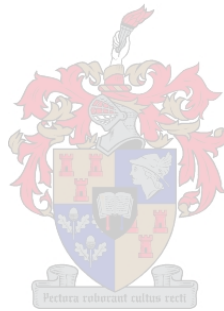# A Molecular Investigation Of A Mixed Ancestry Family Displaying Dementia And Movement Disorders

Fatima Abrahams-Salaam

Thesis presented in partial fulfillment of the requirements for the degree of

Master of

Science in Biomedical sciences at the Faculty of Health Sciences, Stellenbosch University

Supervisor: Dr. Soraya Bardien-Kruger

Co-supervisor: Prof. Jonathan Carr

December 2008

DECLARATION

By submitting this thesis electronically, I hereby declare that the entirety of the work contained therein is my own original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 5 December 2008

OPSOMMING

Hierdie studie ondersoek 'n Suid-Afrikaanse kleurling-familie wat presenteer het met progressiewe dementia en bewegingswanfunksie in 'n aantal individue binne drie generasies. Aanvanklike simptome sluit in persoonlikheidsveranderings en tremors, wat progressief verander het na ernstige dementia en totale immobiliteit. Die gemiddelde aanvangsouderdom van persone met die simptome was in die dertigjare en mortaliteit het binne 10 -15 jaar ingetree. Die doel van die studie was om die genetiese oorsaak van die siektetoestand te bepaal en die patologie te ondersoek.

'n Mutasie-soektog wat moontlik na die fenotipe kon lei, is in die familie gedoen. Dit het mutasies in Huntington-siekte, Parkinson-siekte, Dentatorubral-Pallidoluysian-atrofie, spinoserebrale-ataksie (tipe 1, 2, 3, 6, en 7), Huntington Tipe-2-siekte (HDL2) en verskeie mitokondriale siektes ingesluit. Enkelstring konformasie-polimorfisme analise en direkte DNA-volgorde-bepaling is gebruik om nukleotied-veranderinge te bepaal. Genotipering op 'n ABI genetiese-analiseerder is gebruik om herhalingsvolgorde-verlengings se grootte te bepaal. Deur gebruik te maak van haplogroep- en Kort-Tandem-Herhalings-analise (STRs) van die Y-chromosoom en mitokondriale DNA van 'n aangetaste individu, is die familie se etniese oorsprong bepaal. Ten einde die geen-uitdrukking te bestudeer is Omgekeerde Transkriptasie Polimerase Kettingreaksie (RT-PCR) en komplementêre DNA-analise (cDNA) van die J*unctophilin-3* (*JPH3*)-geen gedoen.

Na uitsluiting van verskeie bekende mutasies is 'n herhalingsvolgorde-verlenging in die HDL2-geen in die familie aangetoon. Huntington Tipe-2 siekte is seldsaam en word veroorsaak deur 'n CAG/CTG-herhalingsverlenging in 'n alternatief-uitgedrukte transkripsie van die *JPH3*-geen. HDL2 kom meestal by swartmense van Afrika-oorsprong voor. Die familie in hierdie ondersoek se etniese oorsprong-bepaling het aangedui dat hulle van kleurling-afkoms is. Hierdie is die eerste beskryfde Suid-Afrikaanse kleurling-familie met 'n herhalingsvolgorde-verlenging in die HDL2-geen. 'n Loods-studie het die voorkoms van die herhalingsvolgorde-verlenging onder drie Suid-Afrikaanse sub-populasies ondersoek, ten einde te bepaal of swart Afrikane meer geneig is om die siekte te ontwikkel. 'n Statisties beduidende verskil (P=0.0014) in die voorkoms van die herhalingsvolgorde-verlenging is onder die swart en kaukasiese Afrikane gevind. Geen gevolgtrekking kon egter gemaak word dat swart Afrikane 'n groter herhalingsvolgorde-verlenging het nie.

Die herhalingsvolgorde-verlenings is geleë in 'n alternatief-uitgedrukte transkripsie van die *JPH3*-boodskapper RNA (mRNA). Ten spyte van die feit dat die *JPH3* geen-omgewing hoogs behoue gebly het tussen mens-, muis-, en sjimpansee-genome, kom die herhalingsvolgorde nie in die muis-homoloog van die geen voor nie. Deur gebruik te maak van fetale brein "cDNA" en PCR met voorvoerders wat spesifiek is vir die twee transkripsie-produkte, het hierdie studie onafhanklik bevestig dat verskillende *JPH3*-"mRNA" transkripsie produkte (die vollengte en 'n korter alternatief) voorkom. As gevolg van die afwesigheid van breinweefsel van HDL2-geaffekteerdes, is die transkripsie-produkte in twee aangetaste individue se limfosiete ondersoek. "Real-time" PCR is gedoen met RNA wat uit limfosiete geïsoleer is van twee HDL2-aangetasdes. Hierdie eksperimente was onvoldoende en vereis verdere optimisering. "Real-time" PCR eksperimente in verskillende weefsels (brein en ander) van HDL2-geaffekteerdes behoort meer inligting te verskaf oor die *JPH3*-geen.

In hierdie studie is die eerste kleurling-familie met 'n CAG/CTG herhalingsvolgorde-verlenging in die HDL2-geen identifiseer. Genetiese raadgewing en pre-simptomatiese toetse in onaangetaste individue binne die familie is nou moontlik. Hierdie studie het sekere eienskappe van die geen onafhanklik bevestig. Verdere navorsing op HDL2 is noodsaaklik om die siekte beter te verstaan.

ABSTRACT

A South African family of Mixed Ancestry presented with a rapidly progressive dementia and a movement disorder which affected a number of individuals across three generations. The initial symptoms included personality changes and tremors that escalated to severe dementia and eventually a completely bedridden state. It was determined that the mean age at onset was in the third decade of life and affected individuals died within 10-15 years after the onset of symptoms. The aim of the present study was to elucidate the genetic cause of the disorder in this family and to further investigate the patho-biology of the disease.

Mutations that could possibly cause the observed phenotype in this family were screened for. These included loci implicated in Huntington's disease, Parkinson's disease, Dentatorubral-Pallidoluysian Atrophy, Spinocerebellar ataxias (types 1, 2, 3, 6, and 7), Huntington's disease-like 2 (HDL2) and several mitochondrial disorders. Single-strand Conformation Polymorphism (SSCP) analysis and direct sequencing were used to detect possible mutations while genotyping on an ABI genetic analyser was used to detect disorders caused by repeat expansions. Haplogroup and Short Tandem Repeats (STRs) analyses of the Y-chromosome and mitochondrial DNA of one affected family member was used to determine the family's genetic ancestry. Reverse transcriptase polymerase chain reaction (RT- PCR) and complementary DNA (cDNA) analyses of the *Junctophlin-3* (*JPH3*) gene was performed to provide information on the expression profile of this gene.

After the exclusion of several genetic loci it was shown that this family had HDL2. This is a rare disease caused by a CAG/CTG repeat expansion in an alternatively spliced version of the *JPH3* gene. HDL2 occurs almost exclusively in individuals of Black African ancestry. The genetic ancestry data suggested that the family member was most likely of South African Mixed Ancestry making this the first reported family of South African Mixed Ancestry with HDL2. A pilot study investigated the repeat distribution amongst three South African sub-populations in order to determine whether there was a bias in the repeat distribution that possibly predisposes Black Africans to develop the disease. The results showed a statistically significant difference (P= 0.0014) in the distribution of the repeats between the Black African and Caucasian cohorts. However, no conclusions could be drawn as to whether Black Africans harboured larger repeats that predisposes them to developing HDL2.

The expanded repeat is located in an alternatively spliced version of the *JPH3* mRNA. Interestingly, this repeat is not present in the mouse homologue of the gene although the rest of the genomic sequence is highly conserved across the human, mouse and chimpanzee genomes. Using foetal brain cDNA and PCR primers designed to be specific for different *JPH3* isoforms, independent confirmation of the presence of two *JPH3* mRNA transcripts (the full length and a shorter alternatively spliced version) was provided. In the absence of brain tissue from an HDL2-affected individual, it was investigated whether both *JPH3* mRNA transcripts could be detected in lymphocytes. Using RNA isolated from the transformed lymphocytes of two HDL2-affected family members, real-time PCR was attempted. These experiments produced inconclusive results and required further optimisation. Further RT-PCR experiments for *JHP3* expression in different tissues (brain and other) obtained from HDL2-affected individuals would be of interest.

The present study identified the first Mixed Ancestry family with HDL2. This family will now be able to request genetic counselling and pre-symptomatic testing for all at-risk family members. Aspects of this study provided independent confirmation of characteristics of the mutated gene. More research on HDL2 will be crucial in understanding the pathogenesis of this disease.

INDEX

ACKNOWLEDGEMENTS

I extend my sincere gratitude to my Lord and Creator, Allah. It is only through His mercy that I have come this far in life.

I would like to express my whole hearted thanks to my supervisor Dr Soraya Bardien-Kruger for her diligent supervision, her patience and constant encouragement through every aspect of this project. Thank you for everything that you have taught me. I would also like to thank my co-supervisor Prof. Jonathan Carr for his advice and supervision on this project. I sincerely thank you both for giving me this opportunity and supporting me throughout.

My deepest gratitude is extended to the staff of the MAGIC lab, for their kinship and assistance, especially Prof. Moolman-Smook and Prof. Valerie Corfield for allowing me to use their lab. A special thanks to Craig, Lundi, Rowena, Carmen and Melissa.

I am eternally grateful to my parents for their support and encouragement and for being my pillars of strength when I needed it the most.

Sincere thanks are due to my husband for being my voice of reason, for the insurmountable patience, constant encouragement and for believing in me every step of the way.

Not forgetting my brothers for providing laughs when things got a little too serious.

Lastly a big thank you is due to the NRF, Harry Crossley Fund, Post- Graduate Bursary department of Stellenbosch University and the Department of Neurology for their financial assistance.

LIST OF ABBREVIATIONS

3'   :3 prime
5'   :5 prime
3' UTR   :3 prime untranslated region
5' UTR   :5 prime untranslated region
µg   :Micrograms
µl   :Micro litre
°C   :Degree Celsius
A   :Adenosine
AA   :Amino acids
AAO   :Age at onset
ABI   :Automated Bioanalyzer
AD   :Alzheimer's disease
ATP   :Adenosine triphosphate
ATN1   :*Atrophin 1* gene
BLAST   :Basic local alignment search tool
bp   :Base pair
C   :Cytosine
Ca^{2+} :Calcium
cDNA   :Complementary DNA
CJD   :Creutzfeld-Jakob disease
CNS   :Central Nervous System
CP value   :Crossing point value
dATP   :Deoxy-adenosine triphosphate
dCTP   :Deoxy-cytosine triphosphate
dGTP   :Deoxy-guanosine triphosphate
DJ1   :*Oncogene DJ1*
DM   :Diabetes Mellitus
DM1   :Myotonic dystrophy Type 1
DMSO   :Dimethyl Sulphoxide
DNA   :Deoxyribonucleic acid
dNTP   :Deoxy-nucleotide triphosphate
DRPLA   :Dentatorubral-Pallidoluysian Atrophy
dTTP   :Deoxy-thymine triphosphate
EBV   :Epstein-Barr Virus
E. coli   :*Escherichia coli*
EDTA   :Ethylene-diamine-tetra-acetic acid
EEG   :Electroencephalogram
EO   :Early onset
ESTs   :Expressed Sequence Tags
FISH   :Fluorescent *in situ* hybridisation
FENIB   :Familial Encephalopathy with Neuroserpin Inclusion Bodies
FTD   :Frontotemporal Dementia
FFI   :Fatal Familial Insomnia
FGF14   :*Fibroblast growth factor 14* gene
g   :Gram
G   :Guanine
GSS   :Gerstmann-Straussler-Scheinker syndrome

H_2O	:Water
HD	:Huntington's disease
HD-like	: Huntington's disease-like
HDL2	:Huntington's disease-like 2
HGDDRU	:Human Genomic Diversity and Disease Research Unit
IPTG	:Isopropyl-beta-D-thiogalactopyranoside
IT15	:*Interesting transcript 15* gene
JLA	:Jerk-locked averaging
JP	:*Junctophilin* gene (mouse)
JPH	:*Junctophilin* gene (human)
JPH3	:*Junctophilin-3* gene
L	:Litre
LB	:Luria Bertani
LHON	:Leber's Hereditary Optic Neuropathy
M	:Molar
ME	:Mitochondrial encephalopathy
MECM	:Mitochondrial encephalocardiomyopathy
MELAS	:Mitochondrial myopathy, Encephalopathy, Lactic acidosis and Stroke-like episodes
MERRF	:Myoclonic Epilepsy with Ragged-Red Fibres
$MgCl_2$	:Magnesium chloride
MIM	:Mendelian Inheritance of Man
ml	:Millilitre
mM	:Millimolar
MM	:MELAS/MERRF overlap
MMSE	:Mini-Mental State Examination
MRC	:Medical Research Council
mRNA	:Messenger ribonucleic acid
MRI	:Magnetic resonance imaging
mtDNA	:Mitochondrial deoxyribonucleic acid
MT-ATP8	:Mitochondrial adenosine triphosphate synthase subunit
MT-ND	:Mitochondrial NADH de-hydrogenase subunit
MTTF	:Mitochondrial phenyl-alanine transfer ribonucleic acid
MTTH	:Mitochondrial histidine transfer ribonucleic acid
MTTI	:Mitochondrial isoleucine transfer ribonucleic acid
MTTK	:Mitochondrial lysine transfer ribonucleic acid
MTTL	:Mitochondrial leucine transfer ribonucleic acid
MTTM	:Mitochondrial methionine transfer ribonucleic acid
MTTQ	:Mitochondrial glutamine transfer ribonucleic acid
NCBI	:National Centre for Biotechnological Information
ND	:Not determined
ng	:Nanograms
NHLS	:National Health Laboratory Service
OMIM	:Online Mendelian Inheritance in Man
OPRI	:Octapeptide repeat insertion
OXPHOS	:Oxidative phosphorylation
PARK2	:*Parkin* gene
PAS	:Periodic acid Schiff
PCR	:Polymerase chain reaction

PD	:Parkinson's disease
PINK1	:*PTEN induced putative kinase 1* gene
PME	:Progressive myoclonus epilepsy
pmol	:*pica* moles
PolyA	:Polyadenosine
PolyC	:Polycytosine
PolyQ	:Polyglutamine
PRPN	:*Prion protein* gene
Pty Ltd	:Proprietary limited
rCRS	:Revised Cambridge Reference Sequence
RED	:Repeat expansion detection
RNA	:Ribonucleic acid
RRF	:Ragged red fibres
rRNA	:Ribosomal ribonucleic acid
rpm	:Revolutions per minute
RT-PCR	:Reverse transcriptase polymerase chain reaction
Rt-PCR	:Real-time polymerase chain reaction
SB	:Sodium Borate
SCAs	:Spinocerebellar Ataxias
SEP	:Somatosensory evoked potentials
SNPs	:Single nucleotide polymorphisms
STRs	:Short tandem repeats
SSCP	:Single Strand Conformation Polymorphism
T	:Thymine
Ta	:Annealing temperature
Tm	:Melting temperature
TNTC	:Too numerous to count
tRNA	:Transfer ribonucleic acid
U	:Units
UCT	:University of Cape Town
UK	:United Kingdom
USA	:United States
UV	:Ultra violet
V	:Volts
W	:Watts
WT	:Wild type
www	:World Wide Web
X-GAL	:5-bromo-4-chloro-3-indolyl-Beta-D-galactopyranoside

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER THREE: RESULTS

Section A: Investigation of the causative mutation

Section B: Analysis of the *Junctophilin-3 (JPH3)* gene

CHAPTER FOUR: DISCUSSION

APPENDIX

REFERENCES

CHAPTER ONE: INTRODUCTION

As this thesis deals with a disorder that manifests with dementia and movement abnormalities, a brief introduction to the physiology of the brain is provided. This introduction will focus on a description of components of the brain that are commonly affected in neurodegenerative and movement disorders. In addition, it also focuses on disorders that manifest with the specific symptoms present in this family.

1.1   The physiology of the brain

The central nervous system (CNS) is the centre for the integration and reception of all nerve impulses generated in the body. It consists of two main components, the brain and the spinal cord and contains two types of cells, namely neurons and neuroglia (glial cells) [Tortora and Grabowski, 1996].

Neurons transmit impulses and are in contact with each other at synapses or junctions. It is through this network of neurons that a stimulus, occurring in any body part, is relayed and processed in the brain.

Unlike neurons, glial cells can multiply and divide. Essentially, the purpose of glial cells is to protect, nurture and repair damaged cells in the CNS. There are four types of glial cells (Figure 1.1). Astrocytes are star shaped and their main functions are relaying impulses and metabolising neurotransmitters. Microglia has a protective function in that they engulf and phagocytosise foreign particles. Ependymal cells provide structural support while oligodendrocytes produce a myelin sheath which is a fibrous layer that protects parts of the neuron. The myelinated portions of the neurons form the white matter of the brain and spinal cord. The grey matter is unmyelinated and contains the cell bodies of neurons [Tortora and Grabowski, 1996].
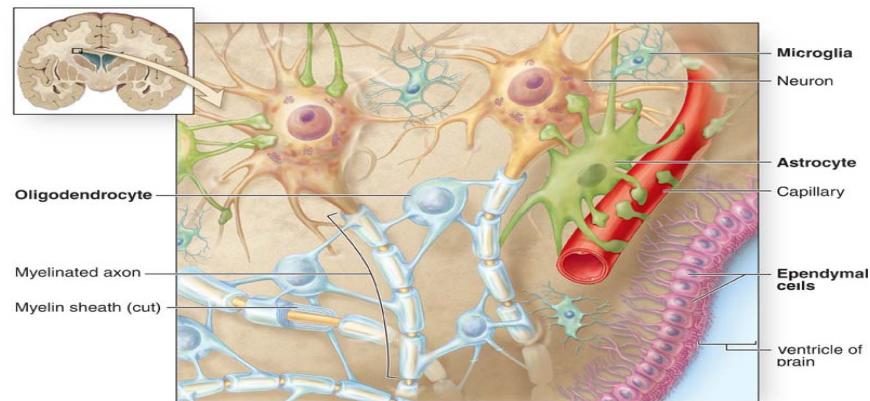
Figure 1.1: A diagram of the four types of glial cells that occur in the brain [Taken from McGraw-Hill online learning centre].

The brain can further be sub-divided into four regional components namely, cerebrum, cerebellum, diencephalon and the brainstem (Figure 1.2).
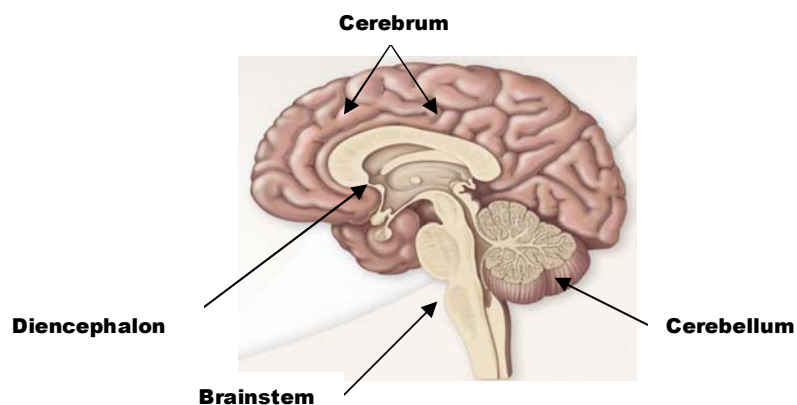
Figure 1.2: A schematic diagram of the four main components of the brain [Adapted from McGraw-Hill online learning centre].

1.1.1 The cerebrum

The cerebrum makes up the bulk of the brain and has highly convoluted grey matter on the surface which is known as the cerebral cortex. The cerebrum is known as the "seat of intelligence" as it regulates and produces almost all processes associated with the brain. The cerebrum is separated into the left and right hemispheres by a central groove called the longitudinal fissure. Each hemisphere further consists of four lobes namely the frontal, occipital, parietal and temporal lobes (Figure 1.3).
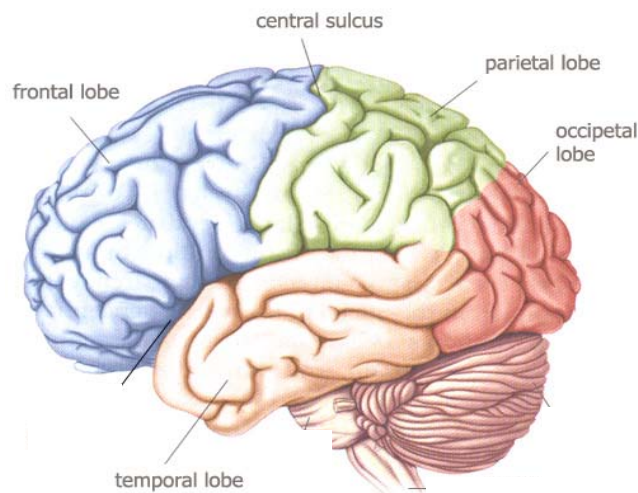
Figure 1.3: A model of the brain depicting the cerebral lobes [Taken from Bear *et al.,* 1996].

Beneath the cortex of the cerebrum lies a dense mass of white matter, the corpus collosum, which connects the left and right halves of the brain. The white matter relays impulses within a hemisphere, between the two hemispheres and from the cerebrum to other parts of the brain or spinal cord [Tortora and Grabowski, 1996]. The cerebrum also contains three functional areas (Figure 1.4). The sensory area receives and interprets sensory information such as touch, pain, temperature, sight and taste. The motor areas control movement of a specific group of muscles. The association areas are concerned with the integration and further processing of sensory information [Silverthorn, 2001].
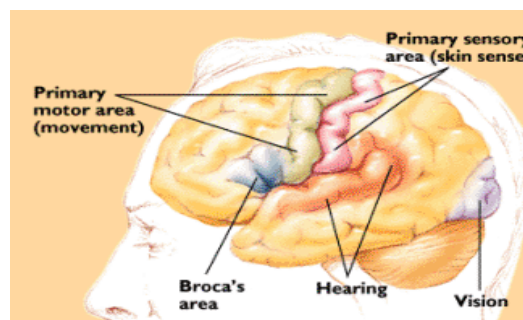
Figure 1.4: A figure of the human brain depicting various functional lobes
[Taken from http://www.gazzaro.it/g/Language%20in%20the%20brain_file/sensory_motor.gif].

The cerebrum contains several clusters of neurons termed the basal ganglia which are connected to each other. The nuclei (clusters of nerve cells) of the basal ganglia consist

of the striatum (caudate and putamen) and the globus pallidus (Figure 1.5). The basal ganglia are involved in the timing and amplitude of movement.
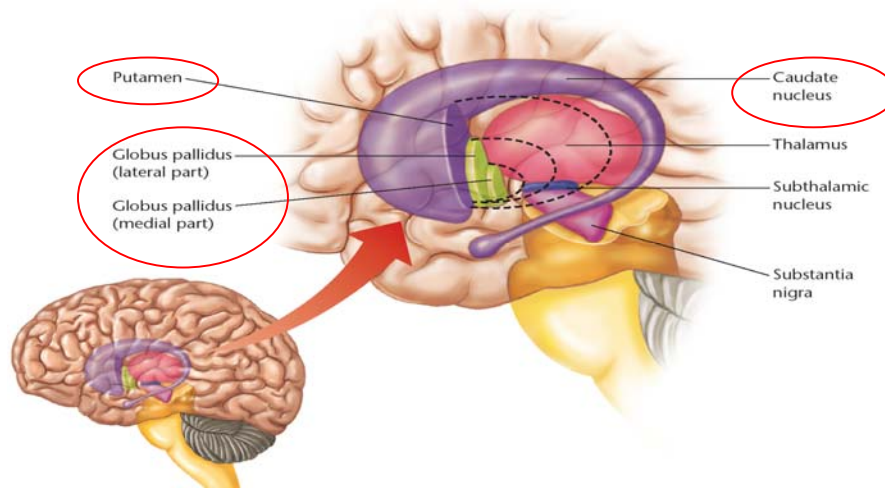
Figure 1.5: A representation of the location and components of the striatum [Taken from Bear *et al.,* 1996].

The limbic system is a group of structures in the cerebrum that includes the hippocampus and amygdala and is involved in emotion and memory. Maintaining the physiological and anatomical structure of the cerebrum is essential for normal functioning. Abnormalities in development or trauma to the cerebrum or structures in the cerebrum cause abnormal or impaired movement, cognitive and sensory dysfunction. Disease of the basal ganglia is associated with movement disorders which are characterised by increased or reduced movements.

1.1.2 The cerebellum

The cerebellum is the second largest part of the brain. Much like the cerebrum, the surface consists of highly convoluted grey matter with deeper white matter arranged like branches of a tree [Van De Graaff, 2001]. The cerebellum has afferent and efferent connections to the pons, medulla, spinal cord and midbrain. The primary function of the cerebellum is to co-ordinate movements by comparing the intended movement with the actual movement being made. It is also an essential contributor to the maintenance of posture and balance. Action tremor in limbs is a common feature in a dysfunctional cerebellum [Watts and Koller, 1997].

1.1.3   The diencephalon

The diencephalon consists predominantly of the thalamus and hypothalamus. The thalamus consists of grey matter organized in clusters and is the principle relaying station for sensory impulses from the efferent regions [Tortora and Grabowski, 1996]. The hypothalamus is located below the thalamus and contains four regions which serve specific functions, thereby contributing to the overall homeostasis of the body. It is essential for homeostasis because it secretes hormones that control the release of other hormones from the pituitary gland [Tortora and Grabowski, 1996]. The hypothalamus plays a role in the integration of the autonomic nervous system such as controlling blood flow and breathing. It also regulates body temperature, thirst, food intake and maintains sleeping patterns.

1.1.4   The brainstem

The brainstem consists of the medulla oblongata, pons and midbrain (mescencephalon). The brainstem relays sensory information to the thalamus and cerebral cortex. It contains nuclei of cranial nerves and therefore receives stimuli for balance, hearing, swallowing, head/shoulder and tongue movements.

All these structures play an important role in the normal functioning of the CNS. Damage or disease in these components may lead to movement and cognitive impairment.

1.2   <u>Dementia</u>

Dementia refers to the progressive decline in cognitive ability that occurs due to loss or impairment of brain cell function. It is not a specific disease but rather a term used to describe a group of symptoms caused by disorders or trauma to the brain. Individuals suffering from dementia have diminished or impaired brain functions beyond what is normally expected in the natural aging process. It has been estimated that when maturity is reached, individuals lose 0.5% of their brain volume per year and this percentage increases every year from the onset. However, there are cases when dementia is severe, rapidly progressive and occurs early in life. In these cases dementia

is normally due to a disease or disorder and is of interest in the present study [Watts and Koller, 1997].

Dementia manifests in affected people as a loss of memory and difficulty with speech and understanding. Other brain functions that are impaired are problem solving, perception of time and place and other cognitive abilities. Dementia occurs predominantly in individuals over 65 years of age. In the year 2000, it was estimated that dementia afflicted some 25 million people worldwide, mostly in the developing countries [Wimo *et al.,* 2006]. The estimated number of people affected by dementia globally is shown in Table 1.1 [Wimo *et al.,* 2006].

Table 1.1: Estimated number of people with dementia worldwide [Wimo *et al.,* 2006].

Continent	Number of cases (million)	% of total
Asia	11.87	46.5
Europe	7.43	29.1
North America	3.08	12.1
Latin America	1.69	6.6
Africa	1.25	4.9
Oceania	0.21	0.8

Dementia is predominantly observed in cases of Alzheimer's disease (AD), vascular dementia, dementia with Lewy Bodies, alcohol/drug related dementia and Frontotemporal Dementia (FTD). Less common causes of dementia are Creutzfeld-Jakob disease (CJD), Parkinson's disease (PD), Huntington's disease (HD) and head trauma [Harvey *et al.,* 2003]. In a majority of cases, there are no effective treatments for the dementia, although drugs blocking acetyl cholinesterase have been shown to improve cognitive function in Alzheimer's disease and Dementia with Lewy Bodies [Flicker, 1999].

Inherited disorders in which dementia is a prominent symptom are of importance in the present study. Furthermore, many of these inherited dementias also manifest with abnormal movements. A summary of a few of these disorders are discussed in the subsequent sections.

1.3   <u>Movement disorders</u>

A movement disorder can refer to two situations, the first being an involuntary or abnormal movement that occurs while an individual is conscious. Secondly, it can also be used to describe a syndrome that has abnormal movements as a prominent symptom. Abnormal movements can be distinguished according to their clinical presentation, i.e. the amplitude of the movement, velocity, posture, rhythm and the ability to suppress the movement. Movement disorders fall into two broad categories namely Hyperkinesias which refer to the excessive movement of body parts and Hypokinesias which is a decrease in voluntary or autonomic movements [Watts and Koller, 1997].

1.3.1   Hyperkinesias

Dystonia refers to an abnormal movement characterised by continuous muscle contraction. It presents as repetitive movements or abnormal posture and can be present in different areas of the body simultaneously but commonly involves a particular body part [Pulst, 2003].

Tremor is defined as involuntary oscillations of a body part. It can occur at different frequencies and is either prominent when the body is at rest or when maintaining a posture. Intention tremor is the most common form of tremor and is characterised by oscillations which increase as the hand or foot reaches a particular target [Pulst, 2003].

Myoclonus is described as a sudden shock-like movement. It is often a sign of cerebral dysfunction and is associated with abnormalities in the cortex of the cerebrum. Cortical myoclonus is associated with epileptic seizures. Action myoclonus refers to myoclonic movements while performing a precise movement [Watts and Koller, 1997].

Chorea manifests as arrhythmic, jerky movements of low amplitude that usually occur in the limbs. It may also present in the face as awkward grimaces and in children, as fidgety movements. It is a prominent symptom in Huntington's disease [Watts and Koller, 1997].

1.3.2 Hypokinesias

Bradykinesia is generally described as a reduction in speed during repetitive movements or general slowness in performing voluntary actions. Other signs include reduced facial expression and blinking [Pulst, 2003].

Rigidity is stiffness in muscles during passive movements. It often occurs in joints, especially in the lower limbs. This is commonly a symptom of Parkinson's disease [Watts and Koller, 1997]. Parkinsonism is a broad term that refers to a range of movement disorders including tremor, rigidity, bradykinesia and loss of postural reflexes [Watts and Koller, 1997].

Movement disorders can occur as the sole symptom in a disease or in conjunction with other symptoms. The latter is usually the case in neurodegenerative diseases where dementia and other neurological signs frequently accompany the movement disorder. There are a number of inherited diseases that manifest with dementia and movements disorders which are relevant to this study and will be discussed in the following sections.

1.4 <u>Mitochondrial DNA and disease</u>

Mitochondria are the powerhouse of the cell, generating energy in the form of adenosine triphosphate (ATP) to drive cellular processes. Mitochondrial DNA (mtDNA) is extrachromosomal DNA that plays an important role in the physiology of the cell and in many different human diseases.

The mitochondrial genome is 16569bp in size and comprised of approximately 93% coding DNA. This genome codes for 37 genes (Figure 1.6) which consists of 13 polypeptides which constitute the mitochondrial respiratory chain (OXPHOS) system as well as the necessary RNA for the translation of these polypeptides (two ribosomal RNAs and 22 transfer RNAs) [Strachan and Read, 1996]. There are hundreds of copies of mitochondria in each cell (polyploidy) and this feature plays an important role in the pathogenicity of mitochondrial diseases.
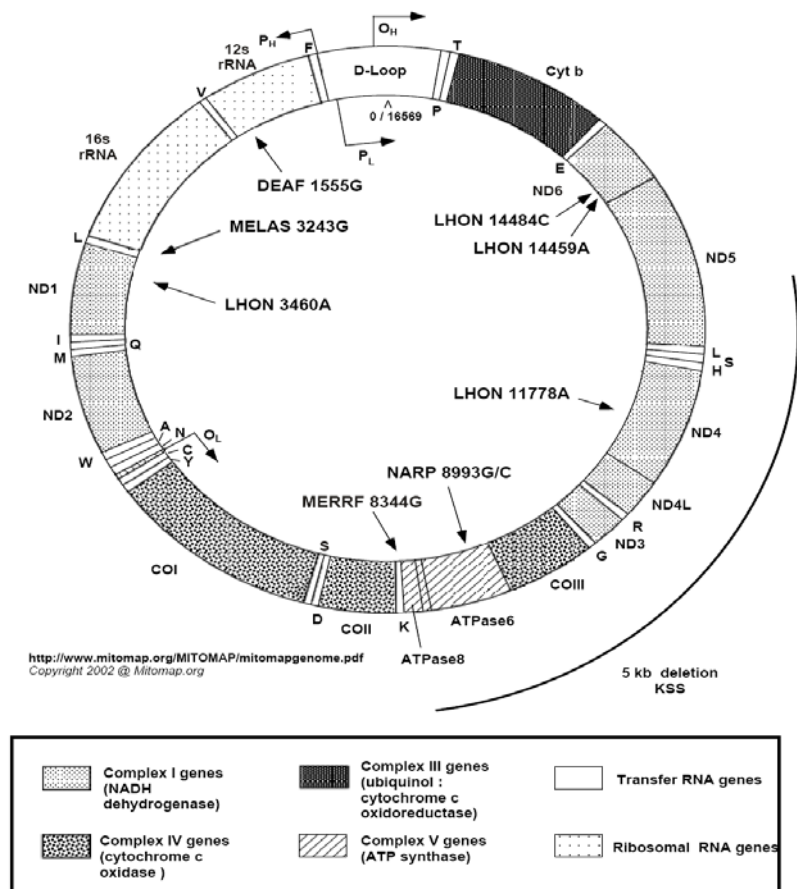
Figure 1.6: A circular representation of the mitochondrial genome indicating the 37 genes [Taken from www.mitomap.org].

Although mtDNA polymerase has a proof-reading mechanism, random polymorphisms occur frequently due to a high rate of replication and the absence of protective histones in the replication process. Furthermore, reactive oxygen molecules that are generated during ATP production causes oxidative damage that result in sequence variants. These variations can be disease-causing mutations or result in predisposition to disease in four ways. Typical mitochondrial syndromes may occur where mutations in mtDNA result in a specific disease. Secondly, a high load of mutated mtDNA in the mother can result in a clinical syndrome in subsequent generations. Thirdly, the natural aging process incorporates mutations into the mtDNA and thereby predisposes aged individuals to disease. Finally, chromosomal mutations that affect mitochondrial ribosomal proteins can result in translational defects in mitochondria [Strachan and Read, 1996].

Mitochondrial mutations can either be homoplasmic (in which all copies of the mitochondria will have the mutation) or heteroplasmic (in which only some copies of the mitochondria will be affected). In most cases of heteroplasmic mutations, the number of mutant copies must exceed a certain threshold before the disease manifests phenotypically [Strachan and Read, 1996]. Due to heteroplasmy, some mutations will only be expressed in certain tissues or systems where the mutational load is highest. Similarly, the disease may only be present in some of the progeny and absent in others. Homoplasmic mutations, however, will be transmitted to all progeny although they may not all display exactly the same phenotypic features of the disease [Strachan and Read, 1996].

Mitochondrial DNA is uniparental and passed through generations via the maternal line. During oocyte development, only some of the mitochondria from the mother are transferred to the egg cells and the progeny will therefore have a limited selection of the mitochondrial load that the mother had. Consequently, a pattern of maternal inheritance of the disorder is essential in recognizing and diagnosing mitochondrial diseases [Taylor and Turnbull, 2005].

1.5 The influence of non-genetic factors in neurodegenerative disorders

The aetiology of many neurodegenerative disorders is multifactorial in that it involves both genetic susceptibility and environmental interactions, which influence the course of the disease. While the genetic mutation may cause the susceptibility to a certain disorder, it is only with the exposure to other risk factors that the disease manifests. Common environmental risk-factors in neurodegenerative diseases include chemical exposure, oxidative stress, vitamin deficiency and exposure to heavy metals. Other risk-factors include age, gender, lifestyle habits such as smoking and drug abuse as well as the presence of metabolic disorders. For example, the increased exposure to heavy metals such as organic aluminium in water was associated with the occurrence of AD [Gauthier *et al.,* 2000]. In addition, Parkinson's disease (PD) has been associated with the exposure to pesticides in rural areas [Abbott *et al.*, 2003]. Another study found an association between PD and rural populations, proposing that the increased incidence of PD may be due to neurotoxins in the water [Priyadarshi *et al.,* 2001].

It is important to recognise these risk factors as avoiding them may reduce or delay the onset of symptoms. Furthermore, these risk factors may alter the clinical phenotype of the disease.

1.6 Inherited disorders displaying dementia and movement disorders

Cases of familial dementia and movement disorders have been found to be due to mutations in both mitochondrial and nuclear genes. The pedigree of the family in the present study showed a tendency for maternal transmission of the disorder. Whether this was a chance occurrence or not was unknown at the time of this study and it was therefore assumed that mutations in either mitochondrial or nuclear genes could be responsible for these symptoms. Given this, both mitochondrial and nuclear neurodegenerative disorders are discussed in the following sections.

1.6.1 Mitochondrial diseases associated with movement abnormalities and dementia

1.6.1.1 Myoclonic Epilepsy with Ragged-Red Fibres (MERRF; MIM 545000)

MERRF is commonly caused by mutations in the transfer RNA (tRNA) genes of the mitochondria resulting in ragged- red fibres in muscle tissue and abnormally shaped mitochondria. The clinical symptoms are myoclonus, seizures, ataxia, and myopathy. The severity and rate of progression of the disease varies amongst different cases but myoclonus is normally the initial symptom [Berkovic *et al.,* 1989]. The age at onset (AAO) ranges from 7-50 years with most cases occurring in childhood after a normal early development [Hirano and DiMauro, 1992].

The prevalence rate is quite low with estimations of 0-1.5/100 000 in Finland [Remes *et al.,* 2005] and 0.25/100 000 in Northern Europe [Chinnery *et al.,* 2000]. However, the disease is clinically heterogeneous, making it difficult to diagnose and these estimates might be biased.

Four point mutations (A8344G, T8356C, G8363A and G8361A) in the *MTTK* gene, encoding tRNA lysine, collectively account for 90% of MERRF cases with the A8344G mutation accounting for approximately 80% of those cases [Shoffner *et al.,* 1990]. The remaining 10% of reported cases is due to other *MTTK* point mutations or single base and larger deletions. Some of the reported mutations may result in additional symptoms not found in typical MERRF. For example in two families, affected individuals harbouring the T8356C mutation showed symptoms of MELAS, another mitochondrial disorder [Zeviani *et al.,* 1993].

Although the mechanism of pathogenesis has not been determined, analysis of cells containing the A8344G mutation showed decreased tRNA lysine production [Enriquez *et al.,* 1995]. It is not clear whether there is a correlation between the amount of mutant mitochondria and the severity of the disease [Berkovic *et al.,* 1989].

1.6.1.2 Mitochondrial myopathy, Encephalopathy, Lactic acidosis and Stroke-like episodes (MELAS; MIM 540000)

MELAS is a well characterised multisystem disorder that manifests phenotypically with subacute-stroke-like episodes, encephalopathy with seizures and dementia. Additional symptoms include short stature, recurrent headaches and vomiting. The initial symptoms may include exercise intolerance and limb weakness that usually presents in early childhood [Hirano *et al.,* 1992]. However, there are rare cases where the AAO is in the fourth decade of life.

Although MELAS is predominantly associated with mutations in the mitochondrial tRNAs, there are several causative mutations in other mitochondrial genes such as the genes encoding the Complex I subunit [Goto *et al.,* 1992] and *MT-ND5* gene [Crimi *et al.,* 2003]. However, it has been estimated that 80% of reported cases of MELAS are due to an A-G transition at position 3243 in the *MTTL1* gene on the mitochondrial genome. A study in Finland found the prevalence of MELAS due to the A3243G mutation to be 16.3 /100 000 [Majamaa *et al.,* 1997].

As with MERRF the symptoms vary in different cases and there are some cases in which the symptoms of MELAS and MERRF overlap, making it difficult to make a distinct diagnosis [Zeviani *et al.,* 1993].

1.6.1.3 Familial multisystem degeneration associated with parkinsonism (LHON-variant) (MIM 516003)

Lebers' hereditary optic neuropathy (LHON) presents in teenagers and adults as acute or subacute loss of vision or complete blindness. The disease is caused by mutations in mtDNA and there are currently 18 allelic variants. However, the most common cause of LHON is the G11778A mutation in the *MT-ND4* gene. In Asia, more than 90% of LHON patients harbour this mutation [Mashima *et al.,* 1993].

The 11778G>A mutation, previously only associated with LHON families, was identified as the cause of a maternally inherited multisystem degeneration disease characterised by parkinsonism in one family [Simon *et al.,* 1999]. The symptoms varied dramatically between affected individuals. In addition to prominent parkinsonism, affected members also displayed akinesia, rigidity, dysarthria, dystonia and dementia.

1.6.2 Non-mitochondrial inherited diseases manifesting with dementia and movement disorders

1.6.2.1 Prion diseases (MIM 1766540)

Prion diseases can occur in both human and animals. Furthermore, the disease can be acquired from animals or may appear sporadically or be inherited. Most cases are sporadic or acquired but about 10-15% of prion diseases are inherited [Windl *et al.,* 1999]. More than 30 point mutations and insertions in the prion protein gene (*PRPN*) have been implicated as the cause of inherited prion diseases [Mead, 2006]. A study of four genes associated with early onset (EO) dementia showed that mutations in the *PRPN* gene were the most frequent cause of EO dementia for those four genes [Finckh *et al.,* 2000].

There are three clinically distinct inherited phenotypes of prion diseases, namely Creutzfeldt-Jakob disease (CJD) [Brown *et al.,* 1994], Gerstmann-Straussler-Scheinker syndrome (GSS) [Hainfellner *et al.,* 1995] and Fatal Familial Insomnia (FFI) [Lugaresi *et al.,* 1986]. Although the symptoms may be markedly heterogeneous, the disease is typically characterised by slowly progressive ataxia, hallucinations, and myoclonus. In the case of CJD, rapidly progressive dementia occurs while in cases of GSS dementia occurs in the later stages of the disease. There are only four frequently occurring mutations in the *PRPN* gene that leads to inherited prion disease, namely E200K, D178N, P102L and OPRI. In a report on 492 cases of prion disease, 350 of the cases were due to these four mutations [Mead, 2006].

The E200K mutation is by far the most commonly occurring mutation for inherited prion disease. Clinically it manifests as typical CJD with muscular rigidity and myoclonus being prominent features [Brown *et al.,* 1986]. The mean AAO is 58 years although this differs in various studies [Brown *et al.,* 1994].

The phenotype of prion disease caused by the D178N mutation has been described as CJD-like. In addition to CJD symptoms, affected individuals also suffer from insomnia and severe myoclonus. In 72 reported cases the AAO ranged from 20-71 years with a median AAO of 50 years [Pocchiari *et al.,* 1998].

The insertion of three octapeptide repeats (OPRI mutation) in the N-terminal of the *PRPN* is also a common cause of inherited prion disease. The AAO is in the fourth decade and the condition presents clinically as myoclonus, ataxia and chorea with cortical dementia as the main feature [Mead, 2006].

The P102L mutation follows a GSS pattern of disease, with slowly progressive ataxia and dementia in later stages of the disease. In 52 reported cases the AAO ranged between 25-70 yrs [Young *et al.,* 1997].

1.6.2.2 Early onset Parkinson's disease (PD; MIM 168600)

Idiopathic Parkinson's disease (PD) is defined as a slowly progressive neurodegenerative disease characterised by loss of dopaminergic neurons and the presence of Lewy bodies (eosinophilic inclusion bodies containing ubiquitin and α synuclein aggregations). The degeneration of dopaminergic neurons causes a decrease of dopaminergic input to the striatum resulting in movement abnormalities [Marras and Tanner, 2003]. The initial symptoms are bradykinesia, tremor and rigidity while symptoms that appear as the disease progresses include dementia and dysarthria. PD is largely a disorder of the aged so the incidence and prevalence of the disease increases dramatically in individuals over 50 years of age. PD that occurs in individuals in which the AAO is younger than 50 years is considered to be early onset PD and is of interest in the present study. PD can occur sporadically or follow a Mendelian or non-Mendelian pattern of inheritance.

PD has been noted to be the second most common neurodegenerative disease globally [Okubadejo *et al.,* 2006]. The prevalence of PD has been estimated in various population groups but to date has not yet been established in South Africa. The prevalence differs in different population groups. A higher prevalence rate was reported in Caucasians compared to the Asian or African populations [Wood *et al.,* 2005]. A gender bias has also been implicated with males having a slightly higher burden than females. However, all these estimates can be biased as there is no diagnostic test for PD during life and clinical diagnoses could be inaccurate [Marras and Tanner, 2003].

Although inherited PD may only contribute to about 5% of PD cases worldwide [Oliveri *et al.,* 2001], great strides have been made in understanding the genetics of PD and has resulted in the identification of a number of mutations in at least five genes that have shown to cause PD while many other genes have been implicated by association studies [Tan and Jankovic, 2006]. To date three genes, namely *DJ1*, *PINK1* and *PARK2*, have been found to be responsible for autosomal recessive early onset (AAO<40years) PD [Gasser, 2005]. Mutations in the *PARK2* (*Parkin*) gene account for 50% of familial and 70% of sporadic cases of early onset PD [Lucking *et al.,* 2000 and Mata *et al.,* 2004]. Similarly, in a study of 73 families which had at least one member with early onset PD, 49% were shown to have mutations in the *PARK2* gene [Lucking *et al.,* 2000].

Clinically, PD caused by mutations in *PARK2* differs from typical idiopathic PD in that dystonia is more common, the progression of the disease may be slower and the AAO can be very young (<20years) but may also be older. Unlike other cases of PD, the presence of Lewy bodies is rare [Takahashi, 1994]. *PARK2* is located at the chromosomal position 6q25.2-q27 and codes for an E3 ubiquitin ligase. Its function is to tag proteins for degradation *via* the ubiquitin pathway [Wood *et al.,* 2005]. There are currently more than 100 mutations in *PARK2* that cause PD. Of these, more than 50 are point mutations while deletions, duplications and exon rearrangements account for the remainder [Tan and Skipper, 2007].

1.6.2.3 Familial Encephalopathy with Neuroserpin Inclusion Bodies
 (FENIB; MIM 604218)

Serine proteases are enzymes that catalyse the hydrolysis of peptide bonds and play a role in intestinal digestion and blood coagulation. During neurogenesis they assist with cell migration and axon development. In adulthood they assist in neuropeptide processing, neural survival, neural structural processing and also play a role in learning and memory processes [Molinari *et al.,* 2003].

Serpins are a family of serine protease inhibitors and neuroserpins form part of this family but are expressed solely in neurons. Neuroserpins are normally expressed in the late stages of neurogenesis where they are postulated to assist with synaptogenesis [Osterwalde *et al.,* 1996].

Mutations in neuroserpins have been identified as the cause of an autosomal dominantly inherited neurodegenerative disease termed Familial Encephalopathy with Neuroserpin Inclusion Bodies (FENIB) [Davis *et al.,* 2002]. The disease is characterised by dementia, myoclonus and the presence of eosinophilic inclusion bodies (Collin's Bodies) in the cerebral cortex [Molinari *et al.,* 2003]. Biochemical analysis of inclusions that were purified from post-mortem brains showed that they were periodic acid Schiff (PAS) positive but ubiquitin and α-synuclein negative and that the major constituent was the neuroserpin protein, *SERPINI1.*

Several point mutations have been identified in the *SERPINI1* gene although only two, S49P and S52R, are common [Davis *et al.,* 2002]. The phenotype of the S49P mutation was first described in a Caucasian family in the United States (US). The affected members presented in the fifth decade of life with initial symptoms of cognitive decline, response regulation difficulties, memory and visio-spatial abnormalities. The later stages were characterised by severe progressive dementia and action myoclonus [Molinari *et al.,* 2003]. The S52R mutation was also reported in a Caucasian family in the US. In addition to the above mentioned symptoms, affected family members also suffered from epilepsy but the AAO in this family was in the third decade of life.

There have been a number of proposed mechanisms of pathogenesis. Thus far, all the point mutations causing FENIB have been found in the "shutter" region of the functional site of the inhibitor. Molecular models of these point mutations show a distinct conformational change in the overall shape of the protein which makes the protein more prone to aggregation. In addition, the degree to which the mutated form tends to aggregate correlated with severity of the disease [Molinari *et al.,* 2003]. As with other neurodegenerative diseases, it is generally accepted that the presence of aggregations plays a detrimental role in neuronal dysfunction. The pathogenesis of FENIB is likely to be due to the precipitation of mutant neuroserpin [Davis *et al.,* 2002].

1.6.3 Repeat expansion disorders

Repeat expansion mutations are known to cause a number of neurodegenerative and neuropsychiatric disorders. Trinucleotide repeat expansions, in particular, are responsible for a vast amount of neurodegenerative diseases with motor disco-ordination. Triplet repeat expansions can be classified as type I or type II depending on where the repeat expansion occurs in the DNA sequence. Type I disorders refers to those conditions in which the repeat lies in a coding region of a gene and therefore codes for functional amino acids. Type II disorders refer to conditions in which the repeat occurs in a non-coding region [Margolis *et al.,* 1997].

Polyglutamine diseases refers to a group of diseases caused by CAG or CAA repeat expansions associated with the production of long polyglutamine (polyQ) tracts. Although there are many proposed mechanisms of pathogenicity for diseases caused by

these mutations, the generally accepted mechanism of pathogenesis is that mutated transcripts containing polyQ tracts aggregate and form inclusions in the cell leading to cellular degradation. These intranuclear inclusions, which are usually immuno-positive for the mutated protein, have become hallmarks for polyglutamine disease [Neri, 2001]. Other theories propose that the mutant protein containing polyQ tracts has a conformational change that leads to abnormal cellular distribution of the protein which is toxic to the cell [Neri, 2001]. Another theory suggests that the mutant polyQ tracts interact with short polyQ tracts that normally occur in transcription factors, thereby affecting transcriptional regulation.

Features of repeat expansion disorders include anticipation and repeat instability [Stevanin *et al.,* 2000]. Anticipation refers to the ability of larger repeats to expand even further in successive generations. Repeat instability is the phenomenon whereby larger numbers of repeats have an unstable transmission in subsequent generations thereby causing repeats to be expanded into the pathogenic range in offspring of parents with large repeats [McInnis*,* 1996].

In the following sections a number of different neurodegenerative disorders caused by repeat expansions will be discussed.

1.6.3.1 Spinocerebellar Ataxias

Spinocerebellar Ataxias (SCAs) are inherited neurodegenerative diseases characterised by limb or gait ataxia and dysarthria. The majority of SCAs are autosomal dominantly inherited. Over the past 14 years more than 28 genetically distinct types have been identified and there are still many cases of SCA for which a genetic cause or the affected gene has not been identified [Pulst, 2003].

In general, the prevalence rate of SCAs has been estimated to be three cases in every 100 000 people but this is a tentative estimation based on studies in isolated regions [Michalik *et al.,* 2004]. A true reflection of the prevalence rate is affected by rare subtypes not being included in many of the population studies. Furthermore, the subtypes in which a chromosomal region was implicated but the exact genetic mutation is unknown, makes it difficult to identify these subtypes [Watts and Koller, 1997]. A direct

comparison between regional occurrences of subtypes is hindered due to founder effects which creates a genetic bias. The presence of founding individuals in a population has a profound effect on the geographic occurrence of a particular subtype, for example SCA type 2 (SCA 2) in Cuba [Estrada *et al.,* 1999]. From a global perspective, SCAs 1, 2, 3, 6 and 7 seem to be the most common types, representing 50-80% of the known cases. In South Africa the most common type is SCA 1 [Ramesar *et al.,* 1997], although cases of SCA types 2, 3, 6, and 7 have been reported [Bryer *et al.,* 2003].

In addition to ataxia and dysarthria, affected persons may also display symptoms such as dementia, oculomotor disturbances, epilepsy, myoclonus and cognitive impairment [Pulst, 2003]. The neuropathological effects of SCAs vary amongst subtypes with a few morphological features that are present in most, if not all, types. These features typically include atrophy and loss of neurons in brain regions involved in movement and co-ordination such as the cerebral cortex, purkinje cells, cerebellum and the striatum. Immunohistochemistry of neurons may show intranuclear aggregations containing polyglutamine tracts in many of the subtypes, while in others cytoplasmic inclusions are present [Schöls *et al.,* 2004]. The clinical overlap between different subtypes makes it difficult to classify the subtypes clinically so generally they are classified using molecular methods.

SCAs are caused by a variety of mutations, predominantly repeat expansions but point mutations have also been implicated in some subtypes [van Swieten *et al.,* 2003]. Six of the characterised types of SCAs (types SCA1, SCA2, SCA3, SCA6, SCA7 and SCA17) are caused by CAG repeat expansions in coding regions of the respective genes and the number of repeats that lead to pathogenesis varies between subtypes [Schöls *et al.,* 2004]. In these cases the CAG repeats are generally thought to form polyglutamine tracts which bind with other proteins to form aggregates that are toxic to the cell thereby causing cellular dysfunction or death. Another sub-type (SCA12) is due to a CAG repeat expansion in the 5' non-coding region of the *PPP2RB* gene. A penta-nucleotide repeat expansion of (ATTCT) has been implicated in SCA 10 whereas point mutations in *FGF14* and *CγPKCγ* have been shown to be causal in two cases of familial SCA.

1.6.3.2 Dentatorubral-Pallidoluysian Atrophy (DRPLA; MIM 125370)

Dentatorubral-Pallidoluysian Atrophy (DRPLA) is a rare neurodegenerative disease characterised by cerebellar ataxia, choreo-athetosis, myoclonic epilepsy and dementia. DRPLA is distinguished from SCAs by the significant neuronal loss in the dentatorubral and pallidoluysian systems [Naito and Oyanagi, 1982]. In addition to the general symptoms, individuals with early onset DRPLA (<20years) also present with signs of progressive myoclonus epilepsy (PME). Late onset cases are clinically different, with most patients displaying choreo-athetosis and psychiatric disturbances [Ikeuchi *et al.,* 1995].

DRPLA is caused by an unstable CAG repeat in the *ATN1* gene located on chromosome 12p13. Unaffected individuals have 3 to 36 repeats whereas in affected individuals the repeats range from 49-88 [Schöls *et al.,* 2004]. The number of repeats is directly associated with the severity and AAO of the disease. Furthermore, anticipation has been observed in several cases [Komure *et al.,* 1995].

DRPLA occurs predominantly in Japan with a prevalence rate of two to four per million although cases in other ethnic groups have been noted [Lee *et al.,* 2001]. The genetic mutation in an African American family affected with a neurodegenerative disorder, termed Haw River Syndrome, was later found to be the CAG expansion in *Atrophin1* (*ATN1*) gene and is thus considered the same disorder. However, the Haw River Syndrome cases showed a slightly different phenotype in that they had calcification of globus palladius and no myoclonic seizures [Burke *et al.,* 1994].

The pathogenic mechanism of the disease is unknown and there are many theories surrounding it. Mutant *ATN1* is expressed at similar levels to the wild type protein indicating that transcription efficiency is not altered. This concurs with the theory that polyglutamine tracts produced by CAG repeat expansions in the *ATN1* gene are pathogenic to the cell and result in cell death [Onodera *et al.,* 1995].

1.6.3.3 Huntington's disease (HD; MIM 143100)

Huntington's disease (HD) is an autosomal dominantly inherited disease caused by a CAG repeat expansion in the *IT15* gene on chromosome 4p16.3. HD manifests as a progressive neurodegenerative disorder involving chorea, impaired cognitive ability and psychiatric disturbances. Affected individuals have an average of 20 years survival following the onset of clinical symptoms. HD accounts for approximately 90% of cases of chorea with a genetic aetiology [Schneider *et al.,* 2007].

HD is a late onset disease where the AAO is usually in the fourth decade of life, more specifically between 35-44 years of age. The early stages of disease are characterised by neurological disturbances, impaired voluntary movements and motor disco-ordination which progresses to a final stage of profound dementia and prominent chorea ultimately leading to death [Schneider *et al.,* 2007]. Neuropathological examination of the brains of deceased HD patients showed severe degeneration of selective neurons in the striatum [Cowan and Raymond, 2006]. Intracellular inclusions containing mutant *IT15* protein were also present.

The prevalence rates varies in different population studies with estimates of four to eight per 100 000 cases in North America and Europe. However, the prevalence rates of HD in Japan, Norway and Finland are much lower [Schneider *et al.,* 2007].

The CAG repeat occurs in Exon 1 of the *IT15* gene. The repeat codes for polyglutamine and HD therefore, falls into the category of Type I polyglutamine disorders. Unaffected individuals may have up to 26 repeats while the effect of 27-35 repeats are considered intermediate in that individuals with this range of repeats are not expected to exhibit the Huntington's phenotype but their mutant allele is prone to expand in to the pathogenic range in subsequent generations [Semaka *et al.,* 2006]. Reduced penetrance was demonstrated for cases in which the repeats were between 36 and 39 while repeats of greater than 40 are considered pathogenic. As with other repeat expansion disorders, the phenomenon of anticipation has been observed in many cases of HD.

1.6.3.4 Huntington's disease-like 2 (HDL2; MIM 606438)

Huntington's disease-like 2 (HDL2) is a rare neurodegenerative disease characterised by movement abnormalities, selective neuronal degeneration and severe, progressive dementia. A hallmark of this disease is the presence of intranuclear inclusion bodies and selective atrophy in the cerebrum, while the cerebellum remains relatively unaffected. Clinically, the disease appears similar to HD but unlike HD it is caused by a repeat expansion in an alternatively spliced exon of the *Junctophilin-3* (*JPH3*) gene [Walker *et al.,* 2003].

The discovery of HDL2

About 1% of HD-like cases are not due to the HD mutation. It has been proposed that the study of diseases that have a HD phenotype but are negative for the HD expansion may provide clues to the pathogenic mechanism of HD.

Given this, in 2001, Margolis *et al.,* described a family (pedigree W) with an HD-like phenotype who were negative for HD, SCA (types 1-3, 6-8 and 12), DRPLA, Huntington's disease-like 1 and mutations in the *TBP* gene [Margolis *et al.,* 2001]. The family was of African American descent and resided in South Eastern United States. The proband, a female, presented at 26 years of age with weight loss and disco-ordination which steadily progressed. After 10 years of disease the proband was severely demented and completely bedridden. The presentation of this disease was similar for all the cases in this family [Margolis *et al.,* 2001].

The MRI scans of the proband showed severe atrophy of the striatum and the cerebral cortex while the cerebellum was relatively unaffected. Microscopical analysis of brain tissue showed selective neuronal degeneration particularly of medium spiny neurons in the striatum. Furthermore, intranuclear inclusions were detected that stained with IC2 antibodies which are proposed to be selective for long polyglutamine tracts but may also stain any form of aggregated protein. A positive reaction was also observed for staining with anti-ubiquitin antibodies but not with α-synuclein and anti-Huntington antibodies. The inclusions were predominantly present in the cortex.

It was suspected that the disease may be due to a repeat expansion because most repeat expansion disorders result in neuropsychiatric disorders [Margolis *et al.,* 1997]. To ascertain the identity of the causative mutation the genomic DNA of several family members was analysed using the repeat expansion detection (RED) assay which detects any CTG/CAG repeats longer than 40 in an individual's genome [Holmes *et al.,* 2001]. The flowchart (Figure1.7) displays the process that led to the identification of the HDL2 mutation.

The RED assay detected a CTG/CAG repeat expansion of 50-60 repeats in all affected family members. Furthermore, this expanded tract was absent in all unaffected family members. In order to characterise this repeat the genomic DNA of one affected family member was subjected to digestion with *EcoRI* and the resulting DNA fragments were separated on a gel and the gel was then cut into 50 slices. The DNA fragments was extracted from the slices and a RED assay was performed on the extracted DNA in order to isolate fragments in which long CAG/CTG repeats were present.

The fragments that were positive for the RED assay were cloned and the resulting library was probed with oligos that were specific for long CAG tracts. Clones that contained long repeats were subjected to direct sequencing. A 6kb clone that contained this 55-60 repeats was identified and the regions flanking the repeat were used to design primers that would amplify the repeats in order to determine the repeat sizes in other family members.

Pedigree W was described with symptoms similar to HD but HD negative.

It was suspected that the family had a repeat expansion disorder like SCAs or HD so genomic DNA (gDNA) was collected from family members and a RED Assay, which detects CAG/CTG repeat expansions, was performed.

An unexplained (not due to SCA, DRPLA *etc*.) CAG/CTG tract of 50-60 repeats was found in all affected and no unaffected family members.

gDNA of an affected family member was digested with *Eco*R1, the fragments were electrophoresed, the gel was sliced into 50 pieces and DNA was extracted. RED ASSAY was performed on DNA from each slice. RED positive slices were cloned and the library was screened with (CAG)^{15} probes. Plasmids with long repeats were extracted and sequenced.

A 6kb region containing 55 CAG/CTG repeats was cloned.

The sequences flanking the repeats were used to design primers to determine repeat length.

METHOD OF REPEAT EXPANSION DETECTION (RED)

gDNA was incubated with (CAG)^{10} primers, underwent electrophoresis and were probed with (CTG)^{8} primers.

When two (CTG)^{8} primers anneal adjacent to each other along the template DNA, they ligate to each other.

After 500 PCR cycles the longest oligo corresponded to the longest CAG/CTG in the genome of the individual.

PCR products underwent electrophoresis and were probed with (CTG)^{8} probes.

Bands produced corresponded to long CAG/CTG tracts. The presence of 30-40 repeats was expected as controls.

The sequence was BLASTed against GenBank database.

The sequence was mapped to *JPH3* gene on chromosome 16q23-24.

Repeats ranged from 51-57 in affected Pedigree W members. Complete segregation of repeat expansion and disease in the family was shown.

The repeat length was determined in 98 unaffected, unrelated individuals as well as in 175 HD individuals and 330 individuals with movement disorders.

Figure 1.7: Schematic diagram of the process leading to the identification of the HDL2 mutation.

The repeat length of 13 individuals from the family (pedigree W) was determined. The range of pathogenic repeats in affected family members was 51-57 triplet repeats. The diversity of repeat length in three sibships indicated that the repeat was most likely unstable during vertical transmission [Holmes *et al.,* 2001]. To further support the pathogenic aetiology of the repeat, the repeat length was determined in 175 individuals with HD, 98 unaffected and unrelated individuals as well as in 330 individuals with unknown movement disorders. In this study [Holmes *et al.,* 2001] the unexpanded alleles for all the individuals ranged from 6 to 27 triplet repeats with a mode of 13 repeats (Figure 1.8).
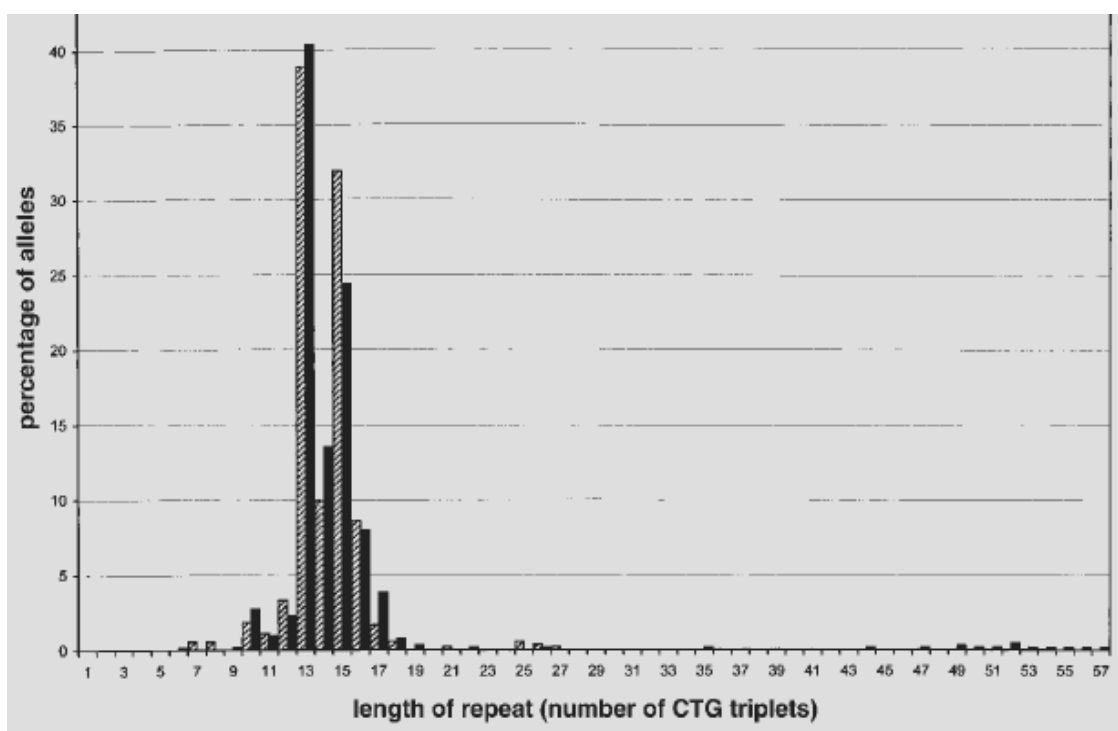
Figure 1.8: Graph representing the repeat length in *JPH3* gene for 603 individuals (Black bars represents individuals with movement disorders of unknown aetiology, grey bars represents the control individuals) [Taken from Holmes *et al.,* 2001].

Expanded alleles, ranging from 44 to 50 repeats, were detected in four individuals of African American descent (unrelated to pedigree W), with a family history of an HD-like disorder but were negative for the HD mutation. In addition, an allele of 49 repeats was found in a Moroccan woman and a repeat expansion of 35 triplets was identified in an individual with a sudden onset of cerebellar signs [Holmes *et al.,* 2001].

Complementary DNA (cDNA) containing this repeat had previously been mapped to chromosome 16 [Margolis *et al.,* 1997]. To further characterise the repeat a 125kb contig containing the repeat was created from DNA of an affected pedigree W member. The contig was compared to the dbEST database and analysed using GENSCAN. The contig, with the repeat in the CAG orientation, did not match any ESTs in the database nor did GENSCAN predict any transcripts for this orientation of the repeat. However, in the CTG orientation the repeat was localised to the *JPH3* gene, between Exons 1 and 2 [Nishi *et al.,* 2000]. The 125 kb contig included Exon 1 which allowed for the localisation of the repeat to 760 nucleotides 3' to the end of Exon 1.

Four factors lend support to the fact that the repeat lies in a variably spliced exon in the *JPH3* gene. Firstly, a polyadenylation signal was identified 281 nucleotides 3' to the repeat. Secondly, GENSCAN predictions displayed a transcript where Exon 1 is spliced to a terminal exon containing the repeat. Thirdly, two Expressed Sequence Tags (EST's) from retinoblastoma and lung carcinoid tissue (BE042890 and BE779067) were identified which show Exon 1 spliced to the terminal exon containing the repeat. Lastly, the authors used RT-PCR on normal brain mRNA and found that this alternatively spliced transcript could exist in different forms due to the orientation of the splice acceptor sites. In these three forms the repeat can either code for polyalanine, polyleucine or lie in the 3' untranslated region (Figure 1.9).
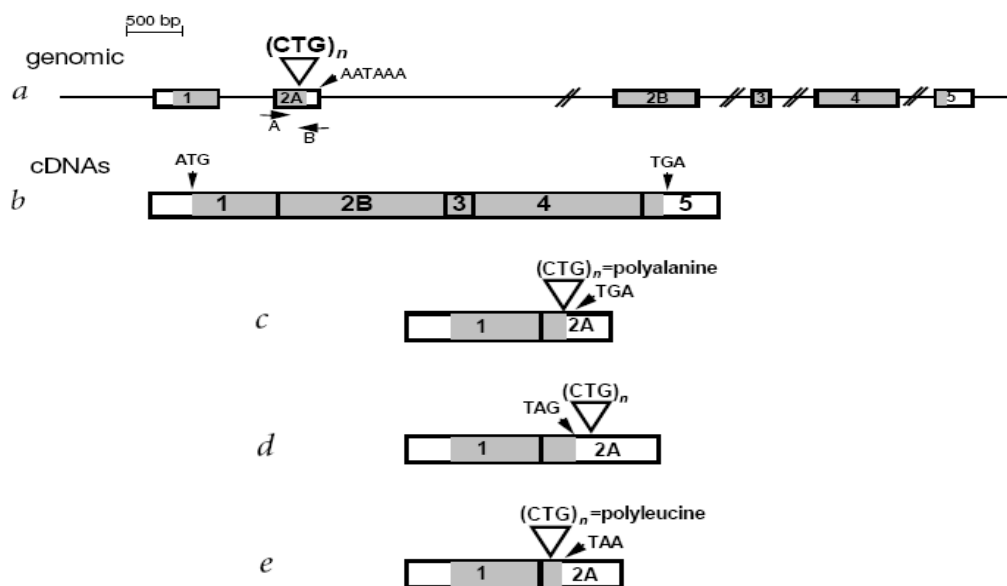
Figure 1.9: (**a**) Graphical structure of full length *JPH3*. The exons are indicated by grey blocks, location of the polyadenylation signal (AATAA) and the repeat are indicated. **(b)** Published *JPH3* mRNA transcript. **(c-e)** Depicts the alternatively spliced versions of the *JPH3* transcripts with exon 1 spliced to exon 2A and the different splice acceptor sites which causes the repeat to code for polyalanine, polyleucine or fall into the 3' UTR [Taken from Holmes *et al.,* 2001].

Population studies of HDL2

Since the discovery of HDL2 a steady increase in cases has been identified worldwide. Thus far, most of the reported cases are in individuals of Black African ancestry [Margolis *et al.,* 2006]. An exception to these cases was a HDL2 positive individual from Morocco [Holmes *et al.,* 2001] and a family from Mexico [Stevanin *et al., 2003*]. However, these individuals were later found to originate from areas in their respective countries that was colonised by Black Africans [Margolis *et al.,* 2006]. Furthermore, a preliminary report had indicated 7 of 20 Black patients from South Africa had tested positive for HDL2 [Krause *et al.,* 2002].

No patients with an HDL2 repeat expansion were found in a study done on 1600 Caucasian patients with HD-like symptoms from Austria and Germany [Bauer *et al.,* 2002]. Of this cohort 147 patients had a family history of HD-like symptoms. The repeats in this cohort ranged from 10 -27 triplets with modes of 14 and 16 repeats (Figure 1.10)
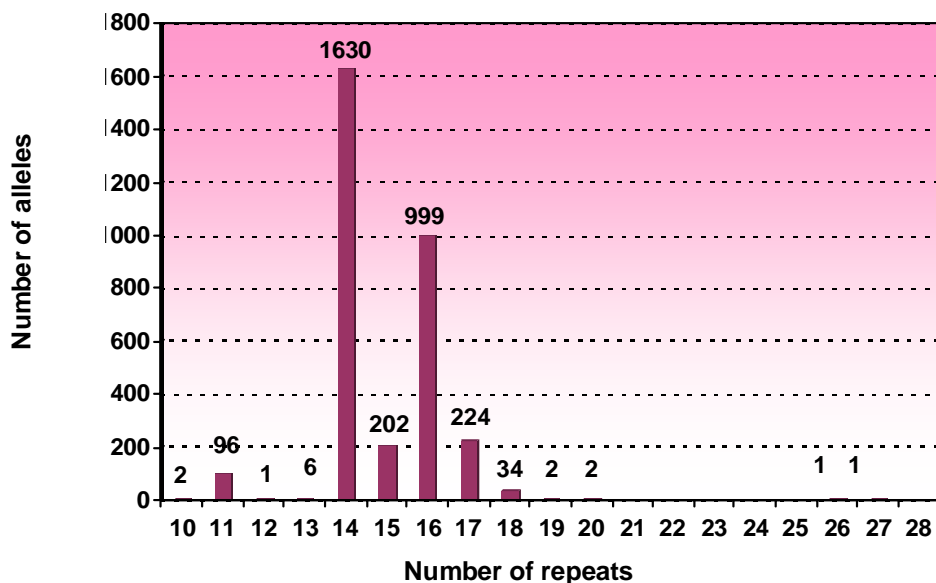
Figure 1.10: Determination of *JPH3* repeat length in 1600 Caucasian individuals from Germany and Austria [Adapted from Bauer *et al.,* 2002].

A study done on individuals from Japan and North America with HD-like symptoms was also carried out. Out of 538 of the North American cases, six were positive for HDL2 but none of the 44 Japanese cases were positive for HDL2 [Margolis *et al.,* 2004].

Characterisation of the HDL2 repeat

In order to determine the risk for developing HDL2, the repeat sizes were determined in a healthy population of Serbia and Montenegro. In this study, a cohort of 198 unrelated "healthy" controls was typed [Keckarevic *et al.,* 2005]. No pathogenic repeat expansions were found and the alleles ranged from 11-18 repeats with a bimodal pattern of distribution with peaks at 14 and 16 repeats, results which were similar to those obtained by Bauer *et al.* [2002].

The repeat length and age of onset is correlated in HDL2 as in other repeat expansion disorders [Margolis *et al.,* 2004]. For example, in 24 HDL2 cases an increase in repeat length resulted in a younger age of onset (Figure 1.11).
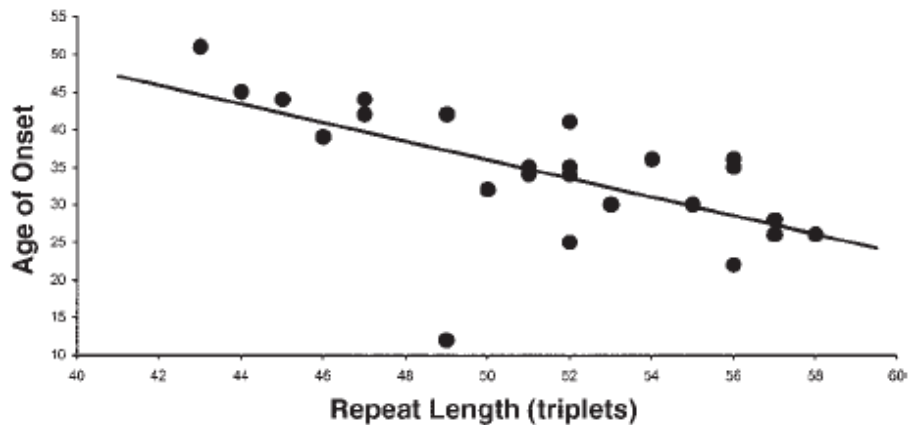
Figure 1.11: The *JPH3* repeat length is correlated with the age of onset [Taken from Margolis *et al.,* 2004].

A repeat expansion of more than 43 repeats has been determined to be pathogenic, whereas repeats of between 40 and 43 are considered to be incompletely penetrant [Margolis *et al.,* 2004]. Moreover, repeats of 33-35 have been reported to be unstable in vertical transmission. The phenotypic effects of repeats in this range have however not been established as an individual with 35 repeats developed Cogan's Syndrome, which is an auto immune disease [Margolis *et al.,* 2004].

Neuropathology of HDL2

From a neuropathological perspective the disease mimics HD in that the caudate and putamen are severely atrophied in the late stages of the disease (Figure 1.12 A and D) [Margolis *et al.,* 2006]. As in HD (Figure 1.12 B and E) the cerebrum was severely atrophied while the cerebellum remained relatively unaffected. The examination of four HDL2 brains showed severe atrophy of the striatum, particularly the caudate and the putamen while mild atrophy of the globus pallidus was observed [Margolis *et al.,* 2006].
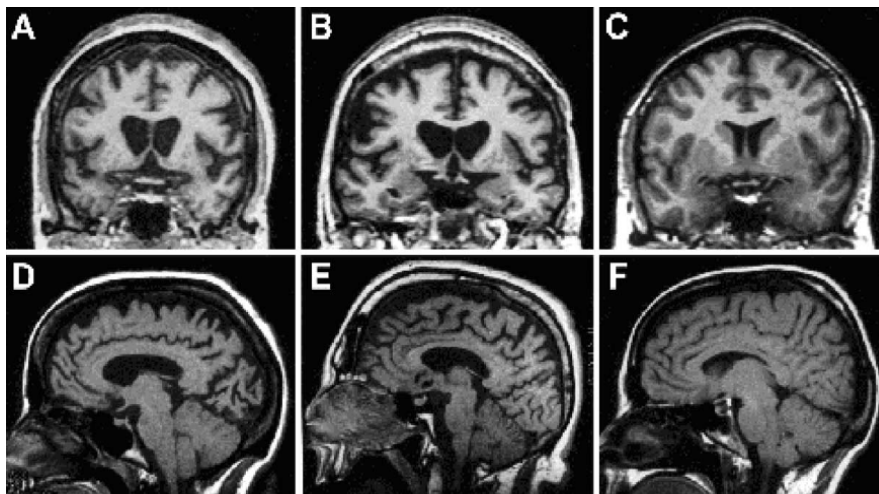
Figure 1.12: (A) and **(D)** MRI scans of an HDL2 case after 10 years disease duration. **(B)** and **(E)** MRI scans of HD brain after 12 years disease duration. **(C)** and **(F)** normal control at 43 years old [Taken from Margolis *et al.,* 2003].

As in HD, a hallmark of this disease is the presence of intranuclear inclusions. However these inclusions did not stain with anti-Huntington antibodies but were positive for staining with IC2 antibodies [Rudnicki *et al.,* 2007].

Summary of the clinical presentation of HDL2

Early clinical signs of the disease include weight loss, personality changes, imbalance and poor motor co-ordination. In later stages of the disease the patients exhibit choreiform movements, rapidly progressive dementia, rigidity and dystonia which culminate in a completely bedridden state within 10 -15 years of onset [Margolis *et al.,* 2004]. The age of onset has been estimated to be in the third to fourth decade of life with the earliest age of onset in a boy of the Mexican pedigree who reported to display choreiform movements and psychiatric disturbances at the age of 12 [Margolis *et al.,* 2004]. An HDL2 family with acanthocytosis has also been reported [Walker *et al.,* 2002]. Although affected family members followed the general disease course, 30- 35% acanthocytosis was detected on peripheral blood smears of the three affected family members.

The Junctophilin-3 gene

The *Junctophilin-3* gene is a member of a conserved family of proteins called Junctophilins. The Junctophilins play a role in the structure of junctional membrane complexes by regulating the Ca^{2+} influx [Takeshima *et al.,* 2000]. Junctional complexes are structures that form between the sarcaoplasmic /endoplasmic reticulum and the plasma membrane in all excitable cells. The complexes are thought to provide functional cross talk between the plasma membrane and Ca^{2+} dependent ion channels.

Four human Junctophilin subtypes have been identified. Blot hybridisation techniques were used to localise the mouse orthologues of the subtypes to their respective expression tissues. *Junctophilin-1* (*JP1*) was localised to skeletal tissue and knock-out mouse models exhibited neonatal death [Nishi *et al.,* 2003]. *Junctophilin-2* (*JP2*) was expressed in both heart and skeletal muscle in mice [Nishi *et al.,* 2003] and also lung and stomach tissue that indicates expression in smooth muscle [Takeshima *et al.,* 2000]. Knock out mouse models showed disruption in normal Ca^{2+} transients in cardiac myocytes which ultimately resulted in embryonic lethality [Nishi *et al.,* 2000].

Junctophilin-3 (*JPH3*) is predominantly expressed in the brain and possibly the testis (Figure 1.13). A northern blot analysis of *JP3* (the mouse ortholgue of *JPH3*) detected a 4.4Kb fragment in mouse brain tissue and a smaller fragment in the testis [Nishi *et al.,* 2003]. The *JP3* probes did not hybridize to any other tissues in this study.
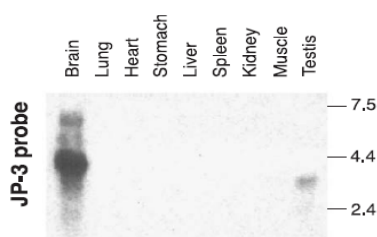
Figure 1.13: Northern Blot of *JP3* in mouse tissues [Taken from Nishi *et al.,* 2003].

Knockout mouse models of *JP3* showed impaired motor co-ordination but no other morphological abnormalities. In addition, normal hearing, emotional and learning abilities were observed [Nishi *et al.,* 2003]. It was noted that the phenotype seen in these mice was, surprisingly, not as severe as the presentation of HDL2 in humans. *Junctophilin-4*

(*JP4*) may have similar functions as *JP3* because *JP4* was also shown to be expressed solely in the brain (Figure 1.14 G) and displayed co-localized expression with *JP3* in mouse brains (Figure 1.14 A–F) [Nishi *et al.,* 2003]. It has been proposed that *JP4* may assist in carrying out *JP3* functions in the knockout mouse models and that may be the reason for the phenotype in the knockout mice being less severe than that observed in HDL2 cases [Margolis *et al.,* 2006].
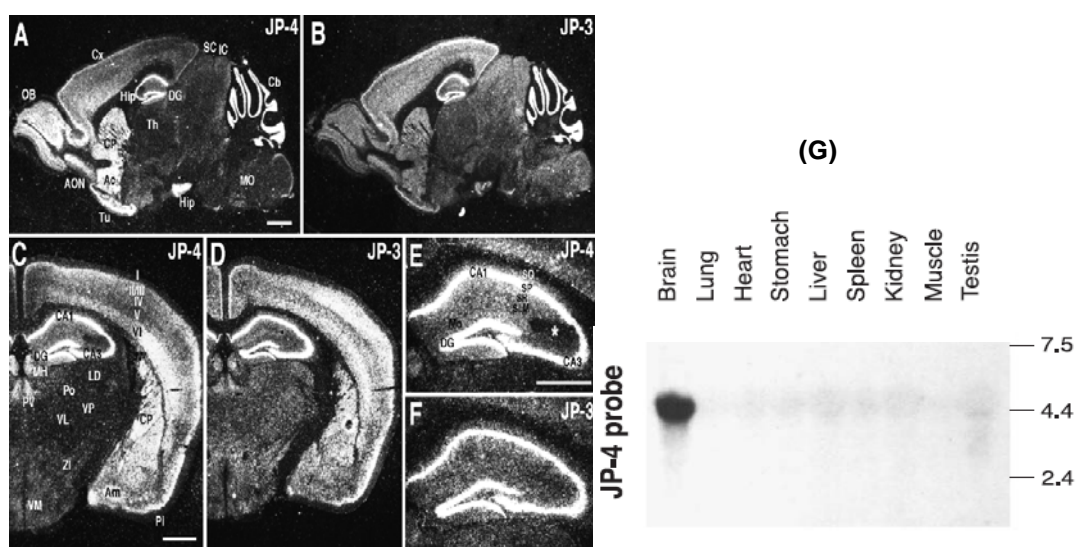
Figure 1.14: (A-F) Dark-field photographs showing similar expression patterns of *JP3* and *JP4* in adult mouse brains. **(G)** Northern blot of *JP4* in mouse tissues [Taken from Nishi *et al.,* 2003].

The mechanism of pathogenesis for HDL2 is unknown although a number of theories have been postulated. One theory explores the possibility that RNA foci are formed in the brains of HDL2 patients as in myotonic dystrophy type 1 (DM1) [Brook *et al.,* 1992]. As in DM1, RNA containing repeat expansions form aggregations which are toxic to the cell. In HDL2, it was proposed that an RNA gain of function mechanism contributes to the pathogenesis of the disease because no mutant protein was found in the intranuclear inclusion bodies. In addition, the pathogenic repeat may fall into an untranslated region possibly explaining why the aggregation of mutant protein is not solely responsible for the disease [Rudnicki *et al.,* 2007].

To investigate the possibility of RNA–gain of function mutation the authors first attempted to identify the *JPH3* transcript in post-mortem brain using the Fluorescent *in situ* hybridisation (FISH) technique with probes specific to Exon 2A [Rudnicki *et al.,*

2007]. Once the foci were identified, the composition of the foci was determined and was shown to contain RNA transcripts in which Exon 1 was present. Both these results showed negative reactions in post mortem HD brains. Further analysis showed that the foci may contain other splice variants. Secondly, all inclusions were in neurons with the inclusion in the striatum having a brighter appearance in dark-field microscopy than those in the frontal cortex indicating that the striatum may be more severely affected than the frontal cortex [Rudnicki *et al.,* 2007].

Another proposed mechanism suggests that the disease may be due to a loss of function. This theory is based on the fact that the phenotypes of knockout mouse models are not as severe as the disease in humans. Furthermore, it was hypothesised that an increase of alternative spliced transcript expression in affected individuals could decrease the expression of full length *JPH3* thereby causing a more severe phenotype.

1.7   Aims and objectives of the present study

The primary aim of the present study was to identify the pathogenic mutation underlying dementia and movement disorders in a South African family of Mixed Ancestry. The investigation involved a clinical and molecular analysis of affected and unaffected, at-risk family members.

Ultimately the identification of the disease-causing mutation in this family would provide the option of informed genetic counselling and pre-symptomatic genetic testing for at-risk family members. In the long term, the identification of this mutation may contribute to the understanding of the patho-physiology involved in this neurological disorder.

<div align="center">**CHAPTER TWO: MATERIALS AND METHODS**</div>

2.1   Study participants

In the present study, a self-reported Mixed Ancestry family presenting with dementia and movement disorders of unknown aetiology was investigated. The proband (5657) presented at Tygerberg Hospital with severe dementia and a movement disorder. A subsequent review of his family history revealed that several family members spanning three generations had also displayed similar symptoms and had died of the disorder. Figure 2.1 depicts the pedigree of the family investigated (hereafter referred to as Family R). The pedigree was constructed based on data obtained from other unaffected family members, a close family friend as well as hospital records.
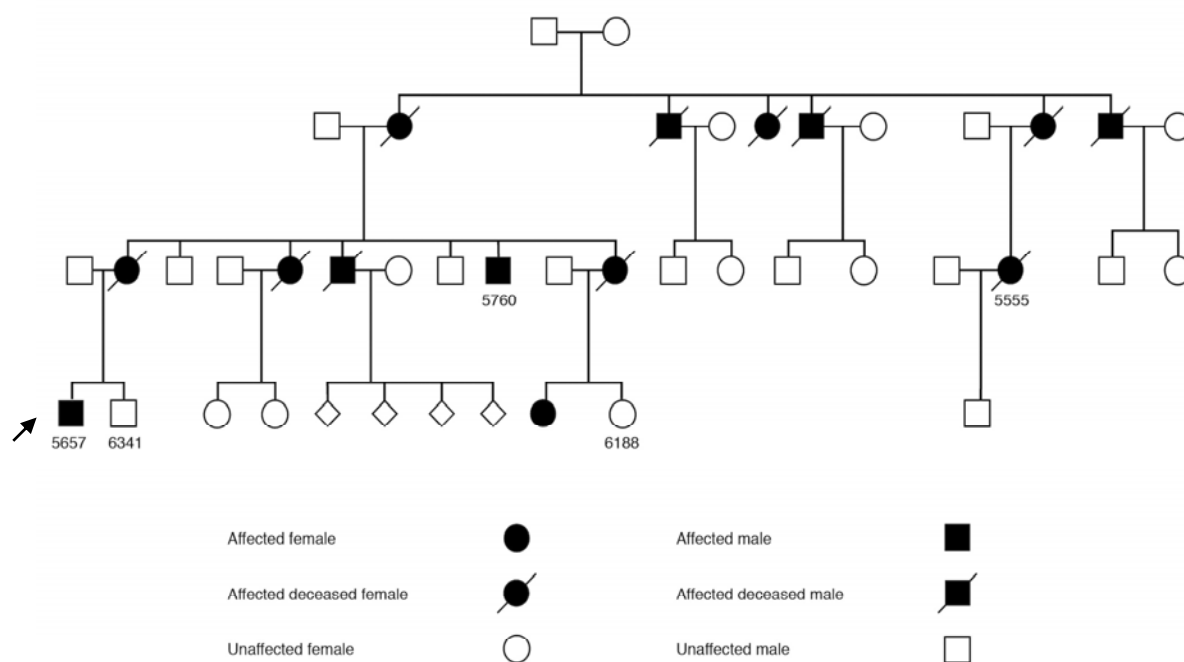
Figure 2.1: Pedigree of the affected family, Family R. Genomic DNA was available for individuals 5760, 5555, 5657, 6341,and 6188. The solid arrow indicates the proband, 5657 and the diamonds symbolise unknown gender.

Later, another family member (5760) was also identified with dementia and a movement disorder. A study was undertaken to identify the disease-causing mutation segregating in this family. In addition, two family members who had affected parents but were clinically unaffected at the time of this study (individuals 6341 and 6188, aged 35 and 36 years respectively) were also recruited. As indicated, most of the affected individuals are deceased hence no DNA was available for them. Due to the late age of onset of the symptoms, many of the fourth generation family members could not partake in this study, as they were too young and asymptomatic. For these reasons, DNA of only five family members was available (5555, 5657, 5760, 6188, and 6341).

In 2006, a complete neurological examination was performed on the proband (5657) and his uncle (5760) (Table 2.1). This examination included a Mini-Mental State Examination (MMSE) which detects and measures the extent of dementia [Folstein *et al.,* 1975]. Once corrected for age and the level of schooling, a MMSE score ranging from 20-26 (out of 30) is indicative of mild dementia while scores between 10-19 indicates moderate dementia. An MMSE score of less than 10 signifies severe dementia [Crum *et al.,* 1993]. An electroencephalogram (EEG) to detect problems in the electrical activity of the brain was also performed. Somatosensory evoked potentials (SEP) measured the electrical activity of the sensory cortex in response to peripheral sensory stimuli. Jerk-locked averaging (JLA) was performed, which measured the brain activity preceding brisk involuntary movements such as those observed in patients with myoclonus. JLA in conjunction with SEP was used to classify this family's movement disorder [Barrett, 1992] Magnetic Resonance Imaging (MRI) was employed to detect any structural alterations in their brains.

According to Tygerberg Hospital's medical records, another affected member (5555) had reported with similar symptoms in the period during 1993-1995 and had then been clinically examined. Unfortunately the examination was not as thorough as that of individuals 5657 or 5760. A summary of the clinical examination is reported (Table 2.1).

Table 2.1: Summary of the neurological examination of affected individuals.

Symptoms	5760	5555	5657
Age of onset	48 years	25 years	39 years
Year of examination	2006	1993-1995	2006
Dementia	Present	Present	Present
MMSE[1]	14	ND	13
Psychiatric symptoms	Dementia only	Echolalia Disinhibited Aggressive	Hallucinations, paranoid delusions
Reduced facial expression	Yes	No	Yes
Hyperreflexia	No	No	Yes
Increased tone	Yes	No	Yes
Bradykinesia	Yes	No	Yes
Postural tremor	Yes (low amplitude)	No	Yes (prominent)
Rest tremor	No	No	Intermittent
Cerebellar signs	Within normal limits	No	Probably normal
Abnormal posture	Yes	No	Yes
Chorea	No	Yes	No
EEG[2]	Normal	ND	Normal
SEP[3]	Normal	ND	Enlarged
JLA[4]	No cortical spike	ND	ND

1) Mini-Mental State Examination (MMSE)

2) Electroencephalogram (EEG)

3) Somatosensory evoked potentials (SEP)

4) Jerk-locked averaging (JLA)

Not determined (ND)

Ethical approval for this study was provided by the University of Stellenbosch, Committee for Human Research (reference number: N06/05/089). For Family R, informed consent (Appendix 1) was obtained to collect blood samples for the genetic

analysis. These samples were collected during the time in which the family members were examined at Tygerberg Hospital.

Blood was also collected from a number of control individuals, representing the Mixed Ancestry, Caucasian and Black South African sub-populations. Samples from 60 Caucasian individuals and 51 from each of the Black and Mixed Ancestry populations were obtained.

2.2 Exclusion of known loci by the National Health Laboratory Service

Prior to this study, the DNA of one affected individual had been analysed by the National Health Laboratory Service (NHLS -Groote Schuur Inherited Metabolic Diseases). This laboratory screened for mutations on the mitochondrial genome, which commonly cause MELAS, MERRF and Leigh's Disease (Table 2.2). In addition, other known loci were screened by NHLS/University of Cape Town (UCT) Human Genetics Research Unit, (Institute of Infectious Disease and Molecular Medicine, University of Cape Town). These included the loci for Huntington's disease (HD) and Spinocerebellar Ataxia (SCA) types 1, 2, 3, 6 and 7.

Table 2.2: Known mutations causing neurodegenerative disorders.

Disorder	Mutation
Myoclonus Epilepsy and Ragged-Red Fibres **(MERRF)**	A8344G Deletion 8356-8851
Mitochondrial myopathy ,Encephalopathy, Lactic acidosis and Stroke-like episodes **(MELAS)**	A3243G T3271C
Leigh's disease	G13513A G14459A T8993C T9176C
Huntington's disease **(HD)**	Repeat expansion in the *IT15* gene
Spinocerebellar Ataxia **(SCA)**	Repeat expansions at the SCA types 1, 2, 3, 6 and 7 loci

These loci were all negative for mutations and hence excluded as the cause of the disorder in this family.

2.3 <u>Selection of candidate genes</u>

Several disorders are known to cause the symptoms that Family R presented with (Introduction: Section 1.6). These disorders were therefore likely candidates to screen for and are listed below (Table 2.3).

Table 2.3: Candidate disorders selected to be screened and the symptoms involved in each disease.

DISEASE	Myoclonus	Dementia	Dystonia	Ataxia	Dysphasia	Impaired vision	Impaired hearing	Deafness	Seizures	Myoclonic epilepsy	Dysarthria	Ophthalmoplegia	Chorea	Epilepsy	OTHER
MERRF[1]	*	*	*	*	*	*		*	*	*			*		Lipomas; RRF[3]
MELAS[2]	*		*	*		*		*	*	*		*			Lactic acidosis, stroke-like episodes ;vomiting
LHON VARIANT [4]		*	*	*							*	*		*	Ptosis
PRION DISEASE (GSS)[5]	*	*		*											Hemiparesis; babinskis'sign; hallucinations
PD[6]		*	*		*										Parkinsonism
FENIB[7]	*	*					*			*	*			*	Inclusion bodies in neurons
DRPLA[9]	*	*		*							*		*	*	
HDL2[11]	*	*	*	*	*						*		*		

(1) Myoclonus Epilepsy and Ragged-Red Fibres **(2)** Mitochondrial myopathy, Encephalopathy, Lactic acidosis and Stroke-like episodes **(3)** Ragged red fibres **(4)** Familial multisystem degeneration associated with parkinsonism **(5)** Gerstmann-Straussler-Scheinker syndrome **(6)** Parkinson's disease **(7)** Familial Encephalopathy with Neuroserpin Inclusion Bodies **(8)** Spinocerebellar ataxias **(9)** Dentatorubral-Pallidoluysian Atrophy **(10)** Huntington's disease **(11)** Huntington's disease-like 2 **(*)** symptom present

As mentioned in Section 2.2 the common mutations associated with MELAS, MERRF, Leigh's disease and five loci for SCA and HD were already excluded as being the

mutations underlying the disorder in this family. Candidate genes were selected to be screened for pathogenic mutations on the likelihood of these genes being implicated in neurological diseases (Tables 2.4 A and B). The selected candidate genes were divided into mitochondrial (Table 2.4 A) and nuclear (Table 2.4 B) genes.

Table 2.4 A: List of all mitochondrial genes selected as candidate genes to be screened.

Mitochondrial gene	Coding for	Position on mitochondria	Associated Diseases
MTTF	tRNA phenyl alanine	577-647	MM; MELAS; MERRF: Nephritis
MTTL1	tRNA leucine 1	3230-3304	MM, MELAS, MERRF, Deafness, DM
MTTI	tRNA isoleucine	4263-4331	MM, MECM,
MTTQ	tRNA glutamine	4329-4400	MELAS
MTTM	tRNA methionine	4402-4469	MM,
MTTK	tRNA lysine	8295-8364	MERRF, ME
MT-ATP8	ATP synthase subunit	8366-8572	
MT-ND4	NADH de-hydrogenase subunit	10760-12137	Familial multisystem degeneration with Parkinsonism (LHON variant)
MTTH	tRNA histidine	12138-12206	MERRF, MELAS, cerebral oedema
MTTS2	tRNA serine 2	12207-12265	
MTTL2	tRNA leucine 2	12266-12336	MM

DM, Diabetes Mellitus; MECM, Mitochondrial Encephalocardiomyopthy; ME, Mitochondrial Encephalopathy; MM, MERRF/MELAS overlap

Table 2.4 B: List of nuclear genes selected to be screened and associated diseases.

Gene	Chromosomal location	Type of mutation	Associated disorder
Serpini1 (*PI12*)	3q26.1	S49P and S52R point mutations	FENIB
Atrophin1 (*ATN1*)	12p13	CAG repeat expansion in exon	DRPLA
Parkin (*PARK2*)	6q25.2-q27	Point mutations /deletions / exon rearrangements/ duplications/insertions	Early onset PD
Junctophilin-3 (*JPH3*)	16q24.3	CTG/CAG repeat expansion in exon2 A	HDL2

2.4   DNA extraction

All blood samples were collected in EDTA tubes and stored at 4ºC until further use. DNA was extracted from whole blood using a modified phenol/chloroform method by Mrs Ina Le Roux (MAGIC Lab, Tygerberg Campus, Stellenbosch University) [Corfield *et al.,* 1993] (Appendix 3). The method involves the separation of genomic DNA from cellular debris by a phenol/chloroform solution. All DNA samples were diluted to approximately 200ng/µl and this was the working concentration for all subsequent reactions.

2.5   Primer design

The Human Mitochondrial DNA Revised Cambridge Reference Sequence (rCRS) [www.mitomap.org] was used as a reference sequence to design the six primer sets used to amplify the mitochondrial candidate genes (Table 2.5). In some cases, more than one gene could be screened with one set of PCR primers, for example, the *MTTI; MTTQ* and *MTTM* genes were amplified with one primer set.

Primers were designed to amplify nine genes encoding mitochondrial tRNAs, the *ATP8ase* gene and a region of the *MT-ND4* gene. These genes were selected for screening because they had all been previously noted to harbour mutations causing movement disorders and/or dementia. In some cases the primers amplified genes that were in close proximity to the selected regions. For this reason the *MT- ATP8, MTTS2* as well as regions of the *MT-ND1* and *MT-ND2* genes were also screened. Table 2.5 summarises the details of the primer sets designed to amplify candidate genes, including their positions on the genome, the genes that were screened and the PCR conditions.

Primers to amplify nuclear regions were designed based on the reference sequences in the NCBI database [www.ncbi.nlm.nih.gov]. The primers used in the *PARK2* study were designed for another study in our laboratory. All 12 exons, including the promoter region, of the *PARK2* gene were amplified using PCR primers that are listed in Table 2.6. Exon 12 was too large to PCR-amplify with one set of primers so two primer sets (12 A and B) were designed to amplify this exon. The primers designed for the *JPH3* study (Table 2.7) were based on reference sequences from ensembl database [www.ensembl.org].

All primers (except those used in the *PARK2* investigation) were designed using Primer3 programme V.02 [http://frodo.wi.mit.edu/cgibin/primer3/primer3_www.cgi]. The primers were designed at least 30bp upstream and downstream to the 5' and 3' ends of target regions respectively.

The only settings that were changed in Primer3 for the primer designing were:

- Primer melting temperature (Tm):

  Minimum 50ºC, Optimum 55ºC, Maximum 60ºC

- Maximum Tm difference =10ºC

- Primer GC%: Minimum = 40ºC , Optimum = 50ºC, Maximum = 60ºC

- Minimum primer length was set as 18bp and the maximum was 22bp

All the other parameters were used at the default settings.

The melting point temperature of the primers (Tm) was manually calculated using the following formula:

$$Tm = 4ºC (G+C) + 2ºC (A+T)$$

The primer annealing temperature (Ta) of each primer set was calculated as 2ºC-5ºC less than the primer with the lower Tm. Primers were synthesized by Inqaba Biotechnical Industries (Pty) Ltd. Fluorescently-labelled primers were synthesized at the Synthetic DNA Laboratory (Department of Cellular and Molecular Biology, University of Cape Town, South Africa) using standard phosoramidite techniques with an Oligo 1000M DNA Synthesizer (Beckman Instruments Inc.). The fluorescent label, FAM-6, was covalently attached to the non-reactive 5' end of the primers.

Table 2.5: List of the primers designed for this study including PCR conditions. Mitochondrial primers are listed in blue and primers for nuclear genes are in black.

Primer Set	Sequence	Expected fragment size	Gene	Position on the gene	Annealing Temperature (ºC)	Additives	Labelled
MTTF Forward Reverse	AATACAACCCCCGCCCATC GGGGATGCTTGCATGTGTA	222bp	*MTTF*	486-505 690-708	55	None	None
MTTL1 Forward Reverse	GCCTTCCCCCGTAAATGATA CGTTCGGTAAGCATTAGG	218bp	*Mt-ND1 partially*	3163-3182 3363-3381	52	None	None
MTTI, Q, M Forward Reverse	GATTCCGCTACGACCAACTC GCTTAGCGCTGTGATGAGTG	397bp	*MTTI,MTTQ MTTM Q MT-ND2 (partially)*	4142-4161 4519-4538	60	None	None
MT-ND4 Forward Reverse	CACGGGCTTACATCCTCATT GGAGAACGTGGTTACTAGCACA	207bp	*MT-ND4 (partially)*	11714-11733 11898-11920	60	None	None
MTTH, S2, L2 Forward Reverse	ACACCTATCCCCCATTCTCC AACGAGGGTGGTAAGGATGG	333bp	*MTTH, MTTS2, MTTL2*	12076-12095 12389-12408	58	None	None
MTTK, MT-ATP8 Forward Reverse	ATGCCCATCGTCCTAGAATC TTTGGAGGTGGGGATCAATA	428bp	*MT-Atp8 ase*	8204-8222 8612-8631	55	None	None
PI 12 Forward Reverse	CCTCCCAACATATCCTTCCA GGGTTTCCTAATCCTTCTAGCC	491bp	*SERPINI1*	53301-53320 53770-53791	58	None	None
DRPLA rpt Forward Reverse	 CCTCTTAGCCAACAGCAATG AGACATGGCGTAAGGGTGTG	~250 bp- 350bp	CAG repeat in *Atrophin1*	1586-1606 1817-1836	58	None	6-FAM
HDL2 Forward Reverse	 GGAATCTCGTCTTTCAGTGG TGAGGAGTGGATATCGGAGAG	~250-300bp	CTG/CAG repeat in *Junctophilin-3*	1274-1294 1496-1516	58	DMSO	6-FAM

Table 2.6: Primers and PCR conditions for the screening of the *PARK2* gene.

Region	Forward primer	Reverse primer	Annealing temp (ºC)	Product size (bp)	Additives
PROMOTER	ATCCAGATGTTTGGCAGCTC	CGGCTCTCCTGGGTTAAATC	55	351	Formamide
EXON 1	GAACTACGACTCCCAGCAG	CCCGTCATTGACAGTTGG	55	300	Formamide
EXON 2	CACCATTTAAGGGCTTCGAG	TCAGGCATGAATGTCAGATTG	55	313	None
EXON 3	TCTCGCATTTCATGTTTGACA	GCAGACTGCACTAAACAAACA	55	364	None
EXON 4	GCTTTAAAGAGTTTCTTGTC	TTTCTTTTCAAAGACGGGTGA	55	299	None
EXON 5	GGAAACATGTCTTAAGGAGT	TTCCTGGCAAACAGTGAAGA	55	223	None
EXON 6	CCAAAGAGATTGTTTACTGTG	GGGGGAGTGATGCTATTTTT	55	280	None
EXON 7	CCTCCAGGATTACAGAAATTG	GTTCTTCTGTTCTTCATTAGC	55	232	None
EXON 8	GGCAACACTGGCAGTTGATA	GGGGAGCCCAAACTGTCT	55	232	None
EXON 9	TCCCATGCACTGTAGCTCCT	CCAGCCCATGTGCAAAAGC	55	297	None
EXON 10	CAAGCCAGAGGAATGAATAT	GGAACTCTCCATGACCTCCA	55	272	None
EXON 11	CCGACGTACAGGGAACATAAA	GGCACCTTCAGACAGCATCT	55	300	None
EXON 12A	TCTAGGCTAGCGTGCTGGTT	GCGTGTGTGTGTGTGTTTGA	55	296	Formamide
EXON 12B	ACATCCTGGGGGAGCATAC	GCTCTGCTGTCTTGTGTGGA	55	237	None

Table 2.7: Primers and PCR conditions used in the analysis of mRNA.

Primer Set	Sequence	Expected fragment size	Gene	Annealing Temperature (ºC)	Additives
HBB Forward Reverse	ACACAACTGTGTTCACTAGC ACCGAGCACTTTCTTGCCAT	(mRNA) 248bp (gDNA) 377bp	*β-Haemoglobin*	55	None
JPH3 **Full length set A** Forward Reverse	TACTCGGACGGAGGGACC GGTACTCCAGCCCGTTTTC	928bp	Exon 2-3 *JPH3*	ND*	None
JPH3 **Full length set B** Forward Reverse	AAGTCCAGTACGGGCTCAG CCACTTTCTAAGCTCGTGTC	530bp	Exon 4-5 *JPH3* (partially)	56	None
JPH3 **Alternatively spliced** Forward Reverse	CTACTCGGACGGAGATGC AGGAGTGGATATCGGAGAGT	181bp	Alternatively spliced *JPH3* mRNA	56	None

*ND: Not determined
gDNA: Genomic DNA

2.6 DNA quantification

DNA and primers were quantified and analysed using the NanoDrop® ND 1000 –UV-Vis Spectrophotometer (NanoDrop Technologies, Inc. USA) according to manufacturer's instructions. The NanoDrop® employed UV/VIS spectrophotometry to accurately determine nucleic acid concentration in a sample which is recorded in ng/ul. It also measured the purity and quality of a nucleic acid sample by measuring the ratio of absorbance at 260nm and 280nm. The 260/280 ratio for good quality, purified DNA is ~1.8. A deviation in this ratio is normally indicative of contaminants, such as salts or phenol, in a sample.

[www.nanodrop.com/techsupport/nd-1000-users-manual.pdf]

2.7 Polymerase chain reaction

PCR amplification was performed in a standard reaction mixture of 50μl on a GeneAmp® PCR system 2720 Thermal (Applied Biosystems). Each PCR reaction contained 0.1μg template DNA; 20pmol of each primer; 2.5mM (0.05mM of each) dNTP (dATP; dTTP; dCTP; dGTP) (Bioline Ltd, London, UK), 0.1 Units (U) BIOTAQ™ DNA Polymerase (Bioline Ltd, UK); 1XNH$_4$ Buffer (Bioline Ltd, London, UK);1.5mM MgCl$_2$ and Sterile Injection water (Fresenius Kabi) to a final volume of 50μl.

In some cases (Tables 2.5, 2.6 and 2.7), additives such as Dimethyl sulphoxide (DMSO) or formamide were used to facilitate the amplification process. In such cases 5% of the final volume contained the respective additive.

The PCR cycling regime was as follows:
An initial denaturing step of 94°C for five minutes followed by 35 cycles of denaturation at 94°C for 30 seconds, varying annealing temperature (Table 2.5, 2.6 and 2.7) for 30 seconds and extension at 72°C for 45 seconds. The 35 cycles was followed by a final extension step of 72°C for 5 minutes.

For colony PCR, sterile techniques were used to pick single colonies as template DNA. The initial denaturing step was extended to ten minutes to ensure that the colonies

would burst and release the DNA. Thereafter, PCR amplification was performed as described above. All PCR products were stored at 4°C until further use.

2.8   Gel electrophoresis

All PCR products and DNA was visualised using agarose gel electrophoresis. Gels were made up to 2% (w/v) by dissolving 1g of agarose (Whitehead Scientific, South Africa) in 50ml of 1X Sodium Borate (SB) buffer (Appendix 2) at a high temperature. Once the solution had cooled down, 5µl of 1µg/µl Ethidium Bromide was added to the solution. Ethidium bromide binds to DNA molecules and allows the DNA to be visualised under ultra violet (UV) light. Ethidium bromide is a carcinogen and should be used with caution. Later in the study, SYBR Green I Nucleic Acid Gel Stain® (Roche Applied Sciences) was used according to manufacturer's specifications, to replace the use of Ethidium bromide. SYBR Green I Nucleic Acid Gel Stain® serves the same purpose as Ethidium bromide but is not thought to be carcinogenic and thus minimised the risk of danger to the user.

Prior to loading a sample onto a gel, 1µl of Bromophenol Blue (Appendix 2) loading dye was added to 8µl of each sample. In all cases 2µl of 100bp DNA ladder (Promega, USA) was co-electrophoresed with the samples on the gel as a size standard.

The samples were resolved by electrophoresis at 250 volts (V) for 10-20 minutes in 1XSB running buffer. The samples were viewed under UV illumination (Transilluminator model LMS-26E, USA) and photographed using a Sony graphic system UP-860CE (Sony Corporation, Japan).

2.9   Purification of PCR products

In some cases PCR products were purified for downstream processes. For sequencing, DNA was purified using *Exonuclease I* (USB Corporation, USA) and Shrimp Alkaline Phosphatase (Roche Applied biosciences) according to manufacturer's instructions. 5U of Shrimp Alkaline Phosphatase and 5U of *Exonuclease I* was added to 8µl of PCR

product. The mixture was then incubated at 37ºC for 15 minutes followed by 80ºC for 15 minutes to deactivate the enzymes.

The concentration and purity of the PCR products were checked using the Nanodrop® according to manufacturer's specifications. The samples were then diluted down to the desired concentration using sterile distilled water.

For cloning purposes the appropriate DNA fragment was excised from an agarose gel using a sterile scalpel and placed in a sterile Eppendorf tube. The fragment was then purified using the GFX™ PCR DNA and gel band purification kit (Amersham Pharmacia Biotech, New Jersey, USA) according to manufacturer's instructions.

2.10   Direct sequencing

Automated direct sequencing [Mattocks *et al.,* 2000] was performed using the BigDye Terminator Sequence Ready Reaction kit version3.1 (Applied Biosystems, USA) and resolved on an ABI 3130X ® Genetic Analyzer (Applied Biosystems, USA). The primers used for sequencing were diluted to 1.1μM and the PCR products were diluted to 20ng/μl. The sequencing data was analysed using BioEdit version 7.0.5 software [www.mbio.ncsu.edu/BioEdit/page2.html].

2.11   Genotyping on an ABI 3130xl® Genetic Analyzer

For parts of this study, fluorescently labelled primers were used for automated genotyping of samples on the ABI Prism 3130XF® Genetic Analyzer (Applied Biosystems, USA). GeneScan-500 LIZ size standard was used as a size marker, which can determine fragments of 75- 500bp in size. The data was analysed using GeneMapper ® V3.7 software (Applied Biosystems, USA). The software sizes the peaks based on the internal size standard and analyses the peaks based on user-defined parameters.

2.12   Single Strand Conformational Polymorphism (SSCP) Analysis

Single Strand Conformational Polymorphism (SSCP) Analysis [Orita *et al.*, 1989] was used as a mutation detection method for screening the *PARK2* gene. The method involves separating double stranded DNA and then rapidly cooling it so that the single stranded DNA form secondary structures to create a unique banding pattern when subjected to electrophoresis on a polyacrylamide gel (Appendix 4). The gel is then silver stained [Gottlieb and Chavko, 1987] (Appendix 5) in order to detect the bands formed by DNA. Changes in the DNA sequence will then be seen as a shift in the banding pattern.

Certain conditions may be altered in order to obtain optimal results. In this study, one such optimisation was employed by varying the percentage of polyacrylamide and urea in the gel (Appendix 4). The samples were run on both 8% and 10% polyacrylamide gels at 4ºC, 70 Watts overnight on an 'in-house" manufactured electrophoresis system. Samples in which a shift was observed were sequenced to determine the exact sequence variation.

2.13   Genetic ancestry testing

Genetic Ancestry testing was performed by the MRC/NHLS/WITS Human Genomic Diversity and Disease Research Unit (HGDDRU) Johannesburg, South Africa. The method uses Y–chromosome analysis to determine the paternal lineage and sequencing of regions in the mitochondrial genome to resolve the maternal lineage of the individual.

Family R member, 5760, was used in this study as a representative of the family. The detection of various single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) on the Y-chromosome created a haplogroup for the individual. The resolved haplogroup was compared against a database containing 42000 males worldwide [www.yhrd.org/index.html].

Sequencing of regions of the mitochondrial genome resolved the maternal haplotype for 5760. This profile was compared to 33 haplogroups recognized world wide and then grouped to a particular sub-haplogroup.

2.14 Cloning and transformation

Following DNA quantification, purified PCR products were cloned into the pGEM®-T Easy cloning vector using the pGEM®-T Easy Vector Systems cloning kit (Promega, USA) as specified by the manufacturer's instructions. Three ligation reactions were set up in which 1:1, 1:3 and 1:5 vector to PCR product ratios were used.

Competent *E. coli* DH5α cells were prepared using the method described by Sambrook and Russell (Appendix 6) [Sambrook and Russell, 2001]. 3μl-5μl Of the respective ligation reactions were used to transform 200μl of thawed competent cells *via* heat shocking. This process entailed incubating tubes containing the cells and the ligation mixture on ice for 20-30 minutes. The mixtures were then placed at 42ºC for 45 seconds only, followed by incubation at room temperature for 2 minutes. 2ml Luria Bertani Broth (Appendix 2) was added to the cells and the cells were incubated on a shaker at 37ºC for 60 minutes. Thereafter, 200μl of this mixture was plated onto Luria Bertani (LB)agar (Appendix 2) containing 15mM Isopropyl-beta-D-thiogalactopyranoside (ITPG), 100μg/ml ampicillin and 5mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-GAL). The plates were then incubated overnight at 37°C.

The following day, a number of blue and white single colonies were picked and placed into separate sterile 15ml Falcon tubes containing 5ml LB broth. These tubes were incubated at 37ºC on a shaker overnight to allow cells to express the vector. The following day the tubes were spun down at 15000rpm and the supernatant was removed. The isolation of plasmid DNA from the cells was attained using GeneJET™ Plasmid Miniprep Kit (Fermentas Life Sciences, UK) according to manufacturer's instructions. The plasmids were subjected to colony PCR, with insert specific primers followed by direct sequencing to determine whether they contained the correct insert.

2.15 Lymphocyte separation and transformation

Lymphocytes were cultured by Mrs Ina Le Roux using an "in-house" method. Blood of two affected family members were collected in heparin tubes. Heparin tubes prevent the coagulation of blood by inhibiting thrombin in the coagulation cascade. Lymphocytes were isolated from the whole blood and transformed (immortalised) to provide a

permanent source of DNA and RNA of the patients. This was achieved by initiating transformations of lymphocytes, with Epstein-Barr virus infection using an optimised protocol (Appendix 7). The cell pellet was then spun down and stored at -20ºC until further use.

2.16   RNA isolation and cDNA synthesis

RNA was isolated from the patients' transformed lymphocytes using TRIzol ® reagent (Invitrogen Life Technologies) in a modified protocol (Appendix 8) [Chomczynski and Sacchi, 1987]. In addition, RNA from 21 pooled foetal brains was purchased (Clonetech, USA). The foetuses had spontaneously aborted at 26-40 weeks and the RNA was extracted from fresh brain tissue. The purity of foetal brain RNA was analysed and both the 260/280 and the 28S/18S ratios showed that the RNA was of good standard (Figure 2.2). Generally a ratio of 1.8 is considered good quality.
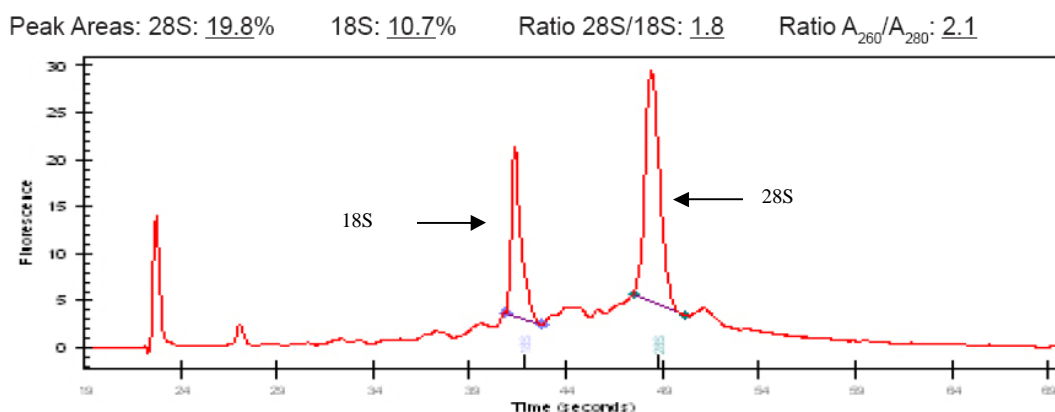
Figure 2.2: Electropherogram indicating the qualitative analysis of the purchased foetal brain RNA.

The quality and quantity of the lymphocyte RNA was appraised by automated electrophoresis on the Experion™ Automated Electrophoresis System using RNA StdSens Analysis Kit (Bio-Rad Laboratories Inc) according to manufacturer's instructions. The kit employs the use of LabChip microfluidic technology to detect, quantify and analyse the quality of RNA. To ensure that the isolated RNA was free of genomic DNA, the samples were treated with TURBO DNA-Free ™ kit (Ambion) according to manufacturer's instructions. Furthermore, the QuantiTect® Reverse

Transcription Kit (Qiagen) also contained a genomic DNA removal buffer which was used according to manufacturer's instructions.

After the successful removal of the majority of genomic DNA contaminants, cDNA synthesis of lymphocyte and foetal brain RNA was carried out as using the QuantiTect® Reverse Transcription Kit according to manufacturer's instructions. The kit provides a mixture of polyA and random oligos that anneals to a number of places on the RNA. The RNA –dependent polymerase then transcribes the first strand of cDNA which is an exact copy of the RNA (Figure 2.3).
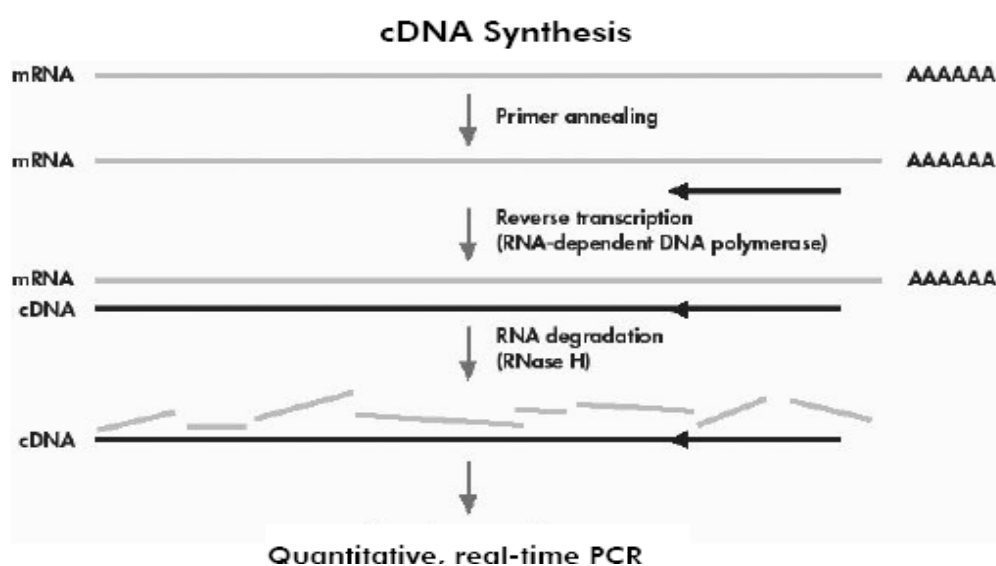
Figure 2.3 Steps of 1st strand cDNA synthesis using polyA oligos [Adapted from www.qiagen.com].

The viability and purity of cDNA was evaluated by PCR with mRNA specific primers designed to amplify the *Beta Haemoglobin* gene (*HBB*). The *HBB* gene was used as a housekeeping gene for subsequent reactions. The primers were designed to amplify both genomic and mRNA *HBB* transcripts. However, in genomic DNA the primers produced a fragment of 377bp whereas in cDNA an amplicon of 248bp was produced (Figure 2.4). This further allowed isolated cDNA to be checked for genomic DNA contamination.

(A)
```
  1 ACATTTGCTT CTG ACACAAC TGTGTTCACT AGC AACCTCA AACAGACACC ATGGTGCATC
 61 TGACTCCTGA GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG
121 TTGGTGGTGA GGCCCTGGGC AGGCTGCTGG TGGTCTACCC TTGGACCCAG AGGTTCTTTG
181 AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG GTGAAGGCTC
241 ATGGCAAGAA AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC AACCTCAAGG
301 GCACCTTTGC CACACTGAGT GAGCTGCACT GTGACAAGCT GCACGTGGAT CCTGAGAACT
361 TCAGGCTCCT GGGCAACGTG CTGGTCTGTG TGCTGGCCCA TCACTTTGGC AAAGAATTCA
421 CCCCACCAGT GCAGGCTGCC TATCAGAAAG TGGTGGCTGG TGTGGCTAAT GCCCTGGCCC
481 ACAAGTATCA CTAAGCTCGC TTTCTTGCTG TCCAATTTCT ATTAAAGGTT CCTTTGTTCC
541 CTAAGTCCAA CTACTAAACT GGGGGATATT ATGAAGGGCC TTGAGCATCT GGATTCTGCC
601 TAATAAAAAA CATTTATTTT CATTGC
```

(B)
```
   1 ACATTTGCTT CTGACACAAC TGTGTTCACT AGCAACCTCA AACAGACACC ATGGTGCATC
  61 TGACTCCTGA GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG
 121 TTGGTGGTGA GGCCCTGGGC AGgttggtat caaggttaca agacaggttt aaggagacca
 181 atagaaactg ggcatgtgga gacagagaag actcttgggt ttctgatagg cactgactct
 241 ctctgcctat tggtctattt tcccacccctt agGCTGCTGG TGGTCTACCC TTGGACCCAG
 301 AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG
 361 GTGAAGGCTC ATGGCAAGAA AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC
 421 AACCTCAAGG GCACCTTTGC CACACTGAGT GAGCTGCACT GTGACAAGCT GCACGTGGAT
 481 CCTGAGAACT TCAGGgtgag tctatgggac gcttgatgtt ttctttcccc ttctttctcta
 541 tggttaagtt catgtcatag gaaggggata agtaacaggg tacagtttag aatgggaaac
 601 agacgaatga ttgcatcagt gtggaagtct caggatcgtt ttagtttctt ttatttgctg
 661 ttcataacaa ttgtttttctt ttgtttaatt cttgctttct ttttttttct tctccgcaat
 721 ttttactatt atacttaatg ccttaacatt gtgtataaca aaaggaaata tctctgagat
 781 acattaagta acttaaaaaa aaactttaca cagtctgcct agtacattac tatttggaat
 841 atatgtgtgc ttatttgcat attcataatc tccctacttt attttcttttt attttttaatt
 901 gatacataat cattatacat atttatgggt taaagtgtaa tgtttttaata tgtgtacaca
 961 tattgaccaa atcagggtaa ttttgcattt gtaattttaa aaaatgcttt cttctttttaa
1021 tatacttttt tgtttatctt atttctaata cttttccctaa tctctttctt tcagggcaat
1081 aatgatacaa tgtatcatgc ctctttgcac cattctaaag aataacagtg ataatttctg
1141 ggttaaggca atagcaatat ctctgcatat aaatatttct gcatataaat tgtaactgat
1201 gtaagaggtt tcatattgct aatagcagct acaatccagc taccattctg cttttattttt
1261 atggttggga taaggctgga ttattctgag tccaagctag gccctttttgc taatcatgtt
1321 catacctctt atcttcctcc cacagCTCCT GGGCAACGTG CTGGTCTGTG TGCTGGCCCA
1381 TCACTTTGGC AAAGAATTCA CCCCACCAGT GCAGGCTGCC TATCAGAAAG TGGTGGCTGG
1441 TGTGGCTAAT GCCCTGGCCC ACAAGTATCA CTAAGCTCGC TTTCTTGCTG TCCAATTTCT
1501 ATTAAAGGTT CCTTTGTTCC CTAAGTCCAA CTACTAAACT GGGGGATATT ATGAAGGGCC
1561 TTGAGCATCT GGATTCTGCC TAATAAAAAA CATTTATTTT CATTGC
```

Figure 2.4: Position of *HBB* primers on mRNA **(A)** and on genomic DNA **(B)** producing differently sized products (Black arrows indicate the positions and orientation of the primers). **(A)** Each exon is in a different colour. **(B)** Exons are in uppercase, while introns are in black lowercase.

2.17 Sequence alignments

Alignments were carried out using the Multiple Sequence Alignment programme by ClustalW [http://align.genome.jp] under the default settings. Alignments between two sequences were also performed using the BLAST (bl2seq) tool on the NCBI database [http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi] with the "filter" tool disabled which allows for repeats to be aligned. Input sequences were obtained from either the NCBI GenBank [http://www.ncbi.nlm.nih.gov/Genbank/] or Ensembl databases [http://www.ensembl.org/index.html] and are listed in Table 2.8.

Table 2.8: List of all sequences and accession numbers used in the analysis of the *JPH3* gene.

Description	Sequence ID	Accession number	Database
Contig from foetal brain cDNA	U80757.1	GI: 2565083	dbEST
Contig from retinoblastoma	BE779067.1	GI: 10200265	dbEST
Contig from lung carcinoid	BE042890.1	GI:8359943	dbEST
Genomic DNA sequence of human *JPH3* gene	NC_000016.8	GI:51511732	GenBank
mRNA of full length *JPH3* in humans	NM_020655.2	GI:21704282),	GenBank
Protein of full length *JPH3* transcript in humans		ENSP00000284262	ensembl
mRNA of alternatively spliced *JPH3* gene in human		ENST00000301008	ensembl
Protein sequence of alternatively spliced *JPH3* in humans	ENSP00000301008	ENSP00000301008	ensembl
Genomic sequence of *Mus musculus Junctophilin-3* gene (mouse *JP3*)	NC_000074.5	GI:149361523	GenBank
Junctophilin-3 (*JP3*), mRNA sequence of mouse *JP3* gene	NM_020605.2	GI:118130503	GenBank
Protein sequence of mouse *JP3*	NP_065630.1	GI:10181142	GenBank
Genomic sequence of *JPH3* in *Pan troglodytes* (Chimpanzee)		ENSPTRG00000008447	ensembl
mRNA sequence of the full length *JPH3* gene of Chimpanzee		ENSPTRT00000015560	ensembl
Protein sequence of the full length *JPH3* gene of Chimpanzee		ENSPTRP00000053372	ensembl
mRNA sequence of alternative spliced *JPH3* in chimpanzee		ENSPTRT00000060083	ensembl
Protein sequence of alternative spliced *JPH3* in chimpanzee		ENSPTRP00000053371	ensembl

2 .18   Statistical analysis

Statistical analysis was performed in consultation with Dr. Lize van der Merwe (Biostatistics Unit, Medical Research Council of South Africa). Statistical tests from packages of the open-source programming environment R (R development core team) were used. The non-parametric Kolmogorov-Smirnov test [www.R-project.org] and Fishers' test was used to analyse the repeat distributions in the population studies. Furthermore, the binomial distribution test was used to appraise the probability of maternal inheritance in the analysis of the pedigree. The observed heterozygosity was calculated as:

$$\frac{\textbf{Number of heterozygotes}}{\textbf{Number of individuals}}$$

2.19   Real-time PCR on LightCycler® 1.5 Instrument

In conjunction with conventional PCR, real-time PCR on the LightCycler® 1.5 Instrument (Roche Applied Sciences, Germany) was used to amplify regions of the *HBB* and *JPH3* genes from foetal brain and lymphocyte cDNA. Real-time PCR is recognised as a much more sensitive and rapid PCR process than conventional PCR and Ethidium Bromide staining. It produces results in real time (as the amplification process is happening). The method uses a fluorescent dye, in this case SYBR Green I, which integrates into double stranded DNA as the amplicons are formed and gives off a fluorescent signal that is detected by the instrument during the extension step of each cycle. The LightCycler®FastStart DNA Master [PLUS] SYBR Green I kit (Roche Applied Sciences) was used to prepare the PCR reactions according to manufacturer's instructions. Real-time PCR was attempted for the amplification of foetal and lymphocyte cDNA and was used as a qualitative and quantative means to determine whether the transcript could be detected in both tissues. LightCycler Software 4.0 was used for the analysis of the results. A melt curve analysis was also performed on the samples. For melt curve analysis the reaction is slowly heated to 95ºC which causes double stranded amplicons to separate and the dye to be released. A rapid decrease in fluorescence is recorded by the software which then creates a characteristic curve for a particular sequence. The melt curve analysis is used to detect any non-specific products in the reaction as each product has a unique curve.

CHAPTER THREE: RESULTS

This section is divided into two sub-sections. The first part deals with identifying the genetic cause of the disease in the family. The second section reports the results of the analysis of the HDL2 locus.

Section A: Investigation of the causative mutation

3.1   Pedigree analysis

Statistical analysis of the pedigree (Figure 2.1) indicated that there was a strong possibility that this disease could have a maternal pattern of inheritance. It was therefore hypothesised that the disease might be due to a mutation on the mitochondrial genome. A statistical analysis using a binomial distribution calculation showed that there was a 1:400 probability that this disorder was maternally linked**.**

Generally the symptoms displayed by affected family members were severe, progressive dementia and movement disorders. A neurological examination of 5657, 5760 and 5555 was performed. It was determined that 5657 and 5760 shared similar mental states but their respective movement disorders differed. The proband, 5657, had cortical myoclonus while his uncle, 5760, had predominantly parkinsonian features. Individual 5555 had prominent chorea on examination but none of the other features were reported.

3.2   Investigation of mitochondrial candidate genes

Due the potentially maternal bias of the disease segregation in Family R, several mitochondrial genes were selected as candidate genes to screen. A literature search on the mitochondrial genome showed that there were many mutations in this genome that could be the cause of dementia and a movement disorder. Most of the mitochondrial mutations associated with dementia occur in genes that encode transfer RNAs (tRNA). For this reason several mitochondrial genes were selected (Table 2.4 A, page 39) to be screened for known or novel mutations.

The amplification of selected regions of the mitochondrial genome was successful for all study participants. A fragment of the expected size was produced for each primer set (Figure 3.1).
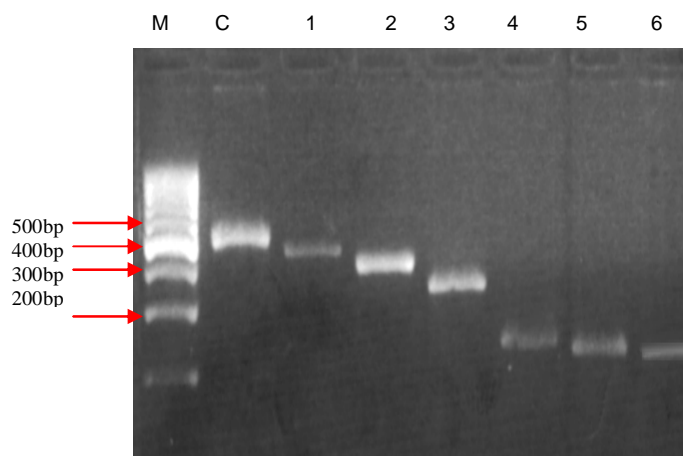
Figure 3.1: A 2% agarose gel representing the electrophoresis of PCR products yielded from the 6 sets of primers used to amplify selected regions of the mitochondrial genome. In many cases more than one gene was amplified with one set of primers. (Lane 1: *MTTK* and *MT-ATP8ase*, lane 2: *MTTI; MTTQ* and *MTTM,* lane 3: *MTTH; MTTS2* and *MTTL2,* lane 4: *MTTF*, lane 5: *MTTL1* and lane 6: *MT-ND4*) .The lane marked M contains a 100bp ladder with the sizes indicated by the red arrows. The lane marked C represents a 450bp fragment as an additional sizing control.

Initially, the PCR products of two affected individuals namely, 5657 and 5760 were sequenced to screen for possible pathogenic mutations. The mitochondrial genome is highly polymorphic and the sequence varies widely between different ethnic groups. For this reason DNA of an ethnically matched unaffected individual was sequenced as a wild type (WT control). The Human Mitochondrial DNA Revised Cambridge Reference Sequence (rCRS) [www.mitomap.org] was used as the reference sequence.

When performing the sequence analysis of these family members, the WT control and the rCRS , five sequence variants in the *MTTF, MT-ND1, MTTI* and *MT-ND2* genes were identified. To confirm whether these variants were indeed disease-causing they would have to segregate in all affected individuals and be absent in unaffected family members. For this reason the PCR products produced by primer sets MTTF and MTTI

were sequenced in additional family members. Figures 3.2-3.6 represents the sequencing of the variants found in the PCR products.

(A) Wild type control
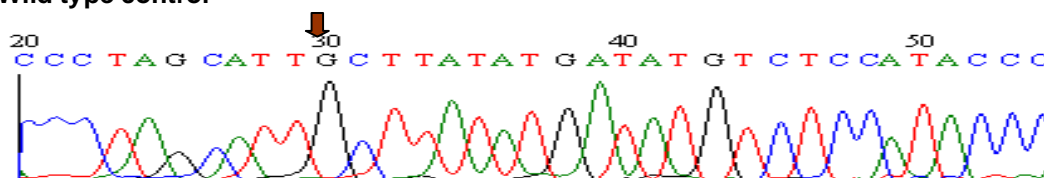
(B) 5657
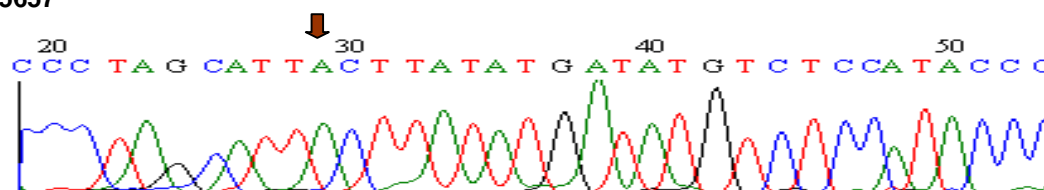
Figure 3.2: Represents a section of the chromatograph produced by the sequencing of the *MTTI, MTTQ, MTTM* and *MT-ND1* genes. The solid red arrow indicates a homoplasmic G to A change in the sequence. This position corresponds to the position 4206 in the *MT-ND1* gene on the mitochondrial genome. **(A)** WT control and **(B)** an affected family member (5657) of Family R.

(A) Wild type control
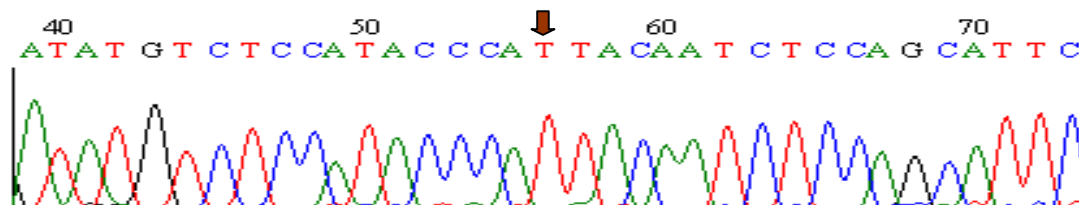
(B) 5657
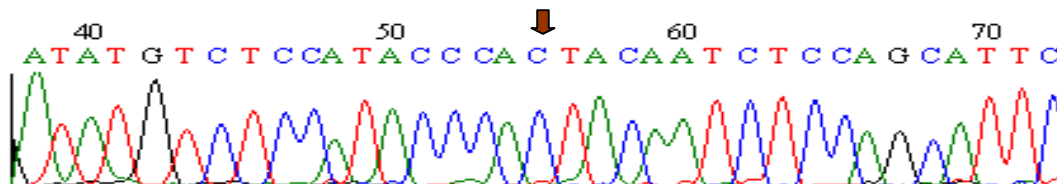
Figure 3.3: Chromatographs depicting the homoplasmic T to C change in the sequence (indicated by the solid red arrows). This corresponds to the position 4232 on the mitochondrial genome, which is in the *MT-ND1* gene. **(A)** WT control and **(B)** member 5657 of Family R.

(A) Wild type control
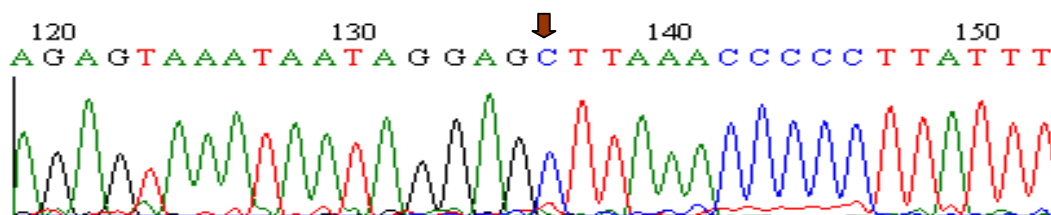
(B) 5657
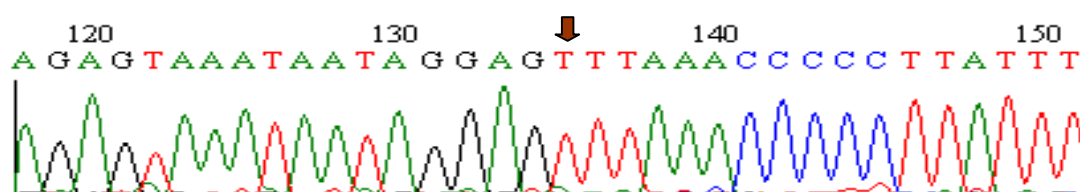
Figure 3.4: Chromatographs depicting the homoplasmic C to T change in the sequence (indicated by the solid red arrows). This corresponds to the position 4312 on the mitochondrial genome which is in the *MT-TI* gene. **(A)** WT control and **(B)** member 5657 of Family R.

(A) Wild type control
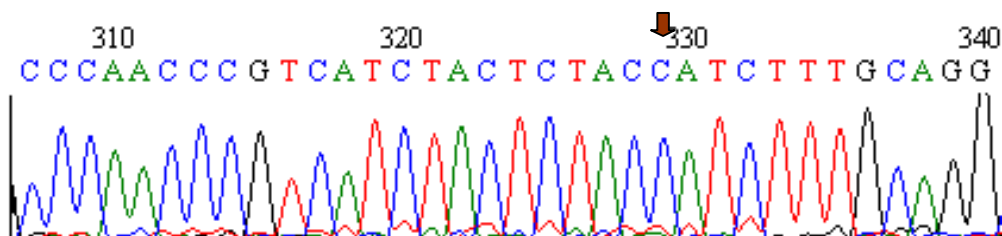
(B) 5657
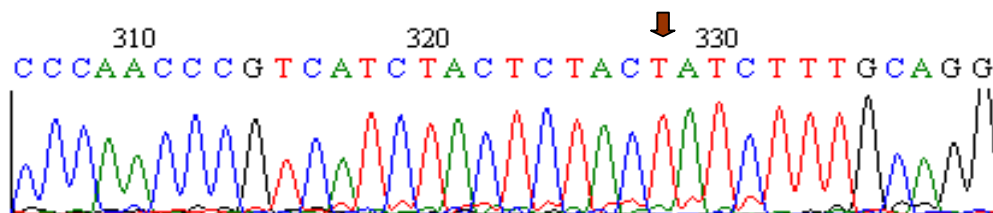
Figure 3.5: Chromatographs depicting the homoplasmic C to T change in the sequence (indicated by the solid red arrows). This change corresponds to the position 4505 on the mitochondrial genome which is in the *MT-ND2* gene. **(A)** WT control and **(B)** member 5657 of Family R.

(A) Wild type control
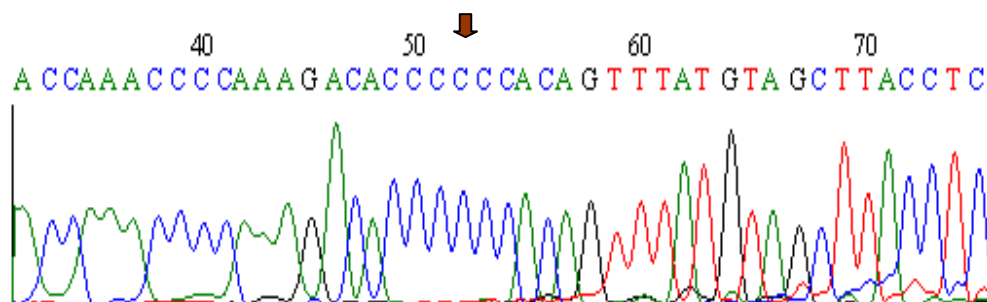
(B) 5657: Forward primer sequence
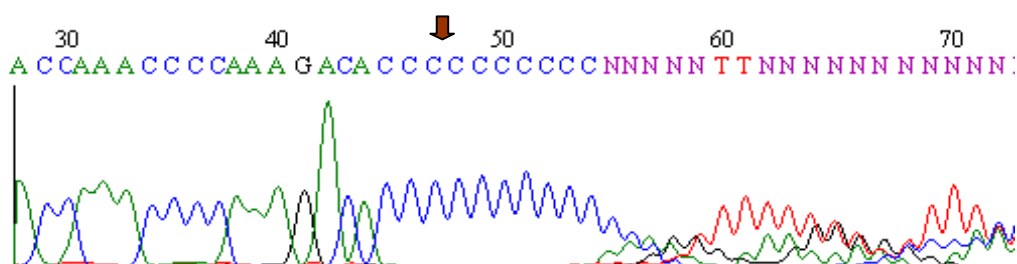
(C) 5657: Reverse primer sequence
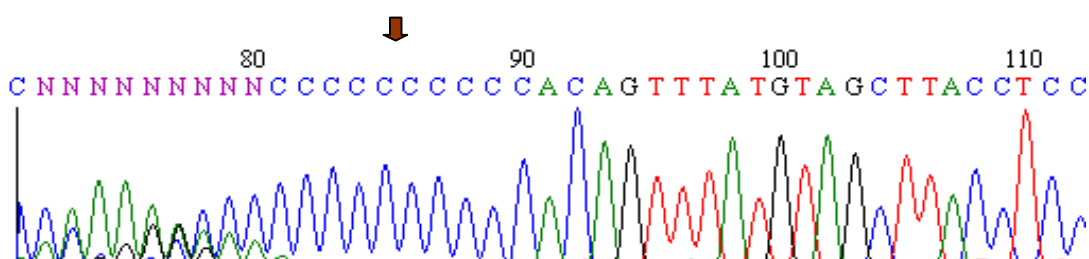
Figure 3.6: Results obtained from automated sequencing displaying the heteroplasmic polyC tract at position 568 which is in the control region on the mitochondrial genome (indicated by the solid red arrows). **(A)** The polyC tract in the WT control consisting of 6C's. **(B)** The polyC tract in member 5657 of Family R where the polyC tract had expanded and was heteroplasmic. **(C)** The same polyC tract of 5657 sequenced with the reverse primer and reverse complemented using the programme BioEdit.

From this investigation three novel variants were determined at positions 4206, 4232 and 4505 on the mitochondrial genome. These variants were deposited in the Mitomap database [www.mitomap.org] under the reference numbers listed in Table 3.1. The other two variants were known variants and have not been associated with disease. All the sequence variants are unlikely to be pathogenic because these variants have not

previously been associated with the symptoms observed in this family. Furthermore, mitochondrial DNA is highly variable and prone to polymorphisms due to the high rate of replication, the absence of protective histones and exposure to reactive oxygen molecules that cause oxidative damage. It would therefore be difficult to prove the pathogenicity of the novel variants. The variant observed at position 4206 occurred only in the WT control. No sequence variants were determined for any of the other candidate genes investigated. Furthermore the G11778A mutation in the *MT-ND4* gene, causing Leber's Hereditary Optic Neuropathy (LHON) was not found in any of the affected family members.

In summary, this investigation indicated that it was unlikely that these mitochondrial genes were the cause for disease in this family.

Table 3.1: A list of sequence variants observed in comparison with the rCRS and WT control sequence.

Position on mitochondrial genome	Gene symbol	rCRS*	WT control	5657	5760	5555	6188	6341	Homoplasmic	Synonymous/ Non-synonymous	Amino acid changes	Known/ Novel	Reference:
568**	*MT-TF*	6C's	6C's	ND	ND	ND	ND	ND	No	NA	NA	Known	Sternberg *et al.,* 2001
4206	*MT-ND1*	A	G	A	A	A	A	A	Yes	Synonymous	L300L TTA-TTG	Novel	20061030001***
4232	*MT-ND1*	T	T	C	C	C	C	C	Yes	Non-synonymous	I 309T ACT-ATT	Novel	20061030002***
4312	*MT-TI*	C	C	T	T	T	T	T	Yes	NA	NA	Known	Moraes *et al.,* 1993
4505	*MT-ND2*	C	C	T	T	T	T	T	Yes	Synonymous	T12T ACC-ACT	Novel	20061030003***

*rCRS: the Human Mitochondrial DNA Revised Cambridge Reference Sequence.
** The variant at position 568 is an expanded polycytosine tract in a non-coding region. The polyC tract is unstable and can therefore not be accurately counted.
***The accession numbers of the variants as provided by the Mitomap database.
ND, not determined.
4206 Change in the control.
4312 Change was in a non-coding region.

3.3   Investigation of the *PARK2* gene

Mutations in the *PARK2* gene have previously been associated with early onset parkinsonism [Tan and Skipper, 2007]. As members of Family R exhibited parkinsonian features, the *PARK2* gene was screened for any known or novel mutations. The PCR amplification was successful for all the exons as well as the promoter region (Figure 3.7).
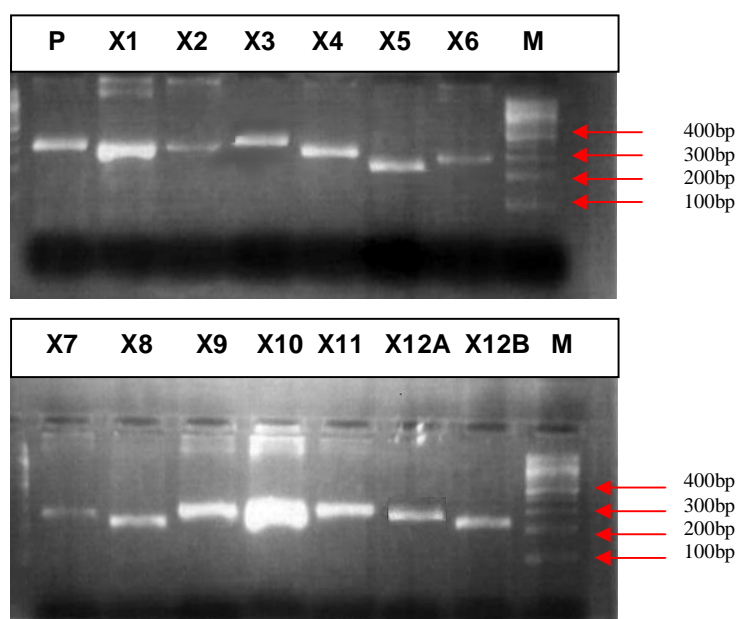
Figure 3.7: Representative 2% agarose gel showing the PCR amplification of the promoter and all the exons of the *PARK2* gene. The lane marked M contains a 100bp ladder with corresponding sizes indicated by the red arrows. The lane marked P contains the promoter region while lanes marked X1-X12B represent Exons 1-12.

SSCP analysis was performed on the PCR-amplified products to screen for sequence variations in the *PARK2* gene. Any shifts in the banding pattern (compared to the controls) were indicative of a variation in the DNA sequence. Two affected family members (5657 and 5760) were screened in conjunction with two control samples (WT1 and WT2) of individuals of Mixed Ancestry. Four mobility shifts were identified and direct sequencing was performed to identify the sequence variants.

Figure 3.8 shows a shift in banding pattern in Exon 8 and subsequent direct sequencing results (Figure 3.9).
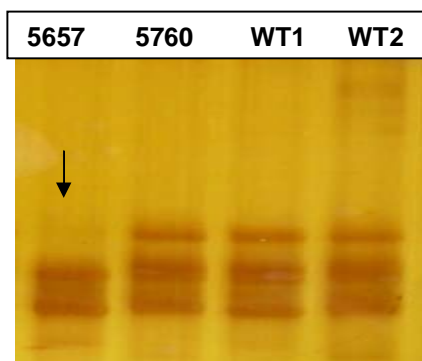
Figure 3.8: SSCP gel representing a shift in the banding pattern in Exon 8. The black arrow indicates a shift in banding pattern in sample 5657.

(A)
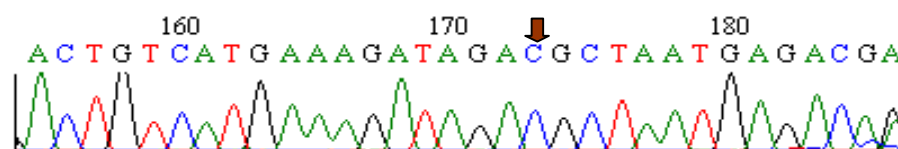
(B)

Figure 3.9: A chromatograph depicting the heterozygous C>T polymorphism in Exon 8 of the *PARK2* gene. **(A)** Represents a WT control while **(B)** represents member 5657 of Family R.

Figures 3.9(A) and (B) depicts the chromatographs produced by sequencing of the Exon 8 from proband 5657 as well as a WT control. In Figure 3.9(B) the solid red arrow indicates the position of the C>T change in individual 5657. In the figure it is clearly shown that both thymine (T) and cytosine (C) are present which was indicative of a heterozygous change. Furthermore, the C>T change has been identified as IVS8 + 48C>T, a known polymorphism in Exon 8 [Kay *et al.,* 2007].

A mobility shift was observed in Exon 10 for individual 5657 (data not shown). The subsequent sequencing result is shown (Figure 3.10).

(A)

(B)
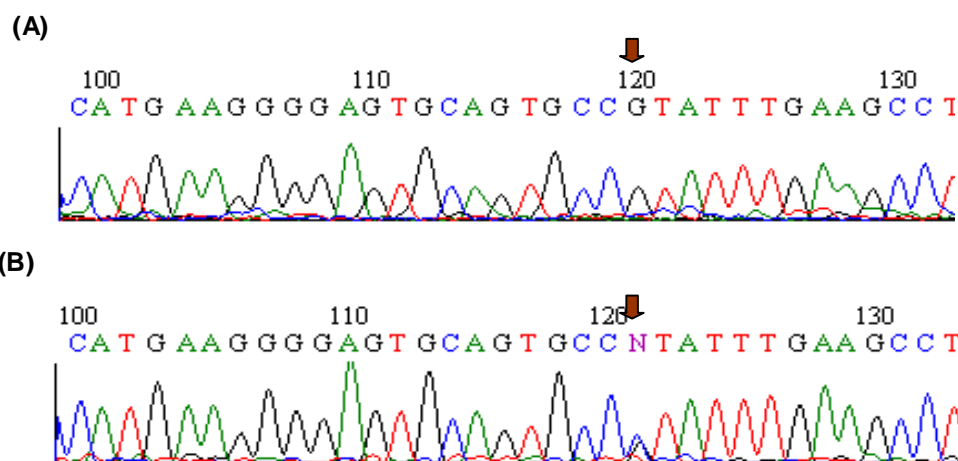
Figure 3.10: A chromatograph depicting the heterozygous G>C polymorphism in Exon 10 of the *PARK2* gene. **(A)** Represents a WT control while **(B)** represents member 5657 of Family R.

Figure 3.10 depicts the chromatographs produced by sequencing of Exon 10 from sample 5657 as well as a WT control. In Figure 3.10(B), the solid red arrow indicates the position of the G>C change in 5657. Both guanine (G) and cytosine were present which was indicative of a heterozygous change. Furthermore, this G>C change also altered the amino acid sequence by substituting a valine for leucine at position 380 and is a known polymorphism in the *PARK2* gene [Kay *et al.,* 2007].

Furthermore, a band shift was observed for individual 5657 in the SSCP analysis of Exon 11 (data not shown) and the PCR product was therefore sequenced (Figure 3.11).

(A)

(B)
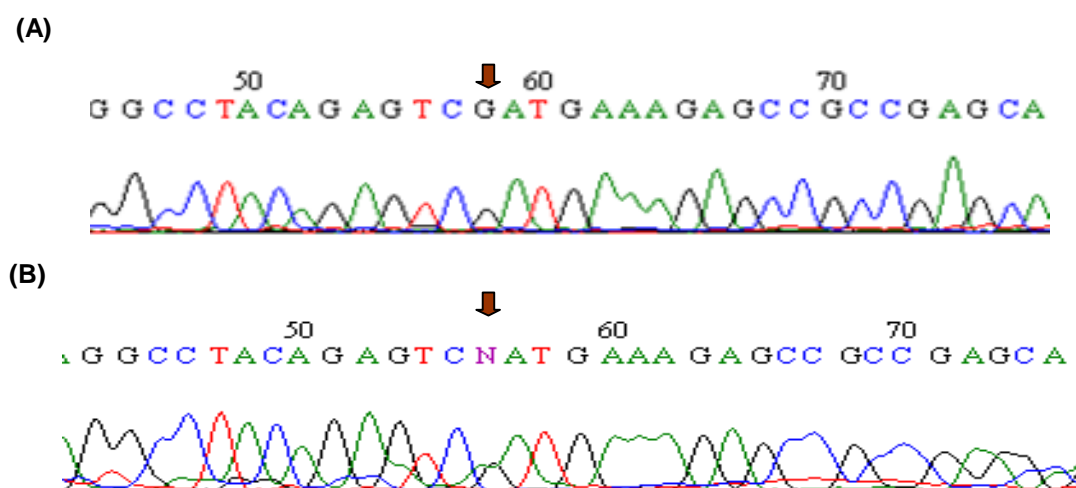
Figure 3.11: A chromatograph depicting the heterozygous G-A polymorphism in Exon 11 of the *PARK2* gene. **(A)** Represents a WT control while **(B)** represents member 5657 of Family R.

In Figure 3.11(B), the solid red arrow indicates a G>A polymorphism. As in the case of the V380L polymorphism, mentioned above, this change was also heterozygous due to the presence of both guanine and adenosine (A) residues at this position. This produced a non-synonymous change in the amino acid sequence, substituting aspartic acid for an asparagine at position 394 (D394N). This has previously been reported as a known polymorphism [Lincoln *et al.,* 2003].

A band shift was also observed for individual 5657 in the SSCP analysis of the promoter region of *PARK2* (data not shown) and the PCR product was analysed by direct sequencing (Figure 3.12).

(A)

(B)
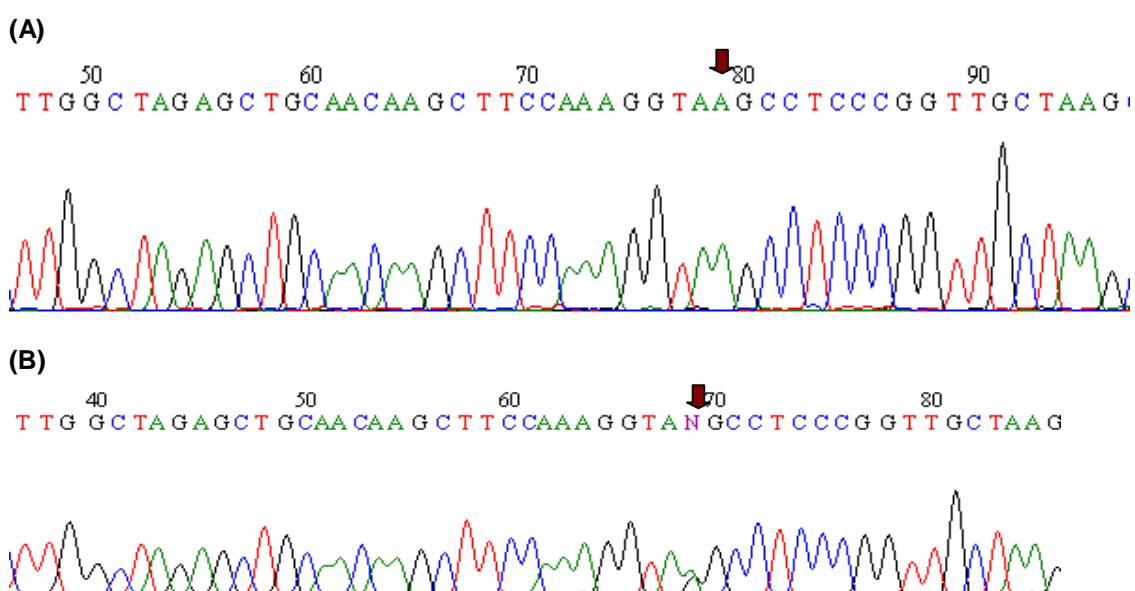
Figure 3.12: A chromatograph depicting the heterozygous A-G polymorphism in the promoter of the *PARK2* gene. **(A)** Represents a WT control while **(B)** represents member 5657 of Family R.

In Figure 3.12 (B), the solid red arrow indicates an A>G polymorphism. This change was heterozygous due to the presence of both guanine and adenosine residues at this position. This polymorphism is in the 5' UTR and is a known polymorphism [West *et al.,* 2002].

Table 3.2 summarizes all the changes observed in *PARK2* gene for family member 5657 by both SSCP analysis and automated sequencing.

Table 3.2: A summary of all the variations observed in *PARK2* gene for individual 5657.

Exon/Region	Sample number	Change in sequence	Known polymorphism	Reference:
Exon 8	5657	C>T	IVS8+48 C>T	Kay *et al.,* 2007
Exon 10	5657	G>C	V380L	Kay *et al.,* 2007
Exon 11	5657	G>A	D394N	Lincoln *et al.,* 2003
Promoter	5657	A>G	-227A>G	West *et al.,* 2002

No other changes in SSCP or sequencing were observed for the other affected family member. Furthermore, the variants found in this investigation were excluded as pathogenic because they are known polymorphisms. In addition, SSCP was proven to be a successful mutation detection method as it enabled the detection of four known polymorphisms. From this investigation it was determined that no pathogenic point mutations or small insertions or deletions were identified in the *PARK2* gene for affected Family R members.

3.4   Investigation of the locus for Familial Encephalopathy with Neuroserpin Inclusion Bodies (FENIB)

Two point mutations in the *SERPINI1* (also known as *PI12*) gene, located at position 3q26.1 have been implicated in FENIB. Patients harboring either of these mutations manifest with dementia and myoclonus, similar to the clinical features observed in Family R. For this reason, two mutations associated with FENIB (S49P and S52R) were screened in affected Family R members.

After successful PCR amplification of the template DNA, automated sequencing of the PCR product was performed in order to detect the presence of two point mutations. The sequencing results of 5657 (Figure 3.13) show the positions at which these mutations occur (indicated by red arrows).
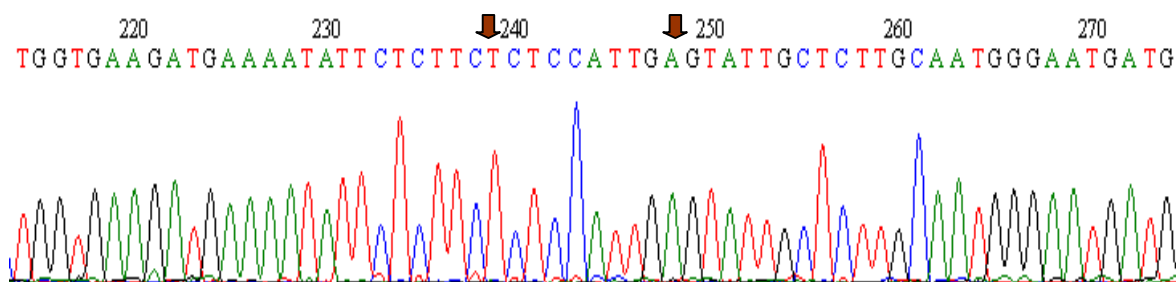
Figure 3.13: Sequencing results of 5657 obtained from automated sequencing with *PI12* forward primer. The solid red arrows indicate the positions of the mutations.

For the S49P mutation the T at position 239 on the chromatograph changes to C while the A at position 248 changes to a C for the S52R mutation (Figure 3.13, indicated by solid red arrows). From the sequencing results it could be determined that neither the S49P nor the S52R mutations were present. Therefore, the disorder in this family is not due to these two mutations in the *PI12* gene.

3.5   Investigation of the locus for Dentatorubral-Pallidoluysian Atrophy (DRPLA)

DRPLA is caused by an unstable expansion of a CAG repeat in the *Atrophin1* (*ATN1)* gene located at the chromosomal position 12p13. The number of repeats in unaffected individuals ranges from 3 to 36 and affected individuals are reported to have repeats ranging from 49 to 88 [Schöls *et al.,* 2004].

Patients harbouring the mutation causing DRPLA and affected Family R members display similar symptoms such as dementia and myoclonus. For this reason Family R members were tested for the DRPLA disease-causing expansion. In addition, an unaffected individual was typed as a negative control. No positive control for DRPLA was available at the time for analysis and comparison.

The DRPLA repeat expansion was screened using fluorescently labelled primers and electrophoresis on the ABI. The primers were designed to amplify the CAG repeat and produce a fragment of approximately 247bp (which corresponds to 14 repeats) in unaffected individuals. In the case of affected individuals, the fragment would be greater

than 356 bp, which corresponds to 49 or more repeats. The genotyping results of the control and 5657 are displayed (Figure 3.14).

(A)

(B)
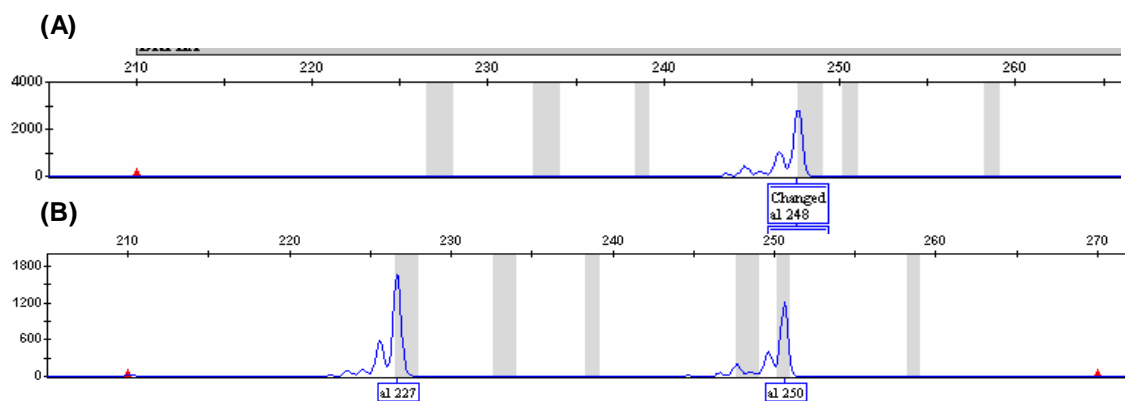
Figure 3.14: (A) The electropherogram of genotyping results of an affected Family R member (5657) and **(B)** represents the results of an unaffected individual.

Initially, DNA of only the proband (5657) of Family R was available for study and genotyping of the DRPLA CAG repeat produced a single peak in the normal range for this individual (Figure 3.14 A) . The unaffected control produced two peaks within the normal range (Figure 3.14 B). It could not be determined whether 5657 was homozygous at this locus or whether 5657 was heterozygous and the repeat expansion was possibly too large to be PCR- amplified. Later, when DNA of another family member (5760) became available, this sample was also genotyped and this showed two alleles in the unaffected range (Figure 3.15).
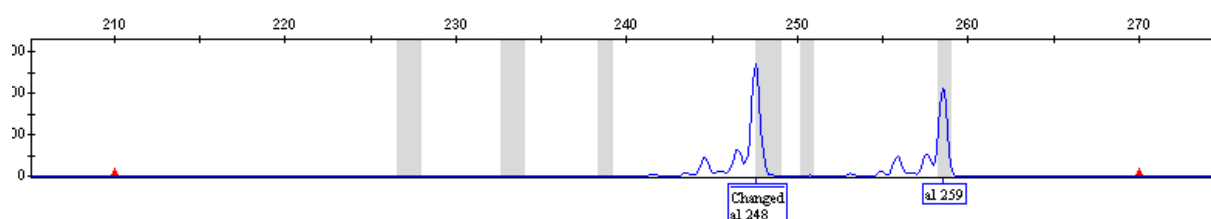
Figure 3.15: The electropherogram of genotyping results of an affected Family R member (5760).

By sizing the PCR fragments on the ABI Genetic Analyzer, the number of repeats could be calculated by applying the following formula:

$$\text{Number of repeats} = \frac{\text{(size of fragment) bp} - 205\text{bp*}}{3}$$

*where 205bp is the length of the PCR fragment excluding the CAG repeat

$$\text{E.g.: number of repeats} = \frac{248bp - 205bp}{3}$$

$$= 14.3 \text{ (rounded off to 14 repeats)}$$

Table 3.3 summarizes the size of the fragments found in two of the affected individuals and an unaffected control.

Table 3.3: The size of the fragments and number of repeats for each of the individuals typed at the DRPLA locus.

Sample	Allele 1		Allele 2	
5657	248bp	14 repeats	248bp	14 repeats
5760	248bp	14 repeats	259bp	18 repeats
Unaffected control	227bp	7 repeats	250bp	15 repeats

As the number of repeats at the DRPLA locus was in the normal range for both affected individuals it was concluded that this family did not have DRPLA

3.6   Investigation of the locus for Huntington's disease-like 2 (HDL2)

Huntington's disease-like 2 (HDL2) is caused by a triplet repeat expansion in the *Junctophilin-3* (*JPH*3) gene. This CTG/CAG repeat occurs in an alternatively spliced exon in the *JPH3* gene on chromosome 16q24.3. In the case of affected individuals the repeat is more than 40 repeats. In the case of unaffected individuals a mean of 14-16 repeats has been reported [Holmes *et al.,* 2001].

HDL2 is characterised by severe progressive dementia, motor disco-ordination, psychiatric symptoms and chorea. As many of these symptoms were present in the affected members of Family R, the locus for HDL2 was investigated.

 Fluorescently- labelled primers were designed to amplify the CAG/CTG repeat which produced an amplicon of approximately 242bp (14 repeats) in unaffected individuals and more than 320bp (more than 40 repeats) in affected individuals (Figure 3.16).

```
1261 tcagtgagag ccca ggaatc tcgtctttca gtggctgcat cgttttcacc attagttgag

1321 ggaatcgatc tgtgccttca ttctaagatg ccaccgcatt cggggcagag ccggggccgg

1381 aagccaggga gctgcctgct gctgctgctg ctgctgctgc tgctgctgct gctgctgtaa

1441 gatggtttct gtgcagggaa ccttggccgg ctctgcagct gcccgcctgc ctggactctc

1501 cgatatccac tcctcagtgc acctgacacg catggagccg gtcctttcct ggaagccaga
```

Figure 3.16: The positions and orientation of the primers for amplification of the CTG/CAG repeat on the *JPH3* gene. The sequence in yellow represents the forward primer while the sequence in grey is the sequence on which the reverse primer is based. The pink region is the CTG triplet repeat sequence.
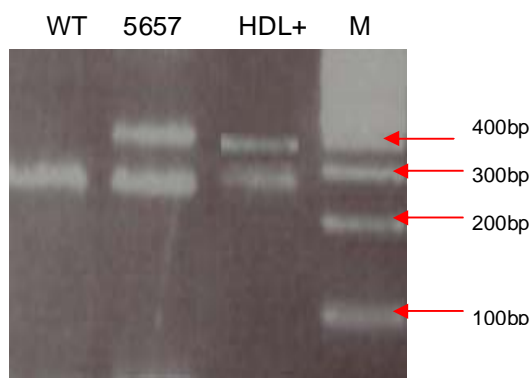
Figure 3.17: An agarose gel representing the amplification of a wild type control (unrelated Mixed Ancestry individual) (WT), a Family R member (5657) and a HDL2 positive control sample (HDL+). Lane M contains 100bp size marker and the sizes are indicated by red arrows.

DNA from an HDL2 positive individual was obtained from the National Health Laboratory Service (NHLS) Molecular Diagnostic Laboratory and used as a positive control in this investigation (Figure 3.17). The PCR product of an affected Family R member as well as an unrelated negative control is also shown (Figure 3.17).

For the HDL2 positive sample, two bands are visible indicating two different sized alleles. The bigger fragment would indicate an expanded allele and the smaller fragment would be indicative of an allele in the normal range. The same pattern is observed for the affected family member (5657), which would be indicative of a possible repeat expansion. Whether this expansion is in the disease-causing range could only be confirmed by accurately sizing the repeats by genotyping on an automated genetic analyser. The fluorescently labelled primers, designed to amplify the repeat, were used

to amplify the DNA of Family R members and genotyped on an ABI 3130xl to determine the number of repeats for each individual (Figure 3.18).
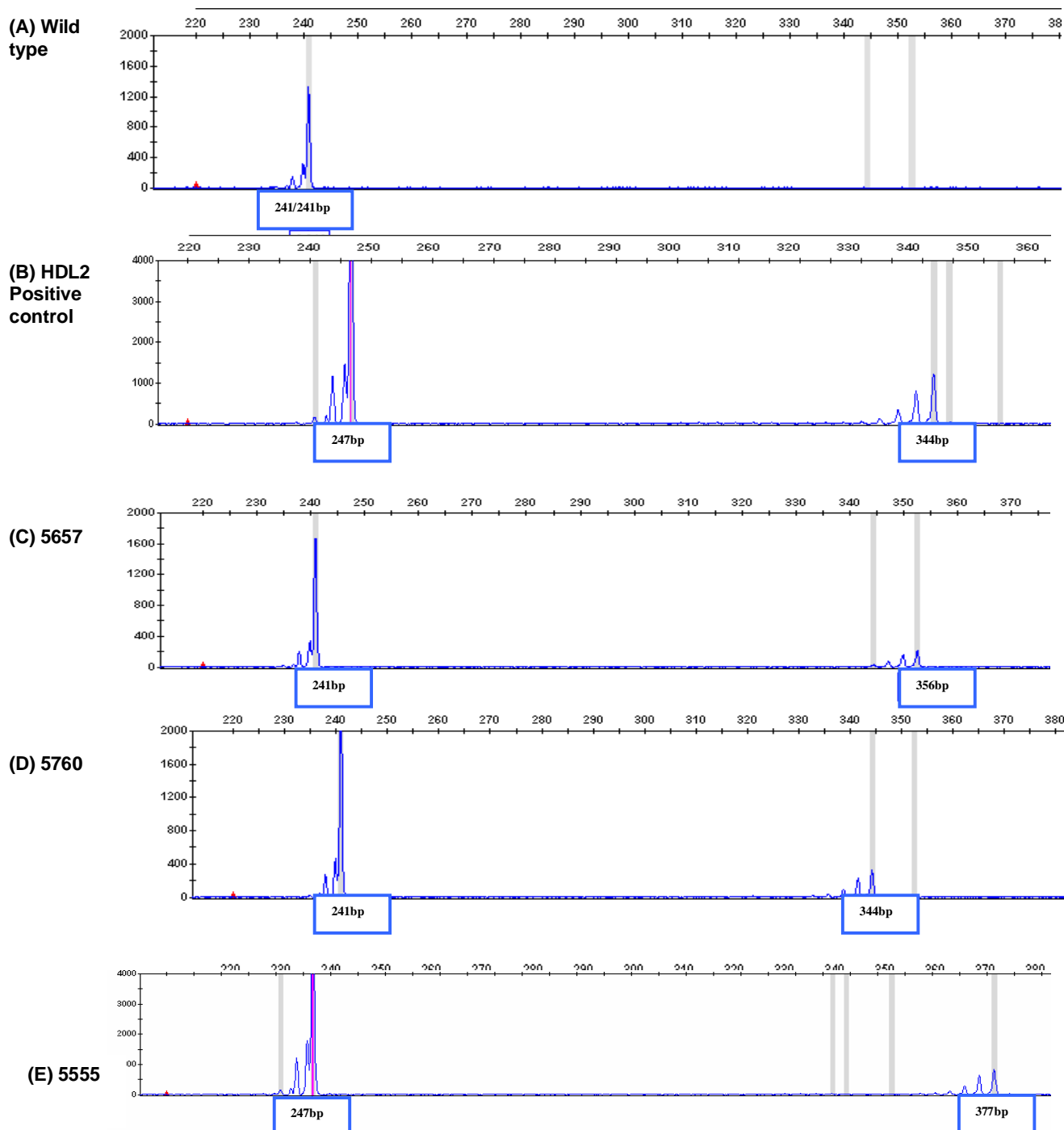
Figure 3.18: Electropherograms indicating the size of the repeats in the *JPH3* gene. (A) Unaffected WT control of Mixed Ancestry, **(B)** HDL2 positive individual, **(C)**, **(D)** and **(E)** are affected members of Family R (5657, 5760 and 5555 respectively).

From the sizing of the PCR fragments the amount of repeats was calculated using the following formula:

$$\text{Number of repeats} = \frac{\textbf{(size of PCR fragment) bp - 200bp*}}{\textbf{3}}$$

*Where 200bp is the length of the PCR product excluding the CTG repeat

$$\text{Example: Number of repeats} = \frac{\textbf{241 - 200bp}}{\textbf{3}}$$

$$= 13.6 \text{ repeats}$$

$$= \underline{14 \text{ repeats}} \text{ (rounded up)}$$

The triplet repeat ranges for Family R members was determined and has been summarised (Table 3.4).The results indicate that three Family R members, namely 5657, 5760 and 5555 had repeats within the pathogenic range. The two unaffected family members (6188 and 6341) had repeats in the normal range. It was therefore concluded that the disease segregating in this family is HDL2.

Table 3.4: A summary of repeat sizes for Family R members.

Sample number	Allele 1		Allele 2		Pathogenic (Y/N)
	Size of PCR fragments*	Number of repeats	Size of PCR fragments*	Number of repeats	
5657	242	14	356	52	Yes
5760	242	14	344	49	Yes
6188	242	14	242	14	No
6341	242	14	245	15	No
5555	248	16	377	59	Yes

* Values are rounded up

In the case of repeat expansion diseases such as HDL2, the causative repeat is usually unstable and prone to expand. Due to this the repeat length may vary between various cells in an individual. Furthermore, the electropherograms produced by the ABI were not very intense for the expanded repeats producing small peaks (Figure 3.18). This may leave room for error in determining the exact length of repeats. It is vital to establish the exact number of repeats in an individual as a difference of one repeat can determine whether an individual is in the pathogenic or intermediate range.

In order to accurately determine the number of CTG repeats, the expanded allele was cloned and sequenced. This was performed by PCR-amplification of sample 5760 and electrophoresing the PCR product on an agarose gel to separate the fragments. The larger fragment (representing the expanded allele) was excised from the gel and purified. The purified fragment was then ligated into a pGem® - T Easy vector. Three different vectors: insert ratios were tested, 1:1; 1:3 and 1:5. These ligations were used to transform competent *E coli* DH5α cells which were grown until blue and white colonies could be clearly observed (Figure 3.19).
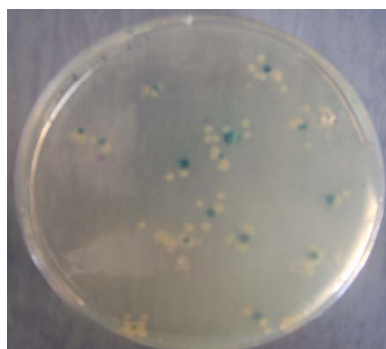
Figure 3.19: A representative LB-agar plate indicating blue and white colonies.

Table 3.5 summarises the amount of colonies produced by 200µl of transformed cells. It was noted that a vector: insert ratio of 1:1 produced no colonies. A vector: insert ratio of 1:3 produced the most colonies but was ineffective as single colonies could not be picked. The 1:5 ratio, however, produced single colonies. Individual colonies were therefore, picked from the 1:5 ratio plates and were subjected to colony PCR.

Table 3.5: Number of colonies produced per 200µl of transformed cells.

Ratio vector : insert DNA	Number of blue colonies/200µl	Number of white colonies/200µl
1:1	None	4
1:3	116	TNTC*
1:5	18	54

*TNTC: too numerous to count.

In order to determine whether the correct insert was taken up by the clone, colony PCR was performed with the HDL2 primers. Colony -PCR of several white colonies, using the HDL2 primers, did not reveal a colony with an appropriately-sized insert. As numerous blue colonies were observed on the plates, (and as the protocol recommends) colony-PCR with HDL2 primers was performed on a few of these blue colonies to determine whether any of these colonies contained the insert (Figure 3.20).
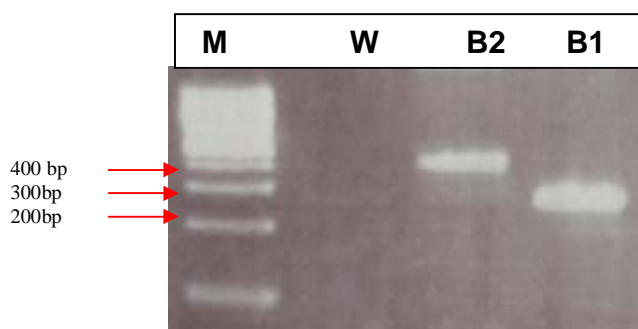
Figure 3.20: Agarose gel representing colony-PCR with HDL2 primers. The lane marked M contains a 100bp ladder with sizes indicated by red arrows. Lane W: random white colony, lane B2: blue colony 2 and lane B1: blue colony 1.

Figure 3.20 is a representative gel of the colony- PCR results for two blue colonies and a randomly selected white colony. One blue colony (B2) produced a fragment of approximately 400bp, which is the expected fragment size and another blue colony (B1) contained a fragment of roughly 250bp while the white colony did not amplify. This indicates that the blue colonies contained the insert which is possible because the pGem® - T Easy vector contains a multiple cloning site with a α-peptide coding region of the β-galactosidase gene. This region is disrupted when the insert is taken up by the plasmid. This process usually results in the formation of white colonies for clones that contain the insert and blue colonies for clones that do not have the insert. However, cases have been reported where blue colonies could also contain insert DNA [www.promega.com]. This occurs when the insert DNA is a multiple of 3bp and is then cloned in-frame to the *Lac-Z* gene thereby not disrupting its function and resulting in blue colonies containing the insert. Moreover, the 250bp fragment found in colony B1 is the smaller allele. When excising DNA from the gel, the 250bp fragment usually moves

faster through the gel but there may be traces of this fragment in the part of the gel that was excised which was expected to only contain the expanded allele.

The colony plasmid containing the 400bp fragment (which roughly corresponds to the expanded allele of 5760) was purified and subjected to direct sequencing using the HDL2 forward primer. Sequencing analysis revealed that the correct fragment was inserted into the pGem® - T Easy at the expected position, and that individual 5760's expanded allele was confirmed to be 49 CTG repeats (Figure 3.21).
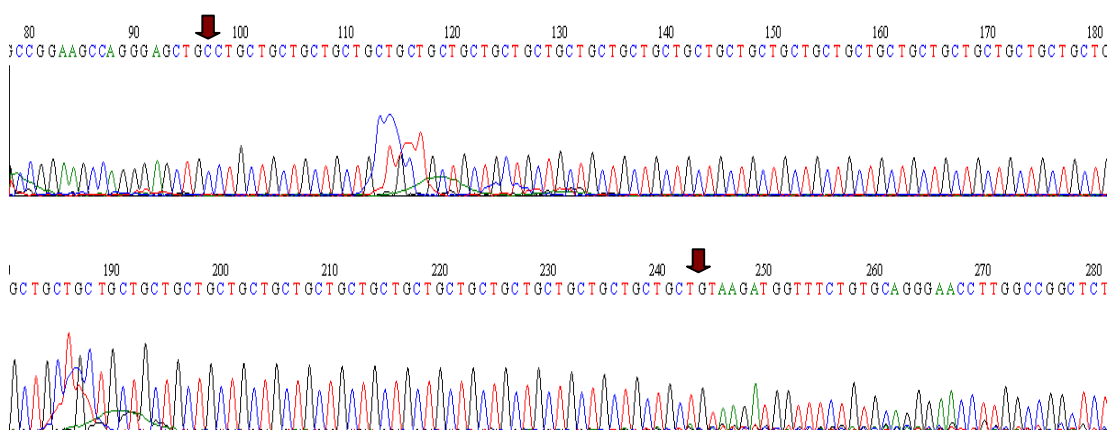
Figure 3.21: Sequencing results of the blue colony (B2) containing ~400bp insert, which corresponds to 49 CTG repeats. The beginning and end of the expanded repeat is indicated by solid red arrows.

These cloning results confirmed that the CTG repeats (Table 3.4) were correctly sized. The irregular peaks observed at positions 110-120 and 180-190 (Figure 3.21) on the chromatograph are a result of dye terminators from the sequencing kit that were not completely removed during the post-sequencing step.

3.7   Genetic ancestry testing

HDL2 has previously only been found in individuals of Black African Ancestry. In order to confirm the families' self-reported ancestry a partial genetic ancestry test was performed on one of the family members. DNA of an affected male Family R member (5760) was analysed by the MRC/NHLS/WITS Human Genomic Diversity and Disease Research Unit (HGDDRU) to determine his lineage.

The genetic analysis of the mitochondrial genome was used to determine the maternal lineage for individual 5760. The resolved haplogroup, L1d1, is associated with the Khoisan population but is also present in the Bantu speaking population. This haplogroup implies an origin in Khoisan population before the arrival of Bantu speakers in South Africa. In addition, this haplogroup was observed at a frequency of 12% in Coloured individuals, 8% in Whites and 6% in Jews of South Africa.

From Y-chromosome analysis to determine the paternal lineage, it was shown that 5760 belonged to the Y-chromosome Haplogroup FG-M213. This haplogroup is mainly found in European and Asian populations [Semino *et al.,* 2000]. Upon further analysis of the y-chromosome it was shown that the genetic profile produced for 5760 exactly matched 40 other individuals from a world wide survey of 42000 males [www.yhrd.org]. Furthermore, one of the 40 individuals was also of South African Mixed Ancestry [Soodyall, personal communication].   This gives us a strong indication that 5760 and other members of Family R can be considered as Mixed Ancestry individuals as this profile is typical of what can be found in the lineage of a South African Mixed Ancestry individual which is a maternal lineage from indigenous Africans and a partial European paternal lineage [Bardien *et al.,* 2007]. This indicates that for clinicians in South Africa, HDL2 should not only be considered in HD-like phenotype patients of Black African ancestry but in cases of Mixed Ancestry as well.

Section B: Analysis of the *Junctophilin-3* (*JPH3*) gene

3.8   Distribution of the CTG/CAG repeat lengths in South African sub-populations.

As previously mentioned HDL2 has thus far only been found in individuals classified as having African ancestry. This led to the question of whether individuals of African ancestry could possibly have larger alleles that are prone to expand into the pathogenic range, thereby predisposing them to develop HDL2.

In this part of the study, the distribution of *JPH3* CTG/CAG repeat length in the Black, Mixed Ancestry and Caucasian South African sub-population groups was investigated. A total of 51 individuals (representing 102 alleles) of each of the Mixed Ancestry and Black groups and 60 (representing 120 alleles) of the Caucasian group was genotyped.

The repeat was in Hardy Weinberg equilibrium for all three subgroups (p=1, Permutation test). However, the distribution of alleles differed amongst the three groups. 11 Different alleles were observed in the Black cohort and they ranged from 8 to 19 repeats. The Mixed Ancestry group had 9 different alleles, ranging from 8 to 21 repeats while the Caucasian cohort had only 7 different alleles ranging from 13 to 27 repeats (Figure 3.22 A and B). This indicates less diversity at this locus in the Caucasian group in comparison to the other sub-populations.

A

	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Black	6%	0%	3%	1%	3%	14%	25%	15%	12%	16%	2%	4%	0%	0%	0%	0%	0%	0%	0%	0%
Mixed Ancestry	3%	0%	1%	0%	0%	10%	33%	12%	24%	12%	4%	0%	0%	2%	0%	0%	0%	0%	0%	0%
Caucasian	0%	0%	0%	0%	0%	1%	55%	9%	25%	8%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1%

B
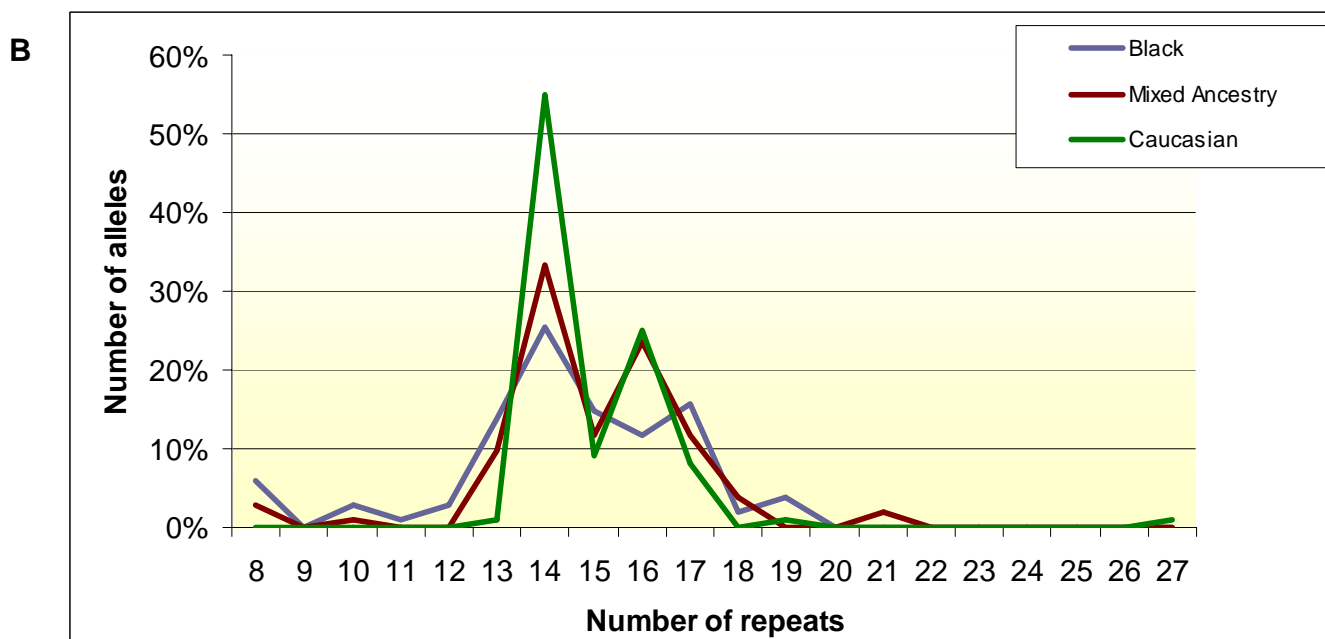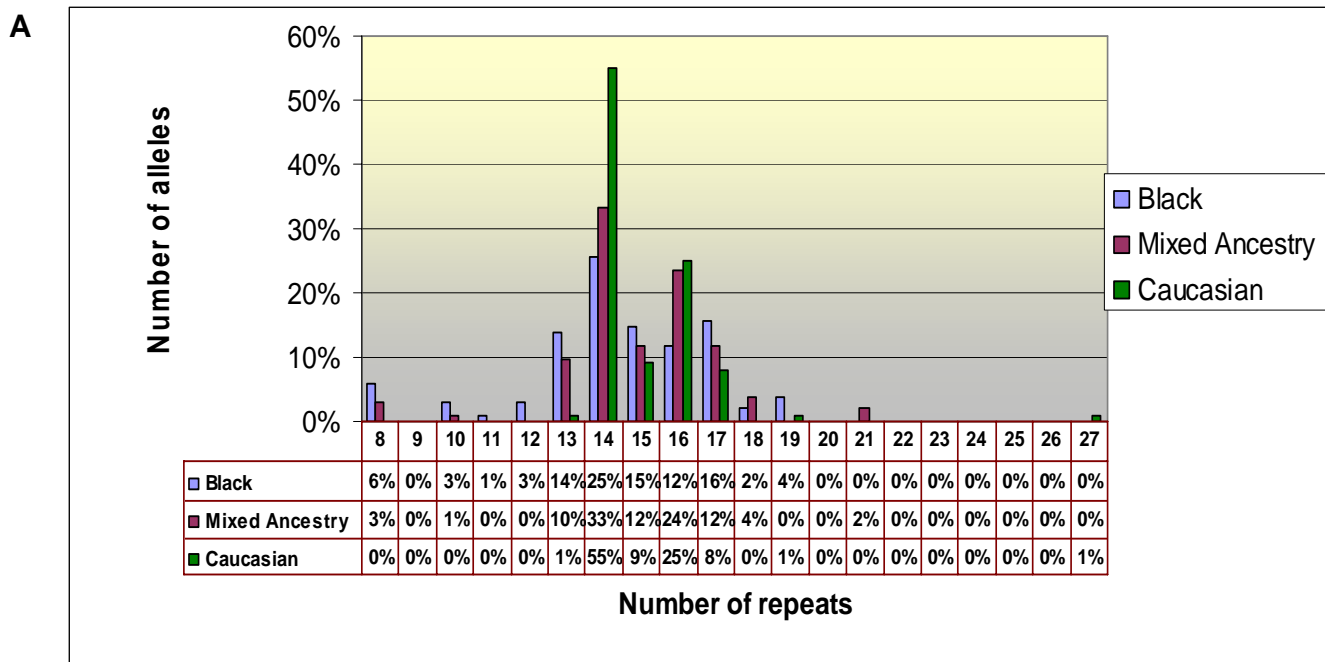
Figure 3.22: Bar graph **(A)** and line graph **(B)** displaying the frequency of alleles in South African sub-populations.

The observed heterozygosity was calculated as 88.4% for Black population, 76.4% for the Mixed Ancestry population and 62.4% for Caucasian. This further establishes that less diversity is observed in the Caucasian cohort compared to the other two groups.

Furthermore, the difference in repeat distribution in this study was statistically significant between the Caucasian and Black group (P= 0.0014) but not significant between the Black and Mixed Ancestry (P=0.3788, two-sided) nor between the Mixed Ancestry and the Caucasian groups (P=0.3186).

The most common allele (mode) for all three groups was the allele containing 14 repeats, as has been shown for other population studies in Germany, France, Montenegro and Scandinavia [Bauer *et al.,* 2002; Keckarevic *et al.,*2005; Stevanin *et al.,* 2003]. The allele containing 16 repeats was also fairly common for all three sub-populations as had been noted in previous population studies. However, the frequency of alleles containing 14 repeats differs dramatically between subgroups. Considering the distribution of the 14 repeats allele in the different population groups can provide a measure of whether this difference is significant.

Figure 3.23 depicts the distribution of alleles containing 14 repeats amongst the different population groups. This graph groups the alleles as 14/14 (homozygous pair of alleles with 14 repeats); 14/other (one allele of 14 repeats) and other/other (alleles that does not contain 14 repeats).
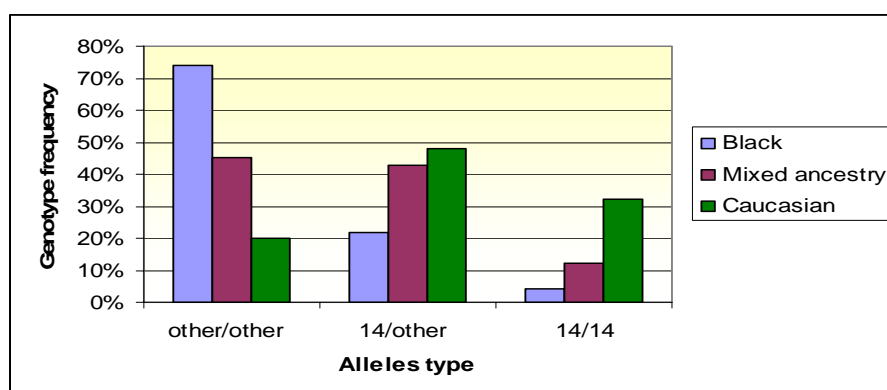
Figure 3.23: Bar graph displaying the frequency of alleles containing 14 repeats.

For the 14 repeat allele, the Caucasian cohort has the highest homozygosity (32%) and also has the highest frequency (55% from Figure: 3.22). The Black cohort, however shows the opposite with the least homozygosity (4%) and least overall frequency (25% from Figure: 3.22) of this allele. The Mixed Ancestry group falls in the middle with a frequency of 12 % for 14/14 genotype and 33% frequency in general (from Figure: 3.22). By applying the Fishers' test it was found that the distribution of the 14 repeat allele differed significantly (P-value=0.0001788) between the three different groups.

3.9   Sequence alignments of the *Junctophilin-3* gene

For many genes, similar gene organisation is observed across species [Guryev *et al.,* 2006]. This implies a functional importance for the regions of the gene that are conserved. In order to determine whether a gene or part of a gene is vital for the survival of a cell it has to further be conserved throughout species. The *JPH3* gene exists in variably spliced forms and the HDL2 pathogenic repeat is thought to be present in an alternatively spliced exon termed Exon 2A. To determine the degree of conservation between the alternatively spliced and full length transcripts around the repeat, the mRNA and protein sequences of the full length and alternatively spliced version were aligned (Figure 3.24). For the alignments "*****" indicates a single conserved residue. A "**:** "indicates a "strongly" conserved group (according to ClustalW score test) while a "**.**" indicates that a "weak" group is conserved.

(A)

```
Full length mRNA        AGAGCAAGGGGAAGTGGGTGTACAAGGGCGAGTGGACGCACGGATTCAAGGGGCGCTACG
Alternatively spliced   AGAGCAAGGGGAAGTGGGTGTACAAGGGCGAGTGGACGCACGGATTCAAGGGGCGCTACG
                        ************************************************************

Full length mRNA        GGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATACGAAGGGACCTGGAGCAACGGGCTGC
Alternatively spliced   GGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATACGAAGGGACCTGGAGCAACGGGCTGC
                        ************************************************************

Full length mRNA        AGGACGGCTACGGGACCGAGACCTACTCGGACGGAGGGACCTACCAGGGCCAGTGGGTCG
Alternatively spliced   AGGACGGCTACGGGACCGAGACCTACTCGGACGGAGAT---------GCCACCGCATTC
                        ********************************* **** * *

Full length mRNA        GTGGCATGCGCCAGGGCTACGGCGTCCGGCAGAGCGTCCCGTATGGCATGGCCGCGGTCA
Alternatively spliced   GGGGCA-GAGCCGGGGC--CGGAAGCCAG-GGAGCTGCCTGC-TGCTGCTGCTGCTGCTG
                        * **** * *** ****   ***   ** * ****  ** *  **    ** ** *

Full length mRNA        TC-CGCTCACCCCTGAGGACGTCCATCAACTCCCTGCGCAGCGAGCACACCAACGGCACG
Alternatively spliced   CTGCTGCTGCTGCTGCTGCTGTAAGATGG-TTTCTGTGCAGGGAACCT-----TGGC-CG
                             *       *  *** *  **          * *** **** ** *     *** **

Full length mRNA        GCGCTGCATCCCGACGCCTCTCCGGCGGTGGCCGGCAGCCCGGCCGTGTCCCGCGGGGGC
Alternatively spliced   GCTCTGCAGCTGCCCGCCTGCCTGG-ACTCTCCGATATCCACTCCTCAGTGCACCTGA--
                        ** ***** *       *****  * **   * ***  * **    **       * *  *

Full length mRNA        TTCGTGCTCGTGGCCCACAGTGACTCCGAGATCCTCAAGAGCAAGAAGAAGGGGCTGTTT
Alternatively spliced   ------------------------------------------------------------
```

(B)

```
Full length protein     MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWSHGFEVLGVYTWPSGNT
Alt spliced protein     MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWSHGFEVLGVYTWPSGNT
                        ************************************************************

Full length protein     YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAGNGAKYEGTWSNGLQDGYG
Alt spliced protein     YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAGNGAKYEGTWSNGLQDGYG
                        ************************************************************

Full length protein     TETYSDGGTYQGQWVGGMRQGYGVRQSVPYGMAAVIRSPLRTSINSLRSEHTNGTALHPD
Alt spliced protein     TETYSDG-------------------------------------------DATAFGAE
                        *******                                          :.**: .:

Full length protein     ASPAVAGSPAVSRGGFVLVAHSDSEILKSKKKGLFRRSLLSGLKLRKSESKSSLASQRSK
Alt spliced protein     PGPEARELPAAAAAAAAAAAAAVRWFLCREPWPALQLPACLDSPISTPQCT---------
                        ..* .    **.: .. . .* :   :*  :     ::  .  . : .:..

Full length protein     QSSFRSEAGMSTVSSTASDIHSTISLGEAEAELAVIEDDIDATTTETYVGEWKNDKRSGF
Alt spliced protein     ------------------------------------------------------------
```

Figure3.24: Alignment of human alternatively spliced and full length mRNA **(A)** and protein **(B)** around the CTG repeat (highlighted in yellow).

It was observed that the protein and mRNA sequences of the full length and alternatively spliced transcript are fully conserved for the first exon (blue sequences, Figure 3.24 A and B). As expected, the second exons (red sequences, Figure 3.24 A and B) were completely different. The second exon of the alternatively spliced transcript (Exon 2A) is a terminal exon which contains the CTG repeats coding for a polyalanine tract (highlighted in yellow). The alternatively spliced version is also much shorter and consists of only 186 amino acids (AA) while the full length transcript is 748 AA in length. As expected this data verifies that the CTG/CAG repeat is not present in the mRNA or protein sequence of the full length *JPH3* but exists in an alternatively spliced version of the gene.

The conservation of a region of a gene across species indicates the importance of that region. For instance the MORN domains of the *JPH3* gene are highly conserved in mouse and nematode [Nishi *et al.,* 2000]. The genomic sequences of human and mouse is highly conserved until the CTG repeat (Figure 3.25 A) because the repeats are interestingly not present in the mouse *JP3* gene. However, the genomic sequence of the *Pan troglodytes* (Chimpanzee) contained the repeats (Figure 3.25 B, highlighted in yellow).

In addition, the chimpanzee genomic sequence in the Ensembl [www.ensembl.org] database has a larger number of repeats repeat (23 repeats) at this position compared to the human consensus sequence (14 repeats). Despite this, the sequences upstream and downstream of the repeat are highly conserved between humans and chimpanzees (Figure.3.25 B).

(A)
```
Human GI:51511732    CCCAGGAATCTCGTCTTTCAGTGGCTGCATCGTTTTCACCATTAGTTGAG
Mouse GI:94471605    CCCAGGAATCTGGCCATTCAGTGGCTGAATCATTTTCACCTTTAGTTTAG
                     ********** * * ********** *** ******* ****** **

Human GI:51511732_>  GGAATCGATCTGTGCCTTCATTCTAAGATGCCACCGCATTCGGGGCAGAG
Mouse GI:94471605    GGAATCGCTCTGCGCCTCCAGTGTAAGATGCCACCACAGTCAGGGCAGAG
                     ****** **** **** ** * ********** ** ** *******

Human_ GI:51511732>  CCGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGCTGCTGCTGCTGCTGC
Mouse GI:94471605    CCGGTGCCGGAAGCCAGGCAACTACCAG--------------------
                     **** ************ * ** ** *

Human_ GI:51511732>  TGCTGCTGCTGCTGCTGTAAGATGGTTTCTGTGCAGGGAACCTTGGCCGG
Mouse GI:94471605    ----------------TGTGATGGTTTCTGTTCAGA-AGTCCCGGTTGG
                      * *********** *** * * ** **

Human GI:51511732_>  CTCTGCAGCTGCCCGCCTGCCTGGACTCTCCGATATCCACTCCTCAGTGC
Mouse GI:94471605    CTTGGCAGCCTTC-----------ACTCCAATGTCACCACTCCTCAGCAG
                      ************        ****     *  **********
```
(B)
```
Human genomic JPH3    CGTTTTCACCATTAGTTGAGGGAATCGATCTGTGCCTTCATTCTAAGATG
Chimp genomic JPH3    CGTTTTCACCATTAGTTGAGGGAATCGATCTGTGCCTTCATTCTAAGATG
                      **************************************************

Human genomic JPH3    CCACCGCATTCGGGGCAGAGCCGGGGCCGGAAGCCAGGGAGCTGCCTGCT
Chimp genomic JPH3    CCACCACATTCGGGGCAGAGCCGGGGCCGGAAGCCAGGGAGCTGCCTGCT
                      ***** ********************************************

Human genomic JPH3    GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG-------------
Chimp genomic JPH3    GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC
                      **********************************

Human genomic JPH3    --------------TAAGATGGTTTCTGTGCAGGGAACCTTGGCCGGCTC
Chimp genomic JPH3    TGCTGCTGCTGCTGTAAGATGGTTTCTGTGCAGGGAGCCTTGGCCGGCTC
                                    **********************  ****************

Human genomic JPH3    TGCAGCTGCCCGCCTGCCTGGACTCTCCGATATCCACTCCTCAGTGC
Chimp genomic JPH3    TGCAGCTGCCCGCCTGCCCGGACTCTCCGATATCCACTCCTCAGTGC
                      ******************  ***************************
```

Figure 3.25: (A) An alignment of human and mouse genomic sequence of the *JPH3* gene around the CTG/CAG repeat. **(B)** Alignment of human and chimpanzee genomic DNA around the repeat. The CTG repeat is highlighted in yellow.

As in humans, an alternatively spliced version of the *JPH3* gene which contained the repeat was identified in the chimpanzee but not in mice. This sequence was aligned to the alternatively spliced human sequence to determine whether the transcripts were similar (Figure 3.26 A and B).

(A)

```
chimp_alt mRNA      GGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATACGAAGGGACCTGGAGCAACGGGCTGC
Hum_alt mRNA        GGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATACGAAGGGACCTGGAGCAACGGGCTGC
                    ************************************************************


chimp_alt mRNA      AGGACGGCTACGGGACCGAGACCTACTCGGACGGAGATGCCACCACATTCGGGGCAGAGC
Hum_alt mRNA        AGGACGGCTACGGGACCGAGACCTACTCGGACGGAGATGCCACCGCATTCGGGGCAGAGC
                    ******************************************* **************


chimp_alt mRNA      CGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
Hum_alt mRNA        CGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
                    ************************************************************


chimp_alt mRNA      CTGCTGTAAGATGGTTTCTGTGCAGGGAGCCTTGGCCGGCTCTGCAGCTGCCCGCCTGCC
Hum_alt mRNA        CTGCTGTAAGATGGTTTCTGTGCAGGGAACCTTGGCCGGCTCTGCAGCTGCCCGCCTGCC
                    **************************** *******************************


chimp_alt mRNA      CGGACTCTCCGATATCCACTCCTCAGTGCACCTGACACGCACGGAGCCGGTCCTTTCCTG
Hum_alt  mRNA       TGGACTCTCCGATATCCACTCCTCAGTGCACCTGA-----------------------
                    *******************************
```

(B)

```
human_alt_prot      MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWSHGFEVLGVYTWPSGNT
chimp_alt_prot      ---------------------------------------SHGFEVLGVYTWPSGNT
                                                           ****************


human_alt_prot      YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAGNGAKYEGTWSNGLQDGYG
chimp_alt_prot      YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAGNGAKYEGTWSNGLQDGYG
                    ************************************************************


human_alt_prot      TETYSDGDATAFGAEPGPEARELPAAAAAAAAAAAAAVRWFLCREPWPALQLPACLDSPI
chimp_alt_prot      TETYSDGDATTFGAEPGPEARELPAAAAAAAAAAAAVRWFLCREPWPALQLPACPDSPI
                    *********:*********************************************  ****


human_alt_prot      STPQCT
chimp_alt_prot      STPQCT
                    ******
```

Figure 3.26: (A) Alignment of human and chimpanzee alternatively spliced *JPH3* mRNA and **(B)** protein. The repeat is highlighted in yellow.

Figure 3.26 confirms that the human and chimpanzee alternatively spliced versions of *JPH3* are well conserved across these two species. In addition the CTG repeats in the mRNA and protein are of equal lengths as opposed to the genomic sequence where the chimpanzee gene had more repeats. Furthermore, Exon1 (blue sequence, Figure 3.26 B) of the human protein is slightly longer than that of the chimpanzee.

Finally, an analysis of the protein sequence for the full length transcripts was needed to show whether the *Junctophilin-3* was conserved across species. This indicates the importance of the gene to the functioning of these organisms. The protein sequences used were the full length transcripts of human, chimpanzee and mouse *Junctophilin-3* gene (Figure 3.27). The protein sequence similarity between the human and chimpanzee sequences were 89%. The human and mouse sequences displayed 90% similarity while the mouse and chimpanzee sequences displayed 99% similarity. From these alignments it was concluded that *JPH3* full length transcript, which does not contain the pathogenic repeat is highly conserved between human, mouse and chimpanzee implying that the full length protein has an important function across diverse species.

```
Chimp_Prot    -------------------------------------------SHGFEVLGVYTWPSGNT
Human_prot    MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWSHGFEVLGVYTWPSGNT
Mouse_prot    MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWSHGFEVLGVYTWPSGNT
                                                         ****************


Chimp_Prot    YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAGNGAKYEGTWSNGLQDGYG
Human_prot    YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAGNGAKYEGTWSNGLQDGYG
Mouse_prot    YQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECTGNGAKYEGTWSNGLQDGYG
              ****************************************:*******************


Chimp_Prot    TETYSDGGTYQGQWVGGMRQGYGVRQSVPYGMAAVIRSPLRTSINSLRSEHTNGTALHPD
Human_prot    TETYSDGGTYQGQWVGGMRQGYGVRQSVPYGMAAVIRSPLRTSINSLRSEHTNGTALHPD
Mouse_prot    TETYSDGGTYQGQWVGGMRQGYGVRQSVPYGMAAVIRSPLRTSINSLRSEHTNGAALHPD
              ******************************************************:*****


Chimp_Prot    ASPAVAGSPAVSRGGFVLVAHSDSEILKSKKKGLFRRSLLSGLKLRKSESKSSLASQRSK
Human_prot    ASPAVAGSPAVSRGGFVLVAHSDSEILKSKKKGLFRRSLLSGLKLRKSESKSSLASQRSK
Mouse_prot    ASPAVAGSPAVSRGGFVLVAHSDSEILKSKKKGLFRRSLLSGLKLRKSESKSSLASQRSK
              ************************************************************


Chimp_Prot    QSSFRSEAGMSTVSSTASDIHSTISLGEAEAELAVIEDDIDATTTETYVGEWKNDKRSGF
Human_prot    QSSFRSEAGMSTVSSTASDIHSTISLGEAEAELAVIEDDIDATTTETYVGEWKNDKRSGF
Mouse_prot    QSSFRSEAGMSTVSSTASDIHSTISLGEAEAELAVIEDDIDATTTETYVGEWKNDKRSGF
              ************************************************************


Chimp_Prot    GVSQRSDGLKYEGEWASNRRHGYGCMTFPDGTKEEGKYKQNILVSGKRKNLIPLRASKIR
Human_prot    GVSQRSDGLKYEGEWASNRRHGYGCMTFPDGTKEEGKYKQNILVSGKRKNLIPLRASKIR
Mouse_prot    GVSQRSDGLKYEGEWVSNRRHGYGCMTFPDGTKEEGKYKQNVLVSGKRKNLIPLRASKIR
              **************.***********************:**.*****************
```

```
Chimp_Prot    EKVDRAVEAAERAATIAKQKAEIAASRTSHSRAKAEAALTAAQKAQEEARIARITAKEFS
Human_prot    EKVDRAVEAAERAATIAKQKAEIAASRTSHSRAKAEAALTAAQKAQEEARIARITAKEFS
Mouse_prot    EKVDRAVEAAERAATIAKQKAEIAASRTSHSRAKAEAALTAAQKAQEEARIARITAKEFS
              ************************************************************

Chimp_Prot    PSFQHRENGLEYQRPKRQTSCDDIEVLSTGTPLQQESPELYRKGTTPSDLTPDDSPLQSF
Human_prot    PSFQHRENGLEYQRPKRQTSCDDIEVLSTGTPLQQESPELYRKGTTPSDLTPDDSPLQSF
Mouse_prot    PSFQHRENGLEYQRPKHQMSCDDIEVLSTGTPLQQESPELYRKGTTPSDLTPDDSPLQSF
              ****************:* *****************************************

Chimp_Prot    PTSPTATPPPAPATRNKVAHFSRQVSVDEERGGDIQMLLEGRAGDCARSSWGEEQTGGSR
Human_prot    PTSPAATPPPAPAARNKVAHFSRQVSVDEERGGDIQMLLEGRAGDCARSSWGEEQAGGSR
Mouse_prot    PASPTSTPPPAPASRTKMAHFSRQVSVDEERSGDIQMLLEGRGGDYARNSWGEEKAGASR
              *:**:.*******:*.*:*:*************.**********.** **.*****::*.**

Chimp_Prot    GVRSGALRGGLLVDDFRTRGSGRKQPGNPKPRERRTESPPVFTWTSHHRASNHSPGGSRL
Human_prot    GVRSGALRGGLLVDDFRTRGSGRKQPGNPKPRERRTESPPVFTWTSHHRASNHSPGGSRL
Mouse_prot    GIRSGALRSGQPTEDFRTRGSGHKQPGNPKPRERRTESPTTFSWTSHHRAGNPCSGGPKL
              *:******.*   .:********:****************..*:*******.* ..**.:*

Chimp_Prot    LELQEEKLSNYRMEMKPLLRMETHPQKRRYSKGGACRGLGDDHRPEDRGFGVQRLRSKAQ
Human_prot    LELQEEKLSNYRMEMKPLLRMETHPQKRRYSKGGACRGLGDDHRPEDRGFGVQRLRSKAQ
Mouse_prot    LEPDEEQLSNYKLEMKPLLRMDACPQDTHPQRRRHSRGAGGDR-----GFGLQRLRSKSQ
              ** :**:****::*******:: **. : .:   .** *.*:     ***:******:*

Chimp_Prot    NKENFRPASSAEPAVQKLASLRLGG-AEPRLLRWDLTFSPPQKSLPVALESDEENGDELK
Human_prot    NKENFRPASSAEPAVQKLASLRLGG-AEPRLLRWDLTFSPPQKSLPVALESDEENGDELK
Mouse_prot    NKENLRPASSAEPTVQKLESLRLGDRPEPRLLRWDLTFSPPQKSLPVALESDEETGDELK
              ****:**********:**** *****. .************************.*****

Chimp_Prot    SSTGSAPILVVMVILLNIGVAILFINFFI
Human_prot    SSTGSAPILVVMVILLNIGVAILFINFFI
Mouse_prot    SSTGSAPILVVMVILLNIGVAILFINFFI
              *****************************
```

Figure 3.27: An alignment of the full length protein transcripts of chimpanzee, human and mouse *Junctophilin-3.*

3.10   Investigation of *JPH3* mRNA isoforms

JPH3 messenger RNA (mRNA) exists in multiple transcripts due to alternative splicing. The pathogenic repeat which causes HDL2 is expressed in an alternatively spliced mRNA transcript which consists of two exons. To validate the presence of this mRNA transcript, primers were design to specifically amplify the alternatively spliced transcript. In addition, primers to amplify the full length predominant transcript were also designed (Figure 3.28 A and B).

(A)

```
GTCTCCAGCGGGAGCGCGAGACGCTGGTCAGGCTCCGCGGCGCAGCTCGAAAAGGAATAA
TCGCCCCCGATTGACTGAAATTCCTCCGGAGCCGGCGCCGCGGCCGCCCGCGCCCGAGAC
CGCGCTCCGGGGCCGCGTCCTCCTCTCCTCCGGAAAACGCTCGCGACCCAGGGCCGCCGG
CGGCCGCGACTCTGCTGTGTCGATCGCCTGAGTCCGTTTTCACCGTTTGCGGGATCTGGA
ACCGAGTTACATGCATGTCCAGTGGGGGCAGGTTTAATTTTGACGACGGAGGGTCCTACT
GTGGAGGCTGGGAGGACGGCAAGGCGCACGGCCATGGCGTCTGCACCGGCCCCAAGGGCC
AAGGCGAATACACCGGCTCGTGGAGCCACGGCTTCGAGGTGCTGGGCGTCTACACCTGGC
CCAGCGGCAACACGTACCAGGGCACCTGGGCGCAGGGCAAGCGCCACGGCATCGGCCTGG
AGAGCAAGGGGAAGTGGGTGTACAAGGGCGAGTGGACGCACGGATTCAAGGGGCGCTACG
GGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATACGAAGGGACCTGGAGCAACGGGCTGC
AGGACGGCTACGGGACCGAGACCTACTCGGACGGAGGGACCTACCAGGGCCAGTGGGTCG
GTGGCATGCGCCAGGGCTACGGCGTCCGGCAGAGCGTCCCGTATGGCATGGCCGCGGTCA
TCCGCTCACCCCTGAGGACGTCCATCAACTCCCTGCGCAGCGAGCACACCAACGGCACGG
CGCTGCATCCCGACGCCTCTCCGGCGGTGGCCGGCAGCCCGGCCGTGTCCCGCGGGGGCT
TCGTGCTCGTGGCCCACAGTGACTCCGAGATCCTCAAGAGCAAGAAGAAGGGGCTGTTTC
GGCGCTCGCTGCTGAGTGGGCTGAAGCTGCGCAAGTCGGAGTCCAAGAGCAGCCTGGCCA
GCCAACGCAGCAAGCAGAGCTCCTTTCGCAGCGAGGCGGGCATGAGCACCGTCAGCTCCA
CGGCCAGCGACATCCACTCCACCATCAGCCTGGGCGAGGCTGAGGCCGAGCTGGCGGTCA
TCGAGGACGACATCGACGCCACCACCACCGAGACCTACGTGGGCGAGTGGAAGAACGACA
AACGCTCCGGCTTCGGCGTGAGCCAGCGCTCGGACGGGCTCAAGTACGAGGGCGAGTGGG
CCAGCAACCGGCGCCATGGCTACGGCTGCATGACCTTCCCGGACGGCACCAAGGAGGAGG
GCAAGTACAAGCAGAACATCCTCGTCGGCGGCAAGCGCAAGAACCTCATCCCCCTGCGGG
CCAGCAAGATCCGCGAGAAGGTGGACCGCGCCGTTGAGGCCGCTGAGCGGGCCGCCACCA
TCGCCAAGCAGAAGGCTGAGATCGCGGCTTCCAGGACCTCCCACTCTCGGGCAAAGGCCG
AGGCAGCCCTCACAGCAGCTCAGAAAGCCCAGGAGGAGGCGCGGATCGCCAGGATCACTG
CCAAAGAGTTCTCCCCTTCCTTCCAGCACCGGGAAAACGGGCTGGAGTACCAGAGGCCGA
AGCGTCAGACCTCCTGTGACGACATCGAGGTGCTGTCCACCGGGACACCCCTGCAGCAGG
AGAGCCCCGAGCTGTACCGCAAGGGCACCACTCCCTCCGACCTGACCCCCGACGACAGCC
CCCTGCAGAGCTTCCCCACCAGCCCCGCGGCCACCCCGCCGCCCGCGCCCGCGCCAGGA
ACAAGGTCGCCCACTTCTCGAGGCAGGTGTCGGTGGACGAGGAGCGGGGCGGGGACATCC
AGATGCTCCTGGAGGGCCGGGCCGGGGACTGCGCCCGCAGCAGCTGGGGCGAGGAGCAGG
CCGGGGGCTCCAGGGGTGTCCGCAGCGGTGCCCTGCGCGGCGGCCTGCTCGTGGATGACT
TCCGCACCCGAGGTTCGGGCCGCAAGCAGCCCGGGAACCCCAAGCCGCGGGAGCGGCGGA
CGGAGTCACCCCCCGTGTTCACGTGGACTTCCCACCACCGGGCCAGCAACCACAGCCCCG
GAGGCTCCAGGCTGCTGGAGCTGCAGGAGGAGAAGCTGAGCAACTACCGGATGGAGATGA
AACCCTTGCTGAGGATGGAGACGCATCCCCAGAAAAGACGCTACAGCAAGGGCGGCGCCT
GCCGGGGCTTGGGGGACGACCACCGCCCCGAGGACCGGGGCTTCGGGGTGCAGAGACTGC
GGTCCAAGGCCCAGAACAAGGAGAACTTCAGGCCGGCCTCCTCCGCGGAGCCCGCCGTGC
AGAAACTGGCGAGCCTGCGGCTGGGCGGGGCCGAGCCCCGGTTGCTGCGTTGGGACTTGA
CCTTCTCCCCGCCCCAGAAATCCTTGCCTGTCGCTCTAGAGTCCGACGAGGAGAATGGGG
ATGAGCTCAAGTCCAGTACGGGCTCAGCGCCTATCCTGGTGGTCATGGTGATCTTGCTCA
ACATCGGAGTCGCCATTCTGTTTATTAACTTTTTCATCTGATGAGATGTCGCGGTAGCAA
AAATAGAGAAAGGGTAGAAAAAAGGGACATTAAAATTAAAAGCAAAACCACAAGAAGGGA
AAGACCGCAACTCGGACAGCCCAGCGACTTCCAAGTCCTCTCACAGAAGAACCACACGAT
TGGGTATCACTCACAGTTTGCCTTTTTTTTCTGGGTAATGTTTTTTGGATTTTAGCCAAAA
TTCTTTGCTTGTATAACACTCTGCTGTGTGGCATGGCAGAAGGAGGCCAGCACGCAGCCC
CTCCAGCTCCACGTGGAGACAGAAGGGATCCCGGCACATCAGTGGTAACAGCGGACGTTG
TCCTCGTGGTCACACGTCCCGTCTTGGGTGTGGATGGAGGGCAGCCCGGGGCAGAGCCTC
AGCCCCGCGGCCCCTGAGTGGCAGGGCTGACTCCCGTCGACACGAGCTTAGAAAGTGGAT
TCACTGCTTTCTCTGTCTAGAACAGACGGGTGACAAGTATGGGCAGGAGGCATGGGGCAG
GGTGGCCCACCCCAGTGGGCAGTAGCCTGGCCTTTTTCTGTGTGAGATCTGTGCTGCACA
CCTGAGGGAGGGGGAGGGATCGGCCACCTCCTCCCTGTGAGACGGATGCAGGTCCTTCCC
```

```
TCTTCTCGGCACTGCCCCCGGCCTTCCATGAGAAGCCGACTCCCCACACCGAGTTTTAAA
GCAAAGCCCTTTTCTTCTGCTGCCCACTCACTGTGGGTCCCATTCGGCTGTTTCCCCCAC
CAGACCCCAGGGAAGCCGGGGCCCACTCCGATCCGCCTGGGCTCAGCTAAGCACGGAAGC
CAAGGGGGCTGTGCCGTGGAGCTGGGCTCGCGCCGGGGCTCTGGGTGTGTGCGCTTGGCG
TGCAGGGTGGACGCGTGGGGTTCCGTGTCCCCAGCAGTGAGGGCCCTAGAGGACGCCTTC
TCCCATGGTTACTGATCTCCACGGGTTTTCACATCTCTGTACTGTGCCTGCCTCAACTTC
CCCTAACAGATATGCATATTCCTTCCAGATGCCTCAGTGCTACACCACAGTGGGCCTGGT
CCCAGGACAGGAATGCGGTTCAAACCCAGTGGCTTGAAACTTCCTGAGAAACTGTAGCAT
ATCCAGCCCCCTAAAATGTACAATGTAACTTGTTCAGTCCAACAAAAACAGGTTCCTTAT
GTTTCTGCCTTCTCCACCAGGGTCGCTCCATCACCCAAACAAAAGAACAAGGTTTGCCAG
GATGTCCGAGTGCCCCCTGGCCCTGGCTCTCGTGTGCATGGACGTGCCTGAGGGGTCCGG
GCACGGCCATACGCAGGACCCCTGTGCCCGGGGAGGCGCTGCAGGGATTCCCCATCCGGT
CGTCTTGGGGCCAGCCCGTCTTATGGACTCTGCCTTGCTTTGCTTATGTTTAGCTGTTTC
TCTGCTACCTTTCGAGCAGACTTCTTTACTACACTGCACTGGATTGCTATATTTTTAACC
AGAAATAAACTAAAGATTAGAGCATGTTCCAGTTAAA
```

(B)
```
GTCTCCAGCGGGAGCGCGAGACGCTGGTCAGGCTCCGCGGCGCAGCTCGAAAAGGAATAA
TCGCCCCCGATTGACTGAAATTCCTCCGGAGCCGGCGCCGCGGCCGCCCGCGCCCGAGAC
CGCGCTCCGGGGCCGCGTCCTCCTCTCCTCCGGAAAACGCTCGCGACCCAGGGCCGCCGG
CGGCCGCGACTCTGCTGTGTCGATCGCCTGAGTCCGTTTTCACCGTTTGCGGGATCTGGA
ACCGAGTTACATGCATGTCCAGTGGGGGCAGGTTTAATTTTGACGACGGAGGGTCCTACT
GTGGAGGCTGGGAGGACGGCAAGGCGCACGGCCATGGCGTCTGCACCGGCCCCAAGGGCC
AAGGCGAATACACCGGCTCGTGGAGCCACGGCTTCGAGGTGCTGGGCGTCTACACCTGGC
CCAGCGGCAACACGTACCAGGGCACCTGGGCGCAGGGCAAGCGCCACGGCATCGGCCTGG
AGAGCAAGGGGAAGTGGGTGTACAAGGGCGAGTGGACGCACGGATTCAAGGGGCGCTACG
GGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATACGAAGGGACCTGGAGCAACGGGCTGC
AGGACGGCTACGGGACCGAGAC`CTACTCGGACGGAG ATGC`CACCGCATTCGGGGCAGAGC
CGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CTGCTGTAAGATGGTTTCTGTGCAGGGAACCTTGGCCGGCTCTGCAGCTGCCCGCCTGCC
TGG`ACTCTCCGATATCCACTCCT`CAGTGCACCTGA
```

Figure 3.28: The positions of the primers designed to amplify the full length **(A)** and alternatively spliced **(B)** *JPH3* transcripts are indicated by highlighted regions. Grey: full length set A, pink: full length set B and yellow: alternatively spliced. The different exons are indicated by the alternating colours.

As a control a third set of primes were designed to amplify a housekeeping gene, namely *β-Haemoglobin* (*HBB*) gene, which was used to test the viability and quality of cDNA. The *HBB* gene is expressed in many tissues at relatively similar rates and is highly conserved, thereby making it a suitable housekeeping gene [Shmueli *et al.,* 2003]. The primers were designed spanning a small intron thereby producing differently sized fragments on genomic DNA and on cDNA, which was useful in controlling for genomic contamination which is a major source of contamination in cDNA (Section 2, Figure 2.4). The primers for *HBB* produced a product of 377bp in genomic DNA and 248bp in cDNA (Figure 3.29).
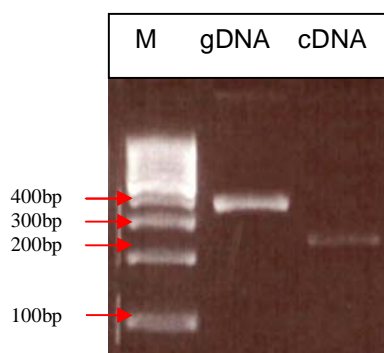
Figure 3.29: Agarose gel depicting the size difference in fragments produced by *HBB* primers for genomic DNA (gDNA) and cDNA.

The *HBB* primers were used in all subsequent PCR reactions as a control to firstly indicate that there was no genomic DNA contaminants in the cDNA samples and secondly that the cDNA was of good quality.

3.10.1   Purchased foetal brain cDNA as template

The *HBB, JPH3* full length and alternatively spliced *JPH3* primers were tested on cDNA generated from foetal brain RNA which was used as a template because *JPH3* is expressed almost exclusively in the brain. The subsequent PCR reactions of foetal brain cDNA using primers for the *HBB* gene are shown (Figure 3.30).
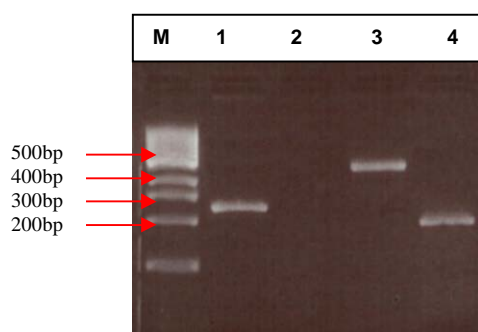
Figure 3.30: Agarose gel depicting PCR products using foetal brain cDNA as template. Lane M contains 100bp marker. Lane 1: amplification with *HBB* primers, lane 2: amplification with *JPH3* full length primers (set A), lane 3: amplification with *JPH3* full length primers (set B), lane 4: amplification with *JPH3* alternatively spliced primers.

Several attempts to optimise the *JPH3* full length set A primers PCR were unsuccessful. Interestingly, the PCR product obtained using primers for the alternatively spliced isoform was correct and verified by sequencing even though, according to Margolis *et al.,* [2006], the alternatively spliced version of the transcript is expressed at lower levels than the full length transcript (Lane 4, Figure 3.30). The full length primers were then redesigned across Exons 4 and 5 of the *JPH3* gene which is closer to the 3' end of the transcript because it was possible that the cDNA did not extend all the way to the 5' end during the generation of the cDNA. The amplification with the redesigned primers (Set B, Lane 3), *HBB* primers (Lane 1) as well as the alternatively spliced primers (Lane 4) was successful for foetal brain cDNA (Lane 3) whereas the full length (Set A) primers were not (Lane 2, Figure 3.30).

These PCR products were subjected to direct sequencing to confirm that the correct fragments were amplified and also to determine the number of repeats in the alternatively spliced product.

(A)
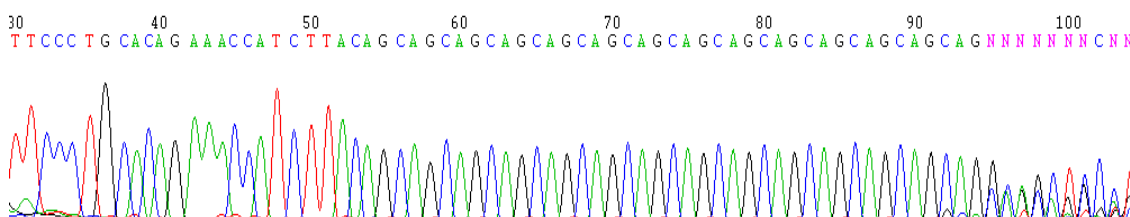
(B)
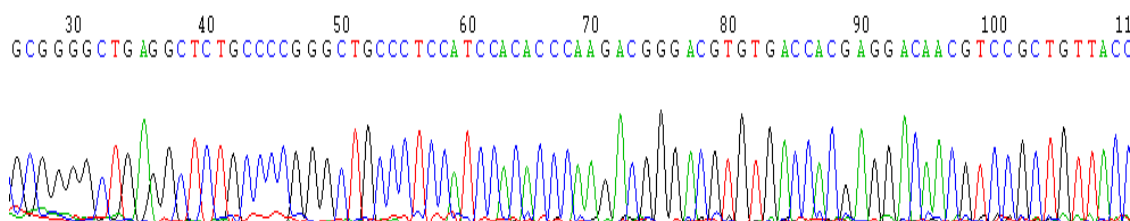
Figure 3.31: (A) Chromatographs produced from the direct sequencing of foetal brain cDNA with *JPH3* alternatively spliced reverse and **(B)** *JPH3* full length set B primers.

The foetal brain RNA was isolated from a pooled sample of foetuses and produced more than one transcript. Due to this the CTG repeat sizes of the different foetuses would differ and cause "messy" sequencing (Figure 3.31 A). The sequencing products were searched against the NCBI database using the BLAST search tool and were shown to

match the reference sequences for *JPH3* full length and alternatively spliced mRNA transcripts.

3.10.2 Patient lymphocyte cDNA as template

An exploratory study to analyse the *JPH3* isoforms in the affected family members was attempted. In this investigation RNA was isolated from transformed lymphocytes of two affected family members. The lymphocytes served as a substitute for brain tissue which was not available at the time of this study. Lymphocytes were used as a surrogate for brain tissue since, due to the process of illegitimate transcription, all forms of transcripts of every gene is expressed as mRNA at a basal level [Chelly *et al.,* 1989].

The Experion™ RNA StdSens Analysis Kit was used to determine the quality and quantity of the extracted RNA. Samples of the RNA from lymphocytes of family members 5657 and 5760 were run on Experion™ automated electrophoresis system which separated and analysed the sample. Figure 3.32 represents the gel matrix with the samples that were applied to it.
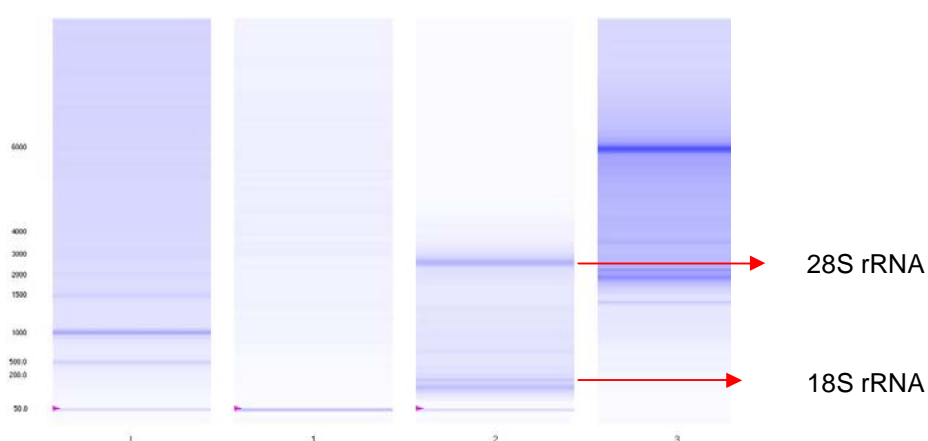
Figure 3.32: Electrophoresis of RNA isolates using the Experion™ automated electrophoresis system.

The lane marked L contains an RNA ladder with the sizes (in bp) along the Y axis. Lane 3 contains RNA-free water as a negative control. Lanes 1 and 2 contain the RNA isolated from 5657 and 5760 respectively. The red arrows indicate the 18S and 28S

ribosomal RNA (rRNA) of sample 5760. Figure 3.33 shows the graphical output of lane 2 as generated by the Experion™ system.
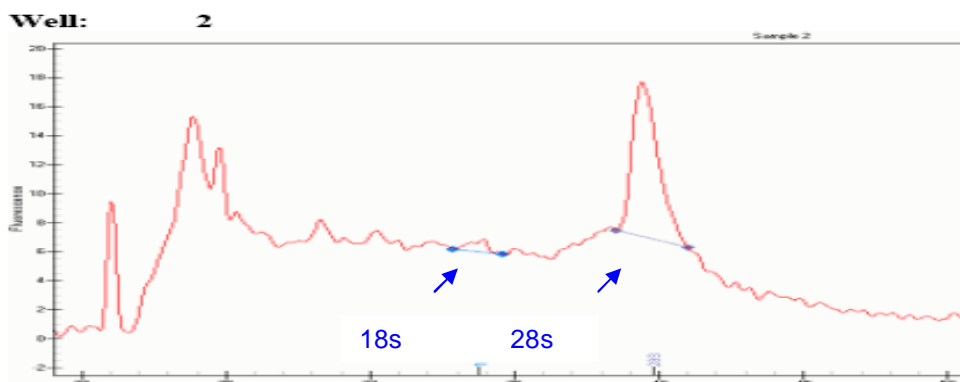
Figure 3.33: Experion™ graphical output of lane 2.

The graph (Figure 3.33) showed that RNA was isolated from 5760 and the peak for 28s was clearly defined. cDNA was generated from the isolated RNA of 5760 and used as template for amplification using the *HBB* primers as a control to determine whether the cDNA was viable and could be amplified. A fragment of the appropriate size was observed for the *HBB* primers (lane 2), which indicated that the cDNA was uncontaminated and was of good enough quality to be used in a PCR reaction. The cDNA was then used as template for amplification with the primers listed in Table 2.7 (Page 44). Subsequent PCR reactions with either full length *JPH3* (Lane 3 and 4) or alternatively spliced primers (Lane 5) were unsuccessful even after extensive attempts at optimising the reaction (Figure 3.34).
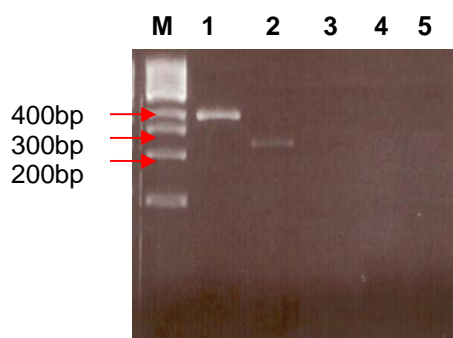
Figure 3.34: Gel electrophoresis of PCR products using lymphocyte cDNA of affected HDL2 patients**.** Lane M contains 100bp marker, lane 1: *HBB* primers on genomic DNA (control), lane 2: *HBB* primers on lymphocyte cDNA, lane 3: full length primers (set A), lane 4: full length primers (set B) and lane 5: alternatively spliced primers**.**

3.10.3 Real-time PCR

To determine whether the patient's lymphocyte cDNA did not amplify with the *JPH3* primers because the transcripts are not being illegitimately transcribed or was just at too low levels to be detected, real-time PCR on a LightCycler® was attempted. Real-time PCR is a more sensitive PCR method which allows one to detect extremely low levels of template in a PCR reaction. Real-time PCR measures the amount of fluorescence generated by SYBR green as it is incorporated into the amplified PCR product.

HBB primers

Initially the LightCycler® reactions were performed using serial dilutions of the genomic DNA (1:1, 1:10 and 1:100) of the genomic DNA, using *HBB* primers.

The genomic DNA amplification was successful with the crossing point (CP) of the reactions between 20-28 cycles. The optimal reaction occurred with the 1:1 template DNA. The melt point analysis produced peaks at 87ºC for all three dilutions indicating that a single product was produced (Figure 3.35). The negative control produced a CP value of 33 and had a different melt-curve to the genomic DNA. The agarose gel (Figure 3.35 C) shows the respective sizes of the bands. The faint, small fragments visible in all the lanes are likely due to the formation of primer dimers.
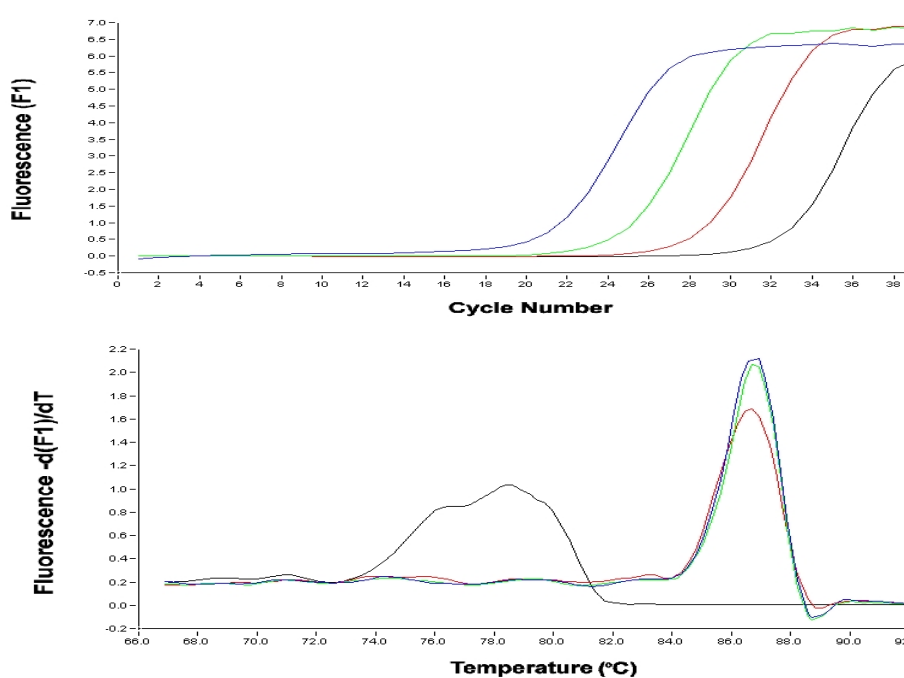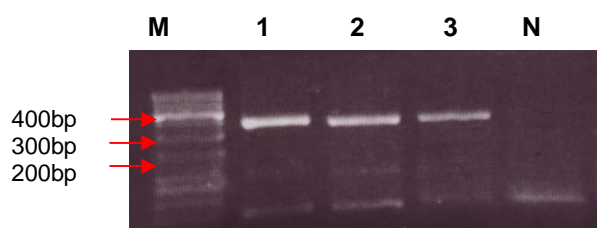
Figure 3.35: (A) The crossing point curve and **(B)** melt curve of genomic DNA (1:1 blue, 1:10 green, 1:100 red and negative control black) amplified with *HBB* primers.**(C)** Agarose gel indicating the size of the PCR products. Lane M contains 100bp marker, lane 1: *HBB* primers on genomic DNA (1:1), lane 2: *HBB* primers on genomic DNA (1:10), lane 3: *HBB* primers on genomic DNA (1:100) and lane N: negative control.

Real-time PCR using *HBB* primers with foetal brain and lymphocyte cDNA as template also produced the correct product which was verified by sequencing (Figure 3.36 A and B).

(A)

(B)
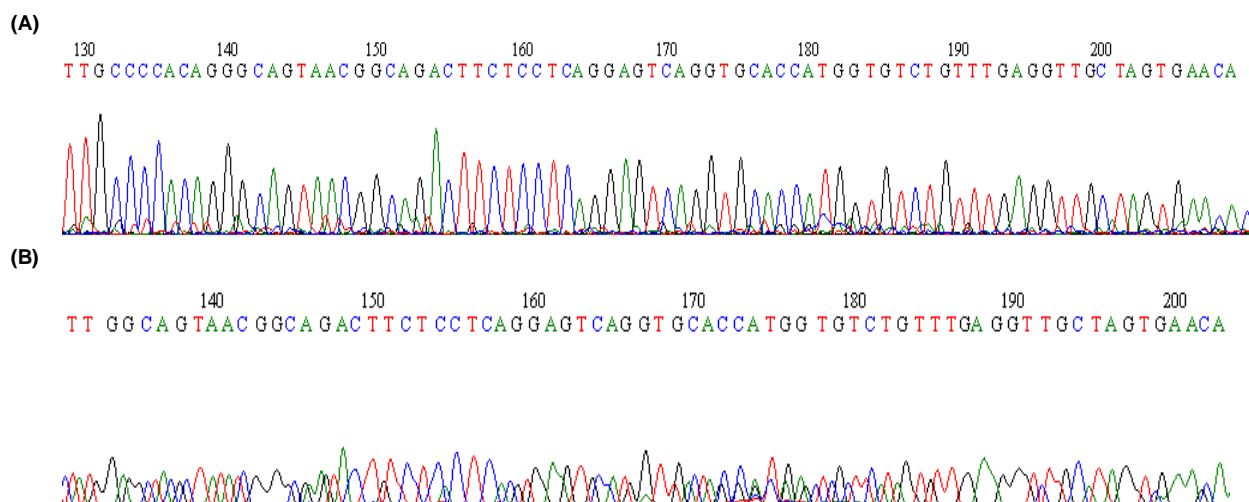
Figure 3.36: Chromatograph depicting the partial sequence obtained from the direct sequencing of **(A)** lymphocyte cDNA and **(B)** foetal brain cDNA amplified with primers for the *HBB* gene.

Based on these promising results, real-time PCR was attempted using primers for the *JPH3* full length and alternatively spliced isoforms.

Full-length and alternatively spliced primers

Using the optimal dilutions from the *HBB* amplification, the amplification of foetal brain and lymphocyte cDNA was attempted. The crossing points of many of the runs were more than 30 cycles and are therefore not optimal. However some runs have been analysed (Figures 3.37 and 3.38).

(A)

(B)
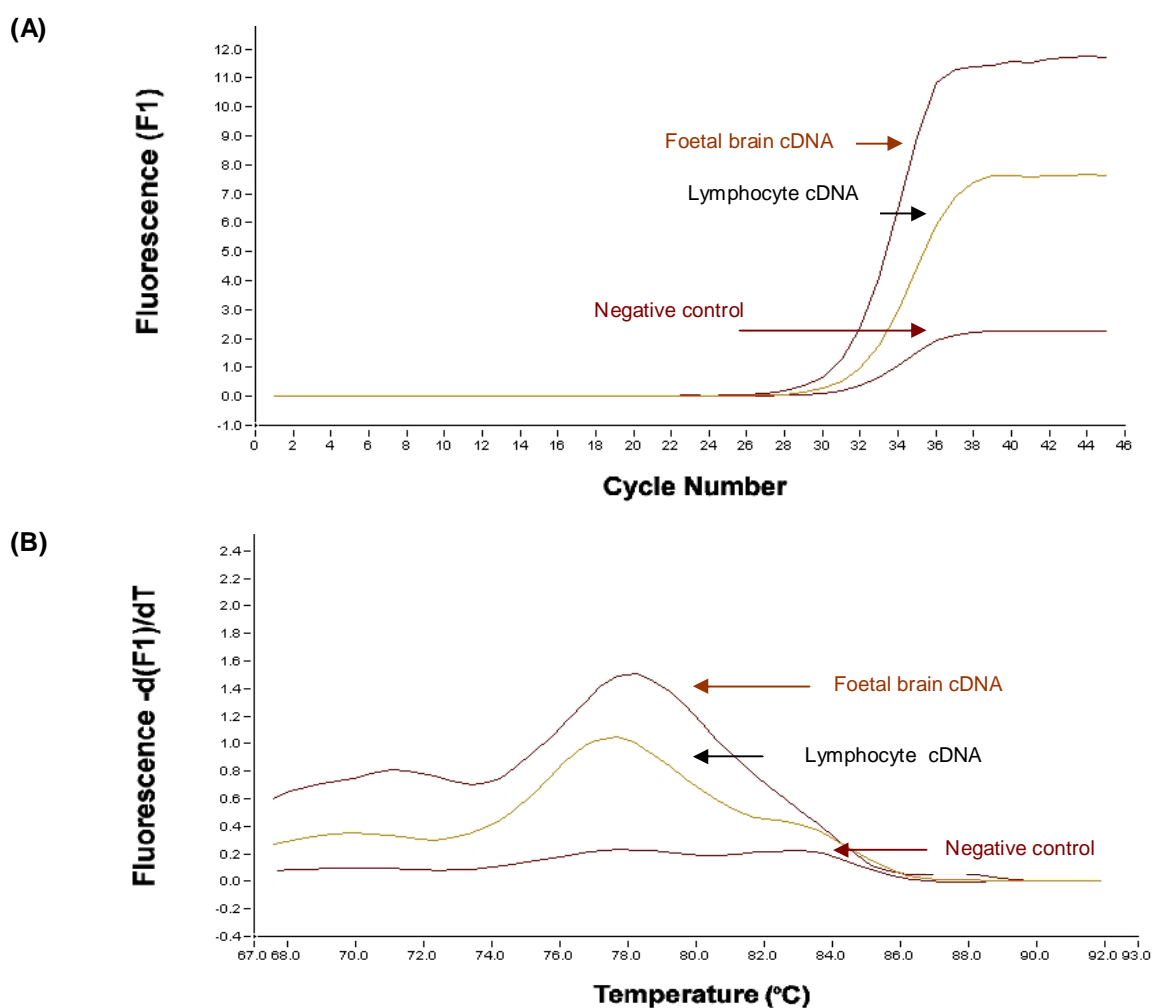
Figure 3.37: Amplification with *JPH3* full length set B **(A)** crossing point curve and **(B)** melt curve analysis of lymphocyte cDNA, foetal brain cDNA and a negative control.

(A)

(B)
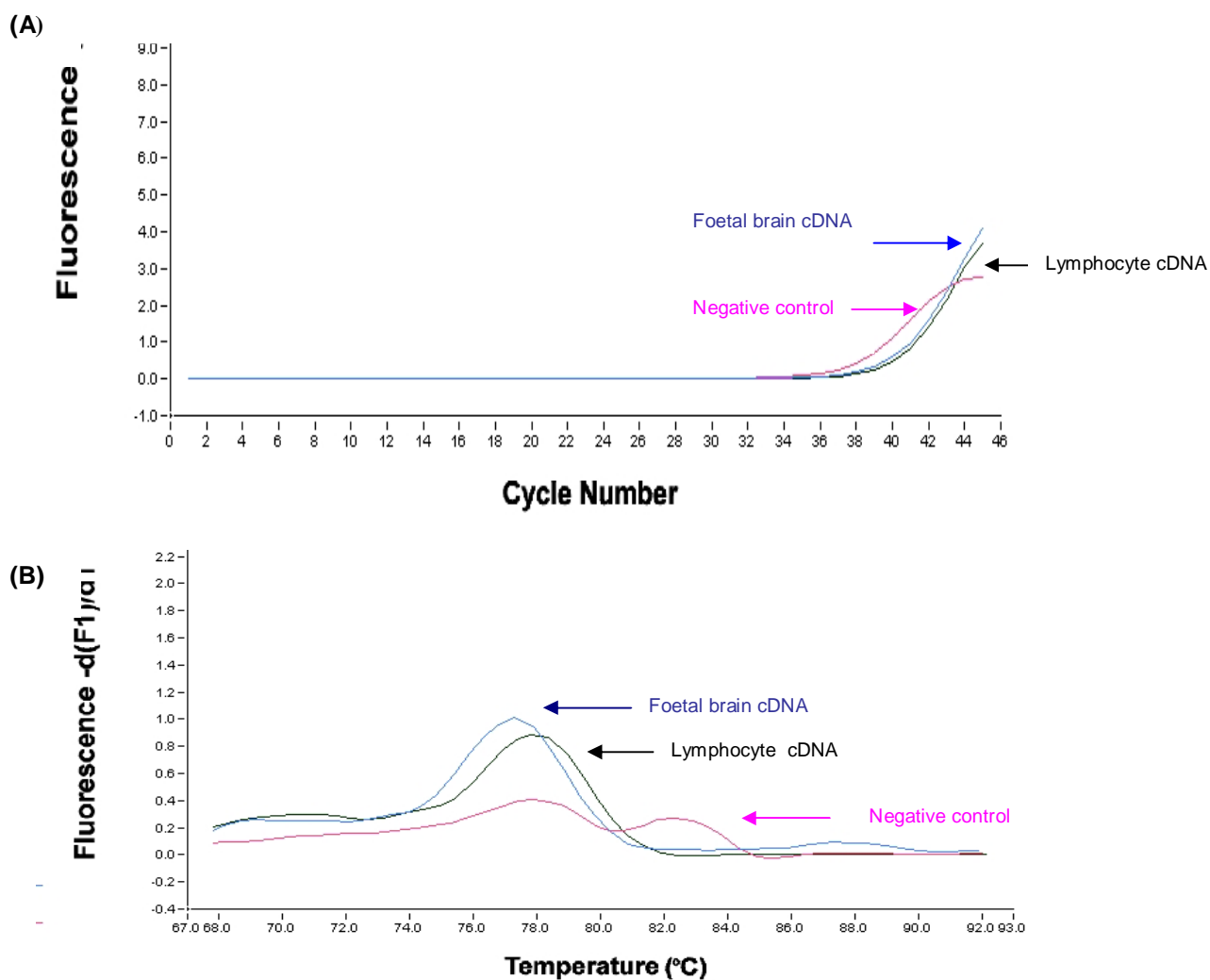
Figure 3.38: Amplification with *JPH3* alternatively spliced primers **(A)** crossing point curve and **(B)** melt curve analysis of lymphocyte cDNA, foetal brain cDNA and a negative control.

 From the graphs (Figures 3.37 and 3.38) it appeared that the negative controls and the cDNA templates have similar CP values. In addition, the melt curve analysis for the negative control does not differ from the cDNA templates. These results indicate that the exploratory real-time PCR reactions were unsuccessful and require further optimisation. At this stage we are unable to determine whether the *JPH3* transcripts are present in lymphocytes.

3. 11 <u>Expressed Sequence Tags alignments</u>

To provide independent confirmation of the existence of the alternatively spliced transcript primers were designed to specifically amplify this repeat in lymphocyte RNA of the affected family and from commercial foetal brain cDNA. Amplification from lymphocyte cDNA was unsuccessful. However, the foetal brain cDNA produced strong fragments of the expected size. This product was then sequenced and searched against ESTs in the dbEST database.

Significant alignment was shown for 17 ESTs (Figure 3.39) however, only three (highlighted in green) of these ESTs contained Exon 1 spliced to an Exon 2A which contains the repeat.

```
Sequences producing significant alignments:               (Bits)  E Value

gb|BI553267.1|   603193587F1 NIH_MGC_95 Homo sapiens cDNA clone...  1175   0.0
gb|BI550632.1|   603195805F1 NIH_MGC_95 Homo sapiens cDNA clone...  1138   0.0
gb|BM548387.1|   AGENCOURT_6573271 NIH_MGC_124 Homo sapiens cDN...   717   0.0
gb|AA912459.1|   om52c01.s1 NCI_CGAP_GC4 Homo sapiens cDNA clon...   560   5e-157
gb|BE042890.1|   ho30c07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clo...   527   5e-147
dbj|DA404387.1|   DA404387 BRTHA3 Homo sapiens cDNA clone BRTHA...   523   7e-146
dbj|DA189798.1|   DA189798 BRAMY3 Homo sapiens cDNA clone BRAMY...   361   6e-97
gb|AI680727.1|   tx40f05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clo...   335   3e-89
gb|U80757.1|    U80757 Human fetal brain (R.L.Margolis) Homo sap...  320   1e-84
gb|AI299726.1|   qm97d09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clon...   292   2e-76
gb|CN309838.1|   17000600179634 GRN_PRENEU Homo sapiens cDNA 5', m   267   1e-68
gb|BE779067.1|   601464841F1 NIH_MGC_67 Homo sapiens cDNA clone...   241   8e-61
gb|AW205682.1|   UI-H-BI1-adu-b-03-0-UI.s1 NCI_CGAP_Sub3 Homo s...   231   5e-58
gb|BI547995.1|   603189433F1 NIH_MGC_95 Homo sapiens cDNA clone...   172   3e-40
dbj|DA414512.1|   DA414512 BRTHA3 Homo sapiens cDNA clone BRTHA...   159   2e-36
dbj|DA123472.1|   DA123472 BRACE3 Homo sapiens cDNA clone BRACE...   141   8e-31
dbj|DA193144.1|   DA193144 BRAMY3 Homo sapiens cDNA clone BRAMY...  84.2   1e-13
```

Figure 3.39: Sequences producing significant alignments in the EST database.

These three ESTs were from different sources. The BE042890.1 EST (Figure 3.40 A) was generated from lung tissue and contains the Exon 1-Exon 2A boundary. The U80757.1 EST (Figure 3.40 B) had been detected in foetal brain cDNA during a study that involved using a RED assay to detect any repeat expansions in brain tissue [Margolis *et al.,* 1997]. The third EST, BE779067 (Figure 3.40 C), was isolated from human retinoblastoma tissue and contained 2 base pairs of Exon1 and the entire Exon 2A.

(A)

```
gb|BE042890.1| ho30c07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3038892
mRNA sequence.
Length=457
CONTAINS EXON 1 /2A BOUNDARY Lung
Score =  527 bits (285),  Expect = 5e-147
Identities = 300/306 (98%), Gaps = 6/306 (1%)
Strand=Plus/Plus

Query  516  ACGCACGGATTCAAGGGGCGCTACGGGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATAC  575
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2    ACGCACGGATTCAAGGGGCGCTACGGGGTGCGGGAGTGCGCGGGCAACGGGGCCAAATAC  61

Query  576  GAAGGGACCTGGAGCAACGGGCTGCAGGACGGCTACGGGACCGAGACCTACTCGGACGGA  635
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  62   GAAGGGACCTGGAGCAACGGGCTGCAGGACGGCTACGGGACCGAGACCTACTCGGACGGA  121

Query  636  GATGCCACCGCATTCGGGGCAGAGCCGGGGCCGGAAGccagggagctgcctgctgctgct  695
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  122  GATGCCACCGCATTCGGGGCAGAGCCGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGCT  181

Query  696  gctgctgctgctgctgctgctgctgctgctg-t-----aagatggtttctgtgcagggAA  749
            ||||||||||||||||||||||||||||||| |     ||||||||||||||||||||||||
Sbjct  182  GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTAAGATGGTTTCTGTGCAGGGAA     241

Query  750  CCTTGGCCGGCTCTGCAGCTGCCCGCCTGCCTGGACTCTCCGATATCCACTCCTCAGTGC  809
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  242  CCTTGGCCGGCTCTGCAGCTGCCCGCCTGCCTGGACTCTCCGATATCCACTCCTCAGTGC  301

Query  810  ACCTGA  815
            ||||||
Sbjct  302  ACCTGA  307
```

(B)

```
gb|U80757.1| U80757 Human fetal brain (R.L.Margolis) Homo sapiens cDNA, mRNA
Length=2107
2 BP OF EXON 1 AND ENTIRE EXON 2A
Score =  320 bits (173),  Expect = 1e-84
Identities = 181/184 (98%), Gaps = 3/184 (1%)
Strand=Plus/Minus

Query  635   AGATGCCACCGCATTCGGGGCAGAGCCGGGGCCGGAAGccagggagctg-c--ctgctgc  691
            ||||||||||||||||||||||||||||||||||||||||||||||||| |  |||||||
Sbjct  2009  AGATGCCACCGCATTCGGGGCAGAGCCGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGC  1950

Query  692   tgctgctgctgctgctgctgctgctgctgctgtaagatggtttctgtgcagggAACC  751
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1949  TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTAAGATGGTTTCTGTGCAGGGAACC  1890
```

```
Query   752   TTGGCCGGCTCTGCAGCTGCCCGCCTGCCTGGACTCTCCGATATCCACTCCTCAGTGCAC   811
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1889  TTGGCCGGCTCTGCAGCTGCCCGCCTGCCTGGACTCTCCGATATCCACTCCTCAGTGCAC   1830

Query   812   CTGA   815
              ||||
Sbjct   1829  CTGA   1826
```

(C) gb|BE779067.1| 601464841F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3867947
5', mRNA sequence.
Length=902
Score = 241 bits (130), Expect = 8e-61
Identities = 179/199 (89%), Gaps = 18/199 (9%)
Strand=Plus/Plus
<mark>2BP OF EXON 1 AND EXON 2A retinoblastoma</mark>

```
Query   635   AGATGCCACCGCATTCGGGGCAGAGCCGGGGCCGGAAGccagggagctgcctgctgctgc   694
              ||||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||
Sbjct   188   AGATGCCACCGGATTCGGGGCAGAGCCGGGGCCGGAAGCCAGGGAGCTGCCTGCTGCTGC   247

Query   695   tgctgctgctgctgctgctgct-gctgc-tgctg-t--aag-atggtttct-gtgcaggg   747
              |||||||||||||||| ||||| ||||| | ||| |||||||||||||||| |||||||||
Sbjct   248   TGCTGCTGCTGCTGCTGCTGCTAGCTGCATGCTGCTGTAAGCATGGTTTCTAGTGCAGGG   307

Query   748   -AACC-TTGGC-CGGCTCTGC-AGCTGCCC-GCCTGCC-TGG-ACTCTCCG-ATATCC-A   798
               |||| ||||| |||||||||| ||||||| | ||||| ||| ||||||||| ||||| |
Sbjct   308   CAACCATTGGCACGGCTCTGCCAGCTGCCCAGACTGCCATGGCACTCTCCGCATATCCCA   367

Query   799   CTCCTCAG-TGCACC-TGA   815
              ||||||||| |||||| |||
Sbjct   368   CTCCTCAGATGCACCGTGA   386
```

Figure 3.40: ESTs that contain Exon 1 and Exon 2A of *JPH3* alternatively spliced mRNA.

This investigation showed that a transcript wherein Exon 1 is spliced to a terminal exon containing the repeat does exist. Furthermore, this provides independent confirmation of the results of a study done by Holmes *et al.,* [2001] which used RT-PCR and gene prediction programmes to show that the repeat expansion which causes HDL2 occurs in an alternatively spliced version of the *JPH3* gene.

Chapter Four: Discussion

The primary objective of the present study was to ascertain the molecular cause of dementia and movement disorders in a South African family of Mixed Ancestry. After screening several novel loci and eliminating several potential disorders, it was determined that this family had Huntington's disease-like 2 (HDL2) [Bardien *et al.,* 2007]. The expanded repeats ranged from 42-59 repeats in three affected family members. HDL2 is a rare, autosomal dominant disease that occurs predominantly, if not exclusively, in individuals of Black African ancestry [Santos *et al.,* 2008]. A possible reason for this may be that the distribution of the CTG repeats in the *JPH3* gene within the Black African population is larger in comparison to other population groups which would potentially predispose this group to repeat expansions in the pathogenic range. This was investigated in a pilot study by determining the repeat distribution amongst three South African sub-populations. The results showed a statistically significant difference in the distribution of the repeats between the Black African and the Caucasian ethnic groups.

Although it is known that HDL2 is caused by a triplet repeat expansion in the *JPH3* gene, the mechanism of pathogenesis is as yet unclear. Therefore any research done on this topic may provide important insights in terms of managing the disease and developing suitable therapies. Given this, the present study also investigated factors that could possibly contribute to understanding the pathogenesis of this disease.

The *JPH3* expanded repeat is located in an alternatively spliced version of the gene. In the present study, it was determined whether the *JPH3* mRNA transcripts could be detected in RNA derived from foetal brain. In addition, even though *JPH3* is thought to be almost exclusively expressed in brain tissue, it was investigated whether the alternatively spliced transcript could be detected in the blood of HDL2 affected individuals due to the process of illegitimate transcription. Neither of the transcripts could be detected in transformed lymphocytes, however both transcripts were detected in foetal brain RNA.

4.1   Exclusion of loci

Inherited dementias are caused by neurodegenerative diseases such as HD or Alzheimer's disease. In studies of families with severe dementia of unknown genetic origin, several diseases need to be excluded before novel loci are screened. HD, SCAs, DRPLA, FENIB and PD are common candidate diseases to exclude in cases of familial dementia where a pathogenic mutation needs to be determined [Filla *et al.,* 2002]. In the present study several of these candidate disorders were excluded. The loci for HD, FENIB, DRPLA, SCAs (types 1, 2, 3, 6 and 7) and the *PARK2* gene were screened. In addition, loci implicated in MELAS, MERRF, Leighs' disease and other novel mitochondrial loci were also screened. All of these loci were excluded as being the disease-causing mutation.

For all the candidate loci, the proband was screened initially and any possible mutation was then assessed in other family members and ethnically matched controls. In the cases where a novel variant was identified, the variant was assessed for possible pathogenicity according to three criteria. Firstly, the sequence variant had to segregate in all affected family members and be absent in clinically unaffected family members and ethnically matched controls. Secondly, if the variant causes a non-synonymous change in the amino acid sequence of the protein, it had a greater potential of being pathogenic. Thirdly, if the variant occurs in a region of the gene which is evolutionarily conserved across species, it has an increased possibility of being pathogenic.

In this study three novel sequence variants were identified in the *MT-ND1* and *MT-ND2* genes in the mitochondrial genome. These three variants occurred in both affected and clinically unaffected members of Family R. These novel variants were deposited into the Mitomap database under the accession numbers 20061030001, 20061030002 and 2006103003. Because mitochondrial DNA is essential in determining lineages of individuals, the reporting of novel variants may be useful in forensics and population genetics studies. Mitochondrial DNA is highly variable and prone to mutations. In fact, it is estimated that the mutation rate in mitochondrial DNA is 10-20 times higher than that of nuclear genes [Strachan and Read, 1996]. It is therefore difficult to prove the pathogenicity of any novel mitochondrial sequence variant.

Several sequence variants were identified in the *PARK2* gene but were excluded as being the causative mutation because they are known polymorphisms and were previously shown to be present at similar frequencies in individuals with Parkinson's disease and "healthy" controls [West *et al.*, 2002, Lincoln *et al.*, 2003 and Kay *et al.*, 2007]. Kay *et al.* [2007] argued that some of the so-called "mutations" in the *PARK2* gene are found in equal distribution among PD patients and healthy controls. The first stage of their study found that the frequency of the IVS8+48 C>T variant was reduced in PD patients compared to the unaffected control group. However, on analysis of a larger sample size this tentative bias fell away and the presence of this polymorphism did not differ significantly between the PD and control groups. In the same study the V380L polymorphism was also assessed and occurred at the same frequency in both PD patients and controls. Similarly, Lincoln *et al.* [2003] considered both the V380L and D394N to be normal variants as both of them occurred at equal frequencies in PD patients and controls. The -227A>G variant is found in the promoter region and was initially thought to influence the transcription of the gene. However, there was no protein binding sight in close proximity to this SNP and it was therefore concluded that this SNP had no effect on the functionality of the gene [West *et al.*, 2002]. This group showed that the distribution of this variant did not differ significantly between 196 control individuals and 319 PD patients.

While many common diseases are not caused by SNPs, the interaction of SNPs in multiple genes, coupled with non-genetic risk factors all contributes to the phenotype of the disease. Furthermore, SNPs may also affect drug therapy and influence the manifestation of the disease [Pirmohamed and Park, 2001]. It is therefore important to identify SNPs in candidate disease genes as it may explain clinical heterogeneity, drug resistance and why certain populations are more susceptible to a particular disease than others.

4.2   Clinical analysis of affected family members

It was determined that Family R had HDL2, however, the clinical presentation of the disease in this family presented differently to other reported HDL2 cases. Firstly, the pedigree showed a strong trend towards a maternal pattern of inheritance as opposed to the HDL2 index family (Pedigree W) which displayed a strictly male-to-male transmission across three generations [Margolis *et al.,* 2001]. Due to this potential bias for maternal inheritance, several candidate loci on the mitochondrial genome were screened. As the disorder was in fact autosomal dominantly inherited, it was concluded that this potential bias was most likely a chance occurrence just as Pedigree W's potential paternal bias was later determined to be a chance event. However, this bias should not be completely disregarded as it is still unknown whether the sex of the transmitting parent plays a role in the aetiology of the disease. More families will need to be studied in order to determine this. Patients should therefore be screened for HDL2 regardless of family history and pattern of inheritance.

Secondly, several groups emphatically state that HDL2 is indistinguishable from HD [Margolis *et al.,* 2004 and Stevanin *et al.,* 2003]. The clinical examination of two of the three affected Family R members showed that they presented with a phenotype atypical to HD. One of the family members had cortical myoclonus while the other had parkinsonism [Bardien *et al.,* 2007]. Moreover, one case showed marked cerebral atrophy and normal caudate nuclei whereas in the advanced stages of HD, the caudate nuclei are usually severely atrophied and are affected to a greater degree than the cerebral cortex [Margolis *et al.,* 2006].

In addition, several features of the disease in Family R differ from those reported in the index pedigree. The autopsy examination of the first available case of Pedigree W showed that the caudate was severely atrophied and the white matter of the cerebral cortex was well myelinated after 15 years of disease duration. Since then four other brains have been examined and all these findings were similar to the first case. In contrast, the proband of Family R showed relative sparing of the caudate nucleus after 15 years of the disease and extensive white matter changes. Furthermore, no acanthocytosis were observed in peripheral blood smears as had been reported in some

cases of HDL2 [Walker *et al.,* 2003]. This finding lends support to the proposition that acanthocytosis might display reduced penetrance in HDL2 cases.

A key point to note from the clinical analysis is that HDL2 appears to be a disease that has a highly variable clinical phenotype and differs considerably even amongst family members. For example, in Family R, (as stated previously) one affected family member had parkinsonism while the other had features of cortical myoclonus, which are two distinct movement disorders. This variability demonstrates that HDL2 is phenotypically heterogeneous. Whether the SNPs identified in the *MT-ND1, MT-ND2* and *PARK2* genes are accountable for the clinical heterogeneity observed in this family could not be determined. A larger number of affected family members would be needed to determine whether a correlation between a particular SNP and the clinical features exists.

The analysis of 26 HDL2 cases, in which the age of onset and repeat length was known, showed that a longer repeat length correlated with an earlier age of onset (anticipation) [Margolis *et al.,* 2004]. This is a common feature of repeat expansion disorders and lends support to the fact that the repeat expansion is the causal mutation. It also suggests that a longer repeat length leads to more severe phenotype [Margolis *et al.,* 2004]. Although the recorded age at onset was estimated and based on verbal accounts from other family members, a similar correlation between age at onset (AAO) and repeat length was observed in affected Family R members. On examination of three affected family members it was found that the individual with 59 repeats had an AAO of 25 years while the individuals with 52 repeats and 49 repeats corresponded to an AAO of 39 and 48 years, respectively. In addition each repeat is roughly equivalent to a 0.4 year earlier AAO. This differs greatly compared to the estimate of 1.24 year earlier AAO per repeat that was observed in the 26 cases described by Margolis *et al.* [2004].

4.3 Genetic ancestry testing

To date, nearly all reported cases of HDL2 have been in individuals of definite or probable Black African ancestry. This tendency has further been supported by the absence of *JPH3* repeat expansions in a large cohort of Caucasians from Germany and Austria as well as France and North America [Bauer *et al.,* 2002 and Stevanin *et al.,* 2003]. Furthermore, individuals from Japan, with an HD-like phenotype, were also

negative for HDL2 expansions. Even two HDL2 cases, initially considered to be of non-African origin, were later found to have possible African origins. One of these cases, an individual from Morocco, was shown to originate from an area in Morocco predominantly occupied by Africans. The other case, a Mexican family, came from an area in Mexico that was colonised by Africans [Margolis *et al.,* 2004].

The ancestry of Family R was determined by genetic testing of one of the family members. The resolved mitochondrial haplogroup, L1d1, represents the maternal lineage, which occurs almost exclusively in populations of sub-Saharan Africa. This haplogroup is usually associated with the Khoikhoi and San clans of early Africa [Barkhan and Soodyall, 2006]. The Y-chromosome haplogroup, representing the paternal lineage, FG-M213, occurs predominantly in populations of Asia and Europe [Semino *et al.,* 2000]. Upon further analysis it was determined that the paternal haplogroup was found in 40 other males from a worldwide survey of 42 000 males. Moreover, this haplogroup exactly matched a male of South African Mixed Ancestry in the survey [Soodyall H, personal communication].

The ancestry testing performed in this study only reveals the partial ancestry of the individual as it does not determine all the possible ancestries that contribute to the lineage of this individual, for example the paternal grandmother. However, these results strongly suggest that this family member is more likely of Mixed Ancestry as opposed to South African Black or Caucasian ancestry. Furthermore, an indigenous African maternal lineage and European paternal lineage is typically found in individuals of Mixed Ancestry in South Africa [Nurse *et al.,* 1985].

This is the first published cases of HDL2 in a South African Mixed Ancestry family [Bardien *et al.,* 2007] although a preliminary report had indicated 7/20 Black patients from South Africa had tested positive for HDL2 [Krause *et al.,* 2002]. A recent report described a Brazilian Mixed Ancestry family with HDL2 that has no apparent African origins [Santos *et al.,* 2008]. As more knowledge about this disease increases there will surely be more reports from different ethnicities. Given this, it is suggested that all patients of Mixed Ancestry, that present with an HD-like phenotype but are negative for the HD mutation, be screened for HDL2, not only Black individuals. Moreover, it will be useful to set up a genetic diagnostic test that will screen for both HD and HDL2

simultaneously in cases where an individual of Black African Ancestry presents with an HD-like phenotype. Thus far no studies have been done to ascertain whether a genetic founder effect in the African population predisposes Africans to developing HDL2.

4.4 Distribution of the *JPH3* CTG repeats in the unaffected South African population

In the previous population studies that determined the distribution of the repeats in the *JPH3* gene, the repeat sizes varied between 8 and 28 consecutive repeats in unaffected individuals [Bauer *et al.,* 2002 and Stevanin *et al.,* 2003]. It is reasonable to assume that larger alleles are prone to expand into the pathogenic range once it reaches a certain threshold as reported for HD and SCA [Gomes-Pereira *et al.,* 2004]. To date, this threshold has not been established in HDL2, although repeats as short as 33 were shown to be unstable in vertical transmission when it expanded to 35 repeats during maternal transmission [Margolis *et al.,* 2004]. This implies that a population with a larger than average number of repeats could be predisposed to developing HDL2. Since the disease only occurred in individuals of Black African ancestry it was hypothesised that individuals of Black African ancestry may have larger repeats than other population groups. Of relevance, a study done in Serbia assessed the repeat distribution in 198 healthy individuals from Serbia and Montenegro to determine the risk of this population developing HDL2. The repeats ranged from 11 to 18 with 14 repeats being the most common. The authors suggested that HDL2 is rare in these populations because the repeat distribution was narrow compared to the range of repeats in the North African population (8-28 repeats) which consequently had a higher incidence of HDL2 [Keckarevic *et al.,* 2005]. Similarly, in the present study, the repeat distribution in three South African sub-populations was assessed to determine which sub-group had a higher risk for developing HDL2.

The pilot study showed a greater diversity in the alleles of Black African individuals (11 alleles ranging from 8 to 19 repeats) than in either the Mixed Ancestry or Caucasian cohorts. Moreover, the diversity in the Mixed Ancestry group (9 alleles ranging from 8 to 21 repeats) exceeded that of the Caucasian group which had only 7 different alleles ranging from 13 to 27 repeats. This can be expected as the Black African population is

an older population than both Caucasian and Mixed Ancestry groups and would therefore have a greater genetic diversity than either.

The repeats were in Hardy-Weinberg equilibrium for all three groups. However, a significant difference in the distribution of the repeats between the Black and Caucasian cohorts was observed. A difference was also observed between the Caucasian and Mixed Ancestry groups but it was not statistically significant.

The mode for all three groups was 14 repeats as observed in all other population studies [Bauer *et al.,* 2002; Keckarevic *et al.,* 2005 and Stevanin *et al.,* 2003]. However, the frequency of the 14 repeat allele differs significantly between the South African sub-populations. The 14 repeat allele was present in 55% of the Caucasian cohort, 33% of the Mixed Ancestry cohort and 25% of the Black African group. Again, this infers decreased allele diversity in the Caucasian group compared to the Black African and Mixed Ancestry groups.

In general, the results of this South African pilot study concurs with other population studies in that the mode is 14 repeats, with 16 repeats being the next most common allele. Although a significant difference in the distribution of the CTG repeat was observed between the Black and Caucasian cohorts, this study did not detect larger repeats in the Black African group. A follow-up study with a larger sample size is needed to determine whether the Black sub-population has the largest number of repeats and are subsequently at a higher risk for developing the disease.

4.5   Independent confirmation of the alternatively spliced *JPH3* mRNA transcript

The expanded repeat in HDL2 patients occurs in an alternatively spliced version of the *JPH3* mRNA transcript and due to variable splice acceptor sites the repeat can code for polyalanine, polyleucine or fall into the 3'UTR [Holmes *et al.,* 2001]. This was confirmed by GENSCAN predictions, the presence of a polyadenylation signal upstream to the repeat, as well as the existence of ESTs which are composed of a similar transcript and finally, by RT-PCR with primers specific for the repeat [Holmes *et al.,* 2001]. This transcript was only detected by one group, so in the present study the occurrence of this

transcript was tested for in commercially available foetal brain RNA as well as transformed lymphocytes of HDL2 affected family members in order to provide independent confirmation of the existence of this alternatively spliced transcript.

Due to the process of illegitimate transcription every mRNA transcript that exists should be present in blood at extremely low levels [Chelly *et al.,* 1989]. Chelly *et al.* [1989] showed that variably spliced transcripts of tissue specific genes could be detected in non-specific cells such as fibroblasts and lymphocytes. *JPH3* is tissue specific and is only expressed in the brain [Takeshima *et al.,* 2000]. In the absence of brain tissue from affected family members, an attempt was made to detect the transcript in lymphocytes because all transcripts should be expressed in blood due to illegitimate transcription. In addition, commercially available foetal brain RNA was used as a positive control because of *JPH3's* brain-specific expression profile.

Both the full length and alternatively spliced versions of *JPH3* mRNA were detected in foetal brain cDNA but not in lymphocytes. This could be due to the extremely low levels at which *JPH3* mRNA is expressed in tissues other than the brain. The possibility that the quality of the lymphocyte cDNA was at fault was excluded because primers designed to amplify a "housekeeping" gene (*HBB*) produced the correct product (verified by sequencing) in a standard PCR reaction in both tissue types. A housekeeping gene is defined as a gene that is expressed in all tissues at relatively similar levels. The *HBB* gene was selected because it is expressed in relatively similar levels in all tissues [Shmueli *et al.,* 2003]. Furthermore, *HBB* is highly conserved throughout different species and contains a relatively small intron which makes it easy to amplify in a PCR reaction. The primers were designed across exons so that differently sized products would be generated for genomic DNA and cDNA. This provided an important control for detecting genomic DNA contamination in the cDNA samples.

The initial set of primers designed to amplify the full length *JPH3* script was designed to amplify across the Exon 2–Exon 3 boundaries. However, the PCR reaction was unsuccessful for both lymphocyte and foetal brain cDNA. After extensive attempts at optimising the reaction, the PCR still failed. It was suspected that this may be because during the cDNA synthesis, the reaction was unable to extend all the way to the 5' end of the mRNA. A second set of primers, closer to the 3' end of the gene, was designed and

amplified the foetal brain cDNA on the first attempt. This implies that fragments closer to the 3' end of the mRNA may be a better target region to amplify when using cDNA as a template. However, the amplification using lymphocyte cDNA was still unsuccessful.

To replicate the results obtained by Holmes *et al.* [2001] the products generated from the PCR fragments with foetal brain cDNA were compared to sequences in the NCBI dbEST database. The database contains ESTs which are short randomly sequenced segments of cDNA. The database provides parts of gene sequences that are expressed in certain tissues, cells or disease states and is useful to confirm the existence of variably spliced mRNA isoforms.

The alternatively spliced transcripts generated in the present study showed significant alignment with 17 ESTs. However, only three of these ESTs contained a sequence wherein Exon 1 was spliced to Exon 2A. These three transcripts were the same as those identified by Holmes *et al.* [2001]. The findings in the present study confirm the existence of an alternatively spliced transcript in foetal brain RNA that contains a CTG/CAG repeat as proposed by Holmes *et al.* [2001].

To further investigate whether *JPH3* transcripts were detectable in lymphocytes, an exploratory study using real-time PCR was attempted. Real-time PCR is a more sensitive method than conventional PCR in that it measures the increase of fluorescence produced by the formation of PCR products in real time. In addition, real-time PCR enables the amplification of very low levels of template DNA. The melt curve analysis helps to determine whether contamination is present as the melt profiles are sequence-specific. The *HBB* primers were used initially to determine the optimal concentration of the cDNA being used. The PCR using primers for the *HBB* gene was successful for genomic DNA, lymphocyte and foetal brain cDNA and this was confirmed with direct sequencing.

Once the optimal cDNA concentrations were attained, PCR with the mRNA specific primers for the alternatively spliced and full length *JPH3* transcripts were attempted. The crossing point (CP) values for most of the reactions were very high (30-35 cycles) and it was therefore concluded that these experiments were not optimised. Selected products were, however, analysed because *JPH3's* expression, especially in the lymphocyte

cDNA, was expected to be at very low levels, and would therefore take a larger number of PCR cycles in order to produce a PCR product.

No conclusions could be drawn about the detection of either the full length or alternatively spliced products in lymphocyte cDNA or whether it is present in blood due to illegitimate transcription. These reactions need to be optimised further as most of the melt curves still showed non-specific peaks. This section of the study was an exploratory endeavour to set up an experimental basis for future real-time assays of HDL2 brain tissue.

4.6   Disease mechanisms of HDL2

Several hypotheses describing the pathogenic mechanisms of unstable repeat expansion disorders exist although none have been conclusively proven. Expansions may occur in non-coding or coding regions and the proposed mechanism of pathogenesis depends on where these expansions occur [Gatchel and Zoghbi, 2005].

Expansions in the non-coding, regulatory regions generally incur a loss of function mechanism in which the gene undergoes transcriptional silencing and the protein is not expressed. A well-known disease caused by a loss of function mechanism is Fragile X syndrome in which a triplet repeat expansion occurs in the 5' UTR of the gene and inhibits transcription of the gene [Li *et al.*, 2001]. *JP3* knock-out mice showed motor abnormalities but no developmental impairment. Furthermore, the symptoms of HDL2 is much more severe than the phenotype in these mice [Nishi *et al.*, 2002] suggesting that a loss of function is not the likely pathogenic mechanism. However, it is possible that reduced *JPH3* expression may contribute to the pathogenic mechanism because untranslated full length *JPH3* transcripts were present in the aggregations found in HDL2 brains thereby insinuating that less transcript was available for translation [Rudnicki *et al.*, 2007].

Expansions in the coding region of a gene, often resulting in long polypeptide tracts, commonly polyglutamine tracts, have been associated with a gain of function mechanism of pathogenesis. A polyglutamine disease usually affects a specific cell type resulting in a disease phenotype that has selective neurodegeneration [Zoghbi and Orr,

2000]. Another feature of polyglutamine diseases is the presence of aggregations of mutant protein and other components of the ubiquitin-proteasome system in the nucleus [Ciechanover and Brundin, 2003]. Thus far there has been no evidence to suggest that HDL2 is a polyglutamine disease, although clinically, it has features of a polyglutamine disease such as the presence of nuclear aggregations and selective neuronal degeneration.

A third proposed mechanism of pathogenesis describes expanded repeats in non-coding regions of the gene resulting in altered function due to mutant RNA and protein interactions. It has been suggested that in this mechanism the RNA containing the expanded repeat binds to RNA-binding proteins which are involved splicing, thereby disturbing normal mRNA splicing and forming aggregations [Gatchel and Zoghbi, 2005]. Increasing evidence suggest that HDL2 is due to a RNA-toxicity. CAG probes detected RNA foci in 30% of neurons in HDL2 brains. Probes specific for the alternatively spliced *JPH3* mRNA co-localized with the CAG probes in specific regions of the brain [Rudnicki *et al.*, 2007]. In addition, the cellular expression *JPH3* mRNA containing expanded CUG repeats produced RNA foci which were toxic to the cell.

Currently it is unknown how the repeat affects the mechanism of pathogenesis but further molecular studies may provide insight into this and may eventually lead to therapeutic intervention. The use of animal models to determine which pathogenic mechanism is involved in HDL2 is crucial.

4.7. <u>Conclusions</u>

Limitations of this study

More information about Family R would have been useful to assess the features of the disease in this family. Recruiting and examining more family members would have increased the significance of the differences in the clinical phenotype between family members. Screening more family members would have been valuable in determining whether the variants found in the *MT-ND1, MT-ND2* and *PARK2* genes were responsible for the clinical heterogeneity observed in this family. Unfortunately many of the affected family members were deceased and the individuals in the fourth generation were too

young and may have been pre-symptomatic. As the disease was a 'sensitive topic' to this family, it was difficult to recruit other family members.

A study using a larger sample size may answer the question of whether Black Africans are predisposed to developing HDL2 due to the repeat size in this population group.

With regards to the mRNA section of this study, precautions were taken to avoid RNA degradation but it may still have occurred because RNA is unstable and easily degraded. Due to time constraints, the real-time PCR reactions were not optimised and therefore these results are equivocal. They could be optimised by varying the PCR cycling parameters and template cDNA concentration or by the addition of PCR additives.

Due to time constraints, the possible significance of the novel polymorphisms identified in the *MT-ND1* and *MT-ND2* genes could not be evaluated. It would have been advantageous to determine the frequency of the observed polymorphisms in a healthy control group, which could have ascertained whether the polymorphisms were significant in the manifestation of the disease.

Future work

It would be interesting to determine whether all HDL2 families worldwide share a common haplotype as the disease may have a genetic founder effect. The analysis of the Y-chromosomes and mitochondrial haplogroups for common SNPs or STRs to trace founder effects has previously been useful [Wise *et al.,* 2005]. The information gathered would be useful in tracing other family members and offering genetic counselling, genetic testing and improved clinical management of the affected family members.

The frequency of HDL2 cases in South Africa should be assessed in order to determine the burden of this disease in South Africa as it has been proposed that HDL2 might be as common as HD amongst Black South African individuals. This could be done by recruiting samples from all the major movement disorder clinics and diagnostic laboratories as was done in the study by Bauer *et al.,* [2002]*.*

Investigating common pathways affected by HD, HDL2 and other neurodegenerative diseases may provide target proteins to which therapeutic drugs could be designed.

At the time of this study, brain tissue of the HDL2 affected Family R members was not available. Recently brain tissue from a deceased affected family member has been acquired and several studies on the tissues from the various brain regions could be proposed:

Although a commercial antibody for the mouse orthologue of *JPH3* is available, it is not known whether it would be effective for detection of the human protein and therefore it is proposed that antibodies against the human *JPH3* full length and alternatively spliced forms be raised. If a successful human antibody can be created, immunohistochemical studies can be used to determine if the expression levels of both forms differ in different regions of the brain. In addition, one would be able to determine which version of the mRNA transcripts is predominantly expressed in those areas of the brain which are more severely affected in HDL2 patients.

In all neurodegenerative diseases, the composition of the aggregates or inclusions have provided evidence of which gene may harbour the disease-causing mutation. The composition, morphology and histology of the intra-nuclear inclusions of the HDL2 brain should be investigated as it may provide vital information about the mechanism of pathogenesis of this disease. Recent studies proposed that the cellular toxicity observed in HDL2 is due to RNA transcripts as has been reported in myotonic dystrophy Type 1 [Rudnicki *et al.,* 2007]. The authors suggest that other transcripts of *JPH3* may also be present and what those transcripts are remains to be determined.

It would be interesting to determine the frequency of the inclusions in different regions of the HDL2 brain as one would expect the number of inclusions to increase in areas of the brain which are more severely affected. In diseases exhibiting anticipation, a larger repeat size may lead to a more severe phenotype. It would be expected that the number of these inclusions will differ between individuals with different repeat sizes and it will therefore be useful to document the number of inclusions as this can be used later in a comparative study once more cases become available.

A question that has thus far been unanswered is whether the repeat expansion disrupts the properties of the junctional complexes. The proposed function of *JPH3* is to facilitate cross-talk across the junctional membranes. It can therefore be assumed that this function will be disrupted in the HDL2 cases. By using electron microscopy, it would be possible to compare the morphology and functionality of the junctional complexes between affected and unaffected regions of the brain. This could shed some light on why the *JPH3*-knockout mice do not exhibit such severe neuronal dysfunction as is seen in HDL2 [Nishi *et al.,* 2002].

Conclusive statements

In summary, the disease-causing mutation underlying the disease phenotype in this Mixed Ancestry family was identified as a repeat expansion in the *JPH3* gene. A genetic pre-symptomatic predictive test with informed counselling can now be made available to at-risk family members. In this regard, the wife of one of the affected family members has approached the genetic counsellor, Prof Greetje de Jong, and requested predictive testing for her 17-year old daughter. She was advised to wait until her daughter is over 18 years.

This molecular diagnosis is of great significance as it may provide insight to the course of the disease thus enabling the affected individuals and their families to make informed decisions on life or family planning. Identifying the genetic cause for a disease can assist the family of affected individuals to prepare themselves financially as medical costs can be exhaustive. Also, knowing the underlying genetic mutation enables clinicians to make informed choices about the treatment and clinical management of the individual instead of spending unnecessary resources on different therapies that are unsuitable. This is especially true for movement disorders and dementias, which are symptoms for a large number of different disorders.

Identifying a family with HDL2 is significant because this is a very rare disorder. By studying this disease in different families, the clinical phenotype can be expanded on and finer description of this heterogenous disease can be formulated. Furthermore, studying HD-like disorders may provide insight into the pathogenic mechanisms of other common neurodegenerative disorders.

Molecular investigations of this kind may provide insight into the disease which may identify disease modifiers or predisposing alleles. It may also lead to the development of disease models and novel therapies. The allele frequency of the HDL2 repeat was determined in order to provide an indication of the risk for developing HDL2 among South Africans. This study showed that the repeat sizes in the South African population was the same as the distribution of repeat sizes worldwide, inferring that individuals from South Africa are not at a higher risk for developing the disease than other nationalities. Furthermore, this study strongly suggests that South African patients presenting with an HD-like disorder, that are negative for the HD mutation, be tested for HDL2 regardless of their family history and ancestry.

It is hoped that some aspects of this study may contribute to a better understanding of this rare and devastating disease and that it will provide data that may contribute towards the design of a possible cure. However, it is acknowledged that a lot more intensive research is needed to shed some light on the enigma of neurodegenerative diseases that appear to affect only certain brain cells.

APPENDIX 1: Consent forms

Informed Consent Form

Title of Study: Genetic analysis of inherited Parkinson's Disease and other related movement disorders.

Principal Investigator:	Dr Jonathan Carr FCP(SA), University of Stellenbosch, Cape Town
Co-investigator:	Dr Soraya Bardien, University of Stellenbosch, Cape Town
Study reference number:	2002/C059

Introduction:

You are invited to take part in a research study at the University of Stellenbosch. It is important that you should read and completely understand the following:

a) Your participation is completely voluntary.

b) You may withdraw from the study at any time without affecting your usual medical treatment.

After you have read the following explanation, please feel free to ask any questions which will help you better to understand and be completely informed about the nature of the study which you have been asked to participate in.

What the study is about:

We do not know what causes Parkinson's disease. It is probably caused by a number of different factors working together. One of these factors may be genes which a person inherits from their parents. (A baby is born with genes which come equally from the mother and the father. They determine many features of the child, for example eye and hair colour and so on). We also know that some diseases occur more commonly in certain people who carry particular genes. For example, a child whose parents have diabetes is more likely to develop diabetes themselves. It is possible that people who get Parkinson's disease and other movement disorders may have inherited a gene from their parents which makes it more likely that they will develop these diseases.

In this study, by analyzing DNA obtained from blood samples of people with Parkinson's disease or other related movement disorders, and their affected and unaffected relatives, we hope to identify the genes or markers responsible for the development of inherited forms of Parkinson's disease or other related movement disorders.

Blood samples collected from patients with Parkinson's disease and other movement disorders will be sent for genetic analysis to Dr. S Bardien at the University of Stellenbosch, Faculty of Health Sciences. Note that all blood samples will be given a number and not your name, so that information about your name is kept secret. The blood sample will be used for genetic studies related to this project only and not for any other purpose.

Requirement to enter the study:

Patients with Parkinson's disease, particularly if they have a family member who also has the disease, or if the disease started at a young age, or familial dystonia, are suitable for this study.

What we need from you:

About 40 cc (8 teaspoons) of blood will be taken from a vein in your arm.

1

Risks, Inconveniences and Discomfort:

You may feel pain when blood is taken. Very rarely a patient may faint when blood is taken from them.

Since some insurance companies may incorrectly assume that your/the participant's taking part in the study is an indication of an increased risk of genetic disease which may adversely affect your/the participant's ability to obtain health or other insurance, no information about your/the participant or your/the participant's family taking part in the study, will be given to any insurance company, since this examination cannot be viewed as a formal genetic test for the presence of absence of certain genes.

Confidentiality:

All information that is obtained from the study will be kept confidential. Documents with information about your personal details will be identified by code number and will only be available to the researchers.

Please note that the results of this study may take many years to complete. When results become available, if you wish, we can send these to your general practitioner. In addition, if you so desire, we can give you the results of the study. However, before you would be able to receive the results, you would need to undergo proper explanation and counselling. In the event that the results of this study are published, your information will remain confidential.

Remuneration/Compensation:

There will be no financial compensation if you decide to take part in this project. The genetic testing of the blood samples will be done at no charge to you.

Results:

The results of the analyses carried out on this material in the current study will be made known to you/the participant in accordance with the relevant protocol, if and when it becomes available. In addition, I/the participant authorise(s) the investigators to make the information available to

....(doctor's name), the doctor involved in my/the participant's care.

The analyses in the current study are specific to the condition or disease mentioned above and cannot determine the entire genetic make-up of an individual.

Genetic analyses may not be successful in revealing additional information regarding some families or some family members.

Even under the best conditions, current technology of this type is not perfect and could possibly lead to unreliable results.

Permission for further studies: Before your/the participant's genetic material is used in further projects in the future, the written approval of the Research Subcommittee C/Ethics Committee, Faculty of Health Sciences, will first be obtained.

IMPORTANT MESSAGE TO PARTICIPANT/REPRESENTATIVE OF PARTICIPANT:

Thank you very much for your/the participant's participation in this study. Should, at any time an emergency arise as a result of the research, or you require any further information with regard to the study, kindly contact:

- **Professor Jonathan Carr at**   **(021) 938 5500 or 084 606 3573.**

- **Debbie Lombard (research nurse)**   **082 428 8693**

2

PARKINSON'S DISEASE RESEARCH PROJECT

DECLARATION BY OR ON BEHALF OF PARTICIPANT:

I, THE UNDERSIGNED, ...

OR, in my capacity as .... of the participant.

A. HEREBY CONFIRM AS FOLLOWS:
1. I/The participant have been invited to participate in the above-mentioned research project of the Neurology Unit, Faculty of Health Sciences, University of Stellenbosch.
2. The information conveyed above was explained to me/the participant by

.... in English

OR was interpreted and explained by ....

3. I/The participant was afforded adequate time to pose any questions and all questions were answered to my/the participant's full satisfaction.
4. I/The participant was not pressurized to participate.
5. I/The participant will not be paid for participation, but reimbursement of travel costs will be considered (if applicable).
6. I/The participant will not incur any additional costs through participation.
7. I/The participant have/has received a copy of the Informed Consent form for my/his/her records.
8. The Research Subcommittee C/Ethics Committee, Faculty of Health Sciences, Stellenbosch University, has approved recruitment and participation of individuals in this study on the basis of:
• Guidelines on Ethics for Medical Research of the SA Medical Research Council;
• Declaration of Helsinki;
• International Guidelines: Council for International Organisations of Medical Sciences (CIOMS);
• Applicable RSA legislation.

B. HEREBY CONSENT VOLUNTARILY TO PARTICIPATE/ ALLOW THE POTENTIAL PARTICIPANT TO PARTICIPATE IN THIS STUDY:

Signed/Confirmed at .... on ....

Signature or right thumb print of participant/ representative of participant

Signature of witness

DECLARATION BY OR ON BEHALF OF INVESTIGATOR(S):

I, ..., declare that I

explained the information in this document to ....

....................and/or his/her representative ....

She/he was encouraged and afforded adequate time to ask me any questions.

This conversation was conducted in English and was translated (if applicable)

into....................(language) by ....

Signed at .... on ....

Signature of investigator/representative of investigator

DECLARATION BY TRANSLATOR (if applicable):

I, ..., confirm that I:

• translated the contents of this document from English into ....(language);
• explained the contents of this document to the participant/the participant's representative;
• also translated the questions posed by ....
• as well as the answers given by ....and
• conveyed a factually correct version of what was related to me.

Signed at .... on ....

Signature of translator

APPENDIX 2: Solutions, media and buffers

Sodium Borate (SB) buffer (20Xstock)

38.137g of $Na_2B_4O_7$ was dissolved in sterile distilled water to make up 1L. For 1XSB running buffer, a 100ml of the 20Xstock was made up to 2L with sterile distilled water (H_2O).

TBE buffer (10Xstock)

0.89M Tris-HCl, 0.89M boric acid and 20mM NaEDTA (at pH 8) was added to sterile distilled water to make up a 10XTBE stock solution.

Bromophenol Blue Loading Dye

0.2%(w/v) bromophenol blue; 50% glycerol and 10mM TRIS was added together and stirred gradually until all the ingredients were properly mixed.

Luria-Bertani Medium

5g of Bacto tryptone; 2.5g of Yeast extract and 5g of NaCl were added to a final volume of 500ml distilled H_2O. The solution was autoclaved at 121ºC and left to cool before antibiotics were added where necessary. For Luria-Bertani Agar, 8g of agar was added to the solution. After autoclaving the solution was cooled to 55ºC before IPTG and X-GAL were added. Approximately 15ml of this solution was poured into Petri dishes and allowed to solidify under sterile conditions.

CAP Buffer

10mM PIPES; 60mM $CaCl_2$ and15% glycerol was made up to a final volume of 250ml with sterile Millipore H_2O. The solution was pH to 7.0 with NaOH and stored at 4ºC.

APPENDIX 3: Phenol Chloroform method for DNA Extraction

[Corfield *et al.,* 1993]

Reagents:

- Cell lysis buffer (4ºC):

  109.5g sucrose; 10ml 100% Triton X; 5ml 1M $MgCl_2$ and 10ml 1M TRIS-HCl were mixed together and stored at 4ºC.

- Phenol:

pH to 8 with buffer. The buffer, phenol and equal volume Chloroform were mixed and stored at 4ºC overnight before use.

- Na-EDTA:

  18.75ml 4M NaCl was added to 250ml 100mM EDTA. This was made up to 1L with sterile distilled H_2O, pH to 8 and autoclaved.

- Octonol-Chloroform:

480ml Chloroform was added to 20ml octonol, mixed well and stored in a dark area.

- 10%SDS:

10g SDS was added to a 100ml of sterile distilled H_2O and incubated at 37ºC to dissolve SDS.

- 3M Sodium-acetate, pH to 6 with acetic acid.
- 10mg/ml *Proteinase K* (Roche')
- 96% ice cold Ethanol
- 1XTE solution
- 70% Ethanol

Method:

1. 5ml EDTA blood samples were transferred to 45ml sterile tubes, cold Lysis buffer was added to the 45ml mark and vigorously shaken. The tubes were incubated on ice for 10 minutes. The cells were then vortexed and pelleted at 3000rpm for 10 minutes after which the supernatant was poured off.

2. The pellet was rinsed with cold lysis buffer a few times until the pellet was light pink.

3. 900µl Na-EDTA and 100µl 10% SDS was added to each pellet. At this point the pellets were either stored at -80ºC or used for DNA extraction as follows.

4. 100µl 10mg/ml *Proteinase K* was added to each tube and the tubes were incubated at 37ºC, overnight.

5. 2ml Distilled H_2O and 500µl 3M Sodium-acetate was added to each tube and the contents were mixed by inverting the tube.

6. 2.5ml Phenol-chloroform was added to each tube. The tubes were placed on a shaker (at a 40º angle to the platform of the shaker) at the speed of 1400rpm for 10 minutes.

7. The contents were then transferred to a 10ml Corex glass tubes and spun at 7000rpm for 12 minutes at 4-10ºC.

8. A 3ml Pasteur pipette was used to transfer the supernatant to a clean Corex tube without disturbing the interphase.

9. 2.5ml Octonol–chloroform was added to the tube containing the supernatant. The tube was sealed with a leak proof top and slowly mixed by inversion until the solution turned milky.

10. The tops were removed from the tubes and the tubes were spun at 7000rpm for 10 minutes at 4ºC.

11. The supernatant was transferred to a 12ml falcon tube. 5-7ml ice cold 96% Ethanol was added to the tube and slowly mixed until the DNA precipitated.

12. Once the DNA precipitate was visible a pipette was used to transfer the DNA to a 1.5ml Eppendorf tube.

13. The tube was filled with 70% Ethanol and spun for 3 minutes at 14 000rpm. The Ethanol was slowly poured off so as not to loose the pellet. This step was repeated once.

14. The pellet was air dried at room temperature.

15. 500µl 1XTE solution was added to each tube and the tubes were incubated at 37ºC overnight or for 10 minutes at 65ºC.

16. The tubes were placed on a rotating wheel at 30rpm for 3 days after which the optical density and the purity of the DNA was checked with the NanoDrop®.

APPENDIX 4: SSCP Gel preparation

Reagents:

Table A. Components for 8% and 10% SSCP gels

	8% Gel	10% Gel
Urea	24g	24g
Sterile distilled H_2O	91.8ml	84ml
10XTBE	8ml	8ml
Glycerol	8ml	8ml
10% APS	1.6ml	1.6ml
TEMED	160µl	160µl

SSCP Loading Dye Reagents:

SSCP Loading Dye contained 95% formamide; 20mM EDTA; 0.01% Bromophenol Blue and 0.05% of Xylene cyanol.

Method:

Glass plates and tanks were cleaned and prepared. A sheet of Gelbond® PAG film (Cambrex Bio science, Rockland, USA), was placed between the glass plates where it covalently bonded to the acrylamide gel during polymerization, thus providing support.

Urea, sterile distilled H_2O, 10XTBE and glycerol was added to a sterile beaker while stirring continuously. 10% APS and TEMED were added last (Table A). The solution was then immediately poured into the cast that had previously been prepared. The gels were kept in the cast until set. Meanwhile, 0.5XTBE running buffer was poured into the tank. The samples were prepared by combining equal volumes of SSCP loading dye to the samples and placing them on a heating block at 94.5ºC for 5 minutes. The samples

were incubated on ice for 5 minutes or until just before loading. The gels were run at 23 Watts at 4ºC overnight.

APPENDIX 5: Silver staining of SSCP gels

Reagents:
- 1% $AgNO_3$ (Solution B) :
  1g $AgNO_3$ was dissolved in distilled sterile H_2O to make up 1L.
- Developing Solution (Solution C):
  15g NaOH; 0.1g of $NaBH_4$ and 4ml of Formaldehyde was dissolved in sterile distilled H_2O to make up 1L.

Method:

500ml Solution B was added to a tray containing the SSCP gel .This tray was placed on a shaker at 3000rmp for 10 minutes. Solution B was then poured off and the gel was rinsed with sterile H_2O. 500ml Solution C was then added to the tray containing the gel and this was placed on a shaker for a further 15 minutes or until the bands could be clearly seen. The solution was than poured off and the gel was once again rinsed with sterile H_2O. The gels were then left to dry and sealed in plastic.

APPENDIX 6: Preparation of *E.coli* DH5α competent cells for transformations

Reagents and materials:
- Sterile inoculation loop
- LB media (Appendix 2)
- 2L Elenmeyer flask
- CAP Buffer

Method:

1. A loop of *E.coli* DH5α was scraped from a frozen stock and was inoculated into a sterile Falcon tube containing 10ml LB media (Appendix 2).

2. The media was incubated overnight at 37ºC with constant shaking at 200rpm. The following day 1ml of this culture was added to 200ml of LB in a 2L Erlenmeyer flask with a cotton wool plug.

3. The flask was then further incubated at 37ºC with gentle shaking at 85rpm until the cells were grown to mid log phase (Optical Density= 0.4-0.6nm).

4. The media containing the cells were aliqouted into 4 X 50ml sterile Falcon tubes and centrifuged for 15 minutes at 4ºC. All the steps thereafter were carried out in a 4ºC cold room.

5. The supernatant was removed and the cells were resuspended in one third the volume of ice cold CAP buffer (Appendix 2).

6. The cells were once again centrifuged at 3000rpm for 15 minutes to pellet the cells and the supernatant was discarded.

7. 5ml Cold CAP buffer was added to the pellet and the pellet was gently resuspended by pippeting. 200µl Aliqouts of the suspension was stored at -70ºC until further use.

APPENDIX 7: Lymphocyte separation and transformation

Materials:
- RPMI medium: RPMI 1640 with 15% foetal calf serum and 1% penicillin or streptomycin.
- EBV medium : Epstein Barr Virus Medium
- Histopaque: Histopaque 1077

Method:
1. RPMI and EBV medium were heated to 37ºC and Histopaque to 22ºC.
2. 3ml Histopaque was syringed into a sterile tube and 4-5ml blood collected in Heparin tubes were slowly added so that it formed a separate layer.
3. The tube was centrifuged at 1800rpm for 20 minutes to further separate the phases.

4. The white blood cell layer was removed using a sterile Pasteur pipette and placed into a clean, sterile tube to which 3ml RPMI was added. The tube was spun for a further 30 seconds at 1000rpm.

5. The cell pellet was resuspended in 0.3ml RPMI medium and poured into a 25cm^3 flask containing 1.5ml EBV media.

6. The flask was incubated in an upright position for a week at 37ºC, with 5% Carbon Dioxide and 90% humidity.

7. 0.5ml RPMI media was added to the solution twice weekly until a final volume of 5ml was reached. At various intervals the solution was viewed under a microscope to monitor cell growth.

8. After a week the cell culture was spun at 1000rpm for 30 seconds. The supernatant was poured off and the cells were regrown to 5ml again in order to concentrate the cells.

9. The cells were transferred to a 75 cm^3 culture flask and 10ml RPMI medium was added twice weekly or until a final volume of 50ml was reached.

10. The cells was frozen in liquid nitrogen by adding with 10% DMSO and RPMI medium in cryovials and stored at -70ºC.

APPENDIX 8: RNA isolation using TRIzol

.

Reagents and materials:

- 2ml Screw cap containing Hybaid ribolyser blue beads (Qiagen)
- 2ml Tube containing Phase-Lock gel heavy (Eppendorf) pelleted at max speed for 20-30s
- Chloroform:isoamyl alcohol (24:1)
- Isopropanol
- TRIzol® Reagent (Invitrogen)
- RNAse free water or 1ml DPEC was dissolved in IL water. This was incubated at 37ºC over night and autoclaved twice.

Method:

1. 200µl Transformed lymphocytes (Appendix 7) were pelleted at 3000rmp for 15 minutes and the supernatant was poured off.

2. The cells were resuspended in 1ml of TRIzol and the suspension was transferred to a 2ml screw cap tube containing ribolysing beads. This process was performed on ice.

4. The cells were mechanically disrupted by vortexing for 20 seconds at a speed of 6 for 3 cycles. The samples were cooled down after each cycle by placing it on ice for 2 minutes between pulses.

5. Thereafter the samples were centrifuged at 13000rpm for 45 seconds.

6. The TRIzol solution above the beads was transferred into a 2ml tube containing phase-lock gel and 300µl Chloroform:isoamyl alcohol.

7. The tubes were inverted gently but rapidly for 15 seconds and incubated at room temperature for 2 minutes.

8. The tubes were centrifuged at 13000rpm for 5-10 minutes after which the top aqueous layer was transferred to a clean sterile tube.

9. An equal volume of isopropanol was added to the tube containing the aqueous layer and mixed by inverting. This solution was stored at -20ºC overnight to precipitate the nucleic acid.

10. The precipitate was collected by centrifugation at 12000rmp for 20 minutes in a 4ºC cold room and the supernatant was discarded.

13. The pellet was washed once with 1ml 70% ethanol and centrifuged at 12000rpm for 5 minutes at 4ºC. The ethanol was removed by aspiration and the pellet was air-dried.

15. The RNA was dissolved in 20-40µl RNAse-free water. The sample was treated with Ambiance Turbo DNAse ™ Free as specified by the manufacturer.

REFERENCES

- Abbott RD, Ross GW, White LR, Sanderson WT, Burchfiel CM, Kashon M, Sharp DS, Masaki KH, Curb JD and Petrovich H: 2003. Environmental, life-style, and physicalprecursors of clinical Parkinson's disease: recent findings from the Honolulu-Asia Aging Study. *J Neurol* 2003; **250**(suppl 3): 30–39

- Bardien S, Abrahams F, Soodyall H, van der ML, Greenberg J, Brink T and Carr J: A South African mixed ancestry family with Huntington disease-like 2: clinical and genetic features. *Mov Disord* 2007; **22**(14): 2083-2089

- Barkhan D and Soodyall H (Ed): Mitochondrial DNA and Y-chromosome DNA variation in African populations: The Prehistory of Africa: Tracing the lineages of modern man South Africa, *Jonathan Ball Publishers*, 2006

- Barrett G: Jerk-locked averaging: technique and application. *J Clin Neurophysiol* 1992; **9**(4): 495-508

- Bauer I, Gencik M, Laccone F, Peters H, Weber BHF, Feder EH, Weirich I, Morris-Rosendahl DJ, Rolfs A, Gencikova A, Bauer P, Wenning GK, Epplen JT, Holmes SE, Margolis RL, Ross CA and Reiss O: Trinucleotide repeat expansions in the *Junctophiln-3* gene are not found in Caucasian patients with a Huntington's disease-like phenotype. *Ann Neur* 2002; **51**(5): 662

- Bear MF, Conners BW and Paradiso MA: Neuroscience: Exploring the brain. Baltimore, *Williams and Wilkins Publishers*, 1996

- Berkovic SF, Carpenter S, Evans A, Karpati G, Shoubridge EA, Andermann F, Meyer E, Tyler JL, Diksic M, Arnold D: Myoclonus epilepsy and ragged-red fibres (MERRF). A clinical, pathological, biochemical, magnetic resonance spectrographic and positron emission tomographic study. *Brain* 1989; **112** (Pt 5): 1231-1260

- Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP and Hudson T: Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a

protein kinase family member. *Cell* 1992; **69**(2): 385

- Bryer A, Krause A, Bill P, Davids V, Bryant D, Butler J, Heckmann J, Ramesar R and Greenberg J. The hereditary adult-onset ataxias in South Africa. *J Neurol Sci* 2003; **216**: 47–54

- Brown P, Cathala F, Castaigne P and Gajdusek DC: Creutzfeldt-Jakob disease: clinical analysis of a consecutive series of 230 neuropathologically verified cases. *Ann Neurol* 1986; **20**(5): 597-602

- Brown P, Cervenàkova L, Boellaard JW, Stavrou D, Goldfarb LG and Gajdusek DC: Identification of a *PRPN* gene mutation in Jakob's original Creutzfeldt-Jakob disease family. *Lancet* 1994; **344**: 130-131

- Burke JR, Wingfield MS, Lewis KE, Roses AD, Lee JE, Hulette C, Pericak-Vance MA and Vance JM: The Haw River syndrome: dentatorubropallidoluysian atrophy (DRPLA) in an African-American family. *Nat Genet* 1994; **7**(4): 521-524

- Chelly J, Concordet JP, Kaplan JC and Kahn A: Illegitimate transcription: transcription of any gene in any cell type. *Proc Natl Acad Sci U S A* 1989; **86**(8): 2617-2621

- Chinnery PF, Johnson MA, Wardell TM, Singh-Kler R, Hayes C, Brown DT, Taylor RW and Bindoff LA and Turnbull DM: The epidemiology of pathogenic mitochondrial DNA mutations. *Ann Neurol* 2000; **48**(2): 188-193

- Chomczynski P and Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987; **162**(1): 156-159

- Ciechanover A and Brundin P: The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron* 2003; **40**: 427-446

- Corfield VA, Moolman JC, Martel R and Brink PA: Polymerase chain reaction-based detection of MN blood group specific sequences in the human genome. *Transfusion* 1993; **33**: 119–124

- Cowan CM and Raymond LA: Selective neuronal degeneration in Huntington's disease. *Curr Top Dev Biol* 2006; **75**: 25-71

- Crimi M, Galbiati S, Moroni I, Bordoni A, Perini MP, Lamantea E, Sciacco M, Zeviani M, Biunno I, Moggio M, Scarlato G and Comi GP: A missense mutation in the mitochondrial *ND5* gene associated with a Leigh-MELAS overlap syndrome. *Neurology* 2003; **60**(11): 1857-1861

- Crum RM, Anthony JC, Bassett SS and Folstein MF: Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* 1993; **269**(18): 2386–2391

- Davis RL, Shrimpton AE, Carrell RW, Lomas DA, Gerhard L, Baumann B, Lawrence DA, Yepes M, Kim TS, Ghetti B, Piccardo P, Takao M, Lacbawan F, Muenke M, Sifers RN, Bradshaw CB, Kent PF, Collins GH, Larocca D and Holohan PD: Association between conformational mutations in neuroserpin and onset and severity of dementia. *Lancet* 2002; **359**(9325): 2242-2247

- Enriquez JA, Chomyn A and Attardi G: MtDNA mutation in MERRF syndrome causes defective aminoacylation of tRNA (Lys) and premature translation termination. *Nat Genet* 1995; **10**(1): 47-55

- Estrada R, Galarraga J, Orozco G, Nordase A and Auburger G: Spinocerebellar ataxia (SCA2): morphometric analyses in 11 autopsies. *Acta Neuropath* 1999; **97**(3): 306-310

- Filla A, De Michele G, Cocozza S, Patrignani A, Volpe G, Castaldo I, Ruggiero G, Bonavita V, Masters C, Casari G and Bruni A: Early onset autosomal dominant dementia with ataxia, extra pyramidal features and epilepsy. *Am Acad Neurol* 2002; **58**: 922-928

- Finckh U, Muller-Thomsen T, Mann U, Eggers C, Marksteiner J, Meins W, Binetti G, Alberici A, Sonderegger P, Hock C, Nitsch RM, Gal A: High frequency of mutations in four different disease genes in early-onset dementia. *Ann N Y Acad Sci* 2000; **920**: 100-106

- Flicker L: Acetylcholinesterase inhibitors for Alzheimer's disease. *BMJ Ed* 1999; **318**: 615-616

- Folstein MF, Folstein SE and McHugh PR: "Mini-mental state" A practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res* 1975; **12**(3): 189-198

- Gasser T: Genetics of Parkinson's disease. *Curr Opin Neurol* 2005; **18**(4): 363-369

- Gatchel JR and Zoghbi HY: Diseases of unstable repeat expansion: mechanisms and common principles. *Nat Genet Rev* 2005; **6**: 743-755

- Gauthier E, Fortier I, Courchesne F, Pepin P, Mortimer J and Gauvreau D: Aluminum forms in drinking water and risk of Alzheimer's disease. *Environ Res A* 2000; **84**: 234-246.

- Gomes-Pereira M, Fortune T, Ingram L, McAbney JP and Monckton DG: Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Gen* 2004; **13**(16): 1815-1825

- Goto Y, Horai S, Matsuoka T, Koga Y, Nihei K, Kobayashi M and Nonaka I: MELAS: A correlative study of clinical features and mitochondrial DNA mutations. *Neurology* 1992; **42**: 545-550

- Gottlieb M and Chavko M: Silver staining of native and denatured eukaryotic DNA in agarose gels. *Anal Biochem* 1987; **165**: 33-37

- Guryev V, Smits B M G, van de Belt J, Verheul M, Hubner N and Cuppen E: Haplotype block structure is conserved across mammals. *PLOS Gen* 2006; **6**: 1111-1118

- Hainfellner JA, Brantner-Inthaler S,Cervenakova L, Brown P, Kitamoto T, Tateishi J, Diringer H, Liberski PP, Regele H, Feucht M, Mayr N, Wessely P, Summer K, Seitelberger F and Budka H: The original Gerstmann-Sträussler-Scheinker family of Austria: divergent clinicopathological phenotypes, but constant *PrP* genotypes. *Brain Pathol* 1995; **5**: 210–211

- Harvey RJ, Skelton-Robinson M and Rossor MN: The prevalence and causes of dementia in people under the age of 65 years. *J Neurol Neurosurg Psychiat* 2003; **74**: 1206-1209

- Hirano M, Ricci E, Koenigsberger MR, Defendini R, Pavlakis SG, DeVivo DC, DiMauro S and Rowland LP: Melas: an original case and clinical criteria for diagnosis. *Neuromuscul Disord* 1992; **2**(2): 125-135

- Hirano M and DiMauro S: Handbook of muscle disease. New York, *Marcel Dekker Inc*, 1992

- Holmes SE, O' Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, Fleisher A, Stevanin G, Brice A, Potter NT, Ross CA and Margolis RL: A repeat expansion on the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat Genet* 2001; **29**: 377-378

- Ikeuchi T, Koide R, Onodera O, Tanaka H, Oyake M, Takano H and Tsuji S: Dentatorubral-pallidoluysian atrophy (DRPLA). Molecular basis for wide clinical features of DRPLA. *Clin Neurosci* 1995; **3**(1): 23-27

- Kay DM, Moran D, Moses L, Poorkaj P, Zabetian C, Nutt J, Factor SA, Yu C, Montimurro JS, Keefe RG, Schellenberg GD and Payami H: Heterozygous *parkin* point mutations are as common in control subjects as in Parkinson's patients. *Anal Neur* 2007; **61**: 47-54

- Keckarevic M, Savic D and Romac S: *JP-3* gene polymorphism in a healthy population of Serbia and Montenegro. *J Genet* 2005; **84**(1): 69-71

- Komure O, Sano A, Nishino N, Yamauchi N, Ueno S, Kondoh K, Sano N, Takahashi M, Murayama N, Kondo I, Nagafuchi S, Yamada M and Kanazawa I: DNA analysis in hereditary dentatorubral-pallidoluysian atrophy: Correlation between CAG repeat length and phenotypic variation and the molecular basis of anticipation. *Neurology* 1995; **45**: 143-149

- Krause A, Temlett J, Van der Meyden K, Ross CA, Callahan C and Margolis RL: CAG/CTG repeat expansions at the HDL2 locus are a common cause of Huntington disease in Black South Africans. *Am J Hum Genet* 2002; **Abstract 2098**

- Lee H, Soong B, Lu Y and Chang Y: Dentatorubropallidoluysian Atrophy in Chinese. *Arch Neurol* 2001; **58**: 1905-1908

- Li Z, Zhang Y, Ku L, Wilkinson KD, Warren ST and Feng Y: The fragile X mental retardation protein inhibits translation via interacting with mRNA. *Nucleic Acids Res* 2001; **29**(11): 2276–2283

- Lincoln SJ, Maraganore DM, Lesnick TG, Bounds R, de Andrade M, PhD, Bower JH, Hardy JA and Farrer MJ: Parkin Variants in North American Parkinson's Disease: Cases and Controls. *Mov Dis* 2003; **18**(11):1306-1311

- Lucking CB, Durr A, Bonifati V, Vaughan J, De Michele G, Gasser T, Harhangi BS, Meco G, Denèfle P, Wood NW, Agid Y and Brice A: Association between early-onset Parkinson's disease and mutations in the *parkin* gene. *N Engl J Med* 2000; **342**(21): 1560-1567

- Lugaresi E, Medori R, Montagna P, Baruzzi A, Cortelli P, Lugaresi A, Tinuper P, Zucconi M and Gambetti P: Fatal familial insomnia and dysautonomia with selective degeneration of thalamic nuclei. *N Engl J Med* 1986; **315**(16): 997-1003

- Majamaa K, Turkka J, Karppa M, Winqvist S and Hassinen IE: The common MELAS mutation A3243G in mitochondrial DNA among young patients with an occipital brain infarction. *Neurology* 1997; **49**(5): 1331-1334

- Margolis RL, Abraham MR, Gatchell SB, Li SH, Kidwai AS, Breschel TS, Stine OC, Callahan C, McInnis MG and Ross CA: cDNAs with long CAG trinucleotide repeats from human brain. *Hum Genet* 1997; **100**(1): 114-122

- Margolis RL, O'Hearn E, Rosenblatt A, Willour V, Holmes SE, Franz ML, Callahan C, Hwang HS, Troncoso JC and Ross CA: A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion. *Ann Neurol* 2001; **50**(6): 373-380

- Margolis RL and Holmes SE: Huntington's disease-like 2: a clinical, pathological and molecular comparison to Huntington's disease. *Cl Neurosc Res* 2003; **3**: 187-196

- Margolis RL, Holmes SE, Rosenblatt A, Gourley L, O'Hearn E, Ross CA, Seltzer WK, Walker RH, Ashizawa T, Rasmussen A, Hayden M, Almqvist EW, Harris J, Fahn S, MacDonald ME, Mysore J, Shimohata T, Tsuji S, Potter N, Nakaso K, Adachi Y, Nakashima K, Bird T, Krause A and Greenstein P: Huntington's disease-like 2 (HDL2) in North America and Japan. *Ann Neurol* 2004; **56**(5): 670-674

- Margolis RL, Holmes SE, Rudnicki DD, O'Hearn E, Ross CA, Pletnikova O, Troncoso JC, Wells RD (Ed) and Ashizawa T (Ed): Genetic instabilities and neurological diseases, 2ⁿᵈ Edition :Huntington's disease-like 2. San Diego, *Academic Press*, 2006

- Marras C and Tanner C in Pulst S M (Ed): Genetics of movement disorders: Parkinson's disease, genetic epidemiology and overview. Amsterdam, *Academic Press*, 2003

- Mashima Y, Hiida Y, Oguchi Y, Kudoh J and Shimizu N: High frequency of mutations at position 11778 in mitochondrial *ND4* gene in Japanese families with Leber's hereditary optic neuropathy. *Hum Genet* 1993; **92**(1): 101-102

- Mattocks C, Tarpey P, Bobrow M and Whitaker J: Comparative Sequence Analysis (CSA): A new sequence-based method for the identification and characterization of mutations in DNA. *Hum Mut* 2000; **16**(5): 437-443

- Mata IF, Lockhart PJ and Farrer MJ: *Parkin* genetics: one model for Parkinson's disease. *Hum Mol Genet* 2004; (Spec No 1): R127–R33

- McInnis MG: Anticipation: an old idea in new genes. *Am J Hum Genet* 1996; **59**: 973-979

- Mead S: Prion disease genetics. *Eur J Hum Genet* 2006; **14**(3): 273-281

- Michalik A, Martin JJ and Van BC: Spinocerebellar ataxia type 7 associated with pigmentary retinal dystrophy. *Eur J Hum Genet* 2004; **12**(1) :2-15

- Molinari F, Meskanaite V, Munnich A, Sonderegger P and Colleaux L: Extracellular proteases and their inhibitors in genetic diseases of the central nervous system. *Hum Mol Genet* 2003; 12 Spec No 2R195-R200

- Moraes CT, Ciacci F, Bonilla E, Janen C, Hirano M, Rao N, Lovelace R.E, Rowland L.P, Schon E.A and DiMauro S: Two novel pathogenic mitochondrial DNA mutations affecting organelle number and protein synthesis. *J Clin Inves* 1993; **92**: 2906-2915

- Naito H and Oyanagi S: Familial myoclonus epilepsy and choreoathetosis, hereditary dentatorubral-pallidoluysian atrophy. *Neurology* 1982; **32**: 798

- Neri C: New light on polyglutamine neurodegenerative disorders: interference with transcription. *Trends Mol Med* 2001; **7**(7): 283-284

- Nishi M, Mizusshima A, Nakagawara K and Takeshima H: Characterization of human *Junctophilin* subtype genes. *Biochem Biophysic Res Comm* 2000; **273**:920-927

- Nishi M, Hashimoto K, Kuriyama K, Komazaki S, Kano M, Shibata S and Takeshima H: Motor discoordination in mutant mice lacking *Junctophilin* Type 3. *Biochemic Biophysic Res Comm* 2002; **292**(2): 318-324

- Nishi M, Sakagami H, Komazaki S, Kondo H and Takeshima H: Coexpression of *Junctophilin* type 3 and type 4 in brain. *Brain Res Mol Brain Res* 2003; **118**(1-2): 102-110

- Nurse GT, Weiner JS and Jenkins T: The growth of hybrid communities. The peoples of Southern Africa and their affinities. Oxford, *Claredon Press*, 1985

- Okubadejo NU, Bower JH, Rocca WA and Maraganore DM: Parkinson's disease in Africa: A systematic review of epidemiologic and genetic studies. *Mov Disord* 2006; **21**(12): 2150-2156

- Oliveri RL, Zappia M, Annesi G, Annesi F, Spadafora P, Pasqua AA, Tomaino C, Nicoletti G, Bosco D, Messina D, Logroscino G, Manobianca G, Epifanio A, Morgante L, Savettieri G and Quattrone A: The *parkin* gene is not a major susceptibility locus for typical late-onset Parkinson's disease. *Neurol Sci* 2001; **22**(1): 73-74

- Onodera O, Oyake M, Takano H, Ikeuchi T, Igarashi S and Tsuji S: Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expanded alleles in the CNS. *Am J Hum Genet* 1995; **57**(5): 1050-1060

- Orita M, Suzuki Y, Sekiya T and Hayashi K. A rapid and sensitive detection of point mutations and genetic polymorphisms using polymerase chain reaction. *Genomics* 1989; **5**: 874-879

- Osterwalder T, Contartese J, Stoeckli ET, Kuhn TB and Sonderegger P: Neuroserpin, an axonally secreted serine protease inhibitor. *EMBO J* 1996; **15**(12): 2944-2953

- Pirmohamed M and Park BK: Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* 2001; **22**(6): 298-305

- Pocchiari M, Ladogana A, Petraroli R, Cardone F and D'Alessandro M: Recent Italian FFI cases. *Brain Pathol* 1998; **8**: 564–566

- Priyadarshi A, Khuderb SA, Schaubb EA and Priyadarshi SS: Environmental risk factors and Parkinson's disease: A meta-analysis. *Env Res* 2001; **86**(2): 122-127

- Pulst S M (Ed): Genetics of movement disorders. Amsterdam, *Academic Press*, 2003

- Ramesar RS, Bardien S, Beighton P and Bryer A: Expanded CAG repeats in spinocerebellar ataxia (SCA1) segregate with distinct haplotypes in South African families. *Hum Genet* 1997; **100**(1): 131-137

- Remes AM, Majamaa-Voltti K, Karppa M, Moilanen JS, Uimonen S, Helander H, Rusanen H, Salmela PI, Sorri M, Hassinen IE and Majamaa K: Prevalence of large-scale mitochondrial DNA deletions in an adult Finnish population. *Neurology* 2005; **64**(6): 976-981

- Rudnicki DD, Holmes SE, Lin MW, Thornton CA, Ross CA, and Margolis RL: Huntington's disease--like 2 is associated with CUG repeat-containing RNA foci. *Ann Neurol* 2007; **61**(3): 272-282

- Sambrook J and Russell DW: Molecular cloning: A laboratory manual, 3rd Edition. USA, *CSHL Press*, 2001

- Santos C, Wanderley H, Vedolin L, Pena SDJ, Jardim L and Sequeiros J: Huntington's disease-like 2: the first patient with apparent European ancestry. *Clin Genet* 2008. 10.1111/j.1399-0004.2008.00981.x

- Schneider S A, Walker RH and Bhatia KP: The Huntington's disease-like syndromes: what to consider in patients with a negative Huntington's disease gene test. *Nature clinical prac, Neurology* 2007; **3**(9): GY517

- Schöls L, Bauer P, Schmidt T and Riess O: Autosomal dominant cerebellar ataxias: clinical features, genetics and pathogenesis. *Lancet Neur* 2004; **3**: 291-304

- Semaka A, Creighton S, Warby S, and Hayden MR: Predictive testing for Huntington disease: interpretation and significance of intermediate alleles. *Clin Genet* 2006; **70**(4): 283-294

- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De BG, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL and Underhill PA: The genetic legacy of Paleolithic *Homo sapiens* in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**(5494): 1155-1159

- Shmueli O, Horn-Saban S, Chalifa-Caspi V, Shmoish M, Ophir R, Benjamin-Rodrig H, Safran M, Domany E, and Lancet D: GeneNote: Whole genome expression profiles in normal human tissues. *Comptes Rendus Biologies* 2003; **326**(10-11): 1067-1072

- Shoffner JM, Lott MT, Lezza AM, Seibel P, Ballinger SW and Wallace DC: Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA (Lys) mutation. *Cell* 1990; **61**(6): 931-937

- Silverthorn DU: Human Physiology: An integrated approach, 2nd Edition. New Jersey, *Prentice Hall*, 2001

- Simon DK, Pulst SM, Sutton JP, Browne SE, Beal MF and Johns DR: Familial multisystem degeneration with parkinsonism associated with the 11778 mitochondrial DNA mutation. *Neurology* 1999; **53**(8): 1787-1793

- Sternberg D, Chatzoglou E, Laforet P, Fayet G, Jardel C, Blondy P, Fardeau M, Amselem S, Eymard B and Lombes A: Mitochondrial DNA transfer RNA gene sequence variations in patients with mitochondrial disorders. *Brain* 2001; **124**(Pt 5): 984-994

- Stevanin G, Durr A and Brice A: Clinical and molecular advances in autosomal dominant cerebellar ataxias: from genotype to phenotype and physiopathology. *Eur J Hum Genet* 2000; **8**(1): 4-18

- Stevanin G, Fujigasaki H, Lebre AS, Camuzat A, Jeannequin C, Dode C, Takahashi J, San C, Bellance R, Brice A and Durr A: Huntington's disease-like phenotype due to trinucleotide repeat expansions in the *TBP* and *JPH3* genes. *Brain* 2003; **126**(7): 1599-1603

- Strachan T and Read AP: Human Molecular Genetics. USA, *BIOS Scientific Publishers Lmtd,* 1996

- Takahashi H: Juvenile Parkinsonism: It's neuropathological aspects. *No To Shinkei* 1994; **46**(6): 523-529

- Takeshima H, Komazaki S, Nishi M, Lino M and Kangawa K: *Junctophilins*: a novel family of junctional membrane complex proteins. *Mol Cell* 2000; **6**(1): 11-22

- Tan EK and Jankovic J: Genetic testing in Parkinson disease: promises and pitfalls. *Arch Neurol* 2006; **63**: 1232–1237

- Tan EK and Skipper LM: Pathogenic mutations in Parkinson's disease. *Hum Mutat* 2007; **28**(7): 641-653

- Taylor RW and Turnbull DM: Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 2005; **6**(5): 389-402

- Tortora G and Grabowski S: Principles of anatomy and physiology, 8th Editon. *John Wiley and Sons Inc*, 1996

- Van de Graaff K: Human anatomy, 6th Edition. *McGraw-Hill Publishers*, 2001

- Van Swieten JC, Brusse E, de Graaf BM, Krieger E, van de Graaf R, de Koning I, Maat-Kievit A, Leegwater P, Dooijes D, Oostra BA and Heutink P: A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebral ataxia. *Am. J. Hum. Genet* 2003; **72**:191–199

- Walker RH, Morgello S, Vidoff-Feldman B, Melnick A, Walsh MJ, Shashidharan P and Brin MF: Autosomal dominant chorea-acanthocytosis with polyglutamine-containing neuronal inclusions. *Neurology* 2002; **58**(7): 1031-1037

- Walker RH, Rasmussen A, Rudnicki D, Holmes SE, Alonso E, Matsuura T, Ashizawa T, Vidoff-Feldman B and Margolis RL: Huntington's disease-like 2 can present as chorea-acanthocytosis. *Neurology* 2003; **61**(7): 1002-1004

- Watts RL and Koller W: Movement Disorders: Neurologic principles and practice. *McGraw-Hill Companies Inc*, 1997

- West AB, Maraganore D, Crook J, Lesnick T, Lockhart PJ, Wilkes KM, Kapatos G, Hardy JA and Farrer MJ: Functional association of the *parkin* gene promoter with idiopathic Parkinson's disease. *Hum Mol Genet* 2002; **11**(22): 2787-2792

- Windl A, Giese W, Schulz-Schaeffer I, Zerr K, Skworc S, Arendt C, Oberdieck M, Bodemer S, Poser HA and Kretzschmar A: Molecular genetics of human prion diseases in Germany. *Hum Genet* 1999; **105**: 244-252

- Wimo A, Jonsson L and Winblad B: An estimate of the worldwide prevalence and direct costs of dementia in 2003. *Dement Geriatr Cogn Disord* 2006; 21:175-181

- Wise CA, Sullivan SG, Black ML, Erber WN and Bittles AH: Y-chromosome and mitochondrial DNA studies on the population structure of the Christmas Island community. *Am J Phys Anthropol* 2005; **128**(3): 670-677

- Wood JM, Worringham C, Kerr G, Mallon K and Silburn P: Quantitative assessment of driving performance in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2005; **76**(2): 176-180

- Young K, Clark HB, Piccardo P, Dlouhy SR and Ghetti B: Gerstmann-Sträussler-Scheinker disease with the *PRNP* P102L mutation and valine at codon 129. *Brain Res Mol Brain Res* 1997; **44**: 147–150

- Zoghbi HY and Orr HT: Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* 2000; **23**: 217-247

- Zeviani M, Muntoni F, Savarese N, Serra G, Tiranti V, Carrara F, Mariotti C and DiDonato S: A MERRF/MELAS overlap syndrome associated with a new point mutation in the mitochondrial DNA tRNA (Lys) gene. *Eur J Hum Genet* 1993; **1**(1): 80-87

Internet resources

- BioEdit: http://www.mbio.ncsu.edu/BioEdit/page2.html

- ClustalW sequence alignment programme: http://align.genome.jp

- Ensembl database: http://www.ensembl.org

- Ensemble sequences: http://www.ensembl.org/index.html

- McGraw-Hill Online Learning Centre:
  http://highered.mcgrawhill.com/sites/0072495855/student_view0/index.html

- Mitomap: http://www.mitomap.org

- NanoDrop® users manual: http://www.nanodrop.com/techsupport/nd-1000-users-manual.pdf

- NCBI: http://www.ncbi.nlm.nih.gov

- NCBI Blast N: http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi

- NCBI GenBank: http://www.ncbi.nlm.nih.gov/Genbank/

- Pimer 3 programme: http://frodo.wi.mit.edu/cgibin/primer3/primer3_www.cgi

- Promega website: http://www.promega.com

- Qiagen website: www.qiagen.com

- R- Project for statistical computing: http://www.R-project.org

- www.gazzaro.it/g/Language%20in%20the%20brain_file/sensory_motor.gif

- Y-chromosome haplotype reference database (YHRD): http://www.yhrd.org/index.html