CRHR1* locus has undergone positive selection and may account for the increase in H1 in human populations, given that H2 is likely the ancestral allele.

5 Linkage Disequilibrium Mapping and Candidate Gene Analysis at PARK10

5.1 Introduction

5.1.1 Late Onset Parkinson's Disease is Linked to 1p32

A recent genome wide scan of 117 Icelandic PD patients and 168 of their unaffected relatives, within 51 families, revealed significant linkage to chromosome 1p32 (*PARK10*) (Hicks *et al.* 2002). A LOD score of 3.9 was initially obtained; additional marker typings in the region increased the LOD score to 4.9. The peak is centred close to D1S231, with markers D1S2874 (telomeric) and D1S475 (centromeric) defining a region with a drop of 1 LOD score from the peak (peak LOD –1 region). Using affected-only, allele sharing methods under an additive model, carriers of one or two 'risk' alleles have 29 and 60 times the risk, respectively, of acquiring PD compared to non-carriers (Hicks *et al.* 2002).

Li and colleagues (Li *et al.* 2002) also reported linkage to 1p32 in US/Australian PD series. In this study, a locus affecting age of disease onset gave a peak LOD score of 3.4 within a 6cM interval, flanked by D1S2134 and D1S200 (Li *et al.* 2002). The US Caucasian gene pool is also predominantly of Northern European extraction, and there was a considerable amount of Norwegian migration to the USA at the turn of the 20th Century (Maraganore *et al.* 2001).

Linked regions from both studies directly overlap and involve typical late onset PD (mean age of onset in US/Australia=60.1 years, Iceland=65.8 years), genetically spanning ~6 cM (a physical distance of ~8 Mb) and contains multiple genes (Hicks *et al.* 2002; Li *et al.* 2002).

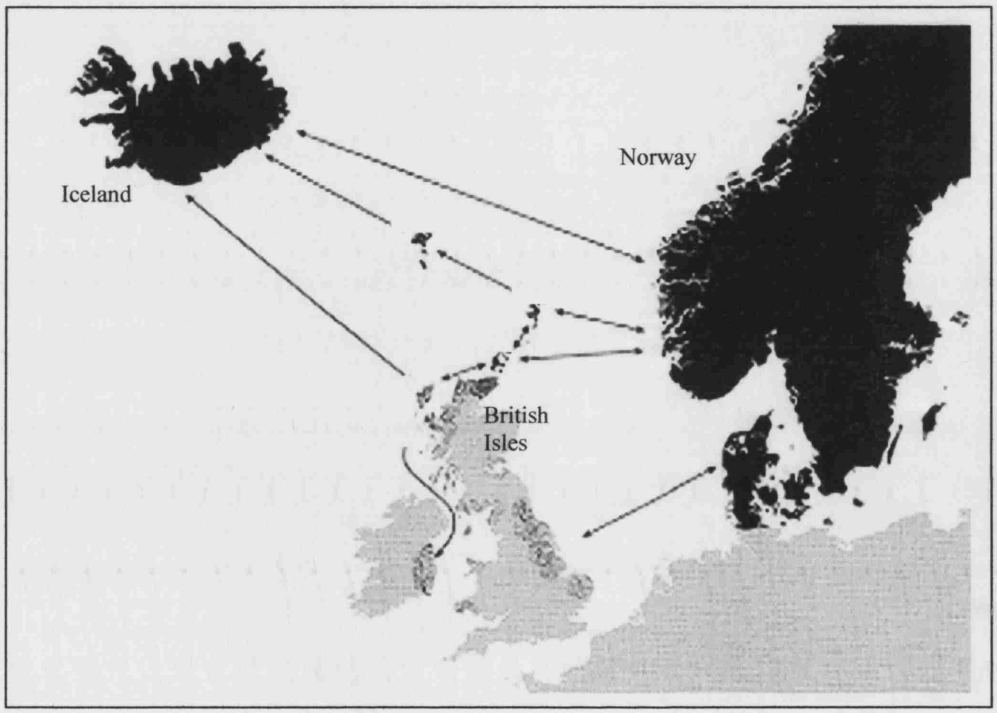
5.1.2 Population History of the Norse Empire

During what is commonly referred to as the Viking Era in European history (8th century through the early 11th century), the Norse peoples occupied large regions of Northern Europe. From their origins in Scandinavia, they expanded in all directions during this time (Helgason *et al.* 2001). The blackened areas of Figure 5.1.2a show the main sailing routes in the North Atlantic between regions of Norse cultural and linguistic dominance. Historical sources indicate that the uninhabited northern Atlantic island of Iceland was one of the regions colonised by the Norse Vikings (warriors), occuring between 870 and 930 AD (Helgason *et al.* 2000).

Genetic analysis in the present day Icelandic population, using sex-specific markers (on the Y chromosome and mtDNA) indicates that 75-80% of Iceland's founder male population (Helgason *et al.* 2000) and ~38% of the females (Helgason *et al.* 2001) are consistent with Norse ancestry and most similar to present day Norwegians.

Figure 5.1.2a Main Sailing Routes Between Regions of Norse Dominance During the Viking Era

The main sailing routes (arrowed) between regions of Norse cultural and linguistic dominance (filled in black) from the 8th century through the early 11th century. Adapted from (Helgason *et al.* 2001).



5.1.3 Multipoint LD Mapping

Disease gene identification through linkage analyses in families and linkage disequilibrium mapping in populations both rely on the same fundamental assumption, that is the co-inheritance of a haplotype in which variants are associated (i.e. in LD) with the disease mutation. Meiotic recombination is one of the key forces which erodes LD over generations (Section 1.2). Linkage analysis focuses upon recent ancestry (within a few generations), in which there have been few opportunities for recombination, and as such regions associated with disease will tend to be larger than those expected at the population level, as historically many more recombination events have occurred, narrowing the size of the disease associated haplotype (reviewed by Cardon and Bell 2001).

When a disease mutation arises, it does so on the background of the surrounding haplotype. The allelic make-up of haplotypes is statistically dependent due to shared ancestry and results from mutation, recombination and the coalescence (merging) of ancestral lineages going back in time (reviewed by Rosenberg and Nordborg 2002). Multipoint LD mapping methods that simultaneously consider a collection of alleles (haplotypes) may provide powerful statistical approaches to the fine-scale mapping of disease loci within linked intervals (Rannala and Reeve 2001; Morris *et al.* 2002; Molitor *et al.* 2003). The key principle underlying these methods is that in the vicinity of a disease locus, case chromosomes will tend to have more recent shared ancestry than controls, because many of them share the disease mutation. The challenge is to detect excess sharing of alleles and distinguish this from background allelic variation (Morris *et al.* 2002).

The 'shattered coalescent' (COLDMAP; COalescent LD MAPping) approach of Morris and colleagues (2002 and 2003) models the underlying genealogy of disease chromosomes, allowing for the possibility of multiple founder disease mutations at the same locus, the presence of sporadic cases, in addition to heterogeneity at the disease locus (the coalescent is 'shattered') (Morris *et al.* 2002; Morris *et al.* 2003). A version that uses haplotype data was retrospectively applied to cystic fibrosis (CF) case and control chromosomes. The location of the Δ508 mutation in the CF gene (*CFTR*) was estimated at 0.85 Mb (95% CI 0.65-1.0Mb); the actual location is 0.88Mb (Morris *et al.* 2002). A more recent version that utilises genotype information narrowed the associated interval of the cytochrome p450 (CYP2D6) poor drug metaboliser (PM) phenotype from 390Kb to a ~190Kb region, with the mutation location estimated <50 Kb from actual mutation within *CYP2D6* (Hosking *et al.* 2002) (Morris *et al.* 2003).

5.1.4 Aims of the Work Described in Chapter 5

Fine-mapping by LD has previously been successful for identifying variability and disease genes within a linked interval (Bobadilla *et al.* 2002) (Rioux *et al.* 2001). Normally, this is carried out in the same population as the original linkage study. However, given the findings in Iceland and the US, we hypothesised that patients from Norway (Iceland's predominant founder population) may have been predisposed to typical late-onset PD by common genetic variability of Norse origin in the 1p32 region (Hicks *et al.* 2002; Li *et al.* 2002). We firstly aimed to assess genetic homogeneity and confirm Norse ancestry in the Trondheim community.

Since extensive levels of LD might be expected in the younger Icelandic population (Varilo *et al.* 2003), the Norwegian population was also considered more suitable for finer resolution mapping. Underlying levels of LD at the genomic region under investigation are thought to impact LD mapping studies (Clark 2003). LD levels are also known to vary between genomic region and populations (Section 1.2.2). Empirical assessment of background LD can provide information about recombination, haplotype diversity and the optimal marker density required to capture association with disease. There is currently no data available regarding the pattern or extent of LD at 1p32 within the Norwegian population. Therefore, we aimed to make a low resolution assessment of LD in the 1p32 region.

The third aim was to prioritise regions within the disease associated interval for further investigation by 1) identification and analysis of candidate genes by direct sequencing and htSNP methods and 2) multipoint LD mapping using SNP genotype data.

5.2 Y Chromosome Haplogroup Analysis

The non-recombining region of the Y chromosome is a widely used tool in the study of human evolution and population diversity. This euchromatic component is ~35 Mb in length, with polymorphisms occuring once every 3-4Kb. They tend to have a unique mutational origin and unambiguously define 'haplogroups' – related groups of chromosomes (Stumpf and Goldstein 2001). Helgason and colleagues (Helgason *et al.* 2000) utilised Y chromosome haplogroups in their studies of the patrilineal heritage of the Icelandic people, and showed that the majority (75-80%) were consistent with Norse origin.

The Trondheim region from which our samples originate is a relatively isolated community (Section 2.1.6). Although Norwegian ethnicity in the country is high, 2% are thought to originate from other ethnic groups (Section 1.4). To discount the possibility that the Trondheim population were derived from founders not consistent with Norwegian ancestry (and therefore not likely to harbour a PD susceptibility locus of Norse origin), we assessed the probability of Norwegian heritage based on genetic variation on the Y chromosome.

The sample group comprised 164 males with a median age of 72 years (range 45-93 years). Twenty one percent of these were from the control group (n=35) and 79% were from the case group (n=129). Y chromosome markers genotyped were M9, SRY1532, TAT, YAP and SRY2627 (assays are shown in Appendix 1). Alleles were coded as either 0 or 1 according to the convention followed by Helgason and colleagues (Helgason *et al.* 2000). Frequencies are shown in Table 5.2a.

Non-recombining region Y chromosome marker data can be treated as haploid data and haplotypes can be easily constructed from genotype data for individuals. The

allelic conformation of haplogroups, according to the nomenclature followed by Helgason and colleagues (Helgason *et al.* 2000) is shown in Table 5.2b.

Marker genotype information allowed 130 individuals to be assigned to one of the 5 haplogroups in Table 5.2b. Helgason and colleagues (Helgason *et al.* 2000) assessed frequencies of these haplogroups in a number of European populations. Greeks were the only other population, in addition to Norwegian, composed of halpogroups 1, 2, 3, 4/21 and 16 only. Trondheim haplogroup frequencies were compared against those reported for Norway and Greece reported by Helgason and colleagues (2000). A Chi-squared test using CLUMP (Section 2.3.10) revealed no significant difference between the Trondheim and Norwegian haplogroup counts ($p=0.07$), in contrast to the Greeks ($p<1.0^{e-6}$; Table 5.2c), consistent with Norse ancestry. However, one control individual (K036) had a haplotype consisting of alleles 00000, which cannot be assigned to a Norwegian haplogroup. Although repeat genotyping/haplogroup reconstruction is needed to verify this, therein lies the possibility that 0.8% of the individuals in the Trondheim series are not consistent with Norse ancestry.

Table 5.2a Y Chromosome Marker Allele Frequencies in Trondheim Males

Allele are coded following the convention of Helgason *et al.* 2000.

Marker	Allele frequencies (n) 0	1	Total n
M9	0.42 (57)	0.58 (79)	136
SRY1532	0.32 (43)	0.68 (91)	134
TAT	0.98 (129)	0.02 (3)	132
YAP	0.99 (139)	0.01 (1)	140
SRY2627	1.00 (146)	0.00 (0)	146

Table 5.2b Allelic Conformation of Chromosome Y Haplogroups

Based on nomenclature used by Helgason *et al.* 2000.

Haplogroup	Marker/Allele M9	SRY1532	TAT	YAP	SRY2627
1	1	1	0	0	0
2	0	1	0	0	0
3	1	0	0	0	0
4 / 21	0	1	0	1	0
16	1	1	1	0	0

Table 5.2c Haplogroup Frequencies in Trondheim, Norway and Greece

Chi-squared analysis shows significant differences between Trondheim and Greek haplogroup counts, but not with Norwegians, as reported by (Helgason *et al.* 2000).

Sample	Haplogroup Frequency (n) 1	2	3	4/21	16	Total	Chi-sq	p
Trondheim	0.24 (31)	0.40 (52)	0.33 (43)	0.01 (1)	0.02 (3)	130		
Norwegian	0.26 (29)	0.51 (57)	0.18 (20)	0.03 (3)	0.03 (3)	112	8.4	0.07
Greek	0.14 (6)	0.46 (19)	0.05 (2)	0.33 (14)	0.03 (1)	42	49.9	$<1.0^{e-6}$

5.3 Linkage Disequilibrium at 1p32

To empirically assess LD within the 1p32 PD-linked region in the Norwegian population, a low-resolution assessment was performed in a subset of control individuals, using SNP genotype data spanning this region.

SNPs were chosen from the CDS (hCV#) or NCBI (rs#) SNP databases. Assays were designed for putative SNPs and were initially tested in 11 Norwegian controls as this may identify ~90% of SNPs with frequencies >0.1. Around 30% of the SNPs genotyped were not present in the initial test sample at useful frequencies (>0.1; (Garner and Slatkin 2003)).

Thirty-six SNPs within the 1p32 LOD-1 linkage interval, spanning a region of ~6.5 Mb (average inter-marker distance ~180 Kb) were genotyped in samples that comprised 29 control individuals, 15 males and 14 females with a median age of 80 years (range 70-99 years) (Section 2.2.6). Assay details are shown in Appendix 1. SNPs, base pair positions and inter marker distances are shown in Table 5.3a.

Table 5.3a SNPs Gentyped to Assess LD at 1p32 in Norwegians

SNPs span ~6.5 Mb of 1p32, with an average marker spacing of ~180 Kb. Positions as documented by CDS (December 2003).

ID	SNP	Position (bp)	Inter SNP Distance (bp)
1	hCV12109344	45,562,613	142,427
2	hCV11872095	45,705,040	74,562
3	hCV1476946	45,779,602	31,804
4	hCV11872065	45,811,406	1,063,414
5	hCV3027948	46,874,820	42,428
6	hCV3027932	46,917,248	38,139
7	hCV3027900	46,955,387	31,108
8	hCV2809720	46,986,495	27,333
9	hCV2371385	47,013,828	1,831,907
10	hCV435984	48,845,735	28,276
11	hCV15886654	48,874,011	731,724
12	hCV9509099	49,605,735	120,553
13	hCV8847088	49,726,288	28,788
14	hCV11870545	49,755,076	15,012
15	hCV277402	49,770,088	149,404
16	hCV1413746	49,919,492	55,944
17	hCV3122409	49,975,436	132,083
18	hCV1591218	50,107,519	5,931
19	hCV11740230	50,113,450	208,132
20	hCV386562	50,321,582	86,017
21	hCV428790	50,407,599	143,002
22	rs1126997	50,552,779	26,138
23	hCV2776338	50,578,917	56,775
24	hCV79313	50,635,692	50,663
25	rs2230325	50,686,355	59,303
26	hCV1805212	50,745,658	40,194
27	hCV1805260	50,785,852	146,508
28	hCV12112305	50,932,360	385,161
29	hCV11286191	51,317,521	168,120
30	hCV11873295	51,485,641	349,773
31	hCV7842707	51,835,414	83,308
32	hCV11869553	51,918,722	117,410
33	hCV1797353	52,036,132	23,813
34	hCV1797412	52,059,945	4,072
35	hCV1797425	52,064,017	20,872
36	hCV1797435	52,084,889	

5.3.1 Extent of LD

Genotype data for the 32 SNPs was used to generate bi-locus haplotypes using Arlequin (Section 2.3.5); these were used to calculate pairwise measures of LD (D' and r^2) with Genotype Transposer (Section 2.3.8). LD is plotted as a function of the physical inter-marker distance, shown in Figure 5.3.1a.

It is apparent that D' can be upwardly inflated with respect to r^2 , especially over large distances in this small sample size. To gauge the marker density required and the resolution at which a susceptibility locus could be fine-scale mapped, higher resolution LD was investigated. $\sqrt{r^2}$ (to better clarify data points) was plotted against physical distance for markers separated by <500Kb, shown in Figure 5.3.1b. A logarithmic regression model best fits the data (F=14.2, p=0.0003), the trendline is shown in white.

The dashed line represents significant LD at the 0.05 level ($\sqrt{r^2}$=0.26) and the dotted line above is Kruglyak's 'useful' LD of r=0.1 ($\sqrt{r^2}$=0.316) (Section 1.2.1). The data suggest that association with disease may be captured with a marker density of ~50 Kb (the distance at which r^2=0.1) and it may be possible to map a locus to within 150Kb (the distance at which significant LD is detected).

Figure 5.3.1a LD as a Function of Physical Distance

Pairwise measures of LD (|D'| and r^2) for 36 SNPs at 1p32 covering ~6.5 Mb, in 58 Norwegian control chromosomes. SNPs are detailed in Table 5.3a.

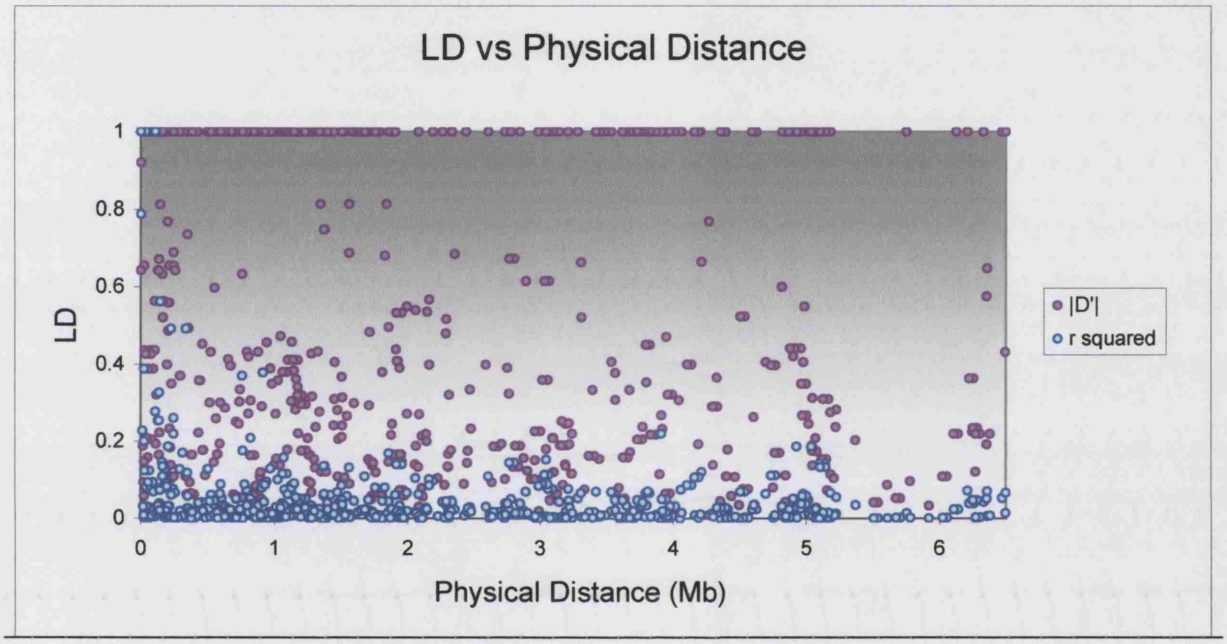
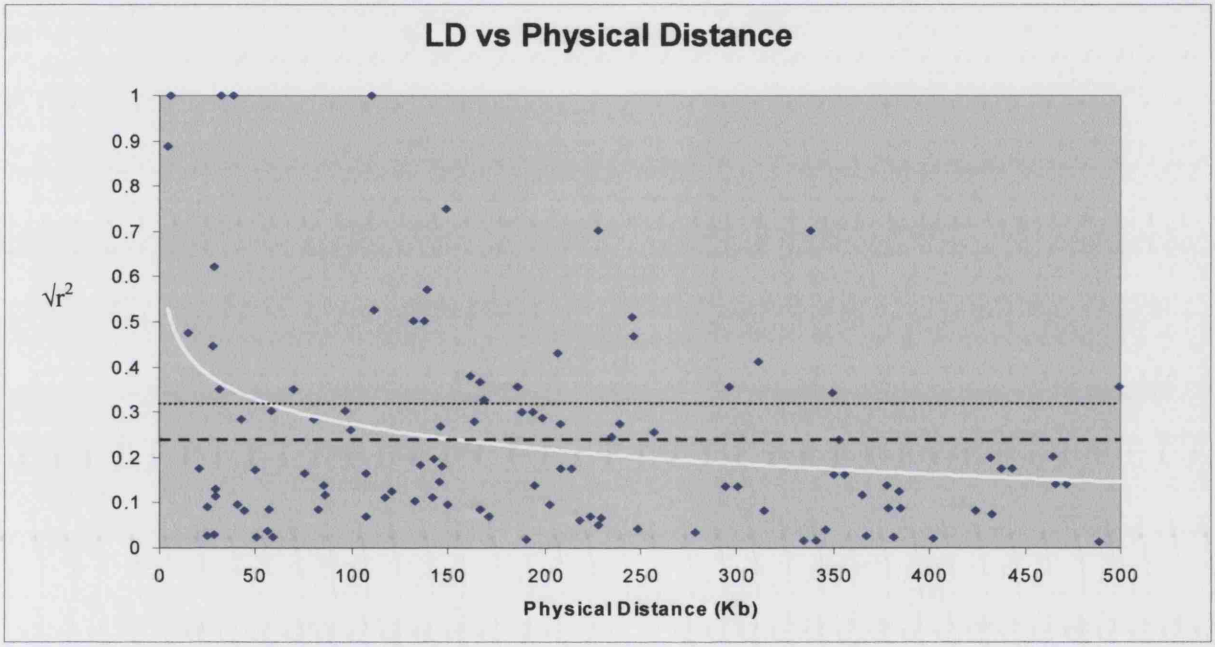


Figure 5.3.1b LD as a Function of Physical Distance for Markers Separated by < 500Kb

$\sqrt{r^2}$ is plotted to better clarify datapoints. A logarithmic regression trendline is shown in white ($F=14.2$, $p=0.0003$). The dashed line represents significant LD at $\alpha=0.05$ ($\sqrt{r^2}=0.26$). Kruglyak's (Kruglyak 1999) useful LD is shown at the dotted line above ($\sqrt{r^2}=0.32$).

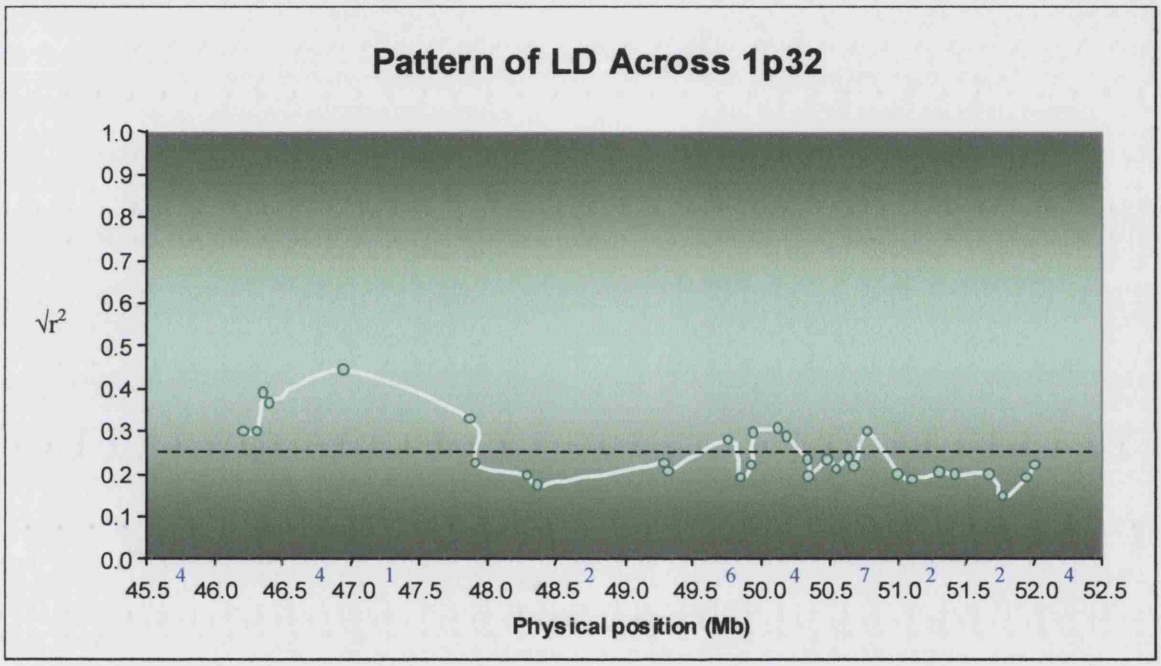


5.3.2 Variation in LD

It is apparent that LD is not uniform across 1p32 (Figure 5.3.1a). To better appreciate variation in the region, sliding averages of 5-marker pairwise measures of $\sqrt{r^2}$ (for clarity) were plotted against physical location and shown in figure 5.3.1c. Areas above the dotted line are regions of significant LD. The interspersion of regions of significant LD, with areas of non-significant LD may be suggestive of a block-like structure of LD within 1p32. However, SNP coverage is not uniform and may not be dense enough to delineate block boundaries.

Figure 5.3.2a Pattern of LD Across 1p32

Sliding averages of 5-marker pairwise measures of $\sqrt{r^2}$ are shown plotted against physical position, across the 1p32 region. The dashed line represents significant LD at α=0.05 ($\sqrt{r^2}$=0.26). The number of SNPs typed in each 0.5 Mb interval are shown on the x-axis in blue.



5.4 LD in Cases and Controls

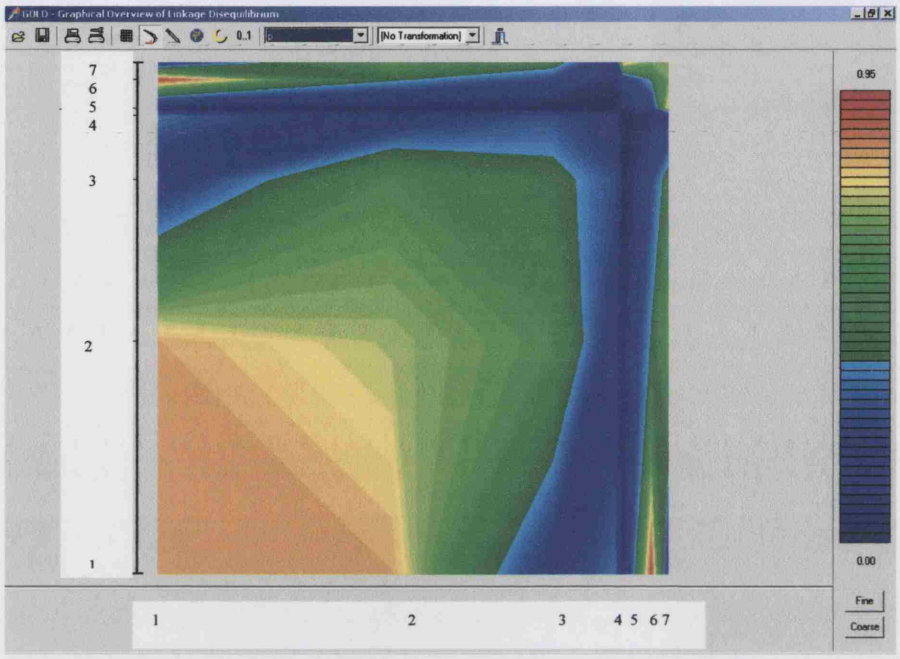
In the study of Hicks and colleagues (Hicks *et al.* 2002), the peak LOD score within the 1p32 PD-linked region was at D1S231. Linkage disequilibrium patterns between 7 SNPs flanking this STRP were initially assessed in a gender/age matched group of 50 cases and 50 controls (median age was 71 years (range 46-94 years) in cases and 81 years in controls (range 69-100 years)). SNPs listed in Table 5.4a were genotyped in cases and controls (Section 2.2.6). Bi-locus haplotypes were generated using Arlequin (Section 2.3.5), significance of LD (p(D')) was calculated using Genotype Transposer and graphical overviews of LD patterns (GOLD maps) were generated using GOLD (Section 2.3.8; Figure 5.4b).

Table 5.4a SNPs Flanking D1S231

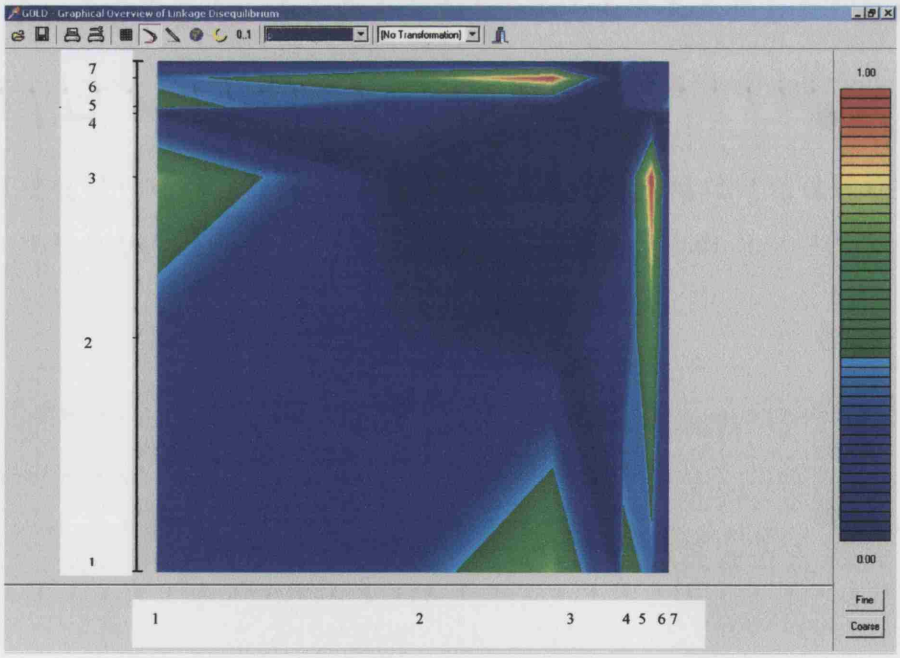
	Celera SNP ID	NCBI (rs) SNP ID	Position (bp)
1	hCV15886654	2494883	48,874,011
2	hCV8847088	1043141	49,726,288
3	hCV386562	1316981	50,321,582
	D1S231		50,489,739
4	hCV2776353	1126997 (now merged w/8375)	50,550,601
5	hCV2776338	856610	50,578,917
6	2230325	2230325	50,686,355
7	hCV1805212	1024313	50,745,658

Figure 5.4a Patterns of LD Around D1S231 in Cases and Controls

Graphical Overview of LD (GOLD) maps show the patterns of LD. Significant regions are dark blue (p(D')<0.01). The positions of SNPs 1-7 are shown on the axes.



Control



Case

There is widespread significant LD in cases, particularly between SNPs 1 and 3. This may be indicative of shared ancestry within cases at this genomic region, and suggests that Norwegian PD cases may harbour 1p32-linked susceptibility. However, this level of disequilibrium may also be prohibitive for fine-scale mapping.

5.5 Candidate Gene Analysis

There are numerous genes with the 1p32-linked interval (Morris *et al.* 2002). Candidate genes were selected for analysis based on selection criteria including proximity to the D1S231 peak LOD marker (Hicks *et al.* 2002) and/or within the region displaying elevated LD in cases compared to controls, for which there was fairly good characterisation data available. Those chosen were *ELAVL4*, *FAF1*, *RNF11*, *EPS15*, *NRD1* and *RAB3b*. Figure 5.5a shows their relative positions within the 1p32-linked interval.

The *FAF1* gene encodes a ubiquitously expressed protein, involved in apoptotic processes in the immune and nervous systems. The N-terminal 201 amino acids contain a ubiquitin homology domain (Ryu *et al.* 1999). *ELAVL4* encodes a putative neuron-specfic RNA processing peptide. It shows homology with two drosophila genes involved in the maturation of neurons (Robinow and White 1988).

The *RNF11* gene encodes three different moieties of RNF11 protein, through alternate splicing of its 3 exons. Two of the RNF11 peptides contain a zinc finger RING domain, which are instrumental in the function of E3 ubiquitin-ligases. It is ubiquitously expressed, including in brain (NCBI locus link ID 26994). A pathogenic mechanism resulting from variation within *RNF11*, analogous to that of Parkin protein dysfunction within the UPP, could be envisaged.

EPS15 was originally discovered as a substrate for the tyrosine kinase activity of the epidermal growth factor receptor (EGFR). It contains a ubiquitin-interacting motif (Klapisz *et al.* 2002) and three copies of the EH domain at the NH_2 termini. EH domains are highly conserved through species and most proteins containing them are implicated in actin dynamics and vesicular transport reactions (Confalonieri and Di

Fiore 2002). EPS15 is thought to play a role in clathrin-mediated endocytosis, including endocytosis of synaptic vesicle membranes. The *C. elegans* orthologue, ehs-1, is involved in nerve terminal synaptic vesicle recycling. Targeted impairment of function induced a depletion of synaptic vesicles and uncoordinated movement in the worm (Salcini *et al.* 1999). Microarray analysis has also shown that expression of the mouse orthologue *eps15* is upregulated in *uch-l1* deficient animals, suggesting a role in pathways leading to PD (Bonin *et al.* 2004).

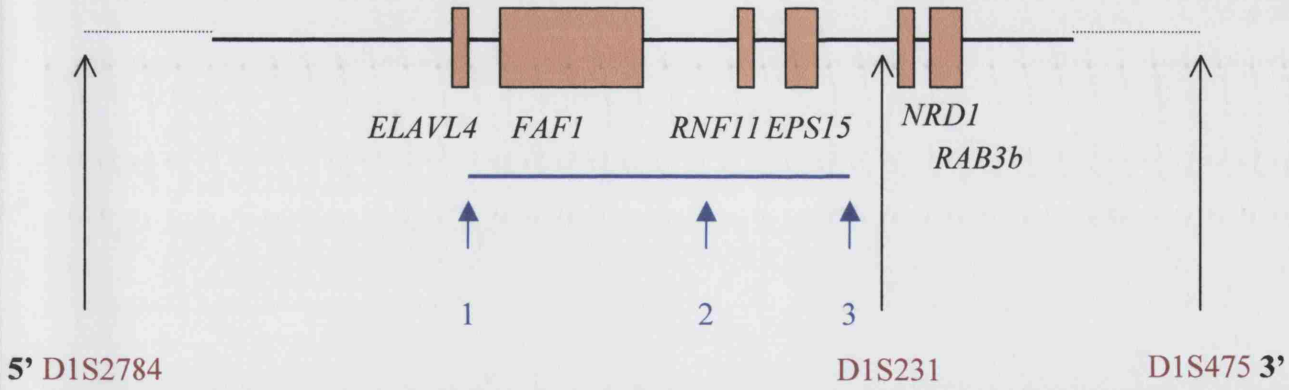
N-arginine dibasic (NRD) convertase (nardilysin) is a metalloendopeptidase that cleaves peptides at the N-terminus of arginines. It was first discovered in rat brain cortex and purified from testes where it is particularly abundant. In testes, it is associated with microtubular structures (Hospital *et al.* 2000). In humans, expression is primarily in heart, skeletel muscle and testes, although in the mouse expression is almost exclusively within neural tissues during early development (Fumagalli *et al.* 1998). This suggests that NRD convertase plays a role in neuronal development. It was hypothesised that genetic variants in *NRD* could play a part in the viable development of neuronal cells susceptible in PD (e.g. within the SN).

RAB3b is a member of the Rab family of GTP-binding proteins, which comprises at least 60 different genes in humans. These monomeric GTPases are key regulators of vesicular transport (Seabra *et al.* 2002). A variety of human diseases are caused by dysfunction of intracellular trafficking; mutations in the Rab27a gene are instrumental in Griscelli Syndrome, and result in accumulation of melanosomes in the perinuclear region of melanocytes (Menasche *et al.* 2000). Although many Rab proteins are ubiquitous, the Rab3 family isoforms (a-d) are enriched in neurons and endocrine cells. Physically, they are associated with secretory vesicles and functional studies point towards a role in neurotransmitter and hormone secretion. By way of analogy

play a role in secretory vesicle mobilisation along the cytoskeleton (Darchen and Goud 2000). The essential functional differences between the four Rab3 moieties is unclear, although RAB3b has been implicated in the calcium-dependent secretion of the dopamine metabolite norepinephrine (Weber *et al.* 1996). It was hypothesised that variation within RAB3b could deleteriously affect dopamine trafficking in dopamine neurons.

Figure 5.5a PARK10-linked PD Susceptibility Candidate Genes

Candidate genes investigated (pink boxes) at their approximate Mb positions. Peak LOD was at D1S231 in Icelandics, D1S2784 and D1S475 flank the peak LOD-1 region (Hicks *et al.* 2002). The blue line represents elevated LD in cases compared with controls (Section 5.4) and shows positions of SNPs 1, 2 and 3 (Table 5.4a).



5.5.1 htSNP Haplotype Analysis

To rapidly screen for variation between cases and controls within candidate genes, haplotype analysis using htSNPs was undertaken. This was carried out in a late onset case control series for genes *RNF11*, *EPS15*, *NRD1* and *RAB3b*. The case group comprised 110 individuals (60% male) with an age of disease onset >60 years (median 68 years, range 61-88 years). Median current ages of cases was 76 years (range 64-94 years). The control group comprised 137 individuals (57% male) with a median age of 81 years (range 69-100 years).

To identify htSNPs, 80 of the control individuals were genotyped for SNPs spanning each gene, at an approximate density of 1 SNP every ~15Kb. SNPs were identified from the CDS SNP database and tested for Norwegian population frequency by genotyping in 11 individuals (Section 2.2.6). SNPs with frequencies >0.1 were chosen for further analysis. Multilocus haplotypes were estimated using Arlequin (Section 2.3.5) and used for htSNP identification (sufficient to capture >90% of haplotype variation) with SNPtagger (Section 2.3.6). Ideograms of the genomic structure of each gene, the positions of SNPs genotyped and those identified as htSNPs are shown in Figure 5.5.1b. SNP identification numbers are given in Table 5.5.1a.

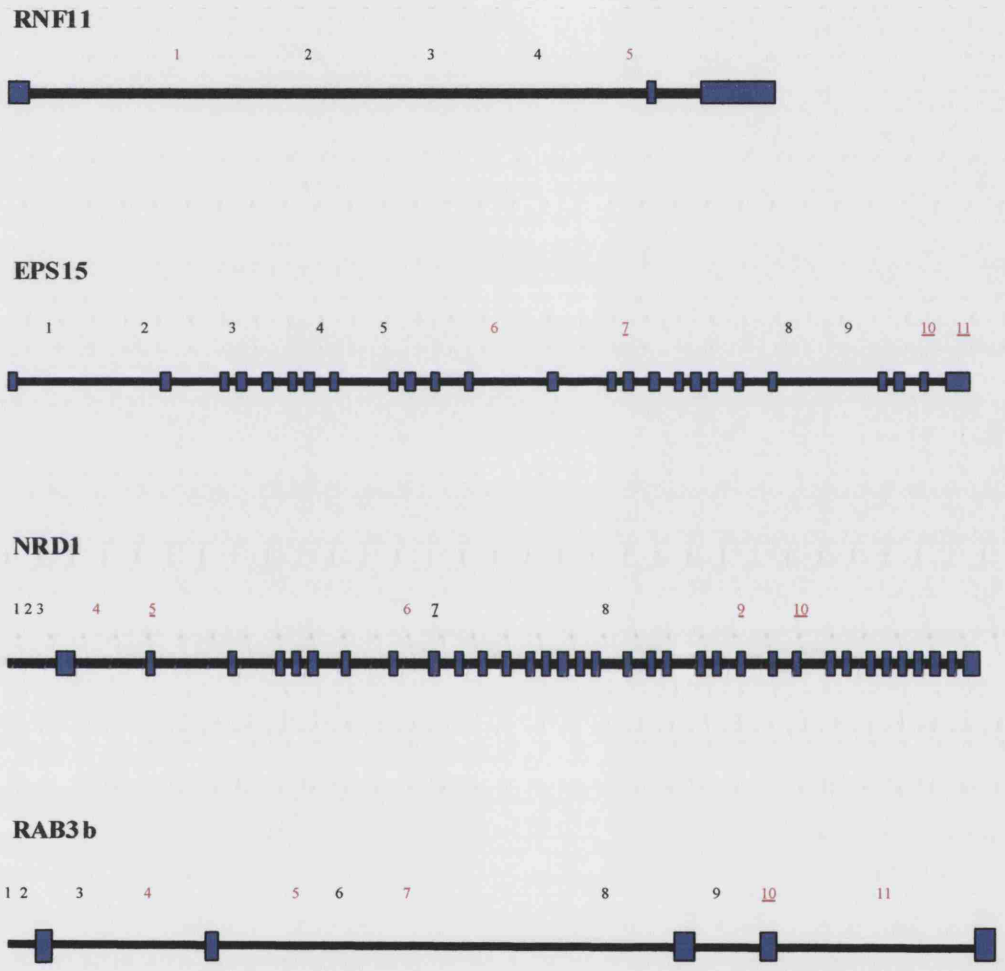
Table 5.5.1a SNPs and htSNPs in 1p32 Candidate Genes

SNPs are listetd 5' to 3'. htSNPs are italisized.

	RNF11	EPS15	NRD1	RAB3b
1	*hCV3122415*	hCV3125022	hCV79313	hCV1805213
2	hCV11304546	hCV3125029	hCV15818300	hCV1805212
3	hCV11304542	hCV3125045	hCV15818304	hCV1805209
4	hCV3122423	hCV1771642	*hCV393617*	*hCV1805202*
5	*hCV937775*	hCV3125034	*Glu indel*	*hCV1805200*
6		*hCV3125031*	*hCV2776338*	hCV315881
7		*hCV1591238*	hCV2776339	*hCV16117609*
8		hCV1591233	hCV2776344	hCV11865899
9		hCV11740229	*hCV2776353*	hCV11865894
10		*hCV11740230*	*hCV8847889*	*2230325*
11		*hCV1591218*		*hCV439172*

Figure 5.5.1b SNPs and htSNPs Positions in Candidate Genes

Exons are depicted as blue boxes. SNPs typed in 80 controls are numbered (as Table 5.5.1a) Of these, htSNPs (sufficient to capture at least 90% of haplotypic variation) are shown in red. Coding SNPs are underlined.



htSNPs were typed in remaining control and case individuals. Multilocus htSNP haplotypes were generated with Arlequin (Section 2.3.5) and frequencies were compared between cases and controls by a χ^2 test using CLUMP (Section 2.3.10). Haplotype counts in cases and controls were not signficantly different (Table 5.5.1c), although a strong association between the '1' allele at hCV2776353 (*NRD1*) and disease was observed when single marker tests of association were carried out (OR [11&12 vs 22] =2.4, 95% CI 1.3-4.4, p<0.005).

Table 5.5.1c htSNP Haplotype Count Comparisons in Cases and Controls

Empirical p values shown are for a χ^2 test (using CLUMP) comparing case and control haplotype counts.

Gene	Haplotype	Control Count	Frequency	Case Count	Frequency	p
RNF11	1 1	192	0.793	144	0.802	0.94
	1 2	36	0.149	27	0.148	
	2 1	14	0.058	9	0.048	
EPS15	1 1 1 1	117	0.492	69	0.411	0.52
	1 1 2 1	63	0.265	57	0.339	
	2 2 1 1	39	0.164	29	0.173	
	2 2 1 2	18	0.076	12	0.071	
	1 2 2 1	1	0.004	1	0.006	
NRD1	1 1 1 1 1	94	0.433	89	0.494	0.42
	1 2 2 2 2	58	0.265	40	0.220	
	2 2 1 2 1	19	0.087	14	0.077	
	2 2 2 2 2	10	0.046	9	0.050	
	2 2 1 2 2	8	0.039	3	0.016	
	1 2 1 2 2	8	0.035	13	0.070	
	2 2 1 1 1	7	0.032	3	0.016	
	2 1 1 1 1	7	0.030	2	0.013	
	1 1 2 1 1	1	0.005	0	0.000	
	1 2 1 1 1	1	0.005	3	0.018	
	1 1 2 2 2	1	0.005	2	0.013	
	1 1 1 2 1	1	0.005	0	0.000	
	2 1 1 1 2	1	0.005	0	0.000	
	2 1 1 2 1	1	0.005	2	0.009	
	2 1 1 2 2	1	0.005	0	0.000	
	2 1 1 1 2	0	0.000	1	0.003	
RAB3b	1 1 1 1 1	122	0.535	104	0.571	0.23
	2 1 1 1 1	74	0.325	54	0.297	
	2 1 2 2 1	10	0.042	6	0.033	
	2 2 1 2 1	6	0.026	4	0.022	
	1 1 1 2 1	5	0.020	2	0.011	
	2 1 2 2 2	4	0.019	9	0.049	
	2 1 2 1 1	3	0.014	1	0.006	
	1 1 1 2 2	3	0.011	0	0.000	
	2 2 1 1 1	1	0.004	0	0.000	
	2 1 1 2 2	1	0.004	2	0.011	

5.5.2 Direct Sequencing

Direct sequencing of candidate genes was also undertaken in some case individuals. *RAB3b* was sequenced in genomic DNA and *ELAVL4*, *FAF1*, *RNF11*, *NRD1* and *EPS15* were sequenced in cDNA. Primers are detailed in Appendix 1.

Coding regions, exon/intron boundaries and 1 Kb of the promoter of *RAB3b* were sequenced in 42 probable and possible cases (Section 2.2.7). Age of disease onset was between 48 and 80 years. No mutations were found.

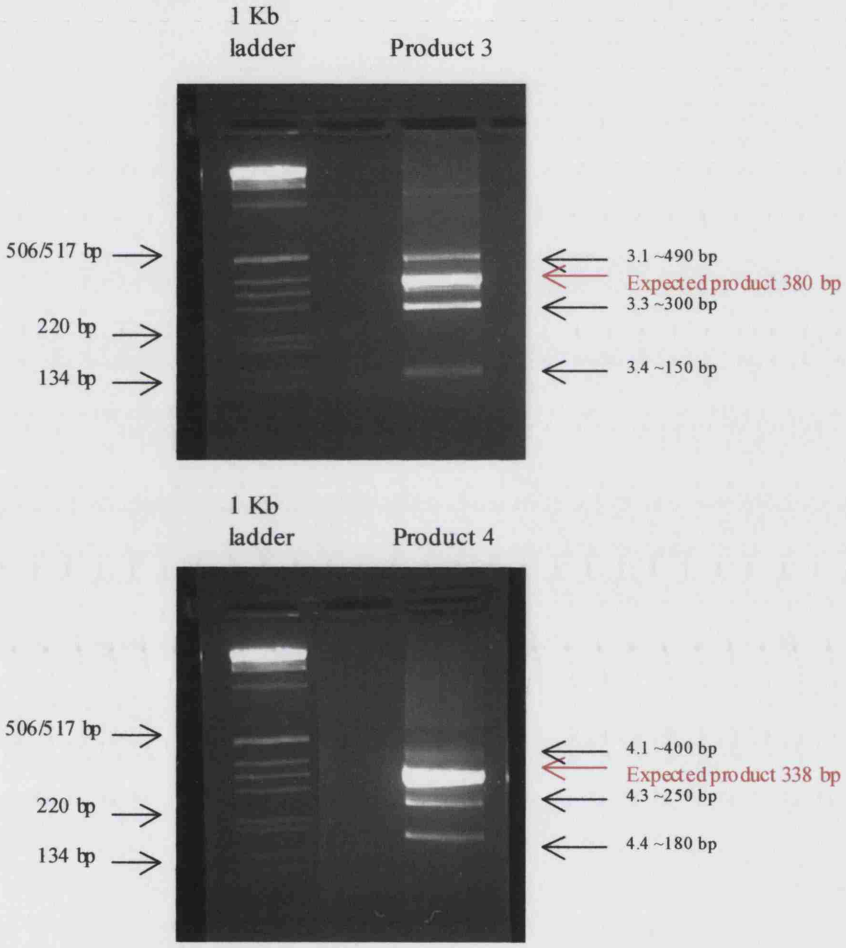
ELAVL4, *FAF1*, *RNF11*, *NRD1* and *EPS15* contain multiple exons and were sequenced in cDNA when this became available from 9 cases (1 probable PD, 8 possible PD; age of onset range 47-63 years). No mutations were found, however irregular sized products from two sequencing primer sets was observed for *EPS15*, so further investigation of these transcripts was undertaken.

5.5.3 EPS15 Transcript Analysis

As EPS15 was sequenced in cDNA, primers designed to amplify and sequence are positioned within exons. As part of the sequencing process, PCR products were agarose gel electrophoresed after purication to check size and intensity (product strength). In sample P023, products for primer sets 3 and 4 displayed additional bands to those expected (Figure 5.5.3a).

Figure 5.5.3a PCR Products from *EPS15* cDNA Primer Sets 3 and 4 in P023

Three additional bands were observed for both primer sets 3 and 4.



The positions of primer sets 3 and 4 within the *EPS15* cDNA are shown in Figure 5.5.3b. All bands were excised from the agarose gel, purifed and sequenced (Section 2.2.7). The majority of sequence anomalies involve an alteration in the copy number of exons 9, 10, 12 or 13 (Table 5.5.3c).

Exons 9 to 14 (including at least 30 base pairs of flanking introns) were subsequently genomically sequenced in P023. Primers are shown in Appendix 1. No variations were found which might affect splicing to produce altered transcripts.

Figure 5.5.3b Positions of Primer Sets 3 and 4 Within *EPS15* cDNA

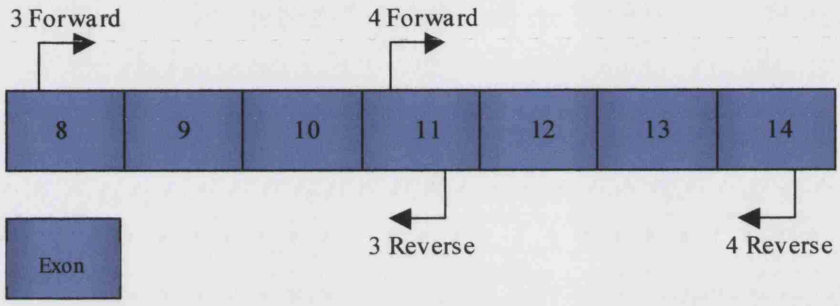


Table 5.5.3c *EPS15* Additional Transcript Sequence in P023

The majority of sequences involve the alteration in copy number of exons 9, 10, 12 or 13

Band	Sequence Anomalies
3.1	Exon 10 duplicated
3.3	Exon 9 not present
3.4	Exon 9 and 10 not present
4.1	Undetermined
4.3	Exon 12 not present
4.4	Exon 12 and 13 not present

5.6 Investigation of the 1p32 LOD-1 region

To further investigate 1p32-linked PD susceptibility over a wider region, a denser SNP marker set, spanning the 1p32 LOD-1 interval (Hicks *et al.* 2002; Li *et al.* 2002) was investigated in a larger case control series. A total of 49 SNPs and 8 STRP markers were genotyped in 233 cases and 136 controls.

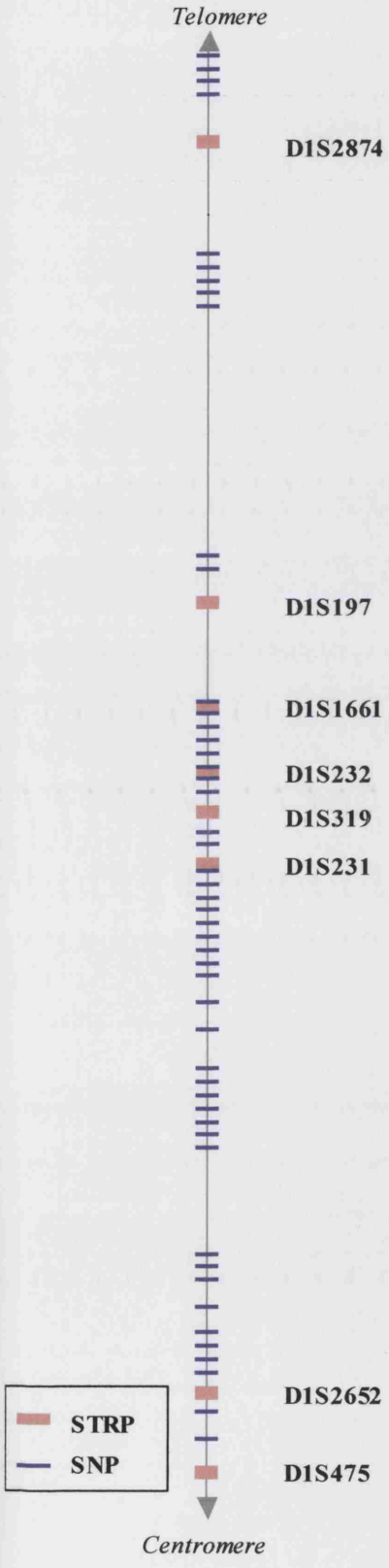
The control group comprised 136 individuals (50% male) with a median age of 81 years (range 69-100). The case group comprised 233 probable and possible PD cases (65% male) with a median age of 70 years (range 40-93). Median age of disease onset was 60 years (range 25-88).

The SNPs genotyped were those used in Section 5.3, in addition to others chosen from the CDS SNP database, with the exception of SNP 25 which was identified through sequencing of the *NRD1* gene (Section 5.5.2). Identification numbers and relative positions within the linked interval are shown in Figure 5.6a. Assay details and allele codes are shown in Appendix 1.

Single marker tests of association were carried out. ORs (11&12 vs 22 and 11vs12 & 22) for SNPs were calculated using SPSS (Section 2.3.11). Chi-squared analyses using CLUMP (Section 2.3.10) was performed for STRP allele frequencies. The $-\text{Log}_{10}$ p values for each of these tests are shown in Figure 5.6b.

Figure 5.6a Markers Analysed at *PARK10*

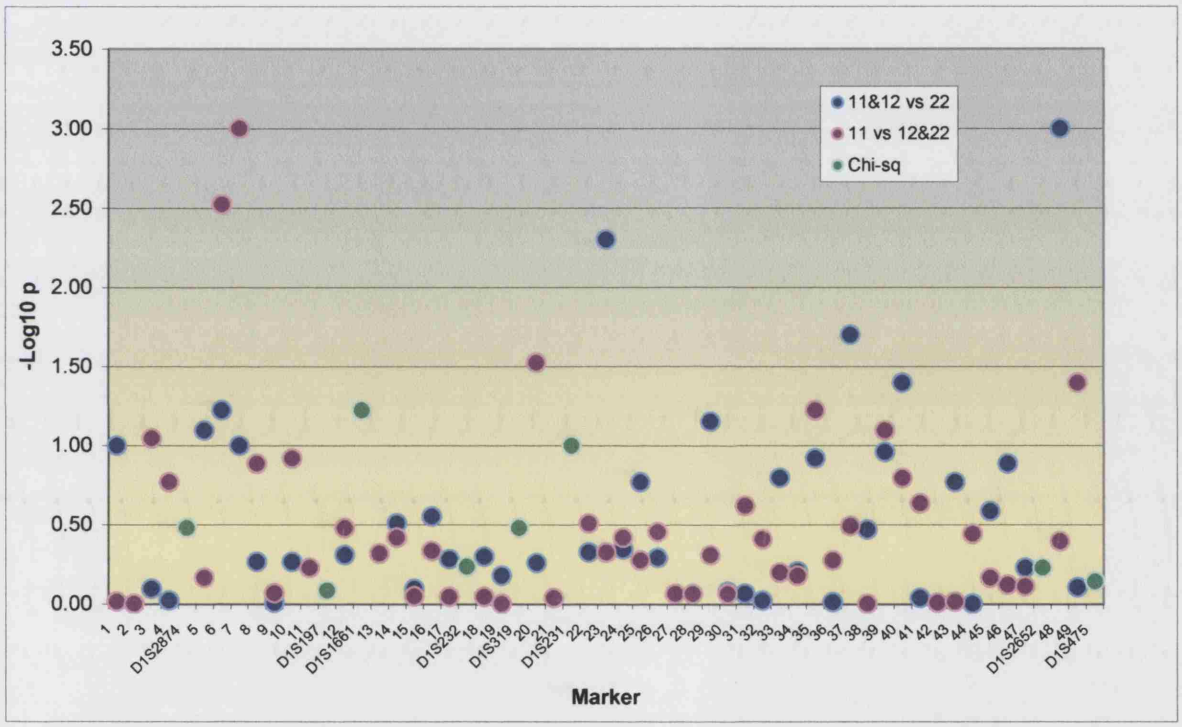
Relative positions of SNPs (purple lines) and STRPs (pink boxes) are shown to the left; names and marker identification numbers are shown to the right. They span ~8.3Mb



	Marker	Position (bp)
1	hCV12109344	45,562,613
2	hCV11872095	45,705,040
3	hCV1476946	45,779,602
4	hCV11872065	45,811,406
	D1S2874	46,194,536
5	hCV3027948	46,874,820
6	hCV3027932	46,917,248
7	hCV3027900	46,955,387
8	hCV2809720	46,986,495
9	hCV2371385	47,013,828
10	hCV435984	48,845,735
11	hCV15886654	48,874,011
	D1S197	49,036,426
12	hCV9509099	49,605,735
	D1S1661	49,717,216
13	hCV8847088	49,726,288
14	hCV11870545	49,755,076
15	hCV277402	49,770,088
16	hCV1413746	49,919,492
17	hCV3122409	49,975,436
	D1S232	50,004,877
18	hCV1591218	50,107,519
19	hCV11740230	50,113,450
	D1S319	50,192,385
20	hCV386562	50,321,582
21	hCV428790	50,407,599
	D1S231	50,489,739
22	hCV8847889	50,550,601
23	hCV2776353	50,550,601
24	hCV2776338	50,578,917
25	Glu indel (ex2)	50,578,917
26	hCV79313	50,635,692
27	rs2230325	50,686,355
28	hCV1805212	50,745,658
29	hCV1805260	50,785,852
30	hCV12112305	50,932,360
31	hCV11286191	51,317,521
32	hCV11873295	51,485,641
33	hCV7842707	51,485,641
34	hCV11869553	51,918,722
35	hCV1797353	52,036,132
36	hCV1797412	52,059,945
37	hCV1797425	52,064,017
38	hCV1797435	52,084,889
39	hCV11871255	52,193,932
40	hCV1452838	52,875,715
41	hCV2431676	52,970,645
42	hCV1761854	53,078,338
43	hCV11870762	53,267,443
44	hCV736440	53,364,006
45	hCV3154981	53,469,892
46	hCV11732075	53,604,083
47	hCV2794215	53,615,152
	D1S2652	53,754,447
48	hCV998739	53,820,346
49	hCV1758290	53,917,746
	D1S475	54,173,693

Figure 5.6b Single Marker Tests of Association at *PARK10*

Single marker tests for association were carried out for 49 SNPs at 1p32 (Figure 5.6a). $-\text{Log}_{10}$ p values for i) ORs (11 & 12 vs 22 blue dots; 11 vs 12 & 22 purple dots) for SNPs and ii) χ^2 analysis of STRP allele counts (green dots) are shown.



Significant p values (when accounting for multiple testing, 0.05/49=0.001) were obtained for carrying the 2 allele at SNP 7 (hVC3027900; p=0.001, $-Log_{10}$ p=3.0) and for carrying the 1 allele at SNP 48 (hCV998739, rs13312; p=0.001, $-Log_{10}$ p=3.0). SNP 7 lies within an intergenic region and SNP 48 is within the 5' UTR of the ubiquitin specific protease 24 gene (*USP24*) and are separated by ~6.8 Mb. The '1' allele at SNP 23 (hCV2776353 in *NRD1*) was also elevated in cases (OR [11&12 vs 22] =2.3, 95% CI=1.3-4.3, p<0.006) a similar association to that detected in late-onset cases (OR [11&12 vs 22] =2.3, 95% CI=1.3-4.4, p<0.005; Section 5.5.1). This SNP lies in the centre of the 6.8Mb region flanked by SNPs 7 and 48.

5.7 Multipoint Linkage Disequilibrium Mapping

Comparison of LD and single-marker analyses pose challenges in mapping genes in populations. Retrospective studies in CF have shown that the major mutation (Δ508) does not lie in the centre of the region displaying highest LD in case chromosomes (Kerem *et al.* 1989; Morris *et al.* 2002). Similarly, SNPs which were most significantly associated with the CYP2D6 poor metaboliser phenotype were located >150Kb either side of the *CYP2D6* gene (Hosking *et al.* 2002; Morris *et al.* 2003). In both cases, COLDMAP analyses have been successful in identifying relatively small regions containing actual disease mutations (Morris *et al.* 2002; Morris *et al.* 2003).

We used genotype data generated in Section 5.6 to perform multipoint LD mapping using the COLDMAP algorithm for genotype data (Morris *et al.* 2003) (Section 5.1.3). This method initially assigns a set of random values for variables including the location of the disease mutation, effective population size, ancestral haplotypes, ordering of coalescent events, time between coalescent events and missing genotype information. Each 'iteration' of the algorithm proposes different values from within a distribution of possible values. Initially, values will be correlated (dependent upon those from the previous iteration) but eventually will converge upon the most likely set of values after many interations, if the data can be resolved (Morris *et al.* 2002).

5.7.1 Runs 1 and 2

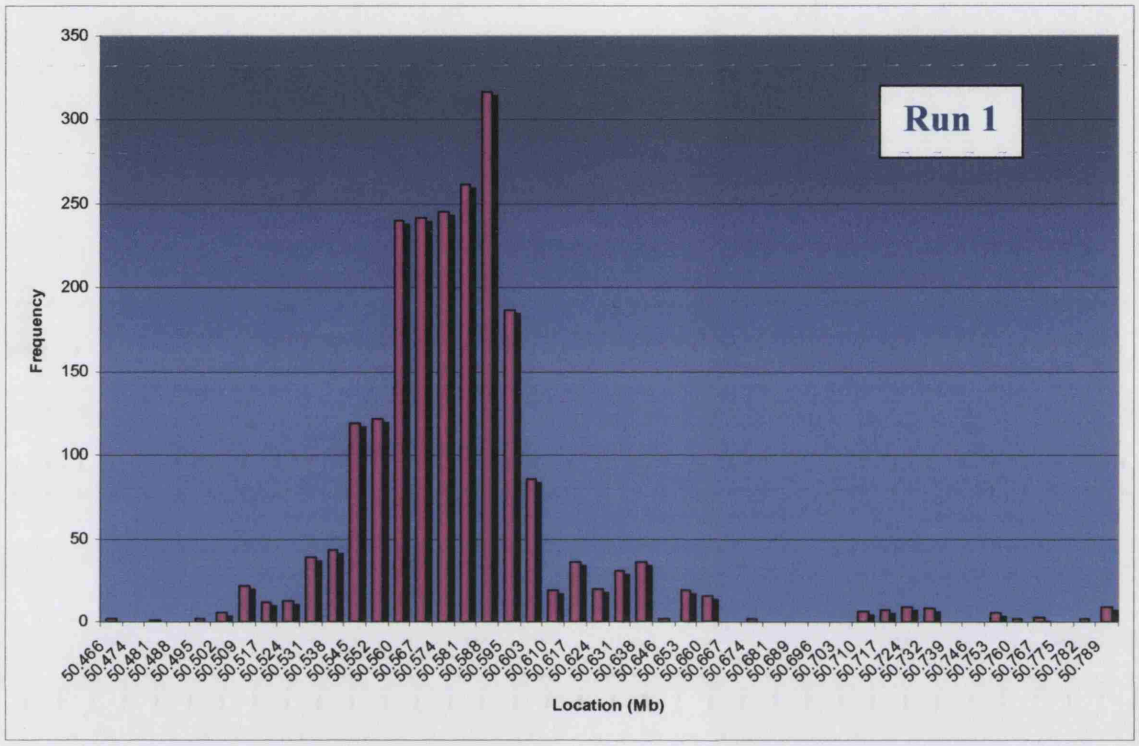
Two runs were initially performed with the same set of input parameters to test whether the algorithm would perform consistently with reproducible results. Run 1 was set to perform 22,000 iterations and the algorithm converged (as shown by stabilisation of the Log posterior probability) at ~10,000 iterations (Appendix 3). Run

2 was set to perform 10,000 iterations initially, then a further 10,000 were performed, starting from the finishing point of the first set of iterations. Run 2 had also converged by the start of the second set of iterations. Each iteration produces one output record with an estimated mutation location. The distribution of these from both runs is shown in Figure 5.7.1a.

The median estimated mutation location from Run 1 is at 50.568 Mb (95% CI=50.504-50.612 Mb) and the median from Run 2 is at 50.558 (95% CI=50.496-50.612 Mb). Both estimates show good agreement; evidence that the algorithm is not producing random results. The combined estimated location interval is ~116 Kb spanning the *NRD1* locus; this is also very close to the peak LOD marker D1S231 in Icelandics (Hicks *et al.* 2002) (Figure 5.7.1b).

Figure 5.7.1a Estimated Mutation Location

Frequency of recorded output for disease mutation is plotted against against the estimated physical position.



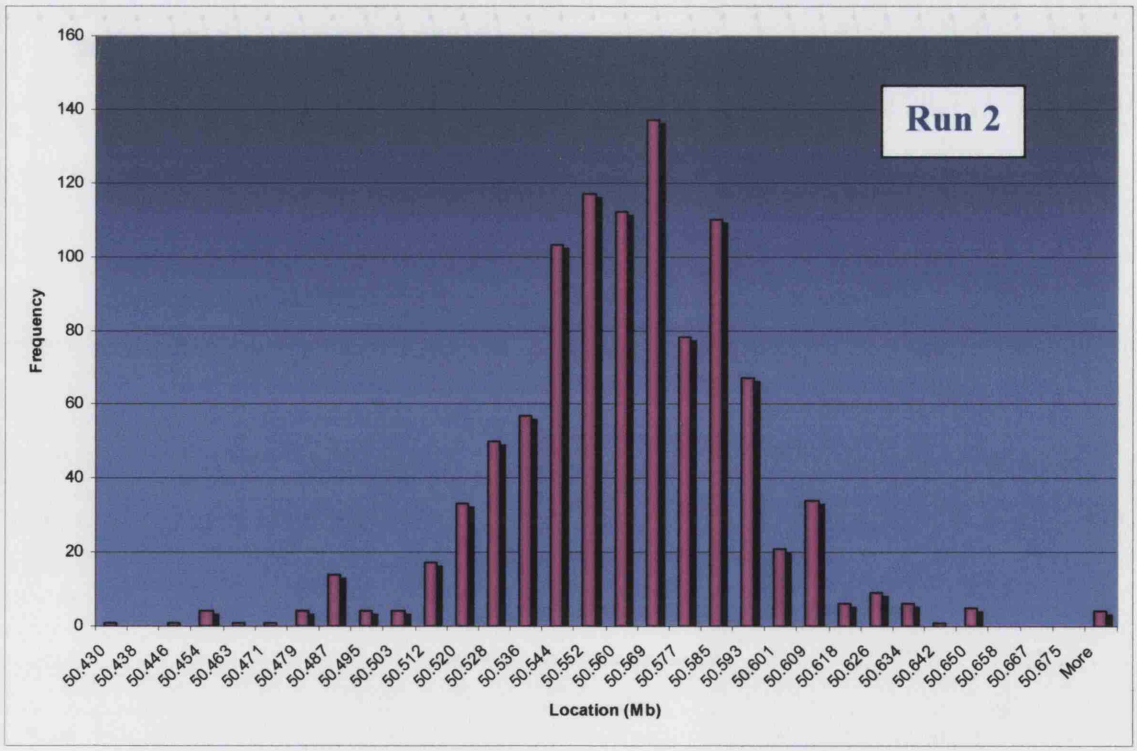
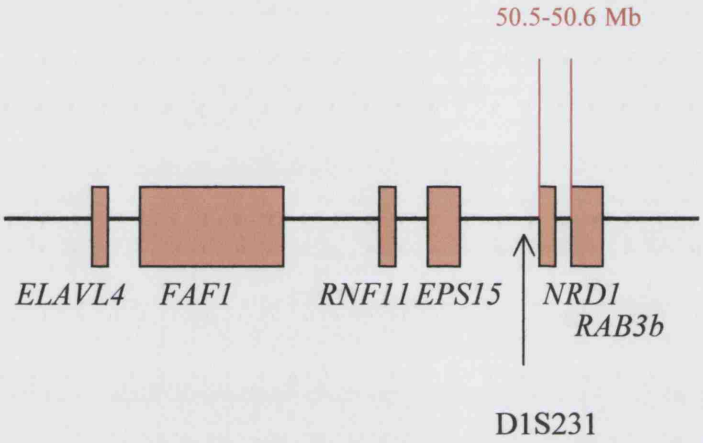


Figure 5.7.1b COLDMAP Estimated Location of a PD Susceptibility Locus in Norwegians

The estimated location is a relatively small region (~116Kb, shown in red) encompassing the *NRD1* locus. D1S231 gave the maximum LOD score for PD susceptibility in Icelandics (Hicks *et al.* 2002).



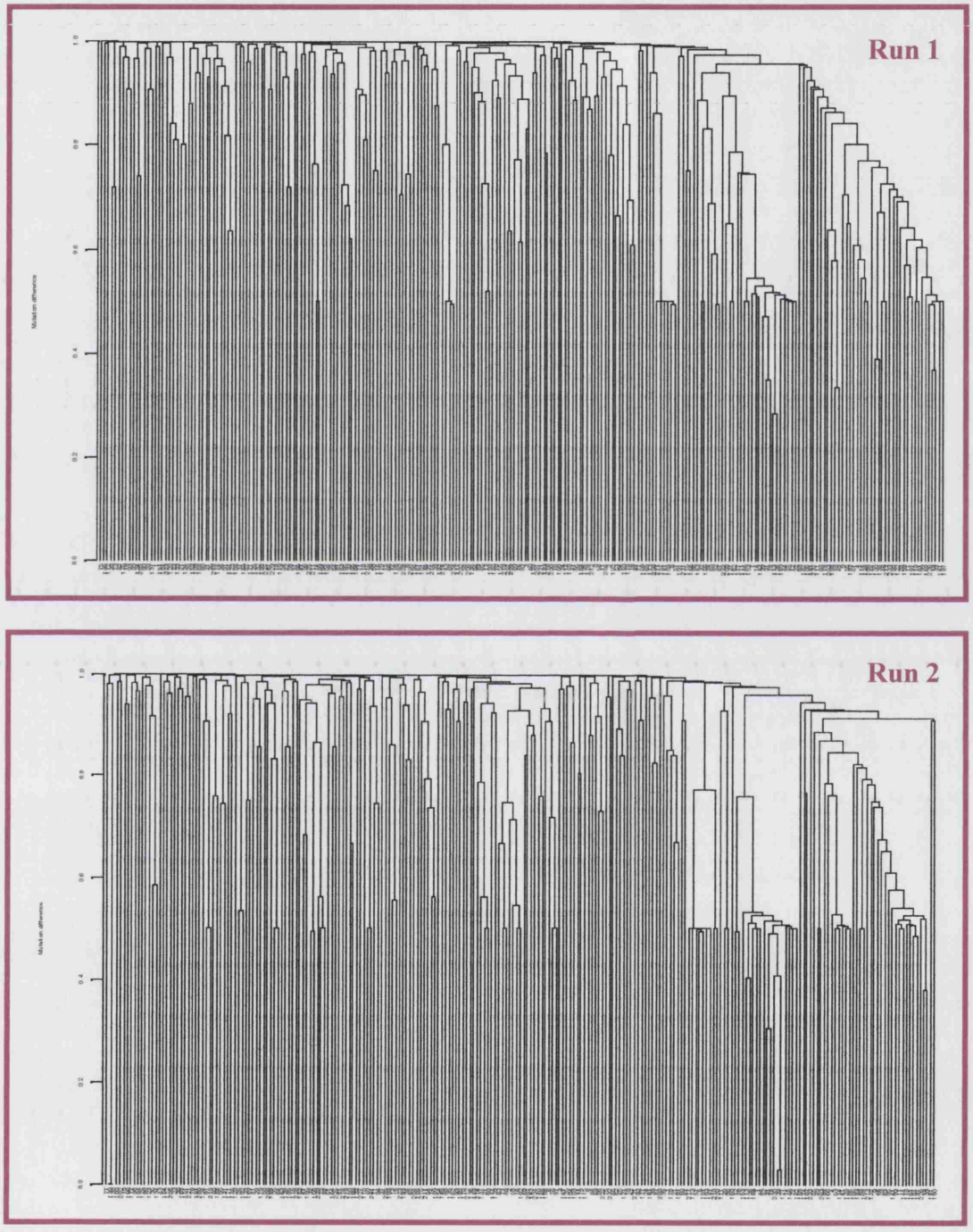
5.7.2 Rho and Cladistic Analysis

The COLDMAP algorithm also produces a 'shattering parameter' referred to as 'ρ' (rho) and refers to the probability that an individual forms part of a genealogically related group. Hence, a value of 1 indicates that a single genealogical tree (coalescent) can be constructed for cases and that the mutation is likely to be homogeneous. Conversely, lower values indicate that the coalescent is shattered and heterogeneity exists at the disease locus (Morris *et al.* 2002). The estimations of ρ from Runs 1 and 2 are 0.632 and 0.626 (95% CIs 0.596-0.662 and 0.574-0.656 respectively), suggestive of a relatively high level of heterogeneity at the disease locus.

COLDMAP also produces a cladogram of disease individuals. Cladograms group together individuals that are expected to share disease mutations in common. Individuals sharing two mutations will cluster together at the bottom of the cladogram. Individuals sharing one mutation will cluster half way up, and those sharing none will cluster towards the top of the cladogram (Morris *et al.* 2002). Cladograms from Runs 1 and 2 are shown in Figure 5.7.2a. Genetic heterogeneity (ρ) is apparent since many individuals are clustered towards the top of the cladograms. However, around ¼ are clustered halfway up, indicating that these individuals may share one susceptibility locus.

Figure 5.7.2a Cladistic Output from COLDMAP

Genetic heterogeneity (ρ) is apparent, although clustering is observed towards the right hand sides of the cladograms.



5.7.3 'Dummy' Run

To test whether the results of Section 5.7.1 may be spurious, a 'dummy' run was performed. Input parameters were as for Runs 1 and 2, except that the case/control status was randomized. The genotype input file, in which all case records are followed by control records, was recoded such that alternating records were assigned the code for 'case' status, and intervening records were assigned 'control' status. This resulted in a random status reassignment of half the samples in each group.

The algorithm was set to perform 20,000 iterations and had not converged upon completion, showing that the algorithm could not resolve the data. The distribution of location estimates was bi-modal (as opposed to unimodal with normal distribution in Runs 1 and 2), a further indication that the input data could not be resolved. Covergence and mutation location estimates are shown in Appendix 4.

5.8 Discussion

The aim of the work described in this chapter was to map a PD susceptibility locus in the Norwegian population. Since the majority of Iceland's founder population were Norse and considerable migration from Scandinavia to the US has occurred, we hypothesised that variability of Norse origin underlies the linkage to 1p32 in Icelandic and US late-onset PD cases (Section 5.1). Although samples used in this study were carefully selected to have Norwegian ancestry dating back at least 4 generations, this was based on self-assessment by those individuals. Analysis of non-recombining Y chromosome markers allowed Trondheim individuals to be assigned to one of the 5 haplogroups present in the Norwegian population (Section 5.2). Haplogroup frequencies did not vary significantly from those reported for Norway ($p=0.07$) in contrast to those reported for Greeks who share Norwegian haplogroups ($p<1.0^{e-6}$, Table 5.2c).

One control individual (K036) carried a haplogroup (00000) not consistent with Norwegian nor Swedish, Danish, Irish, Scottish, British, German, Greek, Italian or Russian ancestry (Helgason *et al.* 2000). Genotyping errors may be to blame and confirmation by repeat genotyping would be able to rule this out. However, if non-Norwegian ancestry was confirmed, it might imply that 0.8% of samples are not Norwegian. This represents only a few individuals (7 out of 774) and is not likely to greatly affect the study. The results indicate that sample ascertainment criteria on the basis of self-reported Norwegian ancestry was effective and that this series represents a useful population in which to map a PD susceptibility locus of Norse origin. In addition, the samples are ethnically well-matched, and not likely to have undergone

significant admixture; favourable population characteristics for mapping population genetic variation associated with complex diseases (Section 1.3).

Elevated LD between SNPs in around the peak LOD marker D1S231, in cases compared with controls, was suggestive of shared ancestry in cases (Section 5.4). Selection for a beneficial variant could account for increased levels of LD at any particular genomic region (Schlotterer 2003), although this does not explain the striking increase in cases compared with controls. We therefore hypothesised that Norwegians harbour 1p32 PD susceptibilty that predates the Icelandic settlement. It may, however, have been useful to determine whether this phenomenon was specific to the 1p32 region, and not a general feature of the case genome. Additional parallel comparisons at other loci not linked to PD would be instructive and is warranted.

Extensive LD in relatively young populations can be of benefit in low resolution mapping, but may be prohibitive for higher resolution studies (Varilo *et al.* 2003). Since the Icelandic population are younger than the Norwegian (Section 1.4), we hypothesised that lower levels of LD within Norway would facilitate high resolution mapping. Using a regression model to predict the decay of LD over physical distance, it was shown that a SNP density of ~50 Kb may be required to map variation to within a ~150Kb interval (Figure 5.3.1b). In theory, the level of LD can also be used to predict the sample size required for detecting disease association. To have the same power to detect the association between the disease locus and the marker locus, the sample size must be increased by $1/r^2$, in comparison to the sample size required to detect association with the disease locus itself (Ardlie *et al.* 2002). This may be beneficial when the LD is uniform across the region under study. However, within 1p32 it is apparent that LD is not uniform. Significant LD can be detected between

markers 1 Mb apart, whereas there is virtually none between some markers that are as close as 25 Kb (Figures 5.3.1a and b).

Sliding averages of 5-marker pairwise measures of LD across the region revealed a possible block-like structure of LD at 1p32. This phenomenon has previously been reported in various genomic regions. If the amount of haplotypic diversity is known, then it should be possible to position SNPs such that most information is captured with minimal marker typings (Section 1.2.3). A contiguous ~3Mb stretch of significant LD is apparent in the Norwegians at the 5' end of the analysed region. An area containing smaller blocks characterises the 3' end. However, marker density in the 5' block is much lower (1.7 SNPs/Mb) than at the 3' end (5.7 SNPs/Mb) (Figure 5.3.2a). Studies using sparser marker sets (Dawson *et al.* 2002; Gabriel *et al.* 2002) have typically revealed longer blocks than those using denser sets (Patil *et al.* 2001; Ke *et al.* 2004) so marker spacing may have impacted the results shown here. Previous studies have used pairwise measures of LD (Daly *et al.* 2001) (Gabriel *et al.* 2002) and blocks are regions where all pairwise statistics are above a certain level. However, this may not adequately reflect haplotype ancestry, due to the non-independence of LD coefficients (Cardon and Abecasis 2003). Other studies define a block as a region where inferred haplotypic diversity is low and a small number of haplotypes account for the majority observed (Daly *et al.* 2001; Zhang *et al.* 2002). Since the completion of this study, an LDU map (Tapper *et al.* 2003) has been constructed for the region, using data from 99 SNPs (Appendix 5). There is evidence of haplotype blocks (horizontal plateaus within the profile), although the value of identifying these, with regard to facilitating multipoint LD mapping with COLDMAP has been disputed (Morris *et al.* 2003).

Assessment of LD at 1p32 provided a cursory assessment in a representative number of samples and revealed that LD is variable within 1p32 and that D' is also inflated compared to r^2 (Figure 5.3.1a); both of which are consistent with that seen at chromosome 22 (Dawson *et al.* 2002). In addition, SNPs were identifed that would be useful for further analyses (Section 5.6).

To identify the 1p32 susceptibility locus, both candidate and postitional approaches were taken. The initial set of candidate genes chosen for analysis in a subset of samples included those close to the 1p32 peak LOD marker, or within the region of elevated LD in cases (Section 5.5). However, roles in the pathogenesis of PD may be envisaged for most. These were *FAF1, ELAVL4, RNF11, EPS15, NRD1* and *RAB3b.*

Multilocus haplotypes composed of htSNPs, sufficient to capture >90% of the variation within *RNF11, EPS15, NRD1* and *RAB3b,* were compared between 110 late-onset cases and 137 controls, closely matched for age (Section 5.5.1). Although no significant differences were found (Table 5.5.1.c), a single marker test of association for one SNP (SNP 23) within *NRD1* was significantly associated with disease. This is a non-synonymous change in exon 23, which may be in LD with a causal variant. The utility of htSNPs is determined largely by the SNP set from which they are chosen. In order to identify htSNPs, it was aimed to type at least one SNP every 15Kb, although many SNPs were not found at useable frequencies and thus the final htSNP set was limited. Also variation in many of the coding regions was not directly asssessed. For example, there were no SNPs genotyped in 30 out of the 34 exons in *NRD1.* Hence there may not have been enough power to capture association with disease with the htSNP set employed and rare haplotypes may have been overlooked.

cDNA sequence analysis of *NRD1* in 9 cases did not reveal any coding mutations (Section 5.5.2). This was undertaken prior to the completion of the multipoint LD mapping in samples for which cDNA was available, representing a cost effective approach over genomic sequencing, given the number of exons within NRD1 (24). However, in light of the results of multipoint LD mapping (Section 5.7) with an estimated mutation location spanning the *NRD1* locus, heterogeneity at the disease locus (average ρ=0.63) and clustering of only ~30 individuals likely to share a disease mutation, it is feasible that those samples included in cDNA analysis do not harbour 1p32-linked PD susceptibility.

The facility to identify individuals within COLDMAP cladistic output was a subsequent addition to the algorithm and time constraints did not allow further investigation in these samples. Hence, it will now be important to identify those individuals estimated to share mutation within the interval spanning *NRD1*. A first pass mutation screen, for example using DNA conformation analysis (Andersen *et al.* 2003) or DHPLC technology (Lilleberg 2003) would readily detect variation in samples for which sequencing or exon dosage analysis could then be carried out. Samples that carry SNP23 should also be included.

cDNA sequence analysis also identified a possible candidate. This was the *EPS15* locus which lies 400Kb 5' of the COLDMAP estimated location 95% CI. Three extra transcripts were observed for two of the cDNA primer sets in one case, P023 who was diagnosed with probable PD at age 59 years (Section 5.5.3). Sequencing of transcripts revealed alterations in copy number of exons 9, 10, 12 and 13. The region between exons 9 and 13 encodes part of the EH domain of EPS15, so variation within this may be functionally important. However, genomic sequencing of exons 9 to 13 (and flanking intronic regions) did not reveal any mutations that might affect

splicing. There is clearly one normal transcript present that appears to be the most abundant. Alterations in exon copy number may account for this and exon dosage analyses using semi-quantitative PCR (Section 2.2.9) would be able to determine this. Brain tissue expression analyses would be useful although post-mortem samples are not available, as we must consider this may be an artefact of the EBV-transformed cell line from which cDNA was derived. This could be ruled out by examining cDNA from fresh cell lines. However, one case represents >10% of all those analysed. It may therefore be possible that other individuals share variation within *EPS15* and late-onset individuals should also be prioritised for exon dosage analyses. This gene is an excellent candidate as a worm orthologue knockout exhibit defects in movement behaviour (Salcini *et al.* 2001) and the mouse orthologue has been implicated in PD pathways (Bonin *et al.* 2004)

Identification of functional conservation within *NRD1* and *EPS15* (using VISTA analysis; Section 4.4.6) may be useful for determining intronic regions that may harbour variation associated with disease. Any such regions should be also prioritised for mutation screening in late-onset patients. It would also be useful to further examine NRD1 and EPS15 in post-mortem brain samples. *In situ* hybridisation would determine the normal expression patterns and levels of these genes and also any differences in PD brains compared with normal controls. Since *NRD* may be neuronally expressed specifically during development only (Fumagalli *et al.* 1998), it is possible that variation within *NRD1* could have a detrimental effect on the development of viable neurons, able to withstand the insults of old age.

Significant association of SNP48 within the ubiquitin-specific protease 24 (*USP24*) gene is also worthy of additional study as the protein product is potentially involved

in the UPP. Examination of *USP24* halpotypes in cases and controls would determine the potential involvement of this locus with PD.

One of the major strengths of this study is that we have employed a large case-control series from a genetic homogenate, which can be beneficial in identifying subtle variation associated with complex disorders. Results of COLDMAP analyses in these samples have identified a relatively small putative mutation containing region at 1p32 which is remarkably close to the most tightly linked marker in the Icelandic population (Hicks *et al.* 2002). In addition, the 95% CI for mutation location (~116Kb) was in good agreement with that expected (150Kb) given the levels of LD at 1p32 (Section 5.3). However, there are factors which may have influenced these results. We employed SNP genotype data from all case and control samples available at the time in the initial COLDMAP runs. Given that 1p32 susceptibility is associated with late-onset PD (Hicks *et al.* 2002) and AAO (Li *et al.* 2002), it may be helpful to stratify data by age of onset and repeat the analyses, although we do not know if a decrease in sample n will result in a decrease in the power to identify mutation location. Marker density may also affect results. Our initial assessment of LD at 1p32 indicated a required marker density of 50Kb to detect association with disease. To minimise genotyping costs, the SNP density used in our initial analyses was ~170 Kb. Lack of suitably frequent SNPs in some regions also meant that considerable regions were not directly assessed. For example, there is ~1.8 Mb between SNPs 9 and 10 (Figure 5.6a). It would therefore be of benefit to generate additional SNP data within the region to include in future analyses, to determine if SNP density would alter the predicted mutation location. Morris and colleagues (2003) suggest that using htSNPs rather than SNP data does not dramatically reduce power (Morris *et al.* 2003),

therefore, additional genotyping of the most informative SNPs may offset the costs associated with additional data generation.

Extra support for linkage to 1p32 in Norwegians might also be achieved by linkage analysis in families. It would also be useful to examine PD cases from the US, in particular the Minnesota region which has been subject to significant Norwegian immigration in the past. It may prove difficult to construct such a well-matched case control series as we have used in this study, although family based methods may be used to overcome potential cryptic population stratification and genetic heterogeneity. Discordant sib-pair analysis may be appropriate, given that parental data may not be available for late-onset cases, and most PD cases are singletons (Rabinowitz 2001). Underlying levels of LD may also be lower than that in Norway (if individuals have not been subject to inbreeding) which may increase the resolution of mapping efforts.

5.9 Summary and Conclusions

PD susceptibility has been linked to 1p32 in both the Icelandic and US populations. Icelandic peoples are largely derived from the same gene pool as present day Norwegians and there has also been significant immigration from Norway to the USA. We hypothesised that 1p32-linked susceptibility may be of Norse origin and sought to identify this through both candidate and positional approaches in a large case control series from central Norway. Y chromosome haplotyping showed the Trondheim case control series to be a suitable population to map a gene of Norse descent and comparisons of LD in cases and controls suggested the possibility of shared ancestry at 1p32. By assessing underlying Norwegian population LD in this region, we predicted that a disease associated variant could be LD mapped to a resolution of ~150Kb. Indeed, multipoint LD mapping estimated a disease associated

allele to lie within a ~116Kb region, containing only the *NRD1* locus. A potential role for *EPS15*, in particular through alterations in normal transcription, was also identified through candidate gene analysis. The potential roles that variation within these genes may have in susceptibility to PD remain to be investigated.

6 General Discussion

6.1 Review of Aims and Experimental Approaches

PD is a complex disorder, for which genetic causes may be many. Families with causal mutations account for only a small percentage of PD cases (Bonifati *et al.* 2004). This study aimed to investigate recessive and susceptibility loci, in a homogeneous population from central Norway. Most cases of disease in this population appear sporadic in nature and may be the result of recessive and susceptibility variants.

Firstly we examined monogenic recessive loci implicated in parkinsonism – *PARK2*, *PARK6* and *PARK7* (Chapter 3). Semi-quantative PCR, DHPLC and sequence analysis was used to identify coding variants and alterations in *Parkin* exon copy number. STRP haplotype reconstruction was used to investigate the possibility of a founder effect at one of the variants identified. At the time of the start of this study, two further regions containing multiple genes, had been linked to monogenic recessively inherited parkinsonism. STRP genotypes, homozygosity screening and assessment of population allele frequencies was used to identify early onset individuals that might also be linked to *PARK6* or *PARK7* and therefore possibly useful for further narrowing candidate regions.

The second aim was to investigate the *MAPT* H1 haplotype associated with parkinsonism (Chapter 4). A case-control analysis was undertaken to investigate the association of H1 and clinically defined PD. Second, SNP association analysis was used in an attempt to identify the extent and boundaries of the *MAPT* extended haplotype. Third, the genetic architecture of H1 haplotypes was explored, by identifying SNPs specific to the H1 haplotype and those which represented htSNPs.

htSNP multilocus haplotype analysis, and case-control analysis of H1 sub-haplotypes was used to identify a specific haplotype and genetic region most associated with PD.

The third aim was to identify a novel parkinsonism susceptibility variant that has been linked to *PARK10* at chromosome 1p32 (Chapter 5). Y-chromosome haplogroup analysis was used to investigate population ancestry in Trondheim and LD analysis was used to investigate shared ancestry in the linked region. The extent and variation of LD at chromosome 1p32 was assessed through analysis of pairwise LD as a function of physical distance, and as a sliding average acoss 1p32. Third, to identify a susceptibility variant, both positional and candidate approaches were taken. Candidate genes were investigated through htSNP haplotype analysis and direct sequencing (genomic and cDNA). Multipoint LD mapping, using "shattered coalescent modelling of genealogies" was used to estimate the location of a PD susceptibility locus and the level of disease heterogeneity expected.

6.2 Summary of Findings

At the time the present study was initiated, two distinct regions of chromosome 1 (*PARK7* and *PARK6*) had been linked to early-onset recessive parkinsonism in European populations. The *PARK7* gene, *DJ-1,* was identified within the time-frame of the present study, therefore the two known mutations at that time, Δ1-5 and L166P, were screened for in individual P113 who showed extended regions of homozygosity for rare alleles in the region. Results were suggestive of a heterozygous deletion including exons 2 and 4. Haplotype analysis in Norwegians identified two cases (P095 and P023) that are homozygous for very rare STRP alleles at *PARK6*; *PINK1* mutation screening should therefore be prioritised in these individuals.

PRKN mutation carrier frequency is comparable between cases and normal controls, although assessment of the contribution of *PRKN* to PD in Norway awaits further screening. Carrier frequency, in particular for A82E, was high compared with an outbred, US case-control series. This was hypothesised to result from a common founder in the relatively isolated population and haplotype analysis about A82E indicated that this was likely.

We demonstrated *MAPT* H1 is a misnomer; it consists of a family of recombining H1 alleles and one *MAPT* H1 sub-haplotype is preferentially associated with Parkinson's disease. The disease associated region lies within a ~90kb interval in 5' *MAPT*, encompassing exons 1-4. We also suggest that the *CRHR1* locus has undergone positive selection and may account for the increase in H1 in human populations, given that H2 is likely the ancestral allele.

At *PARK10,* LD analysis suggested the possibility of shared ancestry in case samples. Multipoint LD mapping estimated a disease associated allele to lie within a ~116Kb region, containing only the *NRD1* locus. A potential role for *EPS15*, possibly through alterations in normal transcription, was also identified.

6.3 Synthesis

Molecular genetics has provided a considerable contribution to understanding not only the pathogenic mechanisms in PD, but also the complex processes in the normal and aging brain. This has been facilitated by the analysis of genes involved in rare inherited forms of the disease, in which variation can also factor in sporadic cases (Singleton *et al.* 2004). It is also apparent that rare genes identified within population isolates can also impact disease in the wider human population (e.g. *PRKN* and *DJ-1*).

In the present study, genetic analysis within a population isolate has shown that multiple genetic factors may be involved with PD susceptibility in Norwegians. Although genetic abnormalities were found in only a proportion of cases, it is possible that subtle variation within any of the genes already implicated in PD (since they can all be linked to some extent) may have a combined and complex effect on a common neurological pathway towards disease outcome. Many cases of apparently sporadic disease may be due to multiple, variably penetrant alleles. Comparison of a complete PD risk 'multigenotype' and detailed phenotypic and pathological analysis, will help to further elucidate the pathways involved in PD.

As additional susceptibility loci are uncovered, the arsenal of potential cellular and animal models will widen and provide powerful tools for the cause. Development of both preventive and palliative therapies is our goal; the molecular understanding acquired through genetic studies, past and future, will play an important role in achieving this.

6.4 Future Direction

6.4.1 Parkin

Complete screening of *PRKN* mutations in Norwegian case and controls will allow a better understanding of 1) the contribution of parkin to PD in Norway and 2) the population frequency of variants in the general population. In *PRKN* cases, collection of family samples and clinical data will allow further analysis of genotype-phenotype correlation and the pathogenesis of mutant *PRKN*. It also would be useful to further investigate the exon 3 proline insertion identified in case P154; *in vitro* studies would begin to shed light on the pathogenic mechanisms of this mutation.

6.4.2 DJ-1

Complete screening of *DJ-1* in P113 will confirm the suspected presence of a deletion that includes exons 2 and 4. It would also be of value to assess the contribution of *DJ-1* in disorders with pathological tau inclusions. Investigation of potential interactions between DJ-1 and tau may shed light on the underlying mechanisms that result in the association of *MAPT* in both tauopathies and PD.

6.4.3 PINK1

PINK 1 mutation screening in P095 and P023 is warranted.

6.4.4 MAPT H1

Support for the association of a particular *MAPT* H1 sub-haplotype in a region encompassing exons 1-4 could be obtained by H1 analysis in an alternate population. The hypothesis that the variability associated with PD affects exon 2 and 3 splicing could be tested by examining 0N/1N/2N ratios in brain samples. PSP cases should also be included in this analysis, and also in H1-SNP analysis, to help address the question of whether distinct mechanisms are responsible for *MAPT* association in PD and tauopathies. Exon 2 and 3 dosage studies in $H1^{PD}$ carriers will be able to rule out the possibility of a heterozygous deletion.

The hypothesis that H1 and H2 are inversion polymorphisms can be tested through macrorestriction mapping pulse field gel electrophoresis. Searching for the signatures of natural selection within other genes contained within the *MAPT* extended haplotype, will maybe shed light on why H1 has become the more common allele in human populations.

6.4.5 PARK10

To show that the association of *PARK10* in cases was not the result of population admixture, LD could be examined in other genomic areas unlinked to PD. Support for this locus harbouring PD susceptibility in Norway, would be gained by the identification of families and traditional linkage analysis. Mutation analysis of *EPS15* and *NRD1* in individuals that 'cluster' together (COLDMAP cladistic analysis) will determine the potential roles that these genes may have in PD. Haplotype analysis, around *USP24* will determine any involvement of this particular locus also.

7 References

Abecasis GR, Cookson WO (2000) *GOLD--graphical overview of linkage disequilibrium.* Bioinformatics **16**:182-3.

Abou-Sleiman PM, Healy DG, Quinn N, Lees AJ, Wood NW (2003) *The role of pathogenic DJ-1 mutations in Parkinson's disease.* Ann Neurol **54**:283-6

Aminoff MJ (2001) *Parkinson's disease.* Neurol Clin **19**:119-28, vi

Andersen PS, Jespersgaard C, Vuust J, Christiansen M, Larsen LA (2003) *High-throughput single strand conformation polymorphism mutation detection by automated capillary array electrophoresis: validation of the method.* Hum Mutat **21**:116-22

Andreadis A, Broderick JA, Kosik KS (1995) *Relative exon affinities and suboptimal splice site signals lead to non-equivalence of two cassette exons.* Nucleic Acids Res **23**:3585-93

Apaydin H, Ahlskog JE, Parisi JE, Boeve BF, Dickson DW (2002) *Parkinson Disease Neuropathology: Later-Developing Dementia and Loss of the Levodopa Response.* Arch Neurol **59**:102-112.

Ardlie KG, Kruglyak L, Seielstad M (2002) *Patterns of linkage disequilibrium in the human genome.* Nat Rev Genet **3**:299-309

Arnason E, Sigurgislason H, Benedikz E (2000) *Genetic homogeneity of Icelanders: fact or fiction?* Nat Genet **25**:373-4

Athanassiadou A, Voutsinas G, Psiouri L, Leroy E, Polymeropoulos MH, Ilias A, Maniatis GM, Papapetropoulos T (1999) *Genetic analysis of families with Parkinson disease that carry the Ala53Thr mutation in the gene encoding alpha-synuclein.* Am J Hum Genet **65**:555-8

Auluck PK, Chan HY, Trojanowski JQ, Lee VM, Bonini NM (2002) *Chaperone suppression of alpha-synuclein toxicity in a Drosophila model for Parkinson's disease.* Science **295**:865-8

Avila J, Lucas JJ, Perez M, Hernandez F (2004) *Role of tau protein in both physiological and pathological conditions.* Physiol Rev **84**:361-84

Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, Hardy J, Lynch T, Bigio E, Hutton M (1999) *Association of an extended haplotype in the tau gene with progressive supranuclear palsy.* Hum Mol Genet **8**:711-5

Bandopadhyay R, Kingsbury AE, Cookson MR, Reid AR, Evans IM, Hope AD, Pittman AM, Lashley T, Canet-Aviles R, Miller DW, McLendon C, Strand C, Leonard AJ, Abou-Sleiman PM, Healy DG, Ariga H, Wood NW, De Silva R, Revesz T, Hardy JA, Lees AJ (2004) *The expression of DJ-1 (PARK7) in normal human CNS and idiopathic Parkinson's disease.* Brain **127**:420-30

Bannon MJ, Pruetz B, Manning-Bog AB, Whitty CJ, Michelhaugh SK, Sacchetti P, Granneman JG, Mash DC, Schmidt CJ (2002) *Decreased expression of the transcription factor NURR1 in dopamine neurons of cocaine abusers.* Proc Natl Acad Sci U S A **99**:6382-5

Betarbet R, Sherer TB, MacKenzie G, Garcia-Osuna M, Panov AV, Greenamyre JT (2000) *Chronic systemic pesticide exposure reproduces features of Parkinson's disease.* Nat Neurosci **3**:1301-6

Bobadilla JL, Macek M, Jr., Fine JP, Farrell PM (2002) *Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening.* Hum Mutat **19**:575-606

Bodis-Wollner I (2003) *Neuropsychological and perceptual defects in Parkinson's disease.* Parkinsonism Relat Disord **9 Suppl 2**:S83-9

Bonifati V, Breedveld GJ, Squitieri F, Vanacore N, Brustenghi P, Harhangi BS, Montagna P, Cannella M, Fabbrini G, Rizzu P, van Duijn CM, Oostra BA, Meco G, Heutink P (2002) *Localization of autosomal recessive early-onset parkinsonism to chromosome 1p36 (PARK7) in an independent dataset.* Ann Neurol **51**:253-6

Bonifati V, Lucking CB, Fabrizio E, Periquet M, Meco G, Brice A (2001) *Three parkin gene mutations in a sibship with autosomal recessive early onset parkinsonism.* J Neurol Neurosurg Psychiatry **71**:531-4

Bonifati V, Oostra BA, Heutink P (2004) *Linking DJ-1 to neurodegeneration offers novel insights for understanding the pathogenesis of Parkinson's disease.* J Mol Med **82**:163-74

Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MC, Squitieri F, Ibanez P, Joosse M, van Dongen JW, Vanacore N, van Swieten JC, Brice A, Meco G, van Duijn CM, Oostra BA, Heutink P (2003) *Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism.* Science **299**:256-9

Bonin M, Poths S, Osaka H, Wang YL, Wada K, Riess O (2004) *Microarray expression analysis of gad mice implicates involvement of Parkinson's disease associated UCH-L1 in multiple metabolic pathways.* Brain Res Mol Brain Res **126**:88-97

Cardon LR, Abecasis GR (2003) *Using haplotype blocks to map human complex trait loci.* Trends Genet **19**:135-40

Cardon LR, Bell JI (2001) *Association study designs for complex diseases.* Nat Rev Genet **2**:91-9

Cardon LR, Palmer LJ (2003) *Population stratification and spurious allelic association.* Lancet **361**:598-604

Cesari R, Martin ES, Calin GA, Pentimalli F, Bichi R, McAdams H, Trapasso F, Drusco A, Shimizu M, Masciullo V, D'Andrilli G, Scambia G, Picchio MC, Alder H, Godwin AK, Croce CM (2003) *Parkin, a gene implicated in autosomal recessive juvenile parkinsonism, is a candidate tumor suppressor gene on chromosome 6q25-q27.* Proc Natl Acad Sci U S A **100**:5956-61

Chartier-Harlin MC (2004) *Alpha-synuclein locus duplication causes familial Parkinson's disease.* Lancet **In press**

Chiba-Falek O, Nussbaum RL (2001) *Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system.* Hum Mol Genet **10**:3101-9

Choi P, Snyder H, Petrucelli L, Theisler C, Chong M, Zhang Y, Lim K, Chung KK, Kehoe K, D'Adamio L, Lee JM, Cochran E, Bowser R, Dawson TM, Wolozin B (2003) *SEPT5_v2 is a parkin-binding protein.* Brain Res Mol Brain Res **117**:179-89

Chung KK, Dawson VL, Dawson TM (2001a) *The role of the ubiquitin-proteasomal pathway in Parkinson's disease and other neurodegenerative disorders.* Trends Neurosci **24**:S7-14

Chung KK, Zhang Y, Lim KL, Tanaka Y, Huang H, Gao J, Ross CA, Dawson VL, Dawson TM (2001b) *Parkin ubiquitinates the alpha-synuclein-interacting protein, synphilin-1: implications for Lewy-body formation in Parkinson disease.* Nat Med **7**:1144-50

Clark AG (2003) *Finding genes underlying risk of complex disease by linkage disequilibrium mapping.* Curr Opin Genet Dev **13**:296-302

Clark LN, Levy G, Tang MX, Mejia-Santana H, Ciappa A, Tycko B, Cote LJ, Louis ED, Mayeux R, Marder K (2003) *The Saitohin 'Q7R' polymorphism and tau haplotype in multi-ethnic Alzheimer disease and Parkinson's disease cohorts.* Neurosci Lett **347**:17-20

Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS (1998) *A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus.* Nat Genet **19**:292-6

Confalonieri S, Di Fiore PP (2002) *The Eps15 homology (EH) domain.* FEBS Lett **513**:24-9

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

Conrad C, Vianna C, Schultz C, Thal DR, Ghebremedhin E, Lenz J, Braak H, Davies P (2004) *Molecular evolution and genetics of the Saitohin gene and tau*

haplotype in Alzheimer's disease and argyrophilic grain disease. J Neurochem **89**:179-88

Consortium TIH (2003) *The International HapMap Project.* Nature **426**:789-96

Cookson MR, Lockhart PJ, McLendon C, O'Farrell C, Schlossmacher M, Farrer MJ (2003) *RING finger 1 mutations in Parkin produce altered localization of the protein.* Hum Mol Genet **12**:2957-65

Cox DG, Canzian F (2001) *Genotype transposer: automated genotype manipulation for linkage disequilibrium analysis.* Bioinformatics **17**:738-9

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) *High-resolution haplotype structure in the human genome.* Nat Genet **29**:229-32

Darchen F, Goud B (2000) *Multiple aspects of Rab protein action in the secretory pathway: focus on Rab3 and Rab6.* Biochimie **82**:375-84

Dauer W, Przedborski S (2003) *Parkinson's disease: mechanisms and models.* Neuron **39**:889-909

Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, *et al.* (2002) *A first-generation linkage disequilibrium map of human chromosome 22.* Nature **418**:544-8

Dawson TM, Dawson VL (2003) *Molecular pathways of neurodegeneration in Parkinson's disease.* Science **302**:819-22

de Silva R, Hardy J, Crook J, Khan N, Graham EA, Morris CM, Wood NW, Lees AJ (2002) *The tau locus is not significantly associated with pathologically confirmed sporadic Parkinson's disease.* Neurosci Lett **330**:201-3

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

Dekker MC, Bonifati V, van Duijn CM (2003) *Parkinson's disease: piecing together a genetic jigsaw.* Brain **126**:1722-33

DeStefano AL, Golbe LI, Mark MH, Lazzarini AM, Maher NE, Saint-Hilaire M, Feldman RG, *et al.* (2001) *Genome-wide scan for Parkinson's disease: the GenePD Study.* Neurology **57**:1124-6

DeStefano AL, Lew MF, Golbe LI, Mark MH, Lazzarini AM, Guttman M, Montgomery E, *et al.* (2002) *PARK3 influences age at onset in Parkinson disease: a genome scan in the GenePD study.* Am J Hum Genet **70**:1089-95

Devlin B, Roeder K, Wasserman L (2001) *Genomic control, a new approach to genetic-based association studies.* Theor Popul Biol **60**:155-66

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

Di Monte DA (2003) *The environment and Parkinson's disease: is the nigrostriatal system preferentially targeted by neurotoxins?* Lancet Neurol **2**:531-8

Djarmati A, Hedrich K, Svetel M, Schafer N, Juric V, Vukosavic S, Hering R, Riess O, Romac S, Klein C, Kostic V (2004) *Detection of Parkin (PARK2) and DJ1 (PARK7) mutations in early-onset Parkinson disease: Parkin mutation frequency depends on ethnic origin of patients.* Hum Mutat **23**:525

Duda JE, Giasson BI, Mabon ME, Miller DC, Golbe LI, Lee VM, Trojanowski JQ (2002) *Concurrence of alpha-synuclein and tau brain pathology in the Contursi kindred.* Acta Neuropathol (Berl) **104**:7-11

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) *The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes.* Nat Genet **25**:320-3

Edland SD, Slager S, Farrer M (2004) *Genetic association studies in Alzheimer's disease research: challenges and opportunities.* Stat Med **23**:169-78

Elbaz A, Levecque C, Clavel J, Vidal JS, Richard F, Correze JR, Delemotte B, Amouyel P, Alperovitch A, Chartier-Harlin MC, Tzourio C (2003) *S18Y polymorphism in the UCH-L1 gene and Parkinson's disease: evidence for an age-dependent relationship.* Mov Disord **18**:130-7

Fahn S (2000) *The spectrum of levodopa-induced dyskinesias.* Ann Neurol **47**:S2-9; discussion S9-11

Fahn S (2003) *Description of Parkinson's disease as a clinical syndrome.* Ann N Y Acad Sci **991**:1-14

Farrer M, Chan P, Chen R, Tan L, Lincoln S, Hernandez D, Forno L, Gwinn-Hardy K, Petrucelli L, Hussey J, Singleton A, Tanner C, Hardy J, Langston JW (2001a) *Lewy bodies and parkinsonism in families with parkin mutations.* Ann Neurol **50**:293-300

Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, Maraganore D, Gwinn-Hardy K, Wszolek Z, Dickson D, Langston JW (2004) *Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications.* Ann Neurol **55**:174-9

Farrer M, Maraganore DM, Lockhart P, Singleton A, Lesnick TG, de Andrade M, West A, de Silva R, Hardy J, Hernandez D (2001b) *alpha-Synuclein gene haplotypes are associated with Parkinson's disease.* Hum Mol Genet **10**:1847-51

Farrer M, Skipper L, Berg M, Bisceglio G, Hanson M, Hardy J, Adam A, Gwinn-Hardy K, Aasly J (2002) *The tau H1 haplotype is associated with Parkinson's disease in the Norwegian population.* Neurosci Lett **322**:83-6

Farrer M, Wavrant-De Vrieze F, Crook R, Boles L, Perez-Tur J, Hardy J, Johnson WG, Steele J, Maraganore D, Gwinn K, Lynch T (1998) *Low frequency of alpha-synuclein mutations in familial Parkinson's disease.* Ann Neurol **43**:394-7

Feany MB, Pallanck LJ (2003) *Parkin: a multipurpose neuroprotective agent?* Neuron **38**:13-6

Foroud T, Uniacke SK, Liu L, Pankratz N, Rudolph A, Halter C, Shults C, Marder K, Conneally PM, Nichols WC (2003) *Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease.* Neurology **60**:796-801

Fumagalli P, Accarino M, Egeo A, Scartezzini P, Rappazzo G, Pizzuti A, Avvantaggiato V, Simeone A, Arrigo G, Zuffardi O, Ottolenghi S, Taramelli R (1998) *Human NRD convertase: a highly conserved metalloendopeptidase expressed at specific sites during development and in adult tissues.* Genomics **47**:238-45

Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F (2002) *A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1.* Ann Neurol **51**:296-301

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) *The structure of haplotype blocks in the human genome.* Science **296**:2225-9

Galvin JE, Lee VM, Baba M, Mann DM, Dickson DW, Yamaguchi H, Schmidt ML, Iwatsubo T, Trojanowski JQ (1997) *Monoclonal antibodies to purified cortical Lewy bodies recognize the mid-size neurofilament subunit.* Ann Neurol **42**:595-603

Garcia-Borreguero D, Larrosa O, Bravo M (2003) *Parkinson's disease and sleep.* Sleep Med Rev **7**:115-29

Garner C, Slatkin M (2003) *On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci.* Genet Epidemiol **24**:57-67

Gasser T, Muller-Myhsok B, Wszolek ZK, Oehlmann R, Calne DB, Bonifati V, Bereznai B, Fabrizio E, Vieregge P, Horstmann RD (1998) *A susceptibility locus for Parkinson's disease maps to chromosome 2p13.* Nat Genet **18**:262-5

Gedde-Dahl T, Jr. (1973) *Population structure in Norway. Inbreeding, distance and kinship.* Hereditas **73**:211-32

Gelb DJ, Oliver E, Gilman S (1999) *Diagnostic criteria for Parkinson disease.* Arch Neurol **56**:33-9

Gill SS, Patel NK, Hotton GR, O'Sullivan K, McCarter R, Bunnage M, Brooks DJ, Svendsen CN, Heywood P (2003) *Direct brain infusion of glial cell line-derived neurotrophic factor in Parkinson disease.* Nat Med **9**:589-95

Goedert M (2003) *Neurodegenerative tauopathy in the worm.* Proc Natl Acad Sci U S A **100**:9653-5

Golbe LI, Di Iorio G, Bonavita V, Miller DC, Duvoisin RC (1990) *A large kindred with autosomal dominant Parkinson's disease.* Ann Neurol **27**:276-82

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

Goldberg MS, Fleming SM, Palacino JJ, Cepeda C, Lam HA, Bhatnagar A, Meloni EG, Wu N, Ackerson LC, Klapstein GJ, Gajendiran M, Roth BL, Chesselet MF, Maidment NT, Levine MS, Shen J (2003) *Parkin-deficient mice exhibit nigrostriatal deficits but not loss of dopaminergic neurons.* J Biol Chem **278**:43628-35

Grammatopoulos DK, Chrousos GP (2002) *Functional characteristics of CRH receptors and potential clinical applications of CRH-receptor antagonists.* Trends Endocrinol Metab **13**:436-44

Guttman M, Kish SJ, Furukawa Y (2003a) *Current concepts in the diagnosis and management of Parkinson's disease.* Cmaj **168**:293-301

Guttman M, Slaughter PM, Theriault ME, DeBoer DP, Naylor CD (2003b) *Burden of parkinsonism: a population-based study.* Mov Disord **18**:313-9

Gwinn-Hardy K (2002) *Genetics of parkinsonism.* Mov Disord **17**:645-56

Hague S, Rogaeva E, Hernandez D, Gulick C, Singleton A, Hanson M, Johnson J, Weiser R, Gallardo M, Ravina B, Gwinn-Hardy K, Crawley A, St George-Hyslop PH, Lang AE, Heutink P, Bonifati V, Hardy J (2003) *Early-onset*

Parkinson's disease caused by a compound heterozygous DJ-1 mutation. Ann Neurol **54**:271-4

Harada A, Oguchi K, Okabe S, Kuno J, Terada S, Ohshima T, Sato-Yoshitake R, Takei Y, Noda T, Hirokawa N (1994) *Altered microtubule organization in small-calibre axons of mice lacking tau protein.* Nature **369**:488-91

Harhangi BS, Farrer MJ, Lincoln S, Bonifati V, Meco G, De Michele G, Brice A, Durr A, Martinez M, Gasser T, Bereznai B, Vaughan JR, Wood NW, Hardy J, Oostra BA, Breteler MM (1999) *The Ile93Met mutation in the ubiquitin carboxy-terminal-hydrolase-L1 gene is not observed in European cases with familial Parkinson's disease.* Neurosci Lett **270**:1-4

Healy DG, Abou-Sleiman PM, Valente EM, Gilks WP, Bhatia K, Quinn N, Lees AJ, Wood NW (2004) *DJ-1 mutations in Parkinson's disease.* J Neurol Neurosurg Psychiatry **75**:144-5

Hedrich K, Djarmati A, Schafer N, Hering R, Wellenbrock C, Weiss PH, Hilker R, Vieregge P, Ozelius LJ, Heutink P, Bonifati V, Schwinger E, Lang AE, Noth J, Bressman SB, Pramstaller PP, Riess O, Klein C (2004) *DJ-1 (PARK7) mutations are less frequent than Parkin (PARK2) mutations in early-onset Parkinson disease.* Neurology **62**:389-94

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B (2001) *mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry.* Am J Hum Genet **68**:723-37

Helgason A, Sigureth ardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K (2000) *Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland.* Am J Hum Genet **67**:697-717

Heng CK, Low PS (2000) *Cladistic analysis: its applications in association studies of complex diseases.* Ann Acad Med Singapore **29**:313-21

Hey J (2003) *Speciation and inversions: chimps and humans.* Bioessays **25**:825-8

Hicks AA, Petursson H, Jonsson T, Stefansson H, Johannsdottir HS, Sainz J, Frigge ML, Kong A, Gulcher JR, Stefansson K, Sveinbjornsdottir S (2002) *A susceptibility gene for late-onset idiopathic Parkinson's disease.* Ann Neurol **52**:549-55

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) *A comprehensive review of genetic association studies.* Genet Med **4**:45-61

Hod Y, Pentyala SN, Whyard TC, El-Maghrabi MR (1999) *Identification and characterization of a novel protein that regulates RNA-protein interaction.* J Cell Biochem **72**:435-44

Hoenicka J, Perez M, Perez-Tur J, Barabash A, Godoy M, Vidal L, Astarloa R, Avila J, Nygaard T, de Yebenes JG (1999) *The tau gene A0 allele and progressive supranuclear palsy.* Neurology **53**:1219-25

Honbou K, Suzuki NN, Horiuchi M, Niki T, Taira T, Ariga H, Inagaki F (2003) *The crystal structure of DJ-1, a protein related to male fertility and Parkinson's disease.* J Biol Chem

Hosking LK, Boyd PR, Xu CF, Nissum M, Cantone K, Purvis IJ, Khakhar R, Barnes MR, Liberwirth U, Hagen-Mann K, Ehm MG, Riley JH (2002) *Linkage disequilibrium mapping identifies a 390 kb region associated with CYP2D6 poor drug metabolising activity.* Pharmacogenomics J **2**:165-75

Hospital V, Chesneau V, Balogh A, Joulie C, Seidah NG, Cohen P, Prat A (2000) *N-arginine dibasic convertase (nardilysin) isoforms are soluble dibasic-specific metalloendopeptidases that localize in the cytoplasm and at the cell surface.* Biochem J **349**:587-97

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

Hutton M (2001) *Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms.* Neurology **56**:S21-5

Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, *et al.* (1998) *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature **393**:702-5

Huynh DP, Scoles DR, Nguyen D, Pulst SM (2003) *The autosomal recessive juvenile Parkinson disease gene product, parkin, interacts with and ubiquitinates synaptotagmin XI.* Hum Mol Genet **12**:2587-97

Ibanez P, De Michele G, Bonifati V, Lohmann E, Thobois S, Pollak P, Agid Y, Heutink P, Durr A, Brice A (2003) *Screening for DJ-1 mutations in early onset autosomal recessive parkinsonism.* Neurology **61**:1429-31

Imai Y, Soda M, Inoue H, Hattori N, Mizuno Y, Takahashi R (2001) *An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin.* Cell **105**:891-902

Imai Y, Soda M, Murakami T, Shoji M, Abe K, Takahashi R (2003) *A product of the human gene adjacent to parkin is a component of Lewy bodies and suppresses Pael receptor-induced cell death.* J Biol Chem **278**:51901-10

Isacson O, Bjorklund LM, Schumacher JM (2003) *Toward full restoration of synaptic and terminal function of the dopaminergic system in Parkinson's disease by stem cells.* Ann Neurol **53 Suppl 3**:S135-46; discussion S146-8

Ishizawa T, Mattila P, Davies P, Wang D, Dickson DW (2003) *Colocalization of tau and alpha-synuclein epitopes in Lewy bodies.* J Neuropathol Exp Neurol **62**:389-97

Itier JM, Ibanez P, Mena MA, Abbas N, Cohen-Salmon C, Bohme GA, Laville M, *et al.* (2003) *Parkin gene inactivation alters behaviour and dopamine neurotransmission in the mouse.* Hum Mol Genet **12**:2277-91

Jeffreys AJ, Kauppi L, Neumann R (2001) *Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex.* Nat Genet **29**:217-22

Jensen PH, Hager H, Nielsen MS, Hojrup P, Gliemann J, Jakes R (1999) *alpha-synuclein binds to Tau and stimulates the protein kinase A-catalyzed tau phosphorylation of serine residues 262 and 356.* J Biol Chem **274**:25481-9

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) *Haplotype tagging for the identification of common disease genes.* Nat Genet **29**:233-7

Jorgensen TH, Degn B, Wang AG, Vang M, Gurling H, Kalsi G, McQuillin A, Kruse TA, Mors O, Ewald H (2002) *Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands.* Eur J Hum Genet **10**:381-7

Kahle PJ, Leimer U, Haass C (2000) *Does failure of parkin-mediated ubiquitination cause juvenile parkinsonism?* Trends Biochem Sci **25**:524-7

Kaplan B, Ratner V, Haas E (2003) *Alpha-synuclein: its biological function and role in neurodegenerative diseases.* J Mol Neurosci **20**:83-92

Ke X, Cardon LR (2003) *Efficient selective screening of haplotype tag SNPs.* Bioinformatics **19**:287-8

Ke X, Hunt S, Tapper W, Lawrence R, Stavrides G, Ghori J, Whittaker P, Collins A, Morris AP, Bentley D, Cardon LR, Deloukas P (2004) *The impact of SNP density on fine-scale patterns of linkage disequilibrium.* Hum Mol Genet **13**:577-88

Kehrer-Sawatzki H, Schreiner B, Tanzer S, Platzer M, Muller S, Hameister H (2002) *Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17.* Am J Hum Genet **71**:375-88

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) *Identification of the cystic fibrosis gene: genetic analysis.* Science **245**:1073-80

King ME, Gamblin TC, Kuret J, Binder LI (2000) *Differential assembly of human tau isoforms in the presence of arachidonic acid.* J Neurochem **74**:1749-57

Kirik D, Georgievska B, Bjorklund A (2004) *Localized striatal delivery of GDNF as a treatment for Parkinson disease.* Nat Neurosci **7**:105-10

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N (1998) *Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.* Nature **392**:605-8

Klapisz E, Sorokina I, Lemeer S, Pijnenburg M, Verkleij AJ, van Bergen en Henegouwen PM (2002) *A ubiquitin-interacting motif (UIM) is essential for Eps15 and Eps15R ubiquitination.* J Biol Chem **277**:30746-53

Klinefelter GR, Welch JE, Perreault SD, Moore HD, Zucker RM, Suarez JD, Roberts NL, Bobseine K, Jeffay S (2002) *Localization of the sperm protein SP22 and inhibition of fertility in vivo and in vitro.* J Androl **23**:48-63

Kruger R, Kuhn W, Muller T, Woitalla D, Graeber M, Kosel S, Przuntek H, Epplen JT, Schols L, Riess O (1998) *Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease.* Nat Genet **18**:106-8

Kruger R, Vieira-Saecker AM, Kuhn W, Berg D, Muller T, Kuhnl N, Fuchs GA, Storch A, Hungs M, Woitalla D, Przuntek H, Epplen JT, Schols L, Riess O (1999) *Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype.* Ann Neurol **45**:611-7

Kruglyak L (1999) *Prospects for whole-genome linkage disequilibrium mapping of common disease genes.* Nat Genet **22**:139-44

Kuopio A, Marttila RJ, Helenius H, Rinne UK (2001) *Familial occurrence of Parkinson's disease in a community-based case-control study.* **7**:297-303

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

Laitinen T (2002) *The value of isolated populations in genetic studies of allergic diseases.* Curr Opin Allergy Clin Immunol **2**:379-82

Lang AE, Lozano AM (1998) *Parkinson's disease. First of two parts.* N Engl J Med **339**:1044-53.

Langston JW, Ballard P, Tetrud JW, Irwin I (1983) *Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis.* Science **219**:979-80

Larsen CN, Krantz BA, Wilkinson KD (1998) *Substrate specificity of deubiquitinating enzymes: ubiquitin C-terminal hydrolases.* Biochemistry **37**:3358-68

Lavedan C, Buchholtz S, Nussbaum RL, Albin RL, Polymeropoulos MH (2002) *A mutation in the human neurofilament M gene in Parkinson's disease that suggests a role for the cytoskeleton in neuronal degeneration.* Neurosci Lett **322**:57-61

Lazzarini AM, Golbe LI, Dickson DW, Duvoisin RC, Johnson WG (1997) *Tau intronic polymorphism in Parkinson's Disease and progressive supranuclear palsy.* Neurology **48**:A427 (abstract)

Le WD, Xu P, Jankovic J, Jiang H, Appel SH, Smith RG, Vassilatis DK (2003) *Mutations in NR4A2 associated with familial Parkinson disease.* Nat Genet **33**:85-9

Lee SJ, Kim SJ, Kim IK, Ko J, Jeong CS, Kim GH, Park C, Kang SO, Suh PG, Lee HS, Cha SS (2003) *Crystal structures of human DJ-1 and Escherichia coli Hsp31, which share an evolutionarily conserved domain.* J Biol Chem **278**:44552-9

Lee VM, Goedert M, Trojanowski JQ (2001) *Neurodegenerative tauopathies.* Annu Rev Neurosci **24**:1121-59

Leroy E, Boyer R, Auburger G, Leube B, Ulm G, Mezey E, Harta G, Brownstein MJ, Jonnalagada S, Chernova T, Dehejia A, Lavedan C, Gasser T, Steinbach PJ, Wilkinson KD, Polymeropoulos MH (1998) *The ubiquitin pathway in Parkinson's disease.* Nature **395**:451-2

Levecque C, Elbaz A, Clavel J, Vidal JS, Amouyel P, Alperovitch A, Tzourio C, Chartier-Harlin MC (2004) *Association of polymorphisms in the Tau and Saitohin genes with Parkinson's disease.* J Neurol Neurosurg Psychiatry **75**:478-80

Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, *et al.* (2002) *Age at onset in two common neurodegenerative diseases is genetically controlled.* Am J Hum Genet **70**:985-93

Lilleberg SL (2003) *In-depth mutation and SNP discovery using DHPLC gene scanning.* Curr Opin Drug Discov Devel **6**:237-52

Lincoln S, Vaughan J, Wood N, Baker M, Adamson J, Gwinn-Hardy K, Lynch T, Hardy J, Farrer M (1999) *Low frequency of pathogenic mutations in the ubiquitin carboxy-terminal hydrolase gene in familial Parkinson's disease.* Neuroreport **10**:427-9

Lincoln SJ, Maraganore DM, Lesnick TG, Bounds R, de Andrade M, Bower JH, Hardy JA, Farrer MJ (2003) *Parkin variants in North American Parkinson's disease: cases and controls.* Mov Disord **18**:1306-11

Litvan I, Bhatia KP, Burn DJ, Goetz CG, Lang AE, McKeith I, Quinn N, Sethi KD, Shults C, Wenning GK (2003) *SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders.* Mov Disord **18**:467-86

Litvan I, Lees AJ (1999) *Progressive supranuclear palsy.* Adv Neurol **80**:341-5

Liu X, Barker DF (1999) *Evidence for effective suppression of recombination in the chromosome 17q21 segment spanning RNU2-BRCA1.* Am J Hum Genet **64**:1427-39

Liu Y, Fallon L, Lashuel HA, Liu Z, Lansbury PT, Jr. (2002) *The UCH-L1 gene encodes two opposing enzymatic activities that affect alpha-synuclein degradation and Parkinson's disease susceptibility.* Cell **111**:209-18

Lockhart PJ, Lincoln S, Hulihan M, Kachergus J, Wilkes K, Bisceglio G, Mash DC, Farrer MJ (2004a) *DJ-1 mutations are a rare cause of recessively inherited early onset parkinsonism mediated by loss of protein function.* J Med Genet **41**:e22

Lockhart PJ, O'Farrell CA, Farrer MJ (2004b) *It's a double knock-out! The quaking mouse is a spontaneous deletion of parkin and parkin co-regulated gene (PACRG).* Mov Disord **19**:101-4

Lonjou C, Zhang W, Collins A, Tapper WJ, Elahi E, Maniatis N, Morton NE (2003) *Linkage disequilibrium in human populations.* Proc Natl Acad Sci U S A **100**:6069-74

Lotharius J, Brundin P (2002) *Pathogenesis of Parkinson's disease: dopamine, vesicles and alpha-synuclein.* Nat Rev Neurosci **3**:932-42

Lowe J, McDermott H, Landon M, Mayer RJ, Wilkinson KD (1990) *Ubiquitin carboxyl-terminal hydrolase (PGP 9.5) is selectively present in ubiquitinated inclusion bodies characteristic of human neurodegenerative diseases.* J Pathol **161**:153-60

Lucking CB, Durr A, Bonifati V, Vaughan J, De Michele G, Gasser T, Harhangi BS, Meco G, Denefle P, Wood NW, Agid Y, Brice A (2000) *Association between early-onset Parkinson's disease and mutations in the parkin gene. French Parkinson's Disease Genetics Study Group.* N Engl J Med **342**:1560-7

Maniatis N, Collins A, Xu CF, McCarthy LC, Hewett DR, Tapper W, Ennis S, Ke X, Morton NE (2002) *The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis.* Proc Natl Acad Sci U S A **99**:2228-33

Maraganore DM, Farrer MJ, Hardy JA, Lincoln SJ, McDonnell SK, Rocca WA (1999) *Case-control study of the ubiquitin carboxy-terminal hydrolase L1 gene in Parkinson's disease.* Neurology **53**:1858-60

Maraganore DM, Hernandez DG, Singleton AB, Farrer MJ, McDonnell SK, Hutton ML, Hardy JA, Rocca WA (2001) *Case-Control study of the extended tau gene haplotype in Parkinson's disease.* Ann Neurol **50**:658-61

Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, *et al.* (2001) *Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease.* Jama **286**:2245-50

Mata IF, Lockhart PJ, Farrer MJ (2004) *Parkin genetics: one model for Parkinson's disease.* Hum Mol Genet **13 Spec No 1**:R127-33

Mattson MP (2003) *Gene-diet interactions in brain aging and neurodegenerative disorders.* Ann Intern Med **139**:441-4

McCormack AL, Thiruchelvam M, Manning-Bog AB, Thiffault C, Langston JW, Cory-Slechta DA, Di Monte DA (2002) *Environmental risk factors and Parkinson's disease: selective degeneration of nigral dopaminergic neurons caused by the herbicide paraquat.* Neurobiol Dis **10**:119-27

McDonald OG, Krynetski EY, Evans WE (2002) *Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation.* Pharmacogenetics **12**:93-9

McDonald WM, Richard IH, DeLong MR (2003) *Prevalence, etiology, and treatment of depression in Parkinson's disease.* Biol Psychiatry **54**:363-75

McKeigue PM (1998) *Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture.* Am J Hum Genet **63**:241-51

McNaught KS, Belizaire R, Isacson O, Jenner P, Olanow CW (2003) *Altered proteasomal function in sporadic Parkinson's disease.* Exp Neurol **179**:38-46

McNaught KS, Mytilineou C, Jnobaptiste R, Yabut J, Shashidharan P, Jennert P, Olanow CW (2002) *Impairment of the ubiquitin-proteasome system causes dopaminergic cell death and inclusion body formation in ventral mesencephalic cultures.* J Neurochem **81**:301-6

Miller RJ, Wilson SM (2003) *Neurological disease: UPS stops delivering!* Trends Pharmacol Sci **24**:18-23

Menasche G, Pastural E, Feldmann J, Certain S, Ersoy F, Dupuis S, Wulffraat N, Bianchi D, Fischer A, Le Deist F, de Saint Basile G (2000) *Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome.* Nat Genet **25**:173-6

Molitor J, Marjoram P, Thomas D (2003) *Fine-scale mapping of disease genes with multiple mutations via spatial clustering techniques.* Am J Hum Genet **73**:1368-84

Momose Y, Murata M, Kobayashi K, Tachikawa M, Nakabayashi Y, Kanazawa I, Toda T (2002) *Association studies of multiple candidate genes for Parkinson's disease using single nucleotide polymorphisms.* Ann Neurol **51**:133-6

Moore DJ, Dawson VL, Dawson TM (2003) *Genetics of Parkinson's disease: what do mutations in DJ-1 tell us?* Ann Neurol **54**:281-2

Mori H, Hattori N, Mizuno Y (2003) *Genotype-phenotype correlation: familial Parkinson disease.* Neuropathology **23**:90-4

Mori H, Oda M, Komori T, Arai N, Takanashi M, Mizutani T, Hirai S, Mizuno Y (2002) *Lewy bodies in progressive supranuclear palsy.* Acta Neuropathol (Berl) **104**:273-8

Morris AP, Whittaker JC, Balding DJ (2002) *Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies.* Am J Hum Genet **70**:686-707

Morris AP, Whittaker JC, Xu CF, Hosking LK, Balding DJ (2003) *Multipoint linkage-disequilibrium mapping narrows location interval and identifies mutation heterogeneity.* Proc Natl Acad Sci U S A **100**:13442-6

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

Mueller RF, Young ID (1996) Emery's Elements of Medical Genetics. Churchill Livingstone

Muller T, Buttner T, Gholipour AF, Kuhn W (2003) *Coenzyme Q10 supplementation provides mild symptomatic benefit in patients with Parkinson's disease.* Neurosci Lett **341**:201-4

Nagakubo D, Taira T, Kitaura H, Ikeda M, Tamai K, Iguchi-Ariga SM, Ariga H (1997) *DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras.* Biochem Biophys Res Commun **231**:509-13

Neumann M, Muller V, Gorner K, Kretzschmar HA, Haass C, Kahle PJ (2004) *Pathological properties of the Parkinson's disease-associated protein DJ-1 in alpha-synucleinopathies and tauopathies: relevance for multiple system atrophy and Pick's disease.* Acta Neuropathol (Berl) **107**:489-96

Oliveira SA, Scott WK, Martin ER, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Ondo WG, Allen FH, Jr., Scott BL, Goetz CG, Small GW, Mastaglia F, Stajich JM, Zhang F, Booze MW, Winn MP, Middleton LT, Haines JL, Pericak-Vance MA, Vance JM (2003) *Parkin mutations and susceptibility alleles in late-onset Parkinson's disease.* Ann Neurol **53**:624-9

Olzmann JA, Brown K, Wilkinson KD, Rees HD, Huai Q, Ke H, Levey AI, Li L, Chin LS (2003) *Familial Parkinson's disease-associated L166P mutation disrupts DJ-1 protein folding and function.* J Biol Chem

Orth M, Tabrizi SJ (2003) *Models of Parkinson's disease.* Mov Disord **18**:729-37

Pals P (2004) *The alpha-synuclein promoter confers susceptibility to Parkinson's disease.* Submitted

Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, Conneally PM, Foroud T (2003) *Significant linkage of Parkinson disease to chromosome 2q36-37.* Am J Hum Genet **72**:1053-7

Parkinson J (1817) An Essay on the Shaking Palsy

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, Underhill PA (2002) *Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms.* Eur J Hum Genet **10**:521-9

Pastor P, Ezquerra M, Munoz E, Marti MJ, Blesa R, Tolosa E, Oliva R (2000) *Significant association between the tau gene A0/A0 genotype and Parkinson's disease.* Ann Neurol **47**:242-5

Pastor P, Ezquerra M, Tolosa E, Munoz E, Marti MJ, Valldeoriola F, Molinuevo JL, Calopa M, Oliva R (2002) *Further extension of the H1 haplotype associated with progressive supranuclear palsy.* Mov Disord **17**:550-6

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR (2001) *Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21.* Science **294**:1719-23

Payami H, Zareparsi S, James D, Nutt J (2002) *Familial aggregation of Parkinson disease: a comparative study of early-onset and late-onset disease.* Arch Neurol **59**:848-50

Pedersen WA, Wan R, Zhang P, Mattson MP (2002) *Urocortin, but not urocortin II, protects cultured hippocampal neurons from oxidative and excitotoxic cell death via corticotropin-releasing hormone receptor type I.* J Neurosci **22**:404-12

Peltonen L, Palotie A, Lange K (2000) *Use of population isolates for mapping complex traits.* Nat Rev Genet **1**:182-90

Periquet M, Latouche M, Lohmann E, Rawal N, De Michele G, Ricard S, Teive H, Fraix V, Vidailhet M, Nicholl D, Barone P, Wood NW, Raskin S, Deleuze JF, Agid Y, Durr A, Brice A (2003) *Parkin mutations are frequent in patients with isolated early-onset parkinsonism.* Brain **126**:1271-8

Periquet M, Lucking C, Vaughan J, Bonifati V, Durr A, De Michele G, Horstink M, Farrer M, Illarioshkin SN, Pollak P, Borg M, Brefel-Courbon C, Denefle P, Meco G, Gasser T, Breteler MM, Wood N, Agid Y, Brice A (2001) *Origin of the mutations in the parkin gene in Europe: exon rearrangements are independent recurrent events, whereas point mutations may result from Founder effects.* Am J Hum Genet **68**:617-26

Peters A (2002) *Structural changes in the normally aging cerebral cortex of primates.* Prog Brain Res **136**:455-65

Petrucelli L, O'Farrell C, Lockhart PJ, Baptista M, Kehoe K, Vink L, Choi P, Wolozin B, Farrer M, Hardy J, Cookson MR (2002) *Parkin protects against the toxicity associated with mutant alpha-synuclein: proteasome dysfunction selectively affects catecholaminergic neurons.* Neuron **36**:1007-19

Piccini P, Brooks DJ (1999) *Etiology of Parkinson's disease: contributions from 18F-DOPA positron emission tomography.* Adv Neurol **80**:227-31

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

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL (1997) *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.* Science **276**:2045-7

Poorkaj P, Bird TD, Wijsman E, Nemens E, Garruto RM, Anderson L, Andreadis A, Wiederholt WC, Raskind M, Schellenberg GD (1998) *Tau is a candidate gene for chromosome 17 frontotemporal dementia.* Ann Neurol **43**:815-25

Poorkaj P, Kas A, D'Souza I, Zhou Y, Pham Q, Stone M, Olson MV, Schellenberg GD (2001) *A genomic sequence analysis of the mouse and human microtubule-associated protein tau.* Mamm Genome **12**:700-12

Powers KM, Smith-Weller T, Franklin GM, Longstreth WT, Jr., Swanson PD, Checkoway H (2003) *Parkinson's disease risks associated with dietary iron, manganese, and other nutrient intakes.* Neurology **60**:1761-1766

Prasher VP, Farrer MJ, Kessling AM, Fisher EM, West RJ, Barber PC, Butler AC (1998) *Molecular mapping of Alzheimer-type dementia in Down's syndrome.* Ann Neurol **43**:380-3

Pritchard JK, Przeworski M (2001) *Linkage disequilibrium in humans: models and data.* Am J Hum Genet **69**:1-14

Pritchard JK, Rosenberg NA (1999) *Use of unlinked genetic markers to detect population stratification in association studies.* Am J Hum Genet **65**:220-8

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) *Association mapping in structured populations.* Am J Hum Genet **67**:170-81

Priyadarshi A, Khuder SA, Schaub EA, Priyadarshi SS (2001) *Environmental risk factors and Parkinson's disease: a metaanalysis.* Environ Res **86**:122-7

Rabinowitz D (2001) *Unbiased discordant sib-pair tests when parental genotypes are missing.* Am J Med Genet **105**:57-9

Rannala B (2001) *Finding genes influencing susceptibility to complex diseases in the post-genome era.* Am J Pharmacogenomics **1**:203-21

Rannala B, Reeve JP (2001) *High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence.* Am J Hum Genet **69**:159-78

Rehman HU (2002) *Role of CRH in the pathogenesis of dementia of Alzheimer's type and other dementias.* Curr Opin Investig Drugs **3**:1637-42

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) *Linkage disequilibrium in the human genome.* Nature **411**:199-204

Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, *et al.* (2001) *Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease.* Nat Genet **29**:223-8

Risch NJ (2000) *Searching for genetic determinants in the new millennium.* Nature **405**:847-56

Rizzu P, Hinkle DA, Zhukareva V, Bonifati V, Severijnen LA, Martinez D, Ravid R, Kamphorst W, Eberwine JH, Lee VM, Trojanowski JQ, Heutink P (2004) *DJ-1 colocalizes with tau inclusions: a link between parkinsonism and dementia.* Ann Neurol **55**:113-8

Robinow S, White K (1988) *The locus elav of Drosophila melanogaster is expressed in neurons at all developmental stages.* Dev Biol **126**:294-303

Rosenberg NA, Nordborg M (2002) *Genealogical trees, coalescent theory and the analysis of genetic polymorphisms.* Nat Rev Genet **3**:380-90

Ross GW, Petrovitch H (2001) *Current evidence for neuroprotective effects of nicotine and caffeine against Parkinson's disease.* Drugs Aging **18**:797-806

Ryu SW, Chae SK, Lee KJ, Kim E (1999) *Identification and characterization of human Fas associated factor 1, hFAF1.* Biochem Biophys Res Commun **262**:388-94

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) *Detecting recent positive selection in the human genome from haplotype structure.* Nature **419**:832-7

Salcini AE, Chen H, Iannolo G, De Camilli P, Di Fiore PP (1999) *Epidermal growth factor pathway substrate 15, Eps15.* Int J Biochem Cell Biol **31**:805-9

Salcini AE, Hilliard MA, Croce A, Arbucci S, Luzzi P, Tacchetti C, Daniell L, De Camilli P, Pelicci PG, Di Fiore PP, Bazzicalupo P (2001) *The Eps15 C. elegans homologue EHS-1 is implicated in synaptic vesicle recycling.* Nat Cell Biol **3**:755-60

Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ, Stephens JC (2003) *SNP and haplotype variation in the human genome.* Mutat Res **526**:53-61

Sambamurti K, Greig NH, Lahiri DK (2002) *Advances in the cellular and molecular biology of the beta-amyloid protein in Alzheimer's disease.* Neuromolecular Med **1**:1-31

Schlotterer C (2003) *Hitchhiking mapping--functional genomics from the population genetics perspective.* Trends Genet **19**:32-8

Schraen-Maschke S, Dhaenens CM, Delacourte A, Sablonniere B (2004) *Microtubule-associated protein tau gene: a risk factor in human neurodegenerative diseases.* Neurobiol Dis **15**:449-60

Schrag A, Ben-Shlomo Y, Quinn NP (2000a) *Cross sectional prevalence survey of idiopathic Parkinson's disease and Parkinsonism in London.* Bmj **321**:21-2

Schrag A, Jahanshahi M, Quinn N (2000b) *How does Parkinson's disease affect quality of life? A comparison with quality of life in the general population.* Mov Disord **15**:1112-8

Schramayr S, Caporossi D, Mak I, Jelinek T, Bacchetti S (1990) *Chromosomal damage induced by human adenovirus type 12 requires expression of the E1B 55-kilodalton viral protein.* J Virol **64**:2090-5

Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, Pahwa R, *et al.* (2001) *Complete genomic screen in Parkinson disease: evidence for multiple genes.* Jama **286**:2239-44

Seabra MC, Mules EH, Hume AN (2002) *Rab GTPases, intracellular traffic and disease.* Trends Mol Med **8**:23-30

Seitz A, Kojima H, Oiwa K, Mandelkow EM, Song YH, Mandelkow E (2002) *Single-molecule investigation of the interference between kinesin, tau and MAP2c.* Embo J **21**:4896-905

Shahani N, Brandt R (2002) *Functions and malfunctions of the tau proteins.* Cell Mol Life Sci **59**:1668-80

Sham PC, Curtis D (1995) *Monte Carlo tests for associations between disease and alleles at highly polymorphic loci.* Ann Hum Genet **59 (Pt 1)**:97-105

Shifman S, Darvasi A (2001) *The value of isolated populations.* Nat Genet **28**:309-10

Shimura H, Hattori N, Kubo S, Mizuno Y, Asakawa S, Minoshima S, Shimizu N, Iwai K, Chiba T, Tanaka K, Suzuki T (2000) *Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase.* Nat Genet **25**:302-5

Shimura H, Hattori N, Kubo S, Yoshikawa M, Kitada T, Matsumine H, Asakawa S, Minoshima S, Yamamura Y, Shimizu N, Mizuno Y (1999) *Immunohistochemical and subcellular localization of Parkin protein: absence of protein in autosomal recessive juvenile parkinsonism patients.* Ann Neurol **45**:668-72

Shimura H, Schlossmacher MG, Hattori N, Frosch MP, Trockenbacher A, Schneider R, Mizuno Y, Kosik KS, Selkoe DJ (2001) *Ubiquitination of a new form of alpha-synuclein by parkin from human brain: implications for Parkinson's disease.* Science **293**:263-9

Siderowf A (2001) *Parkinson's disease: clinical features, epidemiology and genetics.* Neurol Clin **19**:565-78, vi

Siderowf A, Stern M (2003) *Update on Parkinson disease.* Ann Intern Med **138**:651-8

Singleton A, Myers A, Hardy J (2004) *The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases.* Hum Mol Genet **13 Spec No 1**:R123-6

Singleton AB, 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 MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K (2003) *alpha-Synuclein locus triplication causes Parkinson's disease.* Science **302**:841

Smith DI, Huang H, Wang L (1998) *Common fragile sites and cancer (review).* Int J Oncol **12**:187-96

Snyder H, Mensah K, Theisler C, Lee J, Matouschek A, Wolozin B (2003) *Aggregated and monomeric alpha-synuclein bind to the S6' proteasomal protein and inhibit proteasomal function.* J Biol Chem **278**:11753-9

Spielman RS, Ewens WJ (1998) *A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test.* Am J Hum Genet **62**:450-8

Spielman RS, McGinnis RE, Ewens WJ (1993) *Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).* Am J Hum Genet **52**:506-16

Staropoli JF, McDermott C, Martinat C, Schulman B, Demireva E, Abeliovich A (2003) *Parkin is a component of an SCF-like ubiquitin ligase complex and protects postmitotic neurons from kainate excitotoxicity.* Neuron **37**:735-49

Stumpf MP, Goldstein DB (2001) *Genealogical and evolutionary inference with the human Y chromosome.* Science **291**:1738-42

Sveinbjornsdottir S, Hicks AA, Jonsson T, Petursson H, Guomundsson G, Frigge ML, Kong A, Gulcher JR, Stefansson K (2000) *Familial Aggregation of Parkinson's Disease in Iceland.* N Engl J Med **343**:1765-1770

Taira T, Saito Y, Niki T, Iguchi-Ariga SM, Takahashi K, Ariga H (2004) *DJ-1 has a role in antioxidative stress to prevent cell death.* EMBO Rep **5**:430

Takahashi Y, Toh-e A, Kikuchi Y (2001) *A novel factor required for the SUMO1/Smt3 conjugation of yeast septins.* Gene **275**:223-31

Tan EK, Matsuura T, Nagamitsu S, Khajavi M, Jankovic J, Ashizawa T (2000) *Polymorphism of NACP-Rep1 in Parkinson's disease: an etiologic link with essential tremor?* Neurology **54**:1195-8

Tanaka K, Suzuki T, Chiba T, Shimura H, Hattori N, Mizuno Y (2001) *Parkin is linked to the ubiquitin pathway.* J Mol Med **79**:482-94

Tanner CM, Aston DA (2000) *Epidemiology of Parkinson's disease and akinetic syndromes.* Curr Opin Neurol **13**:427-30

Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, Mayeux R, Langston JW (1999) *Parkinson disease in twins: an etiologic study.* Jama **281**:341-6

Tao X, Tong L (2003) *Crystal structure of human DJ-1, a protein associated with early onset Parkinson's disease.* J Biol Chem **278**:31372-9

Tapper WJ, Maniatis N, Morton NE, Collins A (2003) *A metric linkage disequilibrium map of a human chromosome.* Ann Hum Genet **67**:487-94

Templeton AR, Sing CF, Kessling A, Humphries S (1988) *A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations.* Genetics **120**:1145-54

Terry RD (1995) *Biologic differences between early- and late-onset Alzheimer disease.* Alzheimer Dis Assoc Disord **9 Suppl 1**:S26-7

Terwilliger JD, Weiss KM (1998) *Linkage disequilibrium mapping of complex disease: fantasy or reality?* Curr Opin Biotechnol **9**:578-94

Turlejski K, Djavadian R (2002) *Life-long stability of neurons: a century of research on neurogenesis, neuronal death and neuron quantification in adult CNS.* Prog Brain Res **136**:39-65

Twelves D, Perkins KS, Counsell C (2003) *Systematic review of incidence studies of Parkinson's disease.* Mov Disord **18**:19-31

Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, Ali Z, Del Turco D, Bentivoglio AR, Healy DG, Albanese A, Nussbaum R,

Gonzalez-Maldonado R, Deller T, Salvi S, Cortelli P, Gilks WP, Latchman DS, Harvey RJ, Dallapiccola B, Auburger G, Wood NW (2004) *Hereditary Early-Onset Parkinson's Disease Caused by Mutations in PINK1*. Science

Valente EM, Bentivoglio AR, Dixon PH, Ferraris A, Ialongo T, Frontali M, Albanese A, Wood NW (2001) *Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36*. Am J Hum Genet **68**:895-900

Valente EM, Brancati F, Ferraris A, Graham EA, Davis MB, Breteler MM, Gasser T, Bonifati V, Bentivoglio AR, De Michele G, Durr A, Cortelli P, Wassilowsky D, Harhangi BS, Rawal N, Caputo V, Filla A, Meco G, Oostra BA, Brice A, Albanese A, Dallapiccola B, Wood NW (2002) *PARK6-linked parkinsonism occurs in several European families*. Ann Neurol **51**:14-8

van de Warrenburg BP, Lammens M, Lucking CB, Denefle P, Wesseling P, Booij J, Praamstra P, Quinn N, Brice A, Horstink MW (2001) *Clinical and pathologic abnormalities in a family with parkinsonism and parkin gene mutations*. Neurology **56**:555-7

Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leimpeter A, Bloch DA, Nelson LM (2003) *Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity*. Am J Epidemiol **157**:1015-22

van Duijn CM, Dekker MC, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, Testers L, Breedveld GJ, Horstink M, Sandkuijl LA, van Swieten JC, Oostra BA, Heutink P (2001) *Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36*. Am J Hum Genet **69**:629-34

Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, Peltonen L (2003) *The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories*. Hum Mol Genet **12**:51-9

Vaughan JR, Farrer MJ, Wszolek ZK, Gasser T, Durr A, Agid Y, Bonifati V, DeMichele G, Volpe G, Lincoln S, Breteler M, Meco G, Brice A, Marsden

CD, Hardy J, Wood NW (1998) *Sequencing of the alpha-synuclein gene in a large series of cases of familial Parkinson's disease fails to reveal any further mutations. The European Consortium on Genetic Susceptibility in Parkinson's Disease (GSPD).* Hum Mol Genet 7:751-3

Verpillat P, Ricard S, Hannequin D, Dubois B, Bou J, Camuzat A, Pradier L, Frebourg T, Brice A, Clerget-Darpoux F, Deleuze JF, Campion D (2002) *Is the saitohin gene involved in neurodegenerative diseases?* Ann Neurol **52**:829-32

Wang WW, Khajavi M, Patel BJ, Beach J, Jankovic J, Ashizawa T (1998) *The G209A mutation in the alpha-synuclein gene is not detected in familial cases of Parkinson disease in non-Greek and/or Italian populations.* Arch Neurol **55**:1521-3

Weber E, Jilling T, Kirk KL (1996) *Distinct functional properties of Rab3A and Rab3B in PC12 neuroendocrine cells.* J Biol Chem **271**:6963-71

Weiss KM, Clark AG (2002) *Linkage disequilibrium and the mapping of complex human traits.* Trends Genet **18**:19-24

Welch JE, Barbee RR, Roberts NL, Suarez JD, Klinefelter GR (1998) *SP22: a novel fertility protein from a highly conserved gene family.* J Androl **19**:385-93

Wellenbrock C, Hedrich K, Schafer N, Kasten M, Jacobs H, Schwinger E, Hagenah J, Pramstaller PP, Vieregge P, Klein C (2003) *NR4A2 mutations are rare among European patients with familial Parkinson's disease.* Ann Neurol **54**:415

West A, Periquet M, Lincoln S, Lucking CB, Nicholl D, Bonifati V, Rawal N, Gasser T, Lohmann E, Deleuze JF, Maraganore D, Levey A, Wood N, Durr A, Hardy J, Brice A, Farrer M (2002a) *Complex relationship between Parkin mutations and Parkinson disease.* Am J Med Genet **114**:584-91

West AB, Lockhart PJ, O'Farell C, Farrer MJ (2003) *Identification of a novel gene linked to parkin via a bi-directional promoter.* J Mol Biol **326**:11-9

West AB, Maidment NT (2004) *Genetics of parkin-linked disease*. Hum Genet **114**:327-36

West AB, Maraganore D, Crook J, Lesnick T, Lockhart PJ, Wilkes KM, Kapatos G, Hardy JA, Farrer MJ (2002b) *Functional association of the parkin gene promoter with idiopathic Parkinson's disease*. Hum Mol Genet **11**:2787-92

Wintermeyer P, Kruger R, Kuhn W, Muller T, Woitalla D, Berg D, Becker G, Leroy E, Polymeropoulos M, Berger K, Przuntek H, Schols L, Epplen JT, Riess O (2000) *Mutation analysis and association studies of the UCHL1 gene in German Parkinson's disease patients*. Neuroreport **11**:2079-82

Wittmann CW, Wszolek MF, Shulman JM, Salvaterra PM, Lewis J, Hutton M, Feany MB (2001) *Tauopathy in Drosophila: neurodegeneration without neurofibrillary tangles*. Science **293**:711-4

Wright AF, Carothers AD, Pirastu M (1999) *Population choice in mapping genes for complex diseases*. Nat Genet **23**:397-404

Xia Y, Rohan de Silva HA, Rosi BL, Yamaoka LH, Rimmler JB, Pericak-Vance MA, Roses AD, Chen X, Masliah E, DeTeresa R, Iwai A, Sundsmo M, Thomas RG, Hofstetter CR, Gregory E, Hansen LA, Katzman R, Thal LJ, Saitoh T (1996) *Genetic studies in Alzheimer's disease with an NACP/alpha-synuclein polymorphism*. Ann Neurol **40**:207-15

Xu PY, Liang R, Jankovic J, Hunter C, Zeng YX, Ashizawa T, Lai D, Le WD (2002) *Association of homozygous 7048G7049 variant in the intron six of Nurr1 gene with Parkinson's disease*. Neurology **58**:881-4

Yang Y, Nishimura I, Imai Y, Takahashi R, Lu B (2003) *Parkin suppresses dopaminergic neuron-selective neurotoxicity induced by Pael-R in Drosophila*. Neuron **37**:911-24

Zak NB, Shifman S, Shalom A, Darvasi A (2002) *Genetic dissection of common diseases*. Isr Med Assoc J **4**:438-43

Zarranz JJ, Alegre J, Gomez-Esteban JC, Lezcano E, Ros R, Ampuero I, Vidal L, Hoenicka J, Rodriguez O, Atares B, Llorens V, Gomez Tortosa E, del Ser T, Munoz DG, de Yebenes JG (2004) *The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia.* Ann Neurol **55**:164-73

Zhang J, Hattori N, Leroy E, Morris HR, Kubo S, Kobayashi T, Wood NW, Polymeropoulos MH, Mizuno Y (2000a) *Association between a polymorphism of ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) gene and sporadic Parkinson's disease.* **6**:195-197

Zhang K, Deng M, Chen T, Waterman MS, Sun F (2002) *A dynamic programming algorithm for haplotype block partitioning.* Proc Natl Acad Sci U S A **99**:7335-9

Zhang Y, Gao J, Chung KK, Huang H, Dawson VL, Dawson TM (2000b) *Parkin functions as an E2-dependent ubiquitin- protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1.* Proc Natl Acad Sci U S A **97**:13354-9

Zheng K, Heydari B, Simon DK (2003) *A common NURR1 polymorphism associated with Parkinson disease and diffuse Lewy body disease.* Arch Neurol **60**:722-5

Zimprich A, Asmus F, Leitner P, Castro M, Bereznai B, Homann N, Ott E, Rutgers AW, Wieditz G, Trenkwalder C, Gasser T (2003) *Point mutations in exon 1 of the NR4A2 gene are not a major cause of familial Parkinson's disease.* Neurogenetics **4**:219-20

Zondervan KT, Cardon LR (2004) *The complex interplay among factors that influence allelic association.* Nat Rev Genet **5**:89-100

8 Appendices

8.1 Appendix 1 - Marker Assays

3.2.1 Exon dosage *Parkin* exons 3, 4 and 5

Exon	Primer	Sequence (5' - 3')	Product Size (bp)
3	F	CTTGCTCCCAAACAGAATT	314
	R	AGGCCATGCTCCATGCAGACTTGC	
4	F	ACAAGCTTTTAAAGAGTTTCTTGT	261
	R	AGGCAATGTGTTAGTACACA	
5	F	ACATGTCTTAAGGAGTACATTT	227
	R	TCTCTAATTTCCTGGCAAACAGTG	
Control	F	ACGTTCCTGATAATGGGATC	328
	R	CCTCTCTCTACCAAGTGAGG	

PCR Conditions	PCR Mix 1		
	95 °C		5 min
	95 °C		30 sec
	53 °C	23 cycles	45 sec
	68 °C		2 min 30 sec
	68 °C		5 min

3.2.1 *Parkin* point mutation screening exons 3, 7, 9 and 11

Exon	Primer	Sequence (5' - 3')	Product Size (bp)
3	F	CTTGCTCCCAAACAGAATT	314
	R	AGGCCATGCTCCATGCAGACTTGC	
7	F	TGCCTTTCCACACTGACAGGTACT	239
	R	TCTGTTCTTCATTAGCATTAGAGA	
9	F	GGGTGAAATTTGCAGTCAGT	278
	R	AATATAATCCCAGCCCATGTGCA	
11	F	ACAGGGAACATAAACTCTGATCC	303
	R	CAACACACCAGGCACCTTCAGA	

PCR Conditions	PCR Mix 4
	Program 60>50

3.2.1 STRPs about A82E (Parkin exon 3)

Marker	Size Range (bp)	Repeat type	Primer	Sequence
D6S1599	121-165	dinucleotide	Forward	[FAM]-TGTTTTCCACAGGTTCCAG
			Reverse	CTTCAGATGTAGGCTCCACG
D6S980	255-325	dinucleotide	Forward	[FAM]-AGGGAGCCGAGATTGCAC
			Reverse	CTGAAGGGTGAGGAGTTTCT
IVS2-P2	170-190	dinucleotide	Forward	[HEX]-TCAGCCACTGGGTCCATCTG
			Reverse	ATTGGAGTGTTCCTGGAGCTTC
IVS2-P3	250-270	dinucleotide	Forward	[HEX]-ACACATTCCCAGATGCACAG
			Reverse	CAAGAAAGACTCTGCCAATCTG

PCR Conditions	PCR Mix 1
	Program 60>50

3.2.2 PARK7 STRPs

Marker	Size Range (bp)	Repeat type	Primer	Sequence
D1S243	142-170	dinucleotide	F	[HEX]-CACACAGGCTCACATGCC
			R	GCTCCAGCGTCATGGACT
D1S468	125-145	dinucleotide	F	[FAM]-TTAACCGTTTTGGTCCTACC
			R	CTCTGACCAGCATTAAAGATTC
D1S2870	190-212	dinucleotide	F	[HEX]-GATCATGCCAATGCACTAT
			R	CCAGGGTGACACAGCA
D1S1646	130-150	dinucleotide	F	[HEX]-AAGTAGATGGGAGCCACACA
			R	CAAGATCATGCCACTGCCTT
D1S1612	94-130	dinucleotide	F	[FAM]-TCCCATGCCAAAATTCTTAG
			R	GAAAGAAAGAGAAAGAAGGAAGG
D1S244	280-300	dinucleotide	F	[FAM]-GAGCAGCACCGTACAAAT
			R	AGCTCCGCTCCCTGTAAT

PCR Conditions PCR Mix 1
Program 60>50

3.2.2 Exon dosage *DJ-1* exons 2 and 4

Exon	Primer	Sequence (5' - 3')	Product Size (bp)
2	F	[HEX]-CTCTGCTTGAAAATGCTCC	392
	R	GGCAAGACATTAACAAGCG	
4	F	[HEX]-GGCTATCTCCTGTACTTCCC	297
	R	TCACAGCCTCCTCCCGAA	
Control	F	[HEX]-ACGTTCCTGATAATGGGATC	328
	R	CCTCTCTCTACCAAGTGAGG	

PCR Conditions PCR Mix 1

95 °C		5 min
95 °C	23 cycles	30 sec
53 °C		45 sec
68 °C		2 min 30 sec
68 °C		5 min

3.2.3 PARK6 STRPs

Marker	Size Range (bp)	Repeat type	Primer	Sequence
D1S552	244-260	dinucleotide	F	[HEX]-TTCATGCAGCATCATCCC
			R	TGTGGGCAGGTGTAAAGAGT
D1S199	94-116	dinucleotide	F	[HEX]-GGTGACAGAGTGAGACCCTG
			R	CAAAGACCATGTGCTCCGTA
D1S478	155-175	dinucleotide	F	[HEX]-ATGCCCAATACCCCAGT
			R	GCATTCATTTATTCAGCAAGAT
D1S2828	247-279	dinucleotide	F	[FAM]-GGCTCCTGAACCTGGG
			R	AGCTTTGGCTGACCTTCC
D1S2885	217-263	dinucleotide	F	[HEX]-GACATCCATCCCCTGGCTTA
			R	GGGTCCCACTCGGGCT

PCR Conditions PCR Mix 1
Program 60>50

4.2.2 *MAPT* Haplotyping by (intron 9) 238 bp indel

	Sequence	Product Size (bp) H1	H2
Forward primer	GGAAGACGTTTCTCACTGATCTG	498	260
Reverse primer	AGGAGTCTGGCTTCACTCTCTC		
PCR Conditions	PCR Mix 1		
	Program 60>50		

4.3 *MAPT* Extended haplotype boundary analysis

Restriction Enzyme Digest

SNP	F Primer R primer	PCR mix Program	Product size (bp)	Enzyme	Alleles	Fragment Sizes (bp)
hCV15853219	AACCCAAGATGAGTGAGATACATTCC	1	404	Alu I	G	277, 127
	CGTTGGTCCACACTTGTTGAGG	60>50			A	404
hCV11411459	AGCCGATCGAGCCTTTGTCTGG	1	374	Nla IV	G	152, 113, 69, 40
	AGGGTTTGGGTGCTGGGTGTTC	60>50			C	265, 69, 40
hCV2077569	TGCATATAAGAATCACCTGGGAAAC	1	514	Msc I	G	300, 214
	GACATCCTCCACAGACACCAATG	60>50			C	514
hCV7450752	GGCCCTGTTCATAGATGTCCTCAG	1	504	Mwo I	C	251, 204, 49
	GATGATAACTTGGCTTCCTGTGTGG	65>55			G	455, 49
hCV11417996	CTCTTGGAATCTTATTTGTGCAACTG	1	455	Mse I	A	206, 112, 69, 54, 14
	CTCCGATTCAGGTCACACCATC	65>55			C	318, 69, 54, 14
hCV2265270	TGCTACAGCTTACCTAATACAGTGCC	1	536	Hpy 188 I	G	300, 141, 95
	GAGCAAGACATTCAAAGGTAGGAGAC	60>50			A	441, 95
hCV2265267	ACCCTTTAGCAGTCACTCCTCATTTC	1	559	Msl I	T	245, 204, 100
	TCAAATGTCCTCTGCTCTGAGTGC	60>50			C	449, 100
hCV7450783	GTTAGGTCTCATGCCCACTCCC	1	351	Ban II	G	225, 126
	GAGTCAGAGGCTGTCACGAGTTG	60>50			C	145, 126, 80
hCV11936211	TTAAGCTGAAACTGAGGCATGTTCTC	3	301	Aci I	G	126, 112, 63
(rs1816)	CCTTTAGCATGTTTGCACATCCAG	65>55			A	175, 126
hCV11416920	GGGTAGGAAATGGAGCTGGAGC	1	323	Msl I	T	220, 71, 32
	AACTGGCTGTAAATTGGAGGTCCC	60>50			C	220, 103
hCV9473153	CACATAATTGAAATCCCAGGAAATG	3	437	Hha I	G	274, 163
	TTCTTTCAGTCCCAGTATGCTGTTG	65>55			A	437
hCV11411497	TTGGTGCCTTCCTTGGGTGTG	1	490	Fnu 4HI	G	266, 124, 56, 44
	ATGCAAGTCTGAAATCCAATAGGGC	60>50			A	322, 124, 44
hCV11618556	GTCCAGTCTCGCCATGAAATCACC	1	490	Fnu 4HI	C	301, 149, 92
	GCCCACCCATGTAATCCAGGATAATC	60>50			T	393, 149

Sequencing hCG1640620

Primer name	F Primer Sequence R Primer Sequence	PCR mix Program	Product size (bp)	Region
hCG1640620-1F	GTTGAGTGTATAGCAGGCCTCCTAAC	1	517	exon 4
hCG1640620-1R	AGGTGGAGGTCACAGTGAGCTG	65>55		
hCG1640620-2F	CCGTTAGGAGTTCCTGACAGATTGGC	1	899	exon 5
hCG1640620-2R	GAGGATTGCTTGAGCCCGAGAGAC	65>55		
hCG1640620-3F	TGTTAGTCTTTGGCATCCTTTCG	3	845	exon 5
hCG1640620-3R	AAGAACATTTATCTCAGAGCAGGGAG	60>50		
hCG1640620-7F	AATCCACCTTCTCTCTCTCACAAACC	1	865	exon 8
hCG1640620-7R	TACATCATCTTGCTCGCCTCAGG	65>55		

SNP	Primer set	variation	position
17V67981559 (novel)	hCG1640620-1	G > A	Exon 4 -45 bp
hCV2265263	hCG1640620-7	C > T	Exon 8 + 196 bp

4.4.1 *MAPT* H1 SNPs

Restriction Enzyme Digest

ID	SNP	F Primer R Primer	PCR mix Program	Product size (bp)	Enzyme	Alleles	Fragment Sizes (bp)
3	hCV2544832	ACCCACAGACCACGACCTTCCAAC	1	471	Mnl I	G	279, 192
		CCTGCCTCTTTCTGCCCATTGG	60>50			A	471
4	hCV2544799	GGGCCACTGGATCACAAGGTTG	1	528	Tsp509 I	T	258, 181, 59, 30
		CCGGCCCATAATCTGCATTTCTAAC	60>50			C	288, 181, 59
5	hCV2544792	TTAAGGAAGCACCCATGACAGCC	1	461	Mbo II	A	266, 144, 51
		AAACAGTTCTGTGGAATTTCACCCTG	60>50			G	410, 51
7	hCV3202942	GGTAGAGGCCAGGAATGCTGTTAAAC	1	479	Aci I	C	277, 202
		GGTCATGCTCCGATTACAGACTCTTG	60>50			A	479
8	hCV3202957	CAGCCTTCCCTGTCCTTGATTC	1	385	Xho I	G	385
		GCCTTCCCAACAGAGCAACC	60>50			A	287, 98
10	hCV16089259	AGCAAGCTGTGTGACCAG	1	238	Bcl I	G	197, 41
		CCCATTCTCTGACAGATTTG	60>50			A	112, 85, 41
11	hCV16017251	AGACATCCACACGTTCCTC	1	248	Afl III	C	141, 107
		CAAACCACAGCAGAGCAG	60>50			T	248
12	Ex 9iii	CGAGTCCTGGCTTCACTCC	2	370	Bst N I	G	257, 54, 26, 20, 6
		CTTCCAGGCACAGCCATACC	60>50			A	201, 56, 54, 26, 20, 6
13	385342	GGCTGGCCCTGCTCCTTCTCTA	1	352	Tai I	T	247, 105
		TGGCAAGGACGTTGGGGGACAGGG	60>50			C	352
14	385928	GACTGATAGGTGGGAGGTGGCTGC	1	454	Pvu II	CT	228, 226
		CAGCAGCTCGGACGTGAG	60>50			AA	454
16	Ex 3 -347	ACTCTCAGTGCTCCCTCAACAGG	1	732	Dpn II	A	597, 134
		CCTGTGTAAACGCTGCAGTGAAAC	60>50			G	497, 134, 101
19	17V40678709	GCAGCCCAACAGAGAAATACCC	1	606	Bsl EI	C	364, 242
		TCCCAAATTGCTGAGGTTAAGGTC	65>55			T	606
17	Ex 0 +120	TCGACTATCAGGTAAGCG	1	712	BsaWI	G	581, 131
		AAGGGCAACTAAAGTGACAG	60>50			C	712
18	hCV1016016	CTGTTGGATAATGGACATTTGGC	1	684	Apa LI	G	529, 155
		CTGAAGGCTCTTGTTTGACAGTACC	65>55			A	684

SNaPshot

ID	SNP	F Primer R Primer S Primer (+non annealing sequence)	PCR mix Program	Product size (bp)	Alleles
1	rs110402	GTGCACTCTGTACACTCACTGGACC	1	483	C
		GTATGATTCAGGAATAAGGCAGAAGC	60>50		T
		CACAGAGGACTGGTGTTGC			
2	hCV2544834	CTGCACAGAACAAAGTACACGTGAC	1	433	G
		TCCTATGCAAAGAAGACACAAGGG	60>50		A
		aactgaGCGAGGGACCAAGAGAAG			
6	hCV2257661	GGTCATCTCTAGTGGGCATTAACACG	1	381	C
		TGACAAAGGCAAGAGTACACAAAGGG	60>50		T
		aactCCAGGCTGTTCTCGAACT			
9	rs3785883	CCATCACCTTGTCAGAAACTC	1	277	G
		AGCCATGTGGTAGCCTCAG	60>50		A
		aactgactaaCACTGTCACCACTGGGC			
20	17V40666057	GGATTCCAGAAGAGCTAGGTAACCAC	1	738	G
	(Tau47-51 set2)	CTGGCAAGAAAGAAGCACAAAGAG	60>50		A
		GAAGTCTGCCTCTGTGCTA			

4.4.6 Sequencing conserved regions in *MAPT*

Region	F Primer	Sequence	R primer	Sequence	Product size	PCR mix	Program
Promoter	1	GCTTCAAGACAACTTCCCATCTCCC	1	GGTCCGCTGCCCTTTACTTTCAG	684	1	65>55
	2	AAAGCAAGGAGAAAGGAAGCAGC	2	GCGCTTACCTGATAGTCGACAGAG	702	2	65>55
	3	CTACAGGAGGTGGAGAAAGC	3	CTGCAGTCGAGAGTGAAGGG	779	2	60>50
	4	TCGACTATCAGGTAAGCG	4	AAGGGCAACTAAAGTGACAG	711	1	60>50
46.5-50Kb	1	TGTCTTCCGTGAAATGTCCTCCTG	1	CCAAGATGGCTTCCTATTGTTTATGC	708	1	60>50
	2	GGATTCCAGAAGAGCTAGGTAACCAC	2	CTGGCAAGAAAGAAGCACAAAGAG	738	1	60>50
	3	TTTCCAAGCAGAACTTCATTTCCTC	3	TCTCGTTACAGAGTCATGTTCCACTG	535	1	60>50
	4	CTGTTGGATAATGGACATTTGGC	4	CTGAAGGCTCTTGTTTGACAGTACC	683	1	65>55
	5	TCACAGGGTCTGCCATCTGC	5	GCCAGGGTGACAGAGTGAGATTC	698	1	60>50
	6	AGATGTGTATGCTTGCCCTGGTATGG	6	AGGGAAGACAAGGAAGGCGATCTG	737	1	65>55
59-61 Kb	1	GCACAGCTGCGCCATTTGTC	1	CCCAGGTCACACAGCCAGTAAGAGG	772	1	60>50
	2	AGGTTGAGAGAAGTGGCTTGGTTAC	2	GGAAATGCTCGTAGGTTCATGGAC	704	1	60>50
	3	TTTCTTAACATGGCAGGTAGGTCC	3	TAGGCTTCTCCAGTATTTCTTCG	710	1	60>50
	4	GCAGCCCAACAGAGAAATACCC	4	TCCCAAATTGCTGAGGTTAAGGTC	605	1	65>55
	4a	TTATTTATGCTCGATGGCTGCTTTCC	4a	GTTTCACCGTGTTAGCCAAGATGGTC	573	1	65>55
	4b	TGAGGTGGGTGGATCACGAG	4b	AGGGAAGGCAGCTGGACAAG	434	1	60>50
	5	AGGCTGAGGCAGGAGAATGTCTTG	5	CGATTGGATGGGTAGAGAGCAACTG	551	2	65>55
	6	AAGCCCACCTCTAGCATAATACGAG	6	GGAGAATCGCTTGAACCTAGGAG	612	1	60>50
78-80 Kb	1	CCGTGTTCCAGCTGTTTCCAC	1	ACAAACACCGTGGCCAAAGTC	673	1	60>50
	2	AGGTCCCAGAGCAGCCATTTG	2	AATGTCCCAGCCTTGCTCGCAG	642	1	60>50
	3	CTGCTTATGTCTCATTGGCCAGAG	3	TCTTGAGTCTCATGCCTGCACAG	650	1	60>50
	4	ACTCTCAGTGCTCCCTCAACAGG	4	CCTGTGTAAACGCTGCAGTGAAAC	732	1	60>50
	5	GACTTGCCCAAAGCCTCACAG	5	CGAGCTACTCACAACCCAAGATTC	664	1	60>50

4.5.1 *CRHR1* genomic sequencing

Exon	F Primer	Sequence	R primer	Sequence	Product size (bp)	PCR mix	Program
2	CRHR1-2F	TCCAAGAGCAGGGCAACTCCAGAC	CRHR1-2R	AATTGCTTGGAGCCCTGGAGAGG	449	1	60>50
3	CRHR1-3F	TTTAACACCACCTCCAGCCATCTCTC	CRHR1-3R	ACAGGGCACCCAGGAGAATACAGC	558	1	60>50
4	CRHR1-4F	AATGGGCAGAAAGAGGCAGGAAG	CRHR1-4R	GAGAAGGAAGGAAGGAACAAGGACC	532	1	60>50
6	CRHR1-6F	TCCTGGTGGCCTTTGTCCTCTTTC	CRHR1-6R	GGTGCCTTCTACTCCCTGCCATTC	514	1	60>50
7	CRHR1-7F	CTGTGCCTCAGTTTCCTCATCTACAC	CRHR1-7R	TTGTAGGCGGCTGTCACCAAC	465	1	60>50
8	CRHR1-8F	AGAGGAGGAGCCCACAGAACAG	CRHR1-8R	CTCAAACTCACACACCTCCTTCAAG	481	1	60>50
9	CRHR1-9F	GCGGCAGTAGAAGCACCTTGAAGG	CRHR1-9R	TATTTGAACCTGGGCAGTCTGGCTC	486	1	60>50
10	CRHR1-10F	CCGTTCCCAGCCACTCAGTCTTTC	CRHR1-10R	TGAGCACCTGCACATCTCTGACCC	492	1	60>50
12	CRHR1-12F	GCATCCACCACGTCTGAGACCATTC	CRHR1-12R	TCAAAGGTCGGCTGCAGGTGTTG	480	1	60>50
13	CRHR1-13F	CACCTACATGCTGTTCTTCGTCAATC	CRHR1-13R	TCAAAGATACTGGAGCTCTGGCTTC	539	1	60>50
14	CRHR1-14aF	CTGTTCCGACAAATATGCAAAGCAG	CRHR1-14aR	TTTCAATTCATTCCCATGTCCCTTC	601	1	60>50

4.5.2 *CRHR1* A60V

SNaPshot

F Primer R Primer S Primer (+non annealing sequence)	PCR mix Program	Product size (bp)	Alleles	PCR mix	Program
TTTAACACCACCTCCAGCCATCTCTC	1	483	C	1	60>50
ACAGGGCACCCAGGAGAATACAGC	60>50		T		
aactgactaaactaggGAACCACTAGCTGCCCC					

5.2 Y Chromosme Haplotyping

Marker	F Primer	R Primer	Size	Enzyme	Alleles and Fragment sizes		PCR mix	Program
M9	GCAGCATATAAAACTTTCAGG	AAAACCTAACTTTGCTCAAGC	340	Hinf I	0 190, 80, 70	1 26, 80	1	60>50
SRY-1532	TCCTTAGCAACCATTAATCTGG	AAATAGCAAAAAATGACACAAGGC	180	Dra III	0 180	1 110, 70	1	60>50
TAT	GTGAAGTAAGATATCAGATGG	TGCAAGCTTAATTCATAGCAC	1110	Nla III	0 350, 225, 220, 140, 12	1 500, 350, 140, 12	1	60>50
YAP	CAGGGGAAGATAAAGAAATA	ACTGCTAAAAGGGGATGGAT	455/150	-	0 455	1 150	1	60>50
SRY-2627	CGCGGCTTTGAATTTCAAGCTCTG	CCAGGGCCCCGAGGGACTCTT	350	Ban I	0 250, 100	1 350	1	60>50

5.5.1 htSNP discovery in 1p32 candidate genes (*htSNPs*)

** Haplotypic variation captured by htSNPs

	SNP Gene **	F Primer	R Primer (S Primer for SNaPshot)	Variation (1/2)	Enzyme	Alleles and Fragment Sizes		PCR mix	Program
	RNF11 (90%)								
1	*hCV3122415*	AGAGATGAGGTCTTGCTGTGTTGCC	CTTATCCTCCCACTCACTGTCCTTCC	C/A	Mfe I	A 508	C 290, 218	1	60>50
2	hCV11304546	GAAAGTGTGGGAATGAAGAAAGCAAC	GCTGAGGCAGGAGAATCGCTTC	C/T	Stu I	T 485	C 272, 213	1	60>50
3	hCV11304542	CATAGAGTTGAATAGCGTGGGTCCAC	GCCCAAGAAGCTGCTCATGATG	T/C	Hae III	T 300, 88	C 189, 111, 88	1	60>50
4	hCV3122423	CATGCACATGAGAATCGAAAGAAAGAAC	TGGGATTGCCTCTCACGGATTC	T/G	Dde I	G 220, 180	T 180, 114, 106	1	60>50
5	*hCV937775*	CACGTATGTGAACCACAAGACTGAAG	TCTTGGCCTCTGTTGTGATCTCTG	A/G	Hinc II	G 293	A 191, 103	1	60>50
	EPS15 (100%)								
1	hCV3125022	ATGGCAGAGCTACAATTCAATCCAG	AGGATCCAGGGTTGTTACTGTAGAGG	A/G	Bsl I	A 397	G 286, 104	1	60>50
2	hCV3125029	GAAGGGTTTCATCTGCCATTGTAGAG	TCTCTTCTGCCATTGTCAATGTTCAG	T/C	Msl I	T 205, 161	C 366	1	60>50
3	hCV3125045	GCTCAAGAACATTCAAGGCACCAG	CACACTCAAAGACTTGGCTCTATCCC	A/G	Bsa AI	A 390	G 248, 142	3	60>50
4	hCV1771642	AAGAATTCATGGCGATGGATATTGGG	TGACTTCTTGGCTGGGCATGGTG	T/C	Sty I	T 454	C244, 210	1	65>55
5	hCV3125034	CATAAGGGAGGCTCCACTGAAGATTC	ATGATTGGATGCTCCCTGAGGC	G/C	Bgl I	G 460	C 297, 163	3	60>50
6	*hCV3125031*	TAGGAACACAGAATCTGGCTACCTGG	TCCAGATTGCCTGACATGACCTG	A/T	Spe I	A 300	T 247, 53	3	60>50
7	*hCV1591238*	ATTCTGCTGTTTCTTGCTGTAGACGG	CCTGGGCAACATAGCAAGACTCTC (TCGCAGATCTCCACTT)	A/G	(SNaPshot)				
8	hCV1591233	TGGTGTGGGCAATCATAAACG	AAGATGAGATTTAGCAGAGTGCATTG	A/G	Swa I	A 206, 127	G 333	1	60>50
9	hCV11740229	TGATGTCTTGGGCCTTCAC	TGATCCACCACACTGAGCAG	C/T	Rsa I	C 264, 106, 106, 8	T 348, 106, 106	1	60>50
10	*hCV11740230*	ATGGAACAACCGTAGAACTG	CAATCATCCTCTACCCACAG	T/C	Ssp I	C 476	T 263, 213	1	60>50
11	*hCV1591218*	CTTAGGTCACTGGATTCGT	ACACGGTTACTGACTTAGATG	T/G	Bst E II	T 375	G 293, 82	1	65>55
	NRD1 (90%)								
1	hCV79313	AGAAAGCAAGTCAGTATGGC	GGTATCATCTTCTATTGCCT	C/T	Acc I	T 350	C 234, 116	1	60>50
2	hCV15818300	CCTCTGCACTCTAGCCTGGTGACAG	CCAGCAGAATTTCACAGCCTCAGTG	G/A	Rsa I	G 204, 61, 51	A 153, 61, 51, 5	1	65>55
3	hCV15818304	AAAGCAAGGCGTGGTAGAACATCC	TCCCATGTCTCTTGCCTGAATCAC	C/T	Mse I	C 291, 54	T 198, 93, 54	1	60>50
4	*hCV393617*	AAACATTGCTAGAGGGCCACACAAG	AATTACGGGTATGAACCACTGTGCC	A/G	Bsl I	A 385	G 249,136	1	60>50
5	*Glu indel*	[HEX]-CCAGATACATCAAATTACAGAATGGC	TCTTCATCGTCATCTTCTATTTCAGC	169/172	(ABI sizing)	169=1	172=2	1	60>50
6	*hCV2776338*	CTTTATACAGTTTCTAGAGTGG	GAACACTTAGAGACCATATTTG	G/T	Dde I	G 262,128,86,9	T 348,128,9	1	60>50
7	hCV2776339	CGCGTTTGAAAGGCTAACTAGC	TCCTTGCTTTCTGTTCTGTAAGTTTG	T/G	Bsl I	T 330	G 167,163	1	60>50
8	hCV2776344	GTCATCTGCTTAATCCTCTCACTGCC	TGTGTCACGGTGGTTTGTTGTACAG	G/T	Sal I	G 272, 159	T 431	1	60>50
9	*hCV2776353*	TGAGTTCAATTCCACACCAGC	TGCTATATTGCCCAGGCTAGTC	T/C	Dde I	T 345,55,42,39	C 387,55,39	3	60>50
10	*hCV8847889*	AAACTTGACTCATCTGGCAGCTTTAC	TGCTTGTGACATTCCCTTGTACC	T/C	Bst N I	C 308, 81, 36, 24	T 344, 81, 24	1	60>50
	RAB3b (100%)								
1	hCV1805213	GGTGACTTCTTTGCCTGAGCCTC	TGAGGGAGCAGTTCAGCGTAGC	A/C	Mse I	A 218, 153, 65	C 371, 65	1	60>50
2	hCV1805212	TTTGCTGCCCTCTGCTAG	GGAGAAAGCAACCAGAAG	G/A	Rsa I	G 349, 229	A 255, 229, 94	1	60>50
3	hCV1805209	AATGACTCCTCAAAGGGTGGCTG	AACATTTCTTGCCAAGGTCTCATAGG	A/G	Dpn II	A 287, 141, 118	G 405, 141	1	60>50
4	*hCV1805202*	AACAATTTCCACCCTGCTCTTCTCAC	CAGCCTCATGCCCTTTCTTTGC	G/A	Hinf I	G 260,209,40	A 469,40	1	60>50
5	*hCV1805200*	CAGTGACCAACAGAACGCTGTAAGG	CGAACTCCTGACCTTGTGATCCG	A/G	Bst U I	A 308	G 239, 69	1	60>50
6	hCV315881	CCAGGCACTGTGCTAATTCTTACAAC	TCTCCCAGCAGTTAATTTGATCACTC	A/G	Ava I	A 327	G 202, 125	1	60>50
7	*hCV16117609*	TTACACACAGGGATAATCACTGCAAC	GAAATCACAGCCAAGCATAAAGAAG	A/G	Nla III	A 262, 149	G 411	1	60>50
8	hCV11865899	CTGCCCTCAGGAAACCTCAATAGAG	CCTGAAGAACCCAGCCTATCCTAAAG	A/T	Rsa I	A 275, 249	T 524	1	60>50
9	hCV11865894	CACAAGGTCACTCTCAGCTCAAGC	ACCCAGTGTCCCTGCTCAGATG	G/A	Bst N I	G 89, 88, 63	A 177, 63	1	60>50
10	*2230325*	CGAAGACTCCACCTCCAAATAC	CACACTCGTACCCTGTTTACCC	T/C	Dde I	T 132, 112, 87	C 199, 132	1	60>50
11	*hCV439172*	ACATGTCTCCTCCACGGTTCAG	GAGATTAATAACATGAGCAATGGCAC	T/C	Msp I	C 259, 192	T 451	1	60>50

5.5.2 Sequencing candidate genes at 1p32

Gene	Exon	F primer	R primer	Product size (bp)	PCR mix	Program
RAB3b Genomic						
	Promoter	AGAGGAAGGGCAGGGATATGTTAAG	CTCAACTCCTAGCAGGTGTAATCCG	350	1	60>50
	Promoter	ATGACTGGAGACCACAGGAGAAAG	CGTTCTATGGCTCCGTCTTCAC	327	1	60>50
	1	CATGACAATTAAGATAAATCACGAGA	TCTTTCTTGCTCCCCTCCC	572	2	60>50
	2	AGTAAGGAGTGAGTGTGTGAATGGTG	CCCACCTCTAAGCCCTTGTACTG	405	1	60>50
	3	ACTCCAGCCTGGGCAACAAG	CAGCACACAATACAGGACCTAGCAC	338	1	60>50
	4 (5' end)	AAGACTCCACCTCCAAATACCATTAC	GTACATACCAAGCTGCTCTGCAAG	257	1	60>50
	4 (3' end)	TGACATGGAGGAAGAGAGGGTTG	CAGCACAAAGCCCATATGTACCTC	243	1	60>50
	5 (5' end)	GGGTTCCAGCAGAAGGGATG	AGAGACGCGTGTTCTTGGAGG	202	1	60>50
	5 (3' end)	TGGACACAGACCCGTCGATG	AGGCAAAGAATAGCAGCAACTCATC	235	1	60>50
	5 (3' UTR)	GCCCACTCTCCCTGTTACACACTGC	TTCACGCCATTCTCCTGCCTCAG	766	1	60>50
EPS15 Genomic						
	9	GAGAGCCTGACGATGATTTATCATTG	CCTGTCTTAATGGTGCTTTGGTAGG	300	1	60>50
	10	TGGAAGTGATACATTGTTGCTCTGAG	AAGCTTCAGGTAACCCATTCTCAAG	473	1	60>50
	11	TCTACCTCTTCGAACAGTTGAGTGAG	CAAGGCACTCCAGCATATTTATAGG	493	1	60>50
	12	TTTACTGCAAGGTTCTGAACAGGC	TTAGAAACAGAAAGGTGTGTGTGGG	516	1	60>50
	13	AAGCGATTCAGGATTAGGAGGATGG	GCAGGCCAGAATTCCAAGCAAG	337	1	60>50
	14	TGTCCAAGGACAGATTTGAGTTAGC	GCCAACCACCACCATCAATAAC	490	1	60>50
RNF11 cDNA	(5'-3')					
		CACCGCTGCTTTCTCCTC	TTCCAGGGTCATAAACTCC		1	60>50
		GCTGACTGAAGAGGAACAAA	GGCTCTTTGGGCTCAGTG		1	60>50
ELAVL4 cDNA	(5'-3')					
		TTTAACAGAAGAGTCGAAGCTCTGCG	GGCTTTCTCTGCATCCTTTGGATC	358	1	60>50
		AGGACAGAGTTTAGGGTATGGATTTG	CTGCCTCAATCCTCTTATCAAAGC	311	1	60>50
		AAGTCACAGGAGTGTCCAGAGGG	CCACAAGGCTTGTCATTCCATC	357	1	60>50
		CGTAAAGAGACTGATGTCTGGACCAG	AAAGGAAACTTGCAACACTCTGTCTC	352	1	60>50
		TGTCACCATGACCAACTATGATGAG	TGGTACACCTCAGGATAATCCAATG	299	1	60>50
FAF1 cDNA	(5'-3')					
		AGGAGGTGCCGTCTGCC	GAAGTCCAGCATCCGAGG		1	60>50
		CGTTTCGACCTGTAATGC	GGACTTCTCGGTGGGTGA		1	60>50
		GCCACCACCTTCATCATC	GTCATCTCCATCGCTATCAC		1	60>50
		GGGAACAGTCGGAAGAACA	TGGATTCAGCACAAAGCA		1	60>50
		GCTTCTTGCTATCTACCTCC	CTTCCTGTTGTTGGGCTG		1	60>50
		ACAAGGGAACACAACAGTAGA	ATGGCCTCACGTTCCTC		1	60>50
		GATGGCAGAACAGTTTCG	AGCCCTTCTCCTGACGCA		1	60>50
NRD1 cDNA	(5'-3')					
		GTTCAGGCCTGTTCCCC	AGATTCATCCGCTCCTAGAC		1	60>50
		GGAACAAGGCGAAGTCTA	TCATTATCCTCAGTATCAAGA		1	60>50
		GGAGCTGAAATAGAAGATGACG	TTTCACCCTGCTCGTCCTC		1	60>50
		AGCTGAAAGCAGGTCTGCAC	CGGTCAATTGCATCTCTGATC		1	60>50
		ATGCCTCAACTGATTGTGAACG	TGTTGCTGTTGAGGAGGAAGTG		1	60>50
		CGGATCCATTTGACACACC	TCCACCAAACAGTGCAAGAG		1	60>50
		AGGCAGCATTCTTTCTTTCC	GTCCTGCAATGGGTACAGC		1	60>50
		AAATGCTGCAGAAGCTAGGC	TCAGCTGGAAGATGAAGATCTG		1	60>50
		TGGGCTGAACTGTGGAATAG	CACATTTGCTGCAGATTTCTG		1	60>50
		CGTTTCCATCTAATTTCACCG	CCAACGGGCATATTCCAAG		1	60>50
		TTCAATTCCACACCAGCTGTC	ATCACCCTTGTTCAGAGCTTTC		1	60>50
		GCAGTTCCAGGTGGTAGAGC	TTTGGTTGCCTGAGTCCC		1	60>50
		GAACACATCCGGGATTCTAGG	GCGGTCAAAGAGGTACTGCTG		1	60>50
		GAGGTGGATAGGAACTGGAATG	TTCAGGCCAACGTGACTGC		1	60>50
EPS15 cDNA	(5'-3')					
		CCCTTGCATGATGGAAACACC	ACTGGTATCATGAAATCTTGGTGGAG	342	1	60>50
		TGTTGCTTTGCGTCTTGTGGC	AACATGGCAACTGCAAACTCATCTC	348	1	60>50
		TGATGGAATGCTTGACAGAGATGAG	GAACGTGAGGAGGATCAATGCC	380	1	60>50
		CAGTCAGAAGTTAATCAAGGGCATTG	CCAGTTCATCAAGGAGTTCCTGTACC	338	1	60>50
		AAGTGAGGTTCAGGATCTTCAAGATG	TGCATTGAACTAATTTCCTGTTGTG	378	1	60>50
		AATTGGAGGAGAGTGTAGAGTCAGGG	TCAGTCACACCAGAAGGCAGTAGTTC	331	1	60>50
		CACCAGGAATCTCCAGCAAGAAGTAG	TATTGTTGGCTGCACTGAAAGGG	376	1	60>50
		TCAGACTGTTTCTTCAGGCAATCTAC	AGTTCCGATCTTTGGTGGCAG	369	1	60>50
		AGCTCTGTCAGCAACGTAGTGATTAC	CTCTGCTCTTCCTCTCTCTCACTTTC	377	1	60>50
		CCTCTGAAGAAGATATGATCGAATGG	AAGGTGAATTTGTCTGACTGGGTTAC	364	1	60>50
		GAATACAAAAGGTTTGAGATT	CAAAATATTTTTCCTTATCTCT	451	1	60>50
		CTAAAGTGCATTTTTAAATTTC	ACCCACTAAGAACTAAACAGG	418	1	60>50
		AAAGTCTTAACATGGCAGCCATTC	ATTTCAAATTCAGCTTGTCCACTAGG	422	1	60>50
		CTATAAAATTGTCAAGCTAGCAA	CAAGAAGAAATGAATTGGATATT	356	1	60>50

5.6 ALL 1p32 SNPs

	Celera SNP ID	NCBI dbSNP ID (rs#)	F Primer	R Primer	Variation (1/2)	Enzyme	Alleles and Fragment Sizes	
1	hCV12109344	2065996	AAAGCAGGCCCAGCACAG	GGGAGGCATGGTAATGACTAATC	G/A	Tsp 509 I	G 170,36	A 101,69,36
2	hCV11872095		AGACCTGCCACTGACTTCCATCC	TCACCACTTTGGACGTGCCTTC	A/G	Taq I	G 394	A 144, 250
3	hCV1476946		TTAGGTCTTCTGATCCATGAACACAG	AATTAAAGAAAGCCGGGAGTGG	G/T	Hinf I	G 203, 163, 9	T 203, 93, 70, 9
4	hCV11872065		ACATCCTCGATCACCACCTCAC	GGTTGTGGAGGTAAAGAGTGCTAGTC	C/A	Tsp509 I	C 349, 38,14,13	A 241, 108, 38, 14,
5	hCV3027948		TCCATTACGTCACTACTTTCTCAGGG	GAATGACTCCTCAGCACTACTCCTTG	C/T	Nci I	T 445	C 242,203
6	hCV3027932		ACAAACACTCAGATCGCACCTGG	TTCCATCTCTTCTGTGTTTCCTTGTC	C/G	Bsl I	C 310,58	G 216,94,58
7	hCV3027900		CACTGGATTAGGAAGGAGAGACTTTG	CTCTCTGAATCTCAGGACTCTTGGTC	G/T	Bsr B I	G 243, 111	T 354
8	hCV2809720		ACTCTCACCTCCAGCCTCCTCCAATC	TCTTCACCCTGGGCCTTTGCAC	G/A	Bst E II	A 452	C 263, 189
9	hCV2371385		TATTGGGAAGCCAGAATCATCAAATG	GGAATGCACCATATGTCCTCATCTG	A/G	Fok I	A 481, 28	G 252,229,28
10	hCV435984		AATATGCACTGTTTCCCAGGTATGTC	TGCTGCTGCCTGTAAACATGC	G/A	Acc I	A 383	G 219,164
11	hCV15886654	2494883	AGGCACTGAATGGATTTGGGAC	ATCACAGCAGGAGATAGGACAGGC	A/G	Dde I	A 267	G 176,91
12	hCV9509099	1398868	AAACAGCAATGTCAGCACAGGC	AGTGTAGATGATGGCAGGAGAAGATG	A/G	Rsa I	A 198,83	G 128,83,70
13	hCV8847088	1043141	TTTAAGACGTTCCATCATAG	GCTAACAACCTCATTCCTC	C/T	Hae III	T 327,147, 51, 32	C 327,109,51,38,32
14	hCV11870545		CCCTCTCCAGACATCATG	CACTCTCTACCATTATTCTCAC	C/T	Dde I	C 179,66,27	T 245,27
15	hCV277402		AAACACCTAGTACTGCGATG	GGAAATATGTAAACACCTCTG	A/G	Apo I	A 176,76	G 252
16	hCV1413746		AACTATCTGTAAACCAAGCTC	ATGAAGTGAATGTTGTCTGATC	G/A	Rsa I	G 267	A 144,123
17	hCV3122409		CTCATGAAGGCTAGGCTC	GCATCTAGCATTATCTTTCAAG	G/C	(DHPLC)	G=1	C=2
18	hCV1591218		CTTAGGTCACTGGATTCGT	ACACGGTTACTGACTTAGATG	T/G	Bst E II	T 375	G 293, 82
19	hCV11740230		ATGGAACAACCGTAGAACTG	CAATCATCCTCTACCCACAG	T/C	Ssp I	C 476	T 263, 213
20	hCV386562	1316961	CTTTCTGCTTCTAGGTCTCC	GACCTCCTTCTCCTCTTC	A/G	Alu I	A 253,222,24	G 222,134,119,24
21	hCV428790		CTCTTCCAGGCTCATTCAG	CCCCTTTGAGTATGGATGA	G/A	Hae III	A 380	G 228, 152
22	hCV8847889		AAACTTGACTCATCTGGCAGCTTTAC	TGCTTGTGACATTCCCTTGTACC	T/C	Bst N I	C 306, 81, 36, 24	T 344, 81, 24
23	hCV2776353	1126997 [illegible]	TGAGTTCAATTCCACACCAGC	TGCTATATTGCCCAGGCTAGTC	T/C	Dde I	T 345,55,42,39	C 387,55,39
24	hCV2776338	856610	CTTTATACAGTTTCTAGAGTGG	GAACACTTAGAGACCATATTTG	G/T	Dde I	G 262,128,86,9	T 348,128,9
25	Glu indel		[HEX]-CCAGATACATCAAATTACAGAATGGC	TCTTCATCGTCATCTTCTATTTCAGC	169/172	(ABI sizing)	169=1	172=2
26	hCV79313		AGAAAGCAAGTCAGTATGGC	GGTATCATCTTCTATTGCCT	C/T	Acc I	T 350	C 234, 116
27	2230325	2230325	CGAAGACTCCACCTCCAAATAC	CACACTCGTACCCTGTTTACCC	T/C	Dde I	T 132,112,87	C 199,132
28	hCV1805212	1024313	TTTGCTGCCCTCTGCTAG	GGAGAAAGCAACCAGAAG	G/A	Rsa I	G 349,229	A 255,229,94
29	hCV1805260		TCCCTCCAAGGCAGAGTCTCAC	GAGCTCAATGACAAACAATGTATGGG	G/A	Dpn II	A 220,120	G 170,120,50
30	hCV12112305	649676	TTTGTTGCCCAGGGTGGTC	CTCTTCACTCTACATGGTTTCACTGG	T/C	Hph I	C 260,41	T 208,54,41
31	hCV11286191		AACACAAAGGATAAATGCTTGAGGTG	AACGACACTAATCCCATTCATGAGG	G/A	Bst N I	A 458	G 302, 156
32	hCV11873295		CAAACAAGCCTCAGCTAGTTCTCTG	CTCACTATCTTGTCCAGGCTGGTC	A/G	Tsp 509 I	G 328, 66, 15	A 278, 66, 50, 15
33	hCV7842707		TGGGATGAGAAGAGCAGCTGG	CATACCTTTGGTCTGCGGAGAAC	T/C	Hpa II	T 432, 64	C 258, 174, 64
34	hCV11869553		ACTCCAACCTGGGTGACAGAGCAAG	CAAAGGTGCCGCCAATTCGTG	T/C	Nla III	C 410, 78	T 290, 120, 78
35	hCV1797353	1288516	AGGGAAACTGGGAGTTGGGAG	AATAGAGGTGCATGGGAATCCC	G/A	Fok I	G 267	A 150,117
36	hCV1797412	1288489	AGGCCCTCATCTGTCTACATCACC	GCACATCCTGTGTTCTTCCTGTTTC	C/G	Bst N I	G 121,91	C 91,83,38
37	hCV1797425	1288480	TCCCAGGAGGTATGCAAGCAG	GCTACCCTGGATGTTTCCCAAG	C/T	Taq I	C 160,105	T 285
38	hCV1797435	2782491	CACCCTGCCACTCCCAAGTC	CCCAGAGGAGGAGGCATTTG	A/C	Bsp 1286 I	C 94,66,50	A 160,50
39	hCV11871255		GTCTCGAACTCCTGACCTCAAGCG	TCTGCGGATCACACTTGTGAGAATG	C/A	Ase I	C 437	A 293, 144
40	hCV1452838		GCAGTGGGCATCAAATGAGGC	CAGGATAGAGGGAGCTGGAACCAG	G/C	Bso B I	G 203, 144, 50	C 253, 144
41	hCV2431676		CAGTAATATGCACCCTGGAGAGACC	TGCAGAGGTGAGGAGAAGCTGTC	C/G	Ban I	G 273, 144	C 417
42	hCV1761854		CCAAACTTCCCGGCTAGCTGAG	TCTGGCTGGGATTTAGAGGAGTGG	A/G	Hae II	A 318, 57	G 207, 111, 57
43	hCV11870762		GGCTCAAACCCTGACTCAATTCC	GATAAGGACCAGGAGGCAAACTACC	C/G	Xho I	C 294, 188	G 482
44	hCV736440		GAACCAGCCACTCCACATTCAGAC	AACCCAGAAGAGCCGCAGTTAAG	G/A	Hpa II	A 438	G 299, 139
45	hCV3154981		AAGAACATAAGGGCGCTCAATTAATG	CAATCAGAACGCTTCTAACCCACAC	T/A	Taq I	T 277, 166	A 443
46	hCV11732075		ACAAGACCCTGCTAAGACACTGGC	ACGTGGGCAAGTCATTGTCCTC	G/A	Nla III	G 368	A 252, 116
47	hCV2794215		TCCACCATATCAGTCCCTTGTCAGTC	AGGGTGTGGGAGTTCTCTTCTGATTG	T/C	Hpa I	T 435	C 298, 139
48	hCV998739		TAGCTGTCAGCAATCAAGTGGTTACC	TGGTCTTGCAGTTTGAAATGTGATG	C/G	Hind III	G 295,113	C 408
49	hCV1758290		GGATCAGGAAGAAGGATGAGCAAC	TGGCTCAGGATAGATAGATGGGAATC	T/C	Sma I	T 430	C 283, 147

5.6 1p32 STRPs

Marker	Size Range (bp)	Repeat type	Primer	Sequence
D1S2874	230-240	dinucleotide	F	[HEX]-CCTAGACGCCCCTATAAG
			R	GCAAGAATTATGTCTGATGG
D1S197	115-129	dinucleotide	F	[HEX]-TCATGTCCCTCCTCCCAAAG
			R	GAGCAAGCATCCAAAAACGA
D1S1661	89-101	dinucleotide	F	[HEX]-TTCATGGTTATAAAATATGATGCA
			R	GCGAGATTCAGTCTCAAAAA
D1S232	184-202	dinucleotide	F	[HEX]-GAGCAAGACTCTCTGTCCCC
			R	CCATGTTCAAGGGTCAACTG
D1S319	149-163	dinucleotide	F	[HEX]-TGTAAAGCAAAATTACAGCCT
			R	CTTAACCCAAATATTAGGTAGCA
D1S231	158-168	dinucleotide	F	[FAM]-TAACTGAGTGCTTACCTTATGC
			R	CATGTGTTATATTTCAGGGG
D1S2652	94-106	dinucleotide	F	[FAM]-GCAGGTGTGATGCCAGG
			R	TACGGCTGATTGGGAGAAC
D1S475	191-207	dinucleotide	F	[FAM]-CCTAGTGCCTGGNACATAGT
			R	TGACTGGATACTGAGGGTTTGT

PCR Conditions PCR Mix 1
Program 60>50

8.2 Appendix 2 - COLDMAP Input Parameters

The input parameters for the program are given in the file GEN.IN.

These are:

Number of individuals in sample 369

Number of markers 49

Assumed marker mutation rate 0.000025

Assumed recombination rate (cM/Mb) 1

Population growth parameter rate (0=constant) 0

Control parameter for choosing branches to be removed from the shattered genealogy (0=all branches equal rate, 1=branches have weight proportional to their length)

Number of iterations 20000

Thinning parameter for recorded output 10

Random start (0=yes, 1=use final output from previous analysis)

If you specify a random starting point, you then have to specify Starting points for the following parameters:

Shattering parameter (0-1) 0.9

Location of disease locus 52.0

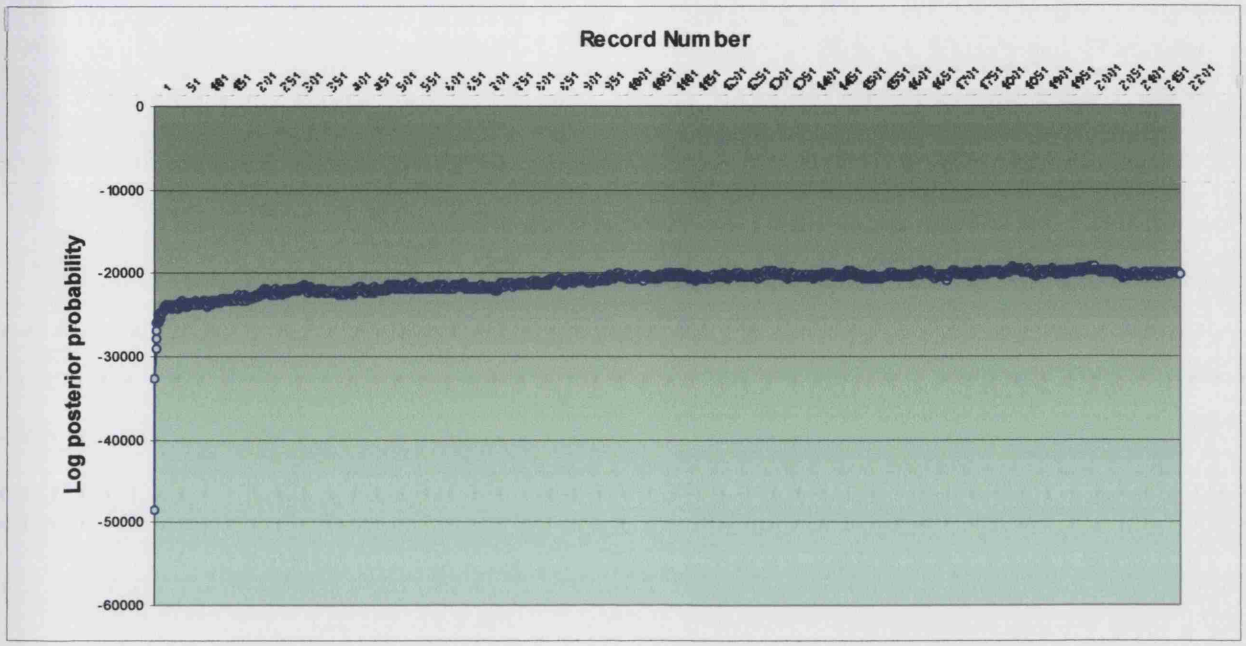
Effective population size 1000

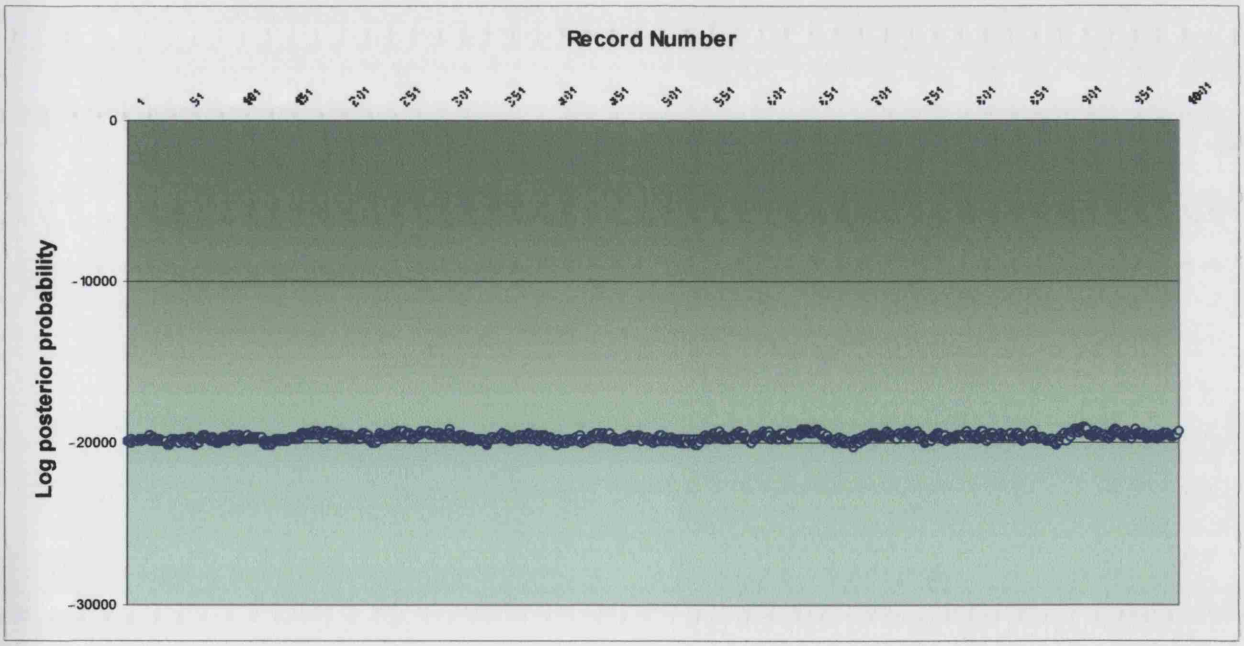
Frequency of allele 1 (initially same at each marker) 0.5

Run the algorithm using the command:

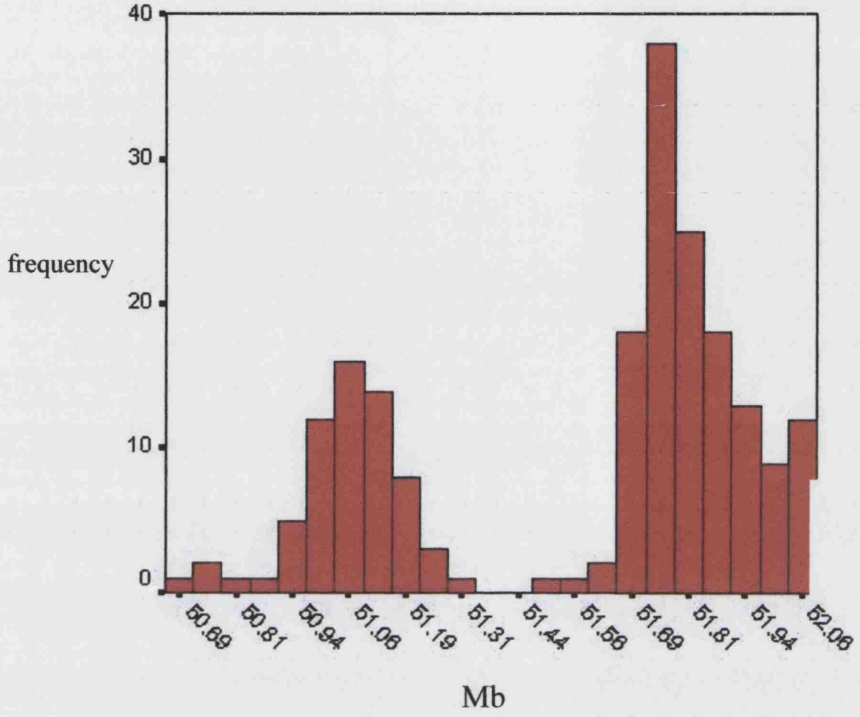
```
./genldco < gen.in > gen.out &
```

8.3 Appendix 3 - COLDMAP Mutation Location Convergence

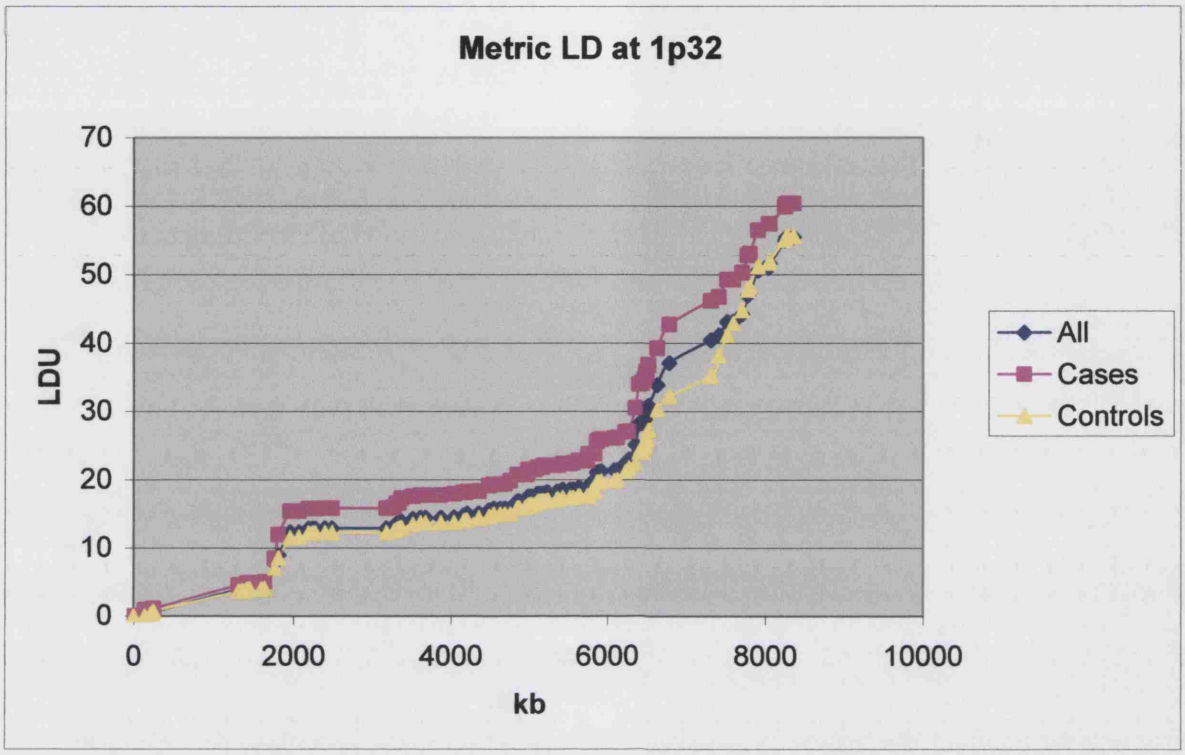




8.4 Appendix 4 - COLDMAP 'Dummy' Run Estimated Mutation Location



8.5 Appendix 5 - Metric LD Map of 1p32



8.6 Appendix 6 - Patient Information

Sample	Age	Sex	Age of Onset	Family History	Classification
P-001	76	m	73	0	Atypical
P-002	64	f	52	1 (cousin)	Probable
P-003	77	m	61	0	Probable
P-004	79	m	72	0	Probable
P-005	64	m	53	0	Probable
P-006	49	m	48	0	Atypical
P-007	77	f	70	0	Possible
P-008	72	m	64	0	Possible
P-009	77	f	74	0	Probable
P-010	59	f	58	0	Possible
P-011	73	m	58	0	Possible
P-012	52	m	49	0	Probable
P-013	76	m	68	0	Probable
P-014	69	m	40	0	Probable
P-015	84	f	72	0	Possible
P-016	52	m	39	0	Probable
P-017	65	f	57	0	Possible
P-018	73	m	70	0	Probable
P-019	66	m	59	1 (cousin)	Probable
P-020	85	f	75	0	Probable
P-021	80	m	67	0	Probable
P-022	60	m	49	0	Probable
P-023	67	m	59	0	Probable
P-024	51	m	44	0	Atypical
P-025	67	f	60	0	Probable
P-026	62	f	50	0	Probable
P-027	81	m	73	1 (sib)	Possible
P-028	65	m	48	0	Probable
P-029	48	m	35	0	Possible
P-030	77	f	67	0	Possible
P-031	78	f	71	0	Atypical
P-032	78	m	77	0	Probable
P-033	62	f	47	0	Probable
P-034	81	f	70	1 (uncle)	Probable
P-035	76	f	70	0	Possible
P-036	71	f	55	0	Probable
P-037	76	m	74	0	Possible
P-038	69	f	46	death at 69	Definite
P-039	75	f	68	0	Probable
P-040	78	f	60	1 (uncle)	Probable

P-041	75	m	71	0	Probable
P-042	70	m	55	0	Probable
P-043	81	m	78	0	Possible
P-044	72	f	65	0	Atypical
P-045	72	m	62	0	Atypical
P-046	79	m	52	0	Probable
P-047	82	m	68	0	Probable
P-048	77	m	72	0	Atypical
P-049	80	m	63	0	Probable
P-050	74	m	65	0	Possible
P-051	74	m	55	0	Probable
P-052	54	m	47	0	Probable
P-054	56	m	42	0	Probable
P-055	62	f	55	0	Probable
P-056	63	m	60	1 (father)	Probable
P-057	82	f	75	0	Possible
P-058	69	f	56	0	Probable
P-059	67	f	62	0	Probable
P-060	63	m	54	0	Probable
P-061	68	f	63	1 (brother)	Probable
P-062	70	f	65	0	Probable
P-063	67	f	53	1 (mother)	Probable
P-064	70	m	58	0	Possible
P-065	72	f	60	0	Probable
P-066	70	m	65	0	Probable
P-067	73	m	60	0	Atypical
P-068	64	m	61	0	Atypical
P-069	64	m	59	0	Probable
P-070	71	f	64	1 (cousin)	Probable
P-071	63	m	53	0	Probable
P-072	69	m	60	0	Probable
P-073	67	m	51	0	Probable
P-077	71	m	61	0	Probable
P-082	72	m	62	0	Possible
P-083	77	m	70	1 (uncle)	Probable
P-084	81	f	64	1 (2 cousins)	Probable
P-085	70	m	64	0	Probable
P-088	66	m	63	1 (cousin)	Probable
P-089	50	m	43	0	Probable
P-090	78	f	75	1 (cousin)	Possible
P-091	69	m	64	1 (cousin)	Possible
P-095	56	m	25	0	Probable
P-096	65	f	62	1 (2 cousins)	Probable
P-097	75	m	62	0	Probable

P-098	56	m	50	1 (aunt)	Probable
P-099	55	m	45	0	Probable
P-100	67	m	49	1 (mother)	Probable
P-101	62	f	57	0	Probable
P-102	75	f	40	0	Probable
P-103	47	f	44	0	Probable
P-104	70	m	60	1 (father)	Probable
P-105	59	m	50	0	Probable
P-112	50	f	36	0	Probable
P-113	62	m	40	0	Probable
P-114	74	f	70	0	Probable
P-115	70	f	51	0	Probable
P-116	79	m	70	0	Possible
P-117	58	m	55	0	Probable
P-120	52	m	49	0	Probable
P-121	79	f	68	1 (2 sibs)	Probable
P-122	76	f	73	1 (mother and grandmother)	Possible
P-123	62	f	33	1 (aunt)	Probable
P-124	80	m	70	0	Probable
P-125	85	m	80	1 (mother and uncle)	Possible
P-126	74	m	67	0	Probable
P-127	52	f	46	0	Possible
P-129	82	m	70	0	Possible
P-130	67	m	61	0	Probable
P-131	76	m	65	0	Probable
P-132	64	m	55	0	Probable
P-133	74	m	67	1 (uncle)	Probable
P-134	72	m	65	0	Probable
P-135	64	f	62	0	Probable
P-136	65	f	53	0	Probable
P-137	73	m	67	0	Probable
P-138	56	f	39	0	Probable
P-139	69	m	63	0	Probable
P-140	77	f	71	0	Atypical
P-141	63	f	54	0	Probable
P-142	67	m	59	0	Probable
P-143	75	f	44	0	Probable
P-144	46	m	38	0	Probable
P-145	71	M	65	0	Possible
P-146	81	f	78	0	Probable
P-147	73	f	60	0	Probable
P-148	51	f	46	0	Probable
P-150	76	m	56	1 (sib)	Probable
P-151	74	f	64	0	Probable

P-152	57	f	55	0	Possible
P-153	75	f	60	0	Probable
P-154	71	f	66	1 (aunt)	Probable
P-155	75	f	69	0	Atypical
P-156	68	f	57	0	Probable
P-158	79	f	72	1 (aunt)	Probable
P-159	70	m	60	0	Probable
P-160	85	m	72	0	Possible
P-161	77	f	62	0	Probable
P-162	68	f	50	0	Probable
P-163	57	m	51	0	Probable
P-164	93	f	88	0	Probable
P-165	77	m	74	0	Possible
P-166	89	f	85	0	Probable
P-167	70	f	62	1 (mother)	Probable
P-168	83	f	70	1 (uncle and cousin)	Probable
P-169	62	f	44	0	Probable
P-170	66	m	63	0	Probable
P-171	49	f	38	0	Probable
P-172	66	f	58	0	Possible
P-173	67	m	59	1 (uncle)	Probable
P-174	55	m	50	0	Probable
P-175	56	m	50	0	Probable
P-176	45	m	40	0	Probable
P-177	93	m	62	1 (brother)	Probable
P-178	75	m	71	0	Possible
P-179	76	m	69	0	Possible
P-180	78	m	73	1 (sib)	Probable
P-181	63	m	32	0	Probable
P-182	81	m	62	0	Probable
P-183	73	m	63	0	Probable
P-184	51	m	46	0	Probable
P-185	67	m	53	0	Possible
P-186	78	m	71	0	Probable
P-187	80	f	77	0	Probable
P-188	72	m	51	0	Probable
P-189	79	f	55	1 (mother)	Probable
P-190	59	m	56	0	Possible
P-192	52	m	48	0	Probable
P-193	71	f	65	1 (mother)	Probable
P-194	76	f	65	0	Probable
P-195	66	m	64	0	Probable
P-196	76	f	73	0	Probable
P-197	76	f	70	0	Possible

P-198	75	m	70	0	Probable
P-199	81	m	64	0	Probable
P-200	61	m	55	1 (uncle)	Probable
P-201	63	f	50	0	Probable
P-202	75	m	73	0	Probable
P-204	75	m	70	0	Possible
P-205	57	f	53	0	Probable
P-206	65	m	63	0	Probable
P-207	57	f	51	0	Probable
P-208	66	f	61	0	Probable
P-209	54	m	48	0	Probable
P-210	71	m	49	1 (3 aunts)	Probable
P-211	70	f	66	0	Probable
P-212	66	m	62	0	Probable
P-213	73	m	62	1 (father)	Probable
P-214	54	m	52	0	Probable
P-215	73	m	70	0	Probable
P-216	78	f	62	1 (brother, sister)	Probable
P-217	65	f	53	0	Probable
P-218	74	f	70	0	Possible
P-219	58	m	51	1 (father)	Probable
P-221	53	f	40	1 (father)	Probable
P-222	60	m	53	1 (mother)	Probable
P-223	73	m	62	0	Possible
P-224	71	f	47	0	Probable
P-225	63	m	58	0	Probable
P-226	65	f	60	1 (aunt)	Probable
P-227	75	m	47	0	Probable
P-228	61	m	48	0	Probable
P-229	70	k	65	0	Probable
P-230	82	m	75	0	Probable
P-234	54	m	44	0	Probable
P-235	63	m	61	0	Possible
P-236	67	m	61	0	Probable
P-237	50	f	35	0	Atypical
P-238	56	m	51	0	Atypical
P-239	71	m	55	1 (brother)	Probable
P-240	74	f	62	0	Probable
P-241	84	f	60	0	Probable
P-242	56	f	50	0	Probable
P-243	63	f	50	1 (father)	Probable
P-244	67	m	60	1 (cousin)	Probable
P-245	80	m	68	0	Possible
P-246	62	f	51	0	Probable

P-247	71	m	50	0	Probable
P-248	71	m	68	0	Possible
P-249	59	m	53	0	Possible
P-250	51	m	47	1 (aunt)	Possible
P-251	76	f	72	0	Probable
P-252	77	m	59	1 (mother)	Probable
P-253	74	m	68	0	Probable
P-254	75	m	70	1 (mother and uncle)	Possible
P-255	77	m	70	0	Probable
P-257	46	f	36	0	Probable
P-258	66	f	56	1 (father)	Probable
P-259	69	m	55	0	Probable
P-260	74	f	63	0	Probable
P-261	46	m	41	0	Possible
P-262	55	m	31	1 (father)	Probable
P-263	55	m	40	0	Possible
P-264	40	m	38	0	Probable
P-265	81	m	69	0	Probable
P-266	63	m	58	1 (uncle)	Possible
P-267	57	f	52	0	Probable
P-269	52	m	53	0	Probable
P-270	80	m	70	0	Possible
P-271	71	f	65	1 (father)	Possible
P-272	60	m	57	1 (sib)	Possible
P-273	76	m	74	1 (cousin)	Possible
P-274	42	m	39	1 (father)	Possible
P-275	45	m	35	1 (uncle)	Probable
P-276	76	m	72	0	Probable
P-277	53	m	50	1 (sib)	Possible
P-278	52	f	40	1 (grandfather)	Probable
P-279	69	m	52	0	Possible
P-280	73	m	54	0	Probable
P-281	77	m	60	0	Possible
P-282	70	f	64	0	Possible
P-283	77	f	71	0	Probable
P-286	45	m	42	0	Probable
P-287	69	m	62	0	Probable
P-288	55	m	46	0	Probable
P-290	74	m	70	0	Atypical
P-291	56	m	44	0	Probable
P-292	47	f	39	0	Probable
P-293	50	m	46	0	Probable
P-294	83	m	82	1 (sib)	Probable
P-295	77	m	65	0	Possible

P-297	67	f	53	0	Probable
P-299	74	m	65	0	Probable
P-300	70	f	67	0	Possible
P-301	72	m	61	0	Probable
P-302	72	f	70	0	Probable
P-304	77	f	60	0	Probable
P-305	73	m	72	0	Probable
P-306	70	m	62	0	Possible
P-307	42	m	39	0	Probable
P-308	76	f	75	1 (father)	Probable
P-309	74	f	69	0	Probable
P-310	51	m	39	0	Probable
P-311	50	m	46	0	Probable
P-312	43	m	55	0	Probable
P-313	43	m	44	0	Probable
P-316	57	f	50	1 (aunt)	Probable
P-317	77	f	72	0	Probable
P-318	74	f	72	0	Probable
P-319	65	m	51	0	Probable
P-320	81	m	78	0	Probable
P-321	71	m	69	0	Probable
P-322	62	m	56	0	Probable
P-323	65	f	52	0	Possible
P-324	54	f	53	0	Possible
P-325	46	f	37	0	Probable
P-326	54	m	51	1 (grandfather)	Probable
P-327	77	m	64	0	Probable
P-328	74	m	64	0	Probable
P-329	59	m	57	0	Probable
P-330	54	m	51	0	Probable
P-331	85	f	72	0	Probable
P-332	82	m	74	0	Probable
P-334	75	m	62	1 (father & uncle)	Probable
P-335	75	m	70	1 (grandmother)	Possible
P-336	78	m	65	0	Probable
P-337	72	f	57	0	Probable
P-338	82	f	72	0	Probable
P-339	71	f	68	0	Probable
P-340	66	m	58	0	Atypical
P-341	68	m	60	0	Probable
P-342	82	m	73	0	Possible
P-343	67	f	67	1 (sib)	Probable
P-344	68	f	49	0	Probable
P-347	63	f	54	1 (aunt)	Probable

P-348	83	m	70	1 (nephew)	Probable
P-349	71	f	60	1 (cousin)	Probable
P-350	65	f	54	0	Probable
P-351	65	m	60	0	Possible
P-352	77	m	49	0	Probable
P-353	76	m	40	0	Probable
P-354	75	m	49	0	Probable
P-355	77	f	72	0	Probable
P-356	51	m	51	0	Atypical
P-357	69	m	48	1 (uncle)	Probable
P-358	79	m	53	1 (cousin)	Probable
P-359	71	m	61	0	Probable
P-360	76	m	54	0	Probable

9 Publications