A MOLECULAR INVESTIGATION OF HUNTINGTON DISEASE; ORIGINS OF THE MUTATION AND CURRENT PREVALENCE IN SOUTH AFRICA

by

FIONA KEBIRUNGI BAINE

BNXFIO003

A thesis presented for the degree of

DOCTOR OF PHILOSOPHY

In the Division of Human Genetics, Department of Clinical Laboratory Sciences

Faculty of Health Sciences, University of Cape Town

SUPERVISORS: PROFESSOR L JACQUIE GREENBERG AND DR MICHAEL R HAYDEN

JANUARY 2015

Declaration

I, Fiona K Baine, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

The referencing style used is based on the journal Human Genetics.

Signed by candidate

Signature:……

Date: ….20 July 2015……..

Acknowledgements

On reflection, I believe it a pity that the blood, sweat and tears that are a part of completing a doctoral project, are reported nowhere in the thesis itself. Nonetheless, there are many individuals that have borne witness to this journey and to them I offer sincere thanks:

To my supervisors; Jacquie Greenberg and Michael Hayden, I have benefited enormously from your kindness and support both personally and professionally. Over several years, you have taught me, challenged me and made this project a reality. Most importantly, I have learned from you what it means to be a scientist connected to humanity.

To Amanda Krause, for your willing participation in and contributions to this project. Thank you too for providing a new home for me, I look forward to many more years of working together.

To Chris Kay and Marlies Ketelaar for their assistance in the molecular aspects of this project performed at the Hayden laboratory, I could not have completed the genotyping and analysis of over 1000 samples without you! Thanks too to Jennifer Collins for your knowledge of the HD database and the assignment of haplotypes.

To Professors Marais, Meissner and Mayosi (UCT),Drs September, Shaboodien (UCT) and Singaraja (CMMT) and Ms Frost (Wits); for sharing control DNA samples. This investigation would not have been complete without this contribution.

To the clinical teams at the Neurogenetics Clinic, Groote Schuur Hospital and the Division of Human Genetics (NHLS/Wits); for access to patient records and many discussions on HD.

To the Division of Human Genetics (UCT); for providing such an incredible work environment and developing me into a scientist. I am indebted to all of you for the support both personal and academic. Special thanks to my friends Lisa Roberts and Danielle Smith for the laughs, hugs and chocolate.

To the Hayden laboratory; for welcoming me with open arms and making sure that I will always feel like Vancouver is yet another home. Special thanks to Safia Ladha and Shaun Sanders, partners in crime and the best of friends.

To Emeritus Professor Peter Beighton, for critical reading of this thesis and providing invaluable insights, along with some choice anecdotes regarding the research.

To Associate Professor Collet Dandara, for your willingness to contribute to my growth and development as an individual and in particular foryour assistance with statistical analyses.

To my friends who have sent their love and support, sometimes over many hundreds of miles, to help me get through the last 4 years;I could not have gotten this far without you and I shall always be grateful.

To my parents, my sisters and my brother; despite the distance between us, I have been constantly aware of your love and support and I do not take it for granted. Thank you for blessing me with almost daily encouragement.

And lastly to Ticha; you have seen the best and the worst of me and chosen to stand by my side. Thank you for cheering me through highs and loving me through lows, and mostly for staying calm through all of it.

I acknowledge with sincere gratitude, financial support from the following: Ripples of Hope Trainee Award, University of Cape Town International Students' Scholarship, the Commonwealth Exchange Students Scholarship and the L'Oreal-UNESCO For Women in Science Fellowship.

Technical aspects of the project (CAG/CCG sizing and haplotyping) were performed at, and funded by, the Hayden laboratory. Additional project funding was obtained from UCT.

This thesis is dedicated to the memories of my grandfathers:

TIBAMANYA MWENE MUSHANGA

and

LAZARUS HOKUKIRIRE

Contents

Section II

List of Figures

List of Tables

Abbreviations and Symbols

°C	degrees Celsius
Δ	delta
µL	microliters
ASOs	antisense oligonucleotides
CAF	Central Analytics Facility, Stellenbosch University
CMMT	Centre for Molecular Medicine and Therapeutics
CTCF	CCCTC-binding factor
DM	Myotonic Dystrophy
DNA	deoxyribonucleic acid
F	forward primer
FXS	Fragile X syndrome
HASA	Huntington's Association of South Africa
HD	Huntington disease
HDL	HD-like
HIV	human immunodeficiency virus
HTT	*Huntingtin*(gene)
IA	intermediate allele
IT15	interesting transcript 15 gene
JHD	juvenile Huntington disease
JPH3	*Junctophilin-3*(gene)
kDa	kilodalton
kb	kilo base pairs
mRNA	messenger RNA
ng	nanograms

NHLS	National Health Laboratory Service
nt	nucleotides
OR	odds ratio
PCR	polymerase chain reaction
polyQ	polyglutamine
Q	glutamine
R	reverse primer
RNA	ribonucleic acid
RNAi	RNA interference
RP	reduced penetrance
SA	South Africa
SBMA	Spinal and Bulbar Muscular Atrophy
SCA	Spinocerebellar Ataxia
SNP	single nucleotide polymorphism
tSNPs	tag SNPs
UCT	University of Cape Town
UBC	University of British Columbia
v	version
Wits	University of the Witwatersrand

Glossary

Genetic terms:

Admixture –used specifically in relation to the coloured subpopulation of South Africa, it refers to the genetic lineages resulting from intermarriages between individuals of different ethnic backgrounds.

De novo – a diagnosis of HD in the absence of a family history and no expanded parental *HTT* alleles.

Haplogroup – a collection of similar haplotypes.

Haplotype – a set of DNA polymorphisms that tends to be inherited together as an intact unit.

HD genocopy – a term that has been proposed as an alternative to the more commonly used 'phenocopy'.

HD phenocopy – a genetic condition that is clinically indistinguishable from HD, but with a different underlying mutation.

Intermediate allele –an allele with a repeat size between 27-35 CAG, with the potential to expand into the pathogenic range in a single transmission.

Reduced penetrance allele – an allele in the lower end of the pathogenic range (36-39 CAG) that may or may not be associated with clinical symptoms in a normal life span.

TagSNPs – SNPs selected to represent the genetic variation across a specific region.

Epidemiology terms:

Incidence – number of new cases affected by a given condition within a specified period of time.

Prevalence – number of people affected by a given condition at a specified point in time.

Statistical terms:

Mean CAG –also average, the mean is calculated as the sum of observations divided by the number of observations. It is a measure used to indicate the point where a population is centred.

Median CAG– the value that divides the distribution of observations in half. It is another measure used to indicate the point where a population is centred.

Mode – also modal CAG, the most frequent observed value.

Odds ratio – the ratio of the probability of occurrence of an event to that of non-occurrence. It expresses the association between a risk factor and a disease by comparing the likelihood of disease under different circumstances.

p-value – used to define statistical significance, it is the estimated probability of rejecting the null hypothesis when that hypothesis is true. The null hypothesis is that there is no difference between the categories. When the *p*-value is < 0.05, the null hypothesis is rejected and this difference is said to be statistically significant.

Range – the distance between the highest and lowest observations, it is a measure of variability in a given population.

Abstract

Huntington disease (HD) is a devastating neurodegenerative condition characterised by a triad of symptoms: behavioural/psychiatric changes, cognitive decline and movement disorder. The dominantly inherited disease-causing mutation is an expanded trinucleotide (CAG) repeat in the *Huntingtin*(*HTT*) gene. Clinical symptoms are believed to be the result of degeneration of specific neuronal populations that are susceptible to the presence of a toxic expanded protein product. The disease is incurable and following the onset of symptoms, is progressively debilitating over 10-20 years and eventually fatal.

Although typical epidemiological studies of prevalence are challenging for a genetic disorder such as HD, family studies and various other methods of ascertainment have been used to estimate its occurrence in different populations. Prevalence is therefore known to vary geographically; population-specific haplotypes have been hypothesised to be the basis of this variation between ethnic groups. High prevalence estimates for populations with European ancestry led to the supposition that the HD mutation was introduced to different regions by Europeans.

In South Africa, a survey in the 1970s estimated that the prevalence of HD in the white and coloured subpopulations was similar at 2 per 100 000 individuals; while that in the black subpopulation was significantly lower, at less than 0.01 per 100 000 individuals. Molecular genetic analyses have since revealed links between the white and coloured subpopulations which would explain the similarity in prevalence; however, our knowledge of the genetics of HD in the black subpopulation, has been sorely lacking.

This study provides, for the first time, a comprehensive analysis of the *HTT* gene in an African population. An evaluation of the normal distribution of CAG-tract sizes highlighted significant differences between the subpopulations. Haplotype analysis identified population-specific disease-associated haplotypes, confirming distinct origins of the HD mutation in the different subpopulations.

In a coloured family with the rare juvenile form of the disease, DNA sequencing revealed no novel variants within the immediate vicinity of the CAG-tract that could be associated with the observed instability. This indicates that genome-wide analyses may be more useful in identifying factors related to repeat instability and future investigations are planned for a cohort of South African patients affected by juvenile onset HD.

A review of all records at the two public laboratories offering a confirmatory molecular test for the diagnosis of HD, provided the opportunity to calculate minimum estimates of frequency based on genetic ascertainment. HD prevalence was estimated at 0.5, 2.2 and 7.8 (per 100 000 individuals) for the black, coloured and white subpopulations respectively. These figures can be associated with the distribution of CAG-tract size and the population-specific haplotypes identified in this study.

Overall, the results reported here emphasise the genetic diversity present in different populations in SA. Knowledge of this diversity is critical to the development of therapeutics. Furthermore, this project has laid the foundation for South Africa to participate in a global prospective study of HD aimed at improving our understanding of disease mechanisms and most importantly, the clinical management and care offered to HD patients.

For the purpose of this thesis, the candidate has elected to adhere to the nomenclature used by the South African government and by its peoples. The peoples' self-identification as Black African, Coloured and White is used throughout. These identities are largely based on ethno-cultural factors and not on socio-political differences; therefore there is no implied negative connotation.

Section I

Chapter 1 provides an introduction to HD and a comprehensive review of available literature on epidemiological and molecular genetic aspects. The subject of HD in South Africa is also addressed as background to the project and the overall aims are outlined.

Chapter 1: INTRODUCTION

1.1 Huntington disease

1.1.1. A history

The first concise description of Huntington disease was made by Dr George Huntington an American physician, in 1872 (Huntington 1872). Following the completion of his medical studies at the age of 21 years, George Huntington presented an essay titled '*On Chorea*' before the Meigs and Mason Academy of Medicine. The essay was later published in The Medical and Surgical Reporter, and Huntington sealed his place in medical history with his astute description of the condition that now bears his name.

Historical records suggest that the disease was recognised as early as the seventeenth century and several scholars may have published descriptions prior to Huntington's seminal publication (Dunglison 1842; Osler 1894). However, Huntington had the benefit of a family practice spanning several generations in a small community, with both his father and grandfather having been physicians in the same East Hampton village. This allowed him to delineate "three marked peculiarities in this disease: 1. Its hereditary nature 2. A tendency to insanity and suicide 3. Its manifesting itself as a grave disease only in adult life" (Huntington 1872).

Chorea, originating in the Greek *khoreia* which means 'dancing in unison', describes a disease of the nervous system characterised by jerky involuntary movements made by afflicted individuals. Prior to George Huntington's description of the particular condition that came to be known as Huntington's Chorea, the term '*chorea*' caused much confusion in the 19th century, with over a dozen conditions

categorised therein (Osler 1894; Hayden 1981). Several forms of chorea had previously been recognised as hereditary; however, it was only three decades later with the re-discovery of Mendel's laws that the mechanism of inheritance described by George Huntington was understood to be of an autosomal dominant pattern (Lanska 2000).

For many decades, Huntington's chorea was synonymous with a vicious stigma, arising from the neurological symptoms in combination with the hereditary aspect. There is evidence that affected individuals, suspected of witchcraft, may have been put to death both in England and the United States (Hayden 1981). In the 20th century, it was one of the conditions targeted by the eugenics movement and afflicted families were subjected to horrendous abuse including forced sterilisations and even death in some societies (Harper 1992a).

It is unfortunate that to this day, a degree of shame and secrecy still encompasses families and individuals living with HD. The personal experience of one HD family has been shared by historian Alice Wexler in a recent perspective (Wexler 2010). The public perceptions of, and theories surrounding how and why certain families are affected, highlight the necessity of addressing history in order to change present and future attitudes.

1.1.2. Clinical presentation

George Huntington's classic delineation of hereditary chorea in 1872 defined many aspects of the adult form of what is now known as Huntington disease (HD). The triad of movement disorder, emotional disturbance and cognitive impairment are the accepted defined clinical features (Walker 2007; Sturrock and Leavitt. 2010).

In addition, a family history of movement disorder or behavioural problems, although undiagnosed, may be an indication of HD. Importantly, the presentation and severity of symptoms, and progression of the condition may vary significantly from one patient to another even within the same family.

The presence of motor symptoms, has conventionally been used to define the onset of disease for the purpose of consistency in making a clinical diagnosis (Walker 2007; Sturrock and Leavitt. 2010). However, subtle changes in behaviour and executive functioning are often evident at an earlier stage.

A large, on-going longitudinal observational study on mutation positive individuals (PREDICT-HD), has estimated that abnormalities may be detectable up to 20 years prior toan official clinical diagnosis (Paulsen et al. 2008). The terms pre-manifest or pro-dromal HD have been used to describe this group of individuals who are asymptomatic, those who exhibit only subtle symptoms or who remain undiagnosed (Sturrock and Leavitt. 2010).

HD is typically a disorder of adulthood with an average age at onset of 35 to 44 years. However, symptoms may manifest anytime from childhood to much later in life. Following onset, the disease is inexorable and patients eventually succumb to complications related to dysphagia, aspiration and infection. Death has been estimated to occur 10-20 years subsequent to diagnosis (Margolis and Ross. 2003; Walker 2007; Roos 2010; Sturrock and Leavitt. 2010).

Disease pathology is limited to the brain and accounts for the major clinical symptoms. Although the spectrum of pathological changes may vary between affected individuals, the HD brain is generally notably smaller by weight and volume (Vonsattel et al. 2008).

Prominent degeneration occurs in the caudate nucleus, the putamen and the cerebral cortex. Nuclear inclusions have been identified in neostriatal and neocortical neurons and can be detected long before the onset of clinical symptoms (Vonsattel et al. 2008); however, their role in disease pathogenesis remains unclear (Walker 2007; Sturrock and Leavitt. 2010).

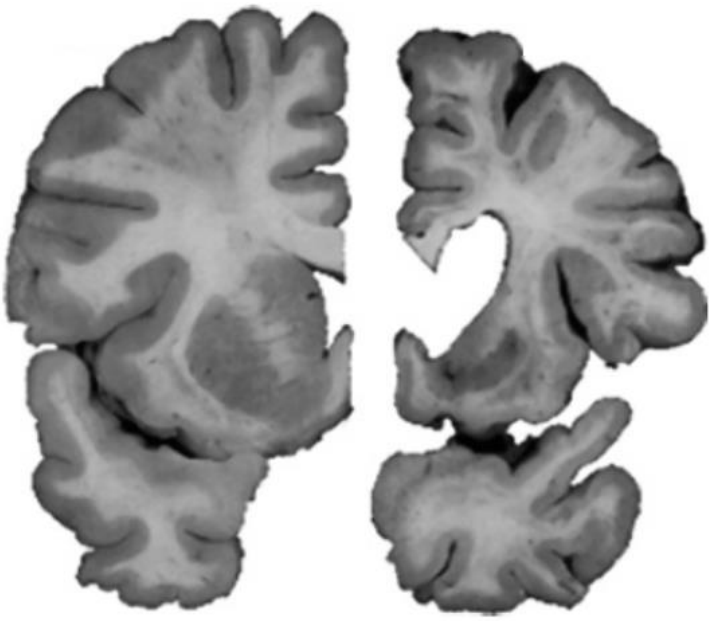


Figure 1. Pathology of the HD brain (right) compared to a normal brain (left). The images depict the severe degeneration and decreased brain volume evident post-mortem (Vonsattel et al. 2008).

The three 'peculiarities' of the condition as delineated by George Huntington are briefly examined below.

1.1.2.1. The hereditary nature

The first and possibly most important characteristic noted by George Huntington was that affected individuals had at least one parent who was affected if they lived to adulthood. However, if the children of an affected individual did not develop the condition even in old age, then the thread was broken and successive generations were free from disease (Huntington 1872).

HD is a classic autosomal dominant monogenic disorder and the molecular genetics of the disease will be addressed extensively in subsequent sections of this thesis.

1.1.2.2. The tendency to insanity and suicide

The second highlighted feature was the "tendency to insanity and suicide" (Huntington 1872). A significant aspect of HD is cognitive dysfunction, resulting in impaired executive functioning, speech and comprehension; any of which may have been viewed as signs of insanity (Margolis and Ross. 2003; Walker 2007). The decline is noticeably progressive over the course of disease, culminating in dementia in the latter stages.

Under the limited understanding of neuropsychiatric disorders in the 1800s, behavioural problems and personality changes in HD patients may well have been perceived as insanity. Psychiatric manifestations are a well-documented phenomenon in HD and known to be widely variable (Anderson and Marder. 2001; Walker 2007; Sturrock and Leavitt. 2010). Common features include apathy,

irritability/aggression, obsessive compulsive behaviour and anxiety. Psychosis has also been reported, albeit less frequently (van Duijn et al. 2014).

Fortunately, psychiatric symptoms are often responsive to treatment, although limited evidence exists for the efficacy of various medications currently in use (Roos 2010; Nance 2012; Killoran and Biglan. 2014). Clinical intervention is critical to manage symptoms and prevent suicidal ideation and/or suicide attempts which often occur as an impulsive act at specific stages of disease (Anderson et al. 2010; Sturrock and Leavitt. 2010).

Numerous studies have been carried out on suicide as it has been shown to be of marked importance within families that carry the mutation (Farrer et al. 1986; Di Maio et al. 1993; Almqvist et al. 1999; Robins Wahlin et al. 2000). The rate of completed suicide in HD mutation carriers is estimated at between four and eight times higher than the general population; while suicidal attempts and ideation occur up to 20% more frequently (Wetzel et al. 2011; Hubers et al. 2012; Hubers et al. 2013).

Both socio-demographic factors and clinical characteristics such as depression, have been associated with suicidal ideation, suicide attempts and completed suicide. Members of HD families may therefore require careful monitoring and regular screening to identify individuals at increased risk of suicide (Hubers et al. 2013).

1.1.2.3. The late onset

The final peculiarity described by George Huntington was the onset of chorea in adult life. Based on his review of the medical records in east Hampton, there were no documented cases of marked chorea presenting before the age of 30 or 40 years

(Huntington 1872; Bates 2005). In addition, it was improbable for an individual with a family history to present with chorea after the age of 40 years, leading to his conclusion that HD was almost exclusively an adult onset condition.

Although no mention was made by Huntington of symptoms occurring in children or adolescents, this rare form of the disease had likely been described by Lyon in 1863 (Hayden et al. 1982). The onset of symptoms before the age of 20 years and in some cases as early as childhood, defines juvenile HD (JHD) which is reported to affect up to 10% of all HD patients (Gonzalez-Alegre and Afifi. 2006; Walker 2007).

The clinical presentation of JHD is different from the adult-onset form, as it is characterised by rigidity and hypokinesia and has been termed the Westphal variant (Töpper et al. 1998; Gonzalez-Alegre and Afifi. 2006; Walker 2007). Seizures are common and a decline in performance at school and delayed motor milestones may be an indication of early onset HD in children with a family history.

1.2 The Huntington disease gene

Between 1983 and 1993, scientists around the world worked tirelessly to isolate and characterise the HD gene. Research efforts were driven by collaborations with a number of non-profit organisations; notably the Committee to Combat Huntington Disease (CCHD) and the Hereditary Disease Foundation (HDF) founded in 1967 and 1978 respectively (Bates 2005).

The success story surrounding the identification of the HD gene and disease-causing mutation draws heavily on the advocacy of patient and family groups such as these, and highlights the importance of their involvement in research. Of utmost significance was the concurrent development and evolution of molecular techniques.

1.2.1. Identification of the gene and mutation

In 1983 a marker linked to the HD locus was reported, assigning the genetic defect to chromosome 4 (Gusella et al. 1983). This was the initial step in a lengthy investigation that eventually led to the isolation of the HD gene. As one of the first successes in mapping an autosomal condition using DNA polymorphisms, this approach fuelled the identification of many disease genes.

Once chromosome 4 was identified as the location of the HD gene, a large international collaborative effort ensued. Making use of novel molecular techniques, new markers were generated and genetic and physical maps were established; these narrowed the region of interest to the short arm of this chromosome (Wasmuth et al. 1988; Whaley et al. 1988; Bates et al. 1991).

Further work and haplotype analysis revealed the most likely candidate region to be a 100 kilobase pairs (kb) fragment, between markers D4S180 and D4S182 on 4p 16.3 (MacDonald et al. 1992). An exon amplification strategy ultimately led to the identification of a novel gene, *IT15* (interesting transcript 15), containing a polymorphic CAG triplet repeat in exon 1 that was found to be expanded on HD chromosomes (The Huntington's Disease Collaborative Research Group 1993).

The *IT15* gene, later renamed the *Huntingtin* (*HTT*) gene, spanned roughly 210 kb and encoded a previously unknown protein of approximately 348 kilodaltons (kDa) (The Huntington's Disease Collaborative Research Group 1993).

This was the fourth condition for which the disease-causing mutation was found to be a triplet repeat expansion, following the identification of Fragile X Syndrome (FXS) (Verkerk et al. 1991), Spinal and Bulbar Muscular Atrophy (SBMA) (La Spada et al. 1991) and Myotonic Dystrophy type 1 (DM1) (Brook et al. 1992); all neuromuscular disorders.

The largest contribution of research material for linkage analyses came from the Maracaibo region of Venezuela, reported to have the highest concentration of HD affected individuals in the world (Young et al. 1986; The U.S.–Venezuela Collaborative Research Project and Wexler. 2004). The now well-known Venezuela kindreds, at the time of their recruitment, already spanned several generations and included thousands of affected individuals. The Venezuela cohort remains an invaluable resource for research and continues to provide insight into various aspects of disease pathogenesis (The U.S.–Venezuela Collaborative Research Project and Wexler. 2004; Brocklebank et al. 2009).

1.2.2. Molecular genetics

Following the identification of the gene and the disease-causing mutation, numerous studies were performed to elucidate the underlying mechanisms. The triplet 'CAG' encodes the amino acid glutamine (Q); an expanded repeat sequencethus translates into an abnormally long stretch of glutamine residues (polyglutamine) in the mutant protein. As a result, HD is classified as one of nine known polyglutamine (polyQ) disorders, all of which result in progressive and selective neurodegeneration (Murphy et al. 2012).

1.2.2.1. Length of the CAG-tract

Of primary importance was the development of robust techniques for PCR amplification of the CAG repeat to enable repeat sizing and establish pathogenic associations (Andrew et al. 1993; Goldberg et al. 1993a). Initially, a constant length was assumed for the adjacent CCG repeat; however, evidence of polymorphism soon led to the development of accurate sizing using primers specific to the CAG repeat (Andrew et al. 1994a; Bates 2005).

The distribution of CAG repeat lengths was examined in different population groups and a clinical diagnosis confirmed for the majority of subjects (Kremer et al. 1994). Disease symptoms were found to be associated with expanded repeats of 36 CAG or more. Alleles with between 36 and 39 repeats indicated reduced penetrance (RP) based on reports of unaffected elderly individuals (Rubinsztein et al. 1996); while fully penetrant alleles had 40 repeats or more.

An intermediate range of alleles with between 27 and 35 CAG repeats has been shown to exhibit instability on transmission, resulting in new mutations and *de novo* cases of HD (Goldberg et al. 1993b; Goldberg et al. 1995).

These new mutations occur in the absence of a family history; the presence of intermediate alleles (IAs) is therefore particularly pertinent to genetic counselling (Semaka et al. 2006; Semaka et al. 2010).

The international standardisation of pathogenic limits across molecular testing laboratories is critical to accurate diagnosis, more so for those individuals that carry alleles at the thresholds. Defined thresholds for repeat sizing in the HD gene (Nance et al. 1998) have been revised as more data became available, and testing guidelines are regularly reviewed for the purpose of quality control and standardisation (Potter et al. 2004; Losekoot et al. 2013).

Current limits are presented in Figure 2 and highlight the range of repeat lengths associated with pathogenesis (positive), reduced penetrance (RP) and IAs.

Figure 2. Categories of CAG repeat lengths relative to HD pathogenicity. IA: intermediate allele, RP: reduced penetrance allele. Image designed by Dr Alice Hawkins and reproduced with permission from www.predictivetestingforHD.com.

Studies of repeat lengths revealed an important inverse correlation between the age of disease onset and the size of the repeat expansion (Andrew et al. 1993; Duyao et al. 1993; Snell et al. 1993). This relationship was found to be particularly strong for repeat lengths of 60 or more, that were inevitably associated with juvenile onset HD. However, the correlation was somewhat weaker for expansions with between 40-50 repeats, suggesting that additional factors may play a role in determining the age of disease onset (The U.S.–Venezuela Collaborative Research Project and Wexler. 2004).

The contribution of other factors to disease onset was investigated in the cohort of Venezuelan HD families. Using linear regression analysis on the log-transformed age of onset, a residual age of onset was calculated. Variance-components methodology applied to this data suggests that approximately 59% of the variability in residual age of onset is due to heritable genetic modifiers and/or shared environmental factors (The U.S.–Venezuela Collaborative Research Project and Wexler. 2004).

These data provided compelling evidence for the existence of genetic loci that influence the onset of symptoms and endorsed the use of genome-wide association studies in HD patients (Li et al. 2003; Li et al. 2006). Although no modifiers have been confirmed in replication studies (Andresen et al. 2007a); the potential for identification of human-validated targets is sufficient justification for on-going analyses (Lee et al. 2012; Ramos et al. 2012; Gusella et al. 2014).

Using the correlation between the size of the repeat and age of onset, several models have been generated for the purpose of predicting disease onset based on the length of the CAG repeat (Langbehn et al. 2004; Andresen et al. 2007b). Attempts to validate these prediction methods emphasise the incomplete understanding of additional factors that contribute to age of onset of HD and the lack of tangible clinical utility offered by the length of the repeat (Langbehn et al. 2010).

1.2.2.2. Instability of the CAG-tract

The instability of the CAG repeat in the HD gene was initially recognised in the expansion (and occasional contraction) of transmitted chromosomes within families used in linkage analyses (Duyao et al. 1993; The Huntington's Disease Collaborative Research Group 1993).

Similar to other repeat expansion disorders, for example DM1 and Spinocerebellar Ataxia type 1 (SCA1), repeat length tends to increase on transmission (Ranen et al. 1995). The result of this is generally an earlier age of onset in the next generation, a phenomenon known as anticipation (Harper et al. 1992; Duyao et al. 1993; Ranen et al. 1995).

The length of the repeat itself influences instability on transmission; longer repeats are more prone to subsequent expansion (Ranen et al. 1995). Significant increases in repeat length are also strongly associated with the sex of the transmitting parent, with large expansions almost exclusively being transmitted paternally (Duyao et al. 1993; Kremer et al. 1995). However, the precise underlying mechanism is still not fully understood. Repeat instability will be examined further in a subsequent section of this thesis.

1.2.3. Genetic testing

Molecular genetic tests for HD are offered by laboratories worldwide using PCR-based methods which have been reliably shown to detect expanded repeats associated with disease pathogenesis (Nance et al. 1998; Potter et al. 2004; Losekoot et al. 2013). There is however, evidence of intra- and inter-laboratory variation and this warrants consideration when results are reported (Quarrell et al. 2012a).

In an effort to minimise variability, laboratories are required to maintain strict standards of quality control and samples are checked against known sizing controls. Where indicated, apparently homozygous results may require additional investigation by Southern blotting and/or triplet repeat primed PCR (TP PCR), in order to exclude the possibility of very large expansions (Guida et al. 1996; Warner et al. 1996; Losekoot et al. 2013).

In practice and subsequent to the identification of the disease-causing mutation, direct mutation analysis facilitates confirmation of a clinical diagnosis of HD. Accurate analysis of carrier status is also available for at-risk family members in order to alleviate uncertainty and enable future planning (Potter et al. 2004; Losekoot et al. 2013).

Predictive or pre-symptomatic testing protocols have been carefully developed and updated over time and are generally recommended for adults (18 years and older) with a known family history of HD (International Huntington Association and the World Federation of Neurology 1994; MacLeod et al. 2013). The process includes extensive counselling and psychological assessment over several months which aims to provide adequate support and to ensure as far as possible, the individual's ability to cope with the result (Harper et al. 2000; Walker 2007).

Pre-symptomatic testing for HD was first made available following the localisation of the gene to the short arm of chromosome 4 (Meissen et al. 1988; World Federation of Neurology Research Group on Huntington's Chorea 1990). As the first dominantly inherited condition for which predictive testing was feasible, protocols for similar late-onset disorders have been largely modelled on the HD experience (Harper et al. 1990; Hayden 2003).

Prenatal testing may be requested by couples who become aware of a familial risk of HD, again with adequate preparation and counselling. Furthermore, pre-implantation genetic diagnosis (PGD) is also available if desired, to circumvent the possibility of having a child who carries the genetic mutation (Van Rij et al. 2011; Losekoot et al. 2013; Morrison 2014).

1.3 The Epidemiology of Huntington disease

Epidemiology is defined as the study of why and how frequently a particular disease occurs within a defined population – *population at risk* (Coggon et al. 2009). Various methods and measures can be used to contribute to the evaluation and development of public health strategies. The *incidence* of a disease refers to the rate at which new cases occur in the population within a specified time period. On the other hand, *prevalence* is the proportion of the population that is affected by a condition at a given time point (Coggon et al. 2009).

Due to the single-gene Mendelian nature of HD, traditional epidemiological methods have not shown much success in determining the incidence or prevalence (Pringsheim et al. 2012). Available information is therefore not generally based on a comprehensive epidemiological analysis of the population at risk and instead classical genetic approaches and analyses have been employed.

Prior to the identification of the genetic mutation and the availability of direct testing, a diagnosis of HD was largely dependent on the clinician's experience with observing and/or diagnosing the disease. With a confirmatory molecular genetic test now widely accessible, diagnosis is no longer subjective. Nevertheless, issues

of variation in study design and methodology still cast doubt on the precision of ascertainment (Harper 1992b; Walker 2007; Pringsheim et al. 2012).

In spite of these limitations, minimum estimates of prevalence have been made for various populations. It is clear from these reports, that although probably underestimated due to incomplete ascertainment, HD prevalence varies geographically and between ethnic groups (Morrison 2012; Pringsheim et al. 2012).

1.3.1. Geographical variation

In most Caucasian populations, HD prevalence has been estimated at between 4 and 8 individuals per 100 000 (Harper 1992b; Walker 2007; Pringsheim et al. 2012). Certain areas, such as the Maracaibo region in Venezuela and the island of Tasmania, exhibit significantly higher prevalence rates due to known founder effects. The genotypic heterogeneity described in the Venezuelan kindreds shows evidence of several different origins for the mutation; nonetheless, the majority of affected individuals can be traced back to a single Caucasian ancestor (Paradisi et al. 2008).

Reports of HD in Asian populations are few and lacking in detail, however, available information suggests that the prevalence is significantly lower than in Caucasian populations (Chang et al. 1994; Nakashima et al. 1996). Although a systematic survey has not been published confirming the current HD prevalence in Asian populations, more recent data indicates similarly low estimates (Shang et al. 2012; Jiang et al. 2014).

Similar to the Asian populations, African populations are underrepresented in the literature. Apart from a single systematic study performed in South Africa (SA) (Hayden et al. 1980a) there are no comprehensive reportson the occurrence of HD in sub-Saharan Africa. Attempts have been made to calculate minimum estimates of prevalence in a few other African countries, however, these are predominantly based on case reports(Harper 1992b). Notably, all these reports pre-date the identification of the gene and thus molecular confirmation of affected status is lacking.

In 2012, a meta-analysis of published studies on HD prevalence and incidence examined data from 17 prevalence studies post-1985 (Pringsheim et al. 2012), only one of which was performed on an African population in Egypt (Kandil et al. 1994). The study methodology, including diagnosis and survey technique, was inadequately described and results were deemed unreliable (Pringsheim et al. 2012).

Importantly, North African populations show evidence of Arab and European admixture and are genetically different from sub-Saharan African populations (Henn et al. 2012). A distinction between the two regions (North Africa and sub-Saharan Africa) is therefore necessary for genetic analyses.

More recently, Lekoubou and colleagues performed a systematic review of the literature available from sub-Saharan Africa on neurodegenerative diseases, and neurocognitive impairment related to HIV infection (Lekoubou et al. 2014). A total of 19 publications reported on HD, of which nine originated in SA. The authors note the lack of information on neurodegenerative disorders in sub-Saharan Africa. The review (Lekoubou et al. 2014) highlights the necessity of determining the

current burden of disease, particularly given the increasing life expectancy of African populations.

An increase in life expectancy suggests that the disease burden for age-related conditions, such as HD, may be on the rise. This has been discussed with regard to European populations (Morrison 2010; Spinney 2010; Morrison 2012; Pringsheim et al. 2012). Unfortunately a similar comparative assessment cannot be performed for African populations due to the lack of data.

Although HD in African populations was originally thought to be the consequence of admixture with Europeans, subtle clinical differences had been reported in black American kindreds (Hayden et al. 1980a; Folstein et al. 1987). The latter authors speculated on the possibility of an additional disease locus but dismissed this as unlikely in the absence of molecular evidence. However, current data confirm the existence of disease phenocopies which also show varying frequency in different ethnic groups.

1.3.2. Huntington disease phenocopies

Following identification of the HD gene, phenocopies were estimated to account for approximately 1% of clinically diagnosed patients (Andrew et al. 1994b; Kremer et al. 1994). An HD phenocopy is a term used to broadly describe any syndrome that manifests similar symptoms in the absence of an *HTT* expansion (Wild and Tabrizi. 2007).

A number of these have since been characterised and include: Dentatorubral-pallidoluysian atrophy (DRPLA), several sub-types of Spinocerebellar Ataxia (SCA) 1, 2 and 3, and the HD-like (HDL) syndromes HDL1, HDL2 and HDL4 (also known

as SCA17) (Schneider et al. 2007; Wild and Tabrizi. 2007; Martino et al. 2013). These conditions exhibit significant overlap with HD on clinical examination. Importantly, their frequency is reported to vary from one ethnic group to another and differential diagnosis should therefore take this into consideration.

Of particular relevance to the epidemiology of HD in SA is HDL2. The causative mutation was first identified in an African American family as a CTG repeat expansion in the *Junctophilin-3* (*JPH3*) gene (Margolis et al. 2001). Subsequent studies have shown that the mutation occurs almost exclusively in patients with African ethnicity or admixture, with only one patient of middle-eastern ancestry reported so far (Margolis and Ross. 2003; Bardien et al. 2007; Wild et al. 2008).

Indeed, black African patients in SA have been reported to exhibit an almost equal distribution between the HD and HDL2 mutations, following a clinical diagnosis of HD (Magazi et al. 2008). Both the *HTT* and *JPH3* expansions are thus routinely investigated in individuals with the HD phenotype and any indication of black African ancestry, including those from the coloured subpopulation (Bardien et al. 2007; Krause and Greenberg. 2008).

1.3.3. The rise in prevalence

The highest estimates of HD prevalence exist for populations of European descent. Populations of African and Asian origin are reported to have significantly lower estimates, less than 1 individual per 100 000 in some cases (Conneally 1984; Harper 1992b; Pringsheim et al. 2012).

Using a meta-analysis approach and service-based assessment, worldwide prevalence of HD has most recently been approximated at 2.71 per 100 000

individuals (Pringsheim et al. 2012), with distinct variation across geographical regions and ethnic groups. Studies of populations with European ancestry revealed an overall prevalence of 5.70 per 100 000 compared to 0.40 per 100 000 for Asia. There were no studies on African populations that met the necessary criteria (Pringsheim et al. 2012).

Recent discussion and commentary around the true prevalence of HD acknowledges the likelihood that *all* published figures are in fact incomplete estimates, even in developed countries (Morrison 2010; Rawlins 2010; Spinney 2010).

A study in Northern Ireland reported a prevalence of 6.4 per 100 000 individuals in 1991 (Morrison et al. 1995); this estimate had risen to 10.6 by 2001 (Morrison et al. 2011). General practice records in the UK revealed a similar increase in adults (older than 21 years) diagnosed with HD (Evans et al. 2013). Between 1990 and 2010, the estimate HD prevalence rate rose from 5.4 individuals to 12.3 individuals per 100 000. This increase was observed across every age range but was most dramatic in individuals of 51 to 60 years.

In the province of British Columbia (BC) in Canada, the prevalence of HD has been recently estimated at 13.7 per 100 000 individuals (Fisher and Hayden. 2014). Using several different sources of ascertainment, this study provides a best estimate of HD prevalence in BC. Importantly, 81.6 per 100 000 individuals were estimated to be at between 25% and 50% risk of having inherited the disease-causing mutation and thus developing the disease later in life. This may have significant implications for future planning and resource allocation.

The studies outlined here indicate that HD prevalence has been largely underestimated, likely due to incomplete ascertainment. The importance of employing a variety of ascertainment tools when assessing prevalence is thus emphasised. However, uncertainty remains as to whether the reportedly higher rates of prevalence are simply the result of improved ascertainment.

Available information suggests that the prevalence of HD is indeed on the rise (Morrison 2012; Pringsheim and Jette. 2012; Pringsheim et al. 2012). In the developed world, patient organisations have played a significant role in raising awareness and combating the stigma associated with HD. The establishment of disease registries and recruitment for clinical trials and studies of HD in Europe and North America, has also encouraged disclosure.

In addition, a longer-living population may account for the higher prevalence of HD observed in parts of the world with effective healthcare management systems (Loy et al. 2010; Rawlins 2010). Clinical management and other interventions have led to an improved quality of life for HD patients and therefore, affected individuals may be more likely to seek medical care.

Overall prevalence of HD in any population has long been associated with differences in the genetic background of the population under investigation. Intermediate alleles that occur in the general population have been shown to occasionally result in new HD mutations (Myers et al. 1993; Semaka et al. 2006; Semaka et al. 2010). Haplotype analysis has provided considerable information regarding mutation origins in different populations and may subsequently reveal clues to the propensity for expansion shown by IAs.

1.3.4. Haplotype differences

Haplotypes constructed using microsatellites and other polymorphic markers prior to the isolation of the gene were the first indication that the HD mutation had several different origins (MacDonald et al. 1991; MacDonald et al. 1992). Subsequently, haplotype analysis using known polymorphisms within the gene, including the CAG-tract and the adjacent CCG polymorphic repeat, revealed evidence of differences in the genetic background of various population groups. These were suggested to be the basis of variation in regional HD prevalence (Andrew et al. 1994a; Squitieri et al. 1994; Rubinsztein et al. 1995).

The availability of single nucleotide polymorphisms (SNPs) has made still more detailed analyses possible. In 2009, a comprehensive analysis of the genetic diversity in the *HTT* region identified a subset of tag SNPs (tSNPs) that were significantly associated with disease chromosomes in a European cohort (Warby et al. 2009). A number of these SNPs were determined to be closely linked to expanded CAG-tracts and constituted a cluster of similar haplotypes (designated haplogroup A) found on 95% of disease chromosomes in the cohort of European patients.

Remarkably, the same haplotypes were found to be enriched on chromosomes in the intermediate range (27-35 CAG) compared to the general population (<27 CAG). The association of specific SNPs with both expanded and intermediate chromosomes suggested that *de novo* mutations occurred on a similar genetic background; supporting the hypothesis that specific haplotypes are predisposed to expansion (Warby et al. 2009).

A subsequent study investigated the genetic background of an East Asian population (China and Japan), in an effort to determine whether the HD mutation occurred on a similar haplotype background to the European population(Warby et al. 2011). Interestingly, whereas expansion was most associated with haplogroup A (variants A1 and A2) in the European HD cohort, expanded chromosomes in the East Asian population were associated with haplogroup C.

The authors proposed two hypotheses to explain why the mutation was more likely to occur on a specific genetic background (or haplotype): 1) Repeat instability is largely due to the size of the CAG repeat and chromosomes with 'high-risk' haplotypes had a larger average CAG-tract size, 2) Genetic *cis*-elements influence the instability of the CAG-tract and increase the propensity for expansion when the repeat is in the intermediate range (Warby et al. 2011).

Notably, the average CAG repeat size in the East Asian population was shown to be lower than that of the European population. However, that difference did not hold for chromosomes on haplogroup C. In fact, haplogroup C in both populations exhibited a similar association with expansion and thus a comparable estimated prevalence; this has been proposed as a 'baseline frequency of HD' (Warby et al. 2011). This study did not include any African populations and no investigationwas made into sequence differences that may influence repeat instability, and thus expansion of specific haplotypes.

1.3.5. Intermediate alleles

The prevalence of HD in any population is thought to be a fine balance between the new mutation rate and the elimination of large expansions (Rubinsztein et al. 1994; Semaka et al. 2010). In HD families, transmission of an allele with over 60

CAG repeats (usually paternal) results in juvenile onset and this allele is generally not passed to the next generation. *De novo* mutations have been reported in individuals with no family history of HD, and shown to arise from a parent with an allele in the intermediate range (Goldberg et al. 1993b; Myers et al. 1993; Almqvist et al. 2001; Semaka et al. 2010).

The new mutation rate may thus be estimated from the proportion of IAs in the general population that result in a pathogenic expansion. Due to the instability of the CAG repeat, these changes may occur sequentially (usually in small increments) from one generation to the next. The frequency of IAs in the general population is reported at between 1 and 3.9% with the rate of new mutations from this pool estimated at 10% (Falush et al. 2001; Semaka et al. 2006).

Susceptibility to instability has been investigated by sperm analyses and exploring familial transmissions (Goldberg et al. 1995; Chong et al. 1997; Brocklebank et al. 2009), however, there is inconclusive evidence regarding the stability of IAs and studies of larger sample sizes are required (Semaka et al. 2010). There has been no assessment of the frequency of IAs or the occurrence of new mutations in the South African population.

Although IAs are considered non-pathogenic, new data suggest that they may confer a behavioural phenotype (Killoran et al. 2013). The Prospective Huntington At Risk Observational Study (PHAROS) is an observational study comprising over 1000 unaffected at-risk participants at 43 different sites in North America (Huntington Study Group PHAROS Investigators 2006). The cohort of individuals with IAs numbered 50 and psychiatric manifestations in this group included apathy and suicidal ideation, commonly associated with prodromal HD (Killoran et

al. 2013). The prodromal stage denotes the period between initial symptoms and clinically defined disease onset.

The authors suggest that this phenotype may be a result of familial associations; the individuals in the study having possibly inherited genetic modifiers of the disease that are believed to alter the disease phenotype in affected relatives. Alternatively, and with evidence from individual case reports of late-onset HD in individuals with IAs, it may be necessary to adjust the limits of the disease-associated repeat of range (Killoran et al. 2013). However, this recommendation would require a definitive association of IAs with disease.

1.4 Huntington disease in South Africa

1.4.1. The people of South Africa

The Republic of South Africa is located at the southern-most tip of the African continent and is divided into nine provinces. The country has a rich diversity stemming directly from its historical background. This diversity is reflected in the fact that SA has 11 official languages and a number of others which are unofficially recognised. As a first language, English ranks only fourth, with Zulu, Xhosa and Afrikaans taking the top three positions (spoken as first language).

The population is estimated at over 50 million with extensive variation in culture and ethnic origin (www.statssa.gov.za). Following the 2011 census, Statistics South Africa (SSA) estimated that just under 80% of the population was comprised of Black Africans; with White (or Caucasian) and Coloured (or mixed ancestry)

people representing significantly smaller proportions and people of Indian/Asian origin making up the smallest proportion (Figure 3).

It is worth reiterating here that the identity of an individual is a complex structure involving language, culture and other non-biological aspects. Genetic ancestry may therefore not reflect the perceived identity of an individual or group.

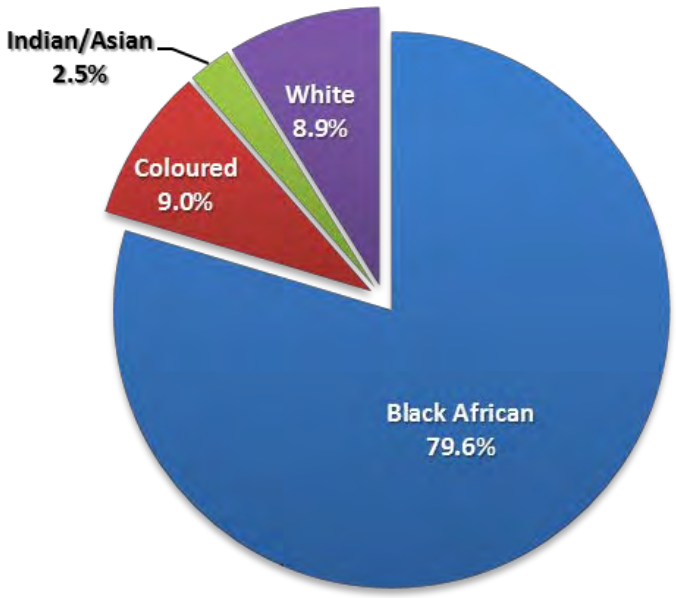


Figure 3. A proportional representation of the South African population. The pie-chart is based on available 2011 census data (www.statssa.gov.za).

1.4.1.1. Black South Africans

As indicated in the figure (Figure 3), the majority of the South African population is of black African ethnicity (also referred to as indigenous African). This group comprises the original people of Southern Africa, the Khoe (formerly Hottentot) and the San (formerly Bushmen); and to a significantly larger extent the Bantu-speaking tribes that migrated into Southern Africa, likely from west and central Africa (Lane et al. 2002; Soodyall et al. 2008; Montano et al. 2011).

The San and Khoe peoples are currently grouped together and known as the Khoe-San. They are represented today by small groups of people in SA, Namibia and Botswana who self-identify as belonging to either of the two groups (Soodyall et al. 2008). Analyses of mitochondrial DNA and Y chromosomes haplogroups, indicate an ancient common origin for the Khoe and the San peoples.

Bantu speakers can be broadly classified on linguistic grounds into Western Bantu who speak an older more diverse language, and the Eastern Bantu (Lane et al. 2002; Montano et al. 2011). The estimated 400 Bantu languages are spoken by over 200 million people across sub-Saharan Africa, and may have originated in central West Africa.

Urbanisation in SA has led to major socio-economic changes that have significantly altered the demographics of the black subpopulation by bringing many different people together. The relationships between genes, languages and geographical location has been the subject of several studies and these have revealed evidence of extensive admixture, amongst different African tribes and with contributions from other peoples (Lane et al. 2002; Soodyall et al. 2008; Montano et al. 2011).

1.4.1.2. Coloured South Africans

The coloured population is a large and ethnically distinct mixed ancestry subpopulation with diverse genetic and cultural contributions. An estimated 9% of the population self-identifies as 'Coloured' and is recognised as a group specific to SA. The majority are historically Afrikaans speakers and largely resident in the Western Cape province based on the most recent census data (www.statssa.gov.za).

As one of the most admixed populations in the world, the historical origins of this group lie in the settlement of Europeans at the Cape and the subsequent enslavement of different peoples. Contributions have been identified from the native Khoe-San, local Bantu speakers and people from Mozambique, Madagascar, India and Indonesia (Mountain 2003; Patterson et al. 2010).

Several investigations have been made into the genetic constitution of the coloured population (de Wit et al. 2010; Quintana-Murci et al. 2010; Chimusa et al. 2013; Daya et al. 2013). Consistent with historical evidence, the results indicate that major components are Khoe-San (largely maternal), Bantu and European; with a smaller less-significant Asian contribution (Quintana-Murci et al. 2010).

Although the coloured population from the Western Cape is known to be different from the groups found in the rest of the country, there is no official distinction made between them (www.statssa.gov.za). Notably, the genetic makeup of the Cape Coloured, as they are commonly known, comprises indigenous African (Khoe-San), European and Indonesian (Javanese) components (de Wit et al. 2010; Daya et al. 2013). Outside of the Western Cape, the coloured groups are predominantly

descended from intermarriages between Bantu-speakers and Europeans who penetrated the interior of SA.

1.4.1.3. White South Africans

In 1652, the Dutch established a refreshment station at the Cape of Good Hope which marked the beginning of European settlement of the Cape (Jenkins 1990). By 1800, the Afrikaans community numbered 15 000, all descendants of a small group of Dutch, German and French immigrants. The 19th century saw the arrival of other European immigrants, although this did not result in much integration with the Afrikaans peoples.

In the 1970s, several genetic conditions were already recognised in this group: Porphyria variegate, Colonis polyposis and Sclerosteosis, to name a few (Beighton 1976). Subsequently, the Afrikaner community has been shown to exhibit a number of founder events including for the HD mutation (Hayden et al. 1980b; Scholefield and Greenberg. 2007). Other founder mutations have been reported for: Long QT Syndrome, Keratolytic Winter Erythema, Hypertrophic Cardiomyopathy and SCA7 (de Jager et al. 1996; Starfield et al. 1997; Moolman-Smook et al. 1999; Greenberg et al. 2006).

1.4.1.4. Indian/Asian South Africans

Comprising the smallest subpopulation in SA, this group is descended largely from people who entered the country from around 1860 (Winship and Beighton. 2011). The majority were Hindus indentured to work in the sugar cane fields of Natal, now the province of KwaZulu-Natal. Predictably, the largest proportion of Indian or Asian peoples anywhere in the country is resident in this province. Smaller groups settled in Johannesburg, Cape Town and other centres.

It is noteworthy that HD is not one of the conditions reported to be prevalent in the Indian community of SA (Winship and Beighton. 2011). To-date, molecular testing records have fewer than five families with a positive diagnosis of HD. Due to the scarcity of known families with HD in this group, the analyses reported on in this thesis did not include Indian/Asian South Africans.

1.4.2. Reports on Huntington disease in South Africa

The proposition of an apparently high occurrence of HD in SA was first made in 1962 by Klintworth in a report on 16 different families diagnosed with HD, including one black family (Klintworth 1962). Of these 16 families, a foreign origin could be traced for nine, with ancestors from England, Austria, Holland, Russia and Germany. Three of the Afrikaner families were determined to have a common Dutch ancestor. In the same year (1962), the first conference on Human and Clinical Genetics was held in Johannesburg; several genetic disorders relevant to SA, including HD, were discussed (Jenkins 1990).

The 1970s marked a period of particular interest in human genetics as a specialty. In 1972, the Department of Human Genetics was established at the University of Cape Town (UCT) (Beighton et al. 2012). This initiative was followed by the creation of a Chair of Human Genetics in 1975, at the University of the Witwatersrand (Wits) in Johannesburg (Jenkins 1990). Human genetics was thus positioned as an important field of research which encouraged the training of a large number of medical geneticists and molecular scientists.

The occurrence of HD across the different ethnic groups of SA, and its presence in both adult and juvenile forms was documented in at least 2 publications in the 1970s (Saffer et al. 1974; Glass and Saffer. 1979).

As a young medical graduate in the mid-1970s, Dr Michael R. Hayden undertook his doctoral research at UCT with the overall aim of better understanding the occurrence of HD in SA. The project investigated the origin of the HD mutation in the South African population based on genealogical analyses and attempted to estimate the prevalence of HD in the country (Hayden 1979).

A pilot study in the Western Cape Province preceded the country-wide investigation, and the published results estimated the frequency of HD in the Cape Coloured population. Minimum prevalence was estimated at 3.5 per 100 000 individuals in this group, with a relatively higher frequency of juvenile HD than reported elsewhere (Hayden and Beighton. 1977; Hayden et al. 1982).

Over several years, all known cases were painstakingly documented and followed up, making it possible to record minimal estimates of disease prevalence. Subsequently, the results of Dr Hayden's completed doctoral thesis were published in a series of manuscripts, three of which appeared in the 1980 August issue of the South African Medical Journal (SAMJ) (Hayden et al. 1980a; Hayden et al. 1980b; Hayden et al. 1980c).

Minimum prevalence estimates were made for the white, coloured and black population groups in South Africa; 22.2, 21.7 and 0.1 per million respectively (Hayden et al. 1980a). These figures were based on a total of 481 cases of HD investigated around the country, including 11 black individuals. The authors discussed the probability that these figures were underestimates due to incomplete ascertainment, believed to be the direct result of a poor understanding of the disorder, both socially and medically.

Another publication reported on an extensive genealogical study in the Afrikaner population,which traced the HD mutation over 14 generations to a single Dutch ancestor who was born in the Cape in 1670 (Hayden et al. 1980b). Additional commentary was made on social aspects of the disease including the imposition of economic and social burdens, and importantly the observed antisocial behaviour such as criminal tendencies (Hayden et al. 1980c).

1.4.3. Molecular investigations

The molecular laboratory at the Division of Human Genetics, UCT (referred to here as 'the division'), has performed a number of studies on HD. The recruitment of HD patients and family members initiated in the 1980s for research into the disease, provided an invaluable resource in biological material (stored blood and DNA samples) (Hayden et al. 1987).

As part of a large international collaboration, a doctoral research project (Greenberg 1990) at the division undertook linkage analysis of several South African HD families. Genetic linkage to the D4S10 locus was established in 10 families from the three main ethnic groups. The haplotypes established using restriction fragment length polymorphisms (RFLPs) were distinct across the different populations of SA (Greenberg et al. 1991).

Using disequilibrium data, a founder effect was indicated in the six coloured families investigated. Two different haplotypes were constructed for each of the black families, neither of which was similar to those identified in the white and coloured families. The results of this study, the first of its kind performing linkage analysis in black African HD families, indicated the unique origin of the HD mutation in this population group (Greenberg et al. 1991).

Following the identification of the HD gene, a master's project within the division constructed haplotypes across a narrower region (September 1999). Using the CAG and CCG repeats, the microsatellite marker D4S127 and the deletion polymorphism Δ2642, haplotypes were constructed for eight families - two each from the black and white subpopulations, and four from the coloured subpopulation.

The results revealed four major haplotypes (and three sub haplotypes) suggesting multiple origins for the mutation. Importantly, the two black African families were shown to have two distinct haplotypes, consistent with the results of the previous study (Greenberg 1990; September 1999).

In 2005, a third study attempted to delineate the origins of the HD mutation in SA based on SNPs (Scholefield 2005). Human genetic variation can be characterised by genotyping SNPs, one of the most abundant polymorphisms in the DNA sequence. The availability and density of SNPs enables researchers to narrow the region under investigation. The (generally) bi-allelic nature of SNPs also makes them more amenable to computational analysis, facilitating the estimation of disease-associated haplotypes and thus mutation origins.

Using the genotypes of six SNPs across the *HTT* gene region, SNP-based haplotypes were constructed for 15 South African HD families representing the black (two families), coloured (six families) and white (seven families) subpopulations (Scholefield and Greenberg. 2007). The white families were further subdivided into Afrikaans and English speakers. Although genotyping of loci was incomplete in the black families (due to depleted DNA stocks), a common core haplotype was identified in two Afrikaner and four coloured families.

The ancestral origins of these six families, implicated a single individual of Dutch origin as a major source of the HD mutation in the SA population as had been reported previously (Hayden et al. 1980b; Scholefield and Greenberg. 2007). Computational analysis of SNP genotypes from control individuals indicated that this founder haplotype was likely to extend into the coloured general population (Scholefield and Greenberg. 2007).

Taken together, these molecular studies provided evidence that the HD gene was not solely the result of immigration as had been suggested (Hayden et al. 1980a). In fact, several distinct origins are apparent in the different population groups. Challenges encountered with microsatellite-based haplotypes were diminished by the more informative SNPs; the abundance of SNP data in combination with high throughput genotyping technologies, has been subsequently exploited by this and other detailed analyses.

1.4.4. Clinical and testing services

In 1978, Dr Hayden established a clinic specialising in the management and care of HD patients and their families at Groote Schuur hospital (GSH) in Cape Town. Importantly, this was the first multiracial clinic at GSH (personal communication, Dr Hayden). Patients were referred to the HD clinic from around the country, and throughout the 1980s biological material was collected with consent for research and stored at the Division of Human Genetics, UCT.

This clinic was later expanded to include the management of all patients diagnosed with neurogenetic conditions, and is operational to date on a monthly basis. Associate Professors Alan Bryer and Jeannine Heckmann are currently the

consultant neurologists, supported by other specialist clinicians, genetic counsellors and nurses.

Subsequent to the identification of the gene in 1993, accurate confirmation of a clinical diagnosis of HD was made possible by direct mutation analysis (The Huntington's Disease Collaborative Research Group 1993). South African DNA samples were initially sent to the Hayden laboratory in Vancouver, Canada for molecular genetic testing. In 1995 a facility was established at UCT with technical support and instruction provided by the Hayden laboratory (personal communication, Dr Hayden). This was followed by a second facility at Wits in 2001 (Krause and Greenberg. 2008; Futter et al. 2009; Sizer et al. 2012).

At the time the present study was initiated, these were (and still are) the only two centres in the country offering a molecular confirmation of diagnosis for HD in the public health care system. The service is provided under the National Health Laboratory Service (NHLS), in affiliation with the two academic institutions: UCT and Wits. The two centres receive all referrals for HD testing from the nine provinces of South Africa, with NHLS/UCT catering largely to the Cape provinces (Western, Northern and Eastern) and NHLS/Wits to the rest of the country (Figure 4).

Figure 4. A provincial map of South Africa showing the locations of the two public testing facilitiesfor HD (indicated by the red stars). Map sourced fromwww.hdwallpapers-3d.com.

A comprehensive service is offered by experienced specialists at both centres. A molecular confirmation of diagnosis is requested, following a clinical assessment performed by a medical geneticist and/or neurologist. Additional services are offered by psychiatrists, clinical psychologists, genetic counsellors and nurses, and social workers. Genetic counselling services are fundamental for the support of patients and families with HD and in SA are recognised under the Health Professions Council of South Africa (HPCSA).

Predictive testing is accessible to family members who would like to know their carrier status. The protocols have been established according to international guidelines and tailored for the specific needs of the population (Greenberg et al. 1996; Krause and Greenberg. 2008; Futter et al. 2009; Sizer et al. 2012). In addition, prenatal testing is available upon request.

Due to the limited resources available in the public health care system, the majority of families have been left to fend for themselves on a daily basis following a diagnosis of HD. There are no existing facilities specialised for the management and care of patients and a number have, on occasion, ended up in psychiatric institutions.

There is therefore an urgent need to determine the medical and social burden pertaining to HD in SA. In addition, understanding the molecular genetics of HD in diverse populations would potentially contribute to global research efforts into a cure and effective therapeutics for HD.

1.5 Project aims and objectives

The project which forms the subject of this thesis was a collaborative effort between the two HD testing centres and institutions in SA: NHLS/UCT and NHLS/Wits; and the Centre for Molecular Medicine and Therapeutics (CMMT)/University of British Columbia (UBC) in Vancouver, Canada. Ethics approval was granted under HREC REF 215/2012 (UCT), M110433/M10745 (Wits), UBC-CREB H05-7032/H06-70467 (UBC); and was renewed annually for the duration.

The overall aim was to investigate the genetic background of the HD gene in the South African population by evaluating three key features: the distribution of CAG and CCG repeats,the occurrence of population-specific *HTT* haplotypes and the potential influence of *cis*-acting elements on repeat instability. It was also necessary to give an overview of current disease prevalence in order to provide context for these results.

Aim 1

Determine the distribution of CAG and CCG repeat lengths in the general population of SA.

Objectives

1. Genotype and size the CAG-tract in a large cohort (at least 1000 individuals) from the general population, with significant representation of each subpopulation
2. Determine mean, median and mode values and the distribution of IAs for each subpopulation
3. Genotype the polymorphic CCG repeat and determine the pattern of allele distribution within each subpopulation

Aim 2

Construct *HTT* haplotypes and determine their association with disease in the different subpopulations of SA.

Objectives

1. Genotype SNPs across the *HTT* gene in a cohort of HD families and individuals from the general population
2. Genotype and size the CAG-tract and CCG repeat for all family members
3. Construct SNP-based haplotypes and identify those associated with disease; sizing data (CAG and CCG repeats) is needed to establish phase

Aim 3

Investigate the presence of *cis*-acting elements within the *HTT* gene that may influence repeat instability in a family with juvenile-onset HD.

Objectives

1. Identify affected individuals with confirmed expanded alleles that have shown unstable transmission across at least one generation
2. Perform direct DNA sequencing and analysis of the region encompassing the CAG-tract, including exon 1 and the promoter region
3. Investigate the presence of sequence variants that may influence observed CAG-tract instability

Aim 4

Estimate the current prevalence of HD in the South African population by means of genetic ascertainment.

Objectives

1. Assess all records of HD testing performed in SA between 1995 and 2010 (a 16-year period)
2. Identify the total number of individuals with a molecularly confirmed diagnosis and calculate minimum estimates of the current prevalence of HD

SECTION II

This section presents the molecular investigations that address the three main aims of the project. Each chapter includes a brief introduction, the methods and materials employed, results and discussion.

The sizing and haplotype studies (Chapters 2 and 3) were performed by the candidate at the Hayden laboratory with support from Chris Kay (CK), Marlies Ketelaar (MK) and Jennifer Collins (JC). CK and MK contributed to molecular assays and data analysis for CAG/CCG sizing and haplotyping. JC performed the manual assignment of haplotypes to specific variants. CK contributed to the statistical analysis of haplogroup variants present in different groups. These results have been published in part, Baine et al. 2013 (included in the Appendices)

The analyses reported in Chapter 4 were performed at the Division of Human Genetics laboratory (UCT) by the candidate. All patient recordsat NHLS/UCT and NHLS/Wits were reviewed by the candidate. The sourcing and aliquoting of DNA control samples used in the study was performed by the candidate.

CHAPTER 2: THE DISTRIBUTION OF CAG AND CCG REPEATS IN THE SOUTH AFRICAN POPULATION

2.1 Introduction

The CAG-tract in the HD gene has been extensively investigated since its identification as the pathogenic mutation (The Huntington's Disease Collaborative Research Group 1993). Publications immediately following the isolation of the gene defined the size limits associated with disease symptoms (Snell et al. 1993; Kremer et al. 1994). A number of polymorphisms around the repeat were also reported and haplotype analysis used to explore their association with expansion (Pecheux et al. 1995; Rubinsztein et al. 1995; Rubinsztein et al. 1996).

Differences in the distribution of unexpanded CAG repeats are evident across ethnically diverse population groups. The mean CAG calculated for general population alleles has been linked to the prevalence of HD in that population (Warby et al. 2011; Pulkes et al. 2014). In populations with a low prevalence of HD, such as East Asia (Japan), the mean CAG-tract size was found to be lower than in those with a high HD prevalence.

In addition, an association has been made between HD prevalence and the proportion of IAs in the general population. Since IAs have the potential to contribute *de novo* mutations (Rubinsztein et al. 1994; Semaka et al. 2010), it can be hypothesised that a lower proportion of IAs may correlate with a lower prevalence of HD in a given population.

In SA, there has been some evidence of differences in the distribution of CAG repeat lengths reported in control cohorts. A previous study at UCT investigated the distribution of CAG-tract size in a total of 300 control chromosomes, with each of the three subpopulations represented by 100 chromosomes (September 1999). The results are summarised below.

Table 1. Overview of the CAG sizing results from a South African cohort investigated in 1999.

	Black	Coloured	White
Total N	100	100	100
Range	12 – 28	14 - 27	11 – 37*
Median	20	20	24
Mean	17.57	18.29	18.73
Mode	16	18	18
% IAs	1	1	-

IAs – intermediate alleles, N – number. *Range includes a single large allele of CAG37 with the second largest allele being CAG25. Data taken from the results of a Master's project (September 1999)

In this sample, although the range for the white subpopulation indicates 11-37, only a single allele (CAG37) in the zone of reduced penetrance was reported, no other alleles larger than CAG25 were present (Table 1). One allele in the IA range was observed in each of the black and coloured subpopulations (CAG28 and CAG27 respectively). This equates to an IA frequency of 1% for the black and coloured groups.

The same study (September 1999), explored the distribution of the polymorphic CCG repeat adjacent to the CAG-tract (data not shown here). In the black subpopulation, CCG10 was the most frequent allele (44%); while the coloured and white subpopulations both had CCG7 as the most frequent allele (43% and 53%

respectively). In all three groups, the CCG allele sizes ranged between 6 and 11 with a median of 8.

Another study on a random sample of 100 individuals (200 chromosomes) from each of the black and white subpopulations was performed at the NHLS/Wits laboratory. A summary of the CAG repeat sizing results is shown in Table 2 below and shows some parallels with the data inTable 1above.

Table 2. Overview of the CAG sizing results from a South African cohort investigated in 2001.

	Black	**White**
Total N	198	200
Range	11 - 32	8 - 28
Median	17	18
Mean	17.24	18.27
Mode	15	17
% IAs (N)	2 (4)	1 (2)

IAs – intermediate alleles , N – number. Data taken from the results of an Honours project (by Claire Hetem).

The mean CAG in the black subpopulation (17.24) is similar to that reported in the previous study (17.57, Table 1). The white subpopulation has a mean CAG of 18.27, comparable to 18.73 (Table 1). The modal CAG values in this second cohort (Table 2) are slightly lower for both the black (15) and the white (17) groups, compared to the modal CAG values (16 and 18, respectively) in the first cohort (Table 1).

An IA frequency of 2% was calculated for the black subpopulation and in this larger sample (N=200), the white subpopulation has an IA frequency of 1% in contrast to the absence of IAs in the first cohort (Table 1).

The studies summarised here were both based on small cohorts which limits the reliability of statistical measures in relation to the general population. The first aim of this current study was thus to genotype the CAG-tract and CCG repeat in a large sample taken from the general South African population, in order to obtain significant representation of the distribution. Of particular interest were the black and coloured subpopulations due to the lack of information regarding the HD gene in African populations.

2.2 Methods

2.2.1. Sample collection and preparation

Individuals representative of the general South African population were identified from various sources. The two molecular genetic laboratories (UCT and Wits) have samples that have been collected and stored with consent for research. These samples were de-linked and anonymised for use as controls. Additional de-linked and anonymised controls were sourced from various researchers with the approval of the Human Research Ethics Committees. A total of 1341 DNA aliquots from the three main subpopulations of SA were utilised in this study.

All samples were identified and transported by the candidate, to the Hayden laboratory at the CMMT in Vancouver for analysis. DNA quantification was performed using a Quant-iT ™ PicoGreen® dsDNA Assay Kit (Life Technologies) according to manufacturer's instructions. The assay is specific to double-stranded DNA (dsDNA) and can be performed in 96-well plates. This enables the quantification of a large number of samples in a single batch. Where appropriate, working dilutions (100 ng/μL) were made.

2.2.2. CAG and CCG sizing

The lengths of the CAG repeat and the adjacent CCG repeat were determined by polymerase chain reaction (PCR) amplification and capillary electrophoresis. Original primer sequences and PCR conditions (Andrew et al. 1994a) have been modified over time for optimal amplification of the 2 repeats. The primers used were available at the Hayden Laboratory (CMMT, Vancouver). The figure below shows the binding positions of these primers.

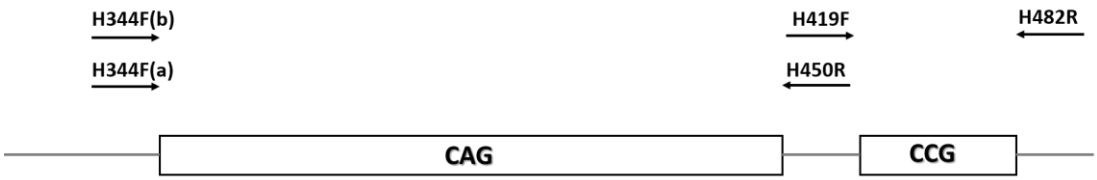


Figure 5. Schematic representation of the CAG and CCG repeats showing primer binding positions. The pure CAG repeat is amplified in one reaction (H344F(a)/H450R); the multiplex reaction amplifies the total CAG and CCG (H344F(b)/H482R), and the pure CCG (H419F/H482R) in a single step. H344F(a) and H344(b) carry different fluorescent tags to enable differentiation of the reaction products (details in Appendices).

Amplification was performed in two separate reactions: the first is specific to the CAG repeat; while the second is a multiplex that amplifies (in the same reaction) the CCG repeat, and the total CAG/CCG repeat. Details of the primer sequences (including modifications), and optimal reaction conditions are given in the appendices.

Each reaction was performed in a 96-well plate and included a negative control (no DNA) and controls with known CAG repeat sizes. Subsequent to amplification, products from the two reactions were pooled and submitted to the analytical facility at UBC. Fluorescent tags on the primers (appendices) allow the separation and recognition of the different PCR products by capillary electrophoresis.

Following capillary electrophoresis, the results were analysed using GeneMapper (v4.0) software (Applied Biosystems) and sizing calls for each sample checked manually to ensure the correct CAG and CCG alleles were recorded.

2.2.3. Statistical analysis

All statistical analyses were performed using Stata/IC (v12.0) (Stata Corporation 2011). Significant differences are reported with *p*-values and where *p* does not reach significance ($p \geq 0.05$), indicated with n.s. (not significant). Mean CAG values were compared using one-way ANOVA, pair-wise comparison of means between the subpopulations was performed.

2.3 Results

This study attempted to address the lack of information regarding the *HTT* gene in black African populations by genotyping as many control samples as could be obtained. In addition, the coloured subpopulation is unique to SA and thus of particular interest. Therefore, the majority of samples collected and reported on here are of black African or coloured ethnicity.

2.3.1. CAG sizes in the general population

An overview of the results for each subpopulation is presented in Table 3.

Table 3. CAG-tract sizing results for the general population cohort investigated in this project.

	Black African	Coloured	White
N	1204	1042	436
Range	11- 33	9 - 36	9 - 31
Median	17	17	17
Mean (SD)	16.88 (±2.73)	17.62 (±2.96)	18.48 (±3.18)
Mode	15	17	17
% IAs (N)	1.50 (18)	1.73 (18)	3.67 (16)
% Large alleles (N)	2.16 (26)	5.18 (54)	6.65 (29)

IAs – intermediate alleles, N – number, SD – standard deviation

The range of normal CAG sizes is widest in the coloured subpopulation (9-33, excluding the two alleles with 36 repeats which lie in the zone of reduced penetrance). The median value for all three subpopulations is 17 (Table 3) and modal CAG is different in the black subpopulation (15) compared to the white and coloured subpopulations (17).

There is a notable difference in the proportion of IAs, which in the white subpopulation (3.67%) is more than twice that in the coloured (1.73%) and black

groups (1.5%) (Table 3). In addition, the proportion of large alleles (with between 23 and 26 CAG repeats) differs across the three subpopulations. Although the proportions of IAs and large normal alleles would be expected to show consistent differences, this is not true of the cohort investigated here and will be discussed further.

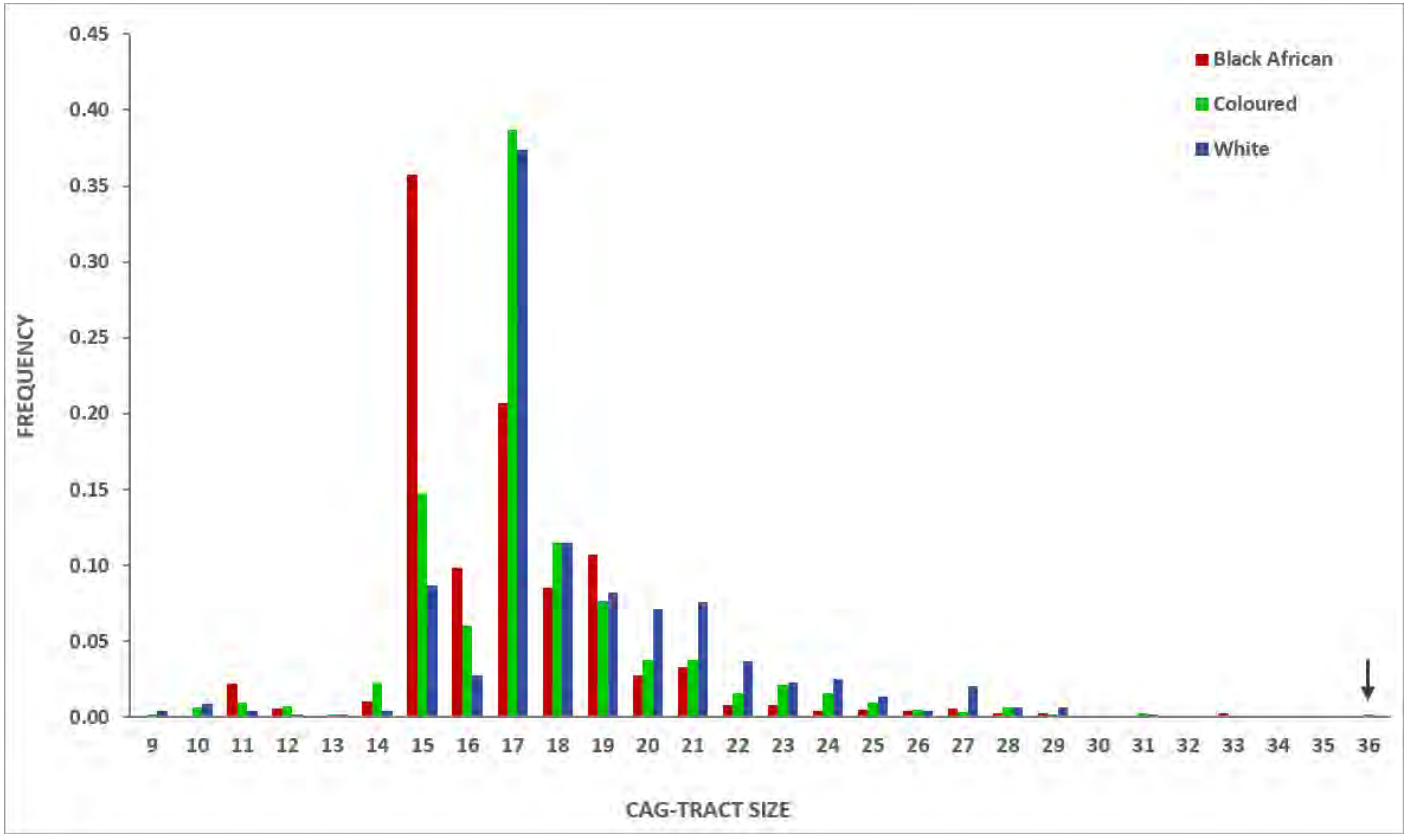


Figure 6. Frequency distribution of CAG repeats on general population alleles. Differences in the modal CAG of the three subpopulations are clearly evident in the peaks of the bar graph; similar in the coloured and white groups (CAG17) and lower for the black subpopulation (CAG15). Two alleles in the zone of reduced penetrance (RP) were observed in the coloured subpopulation (indicated by the arrow).

A one-way ANOVA indicated a significant difference between the three subpopulations ($p < 0.0001$). Pairwise comparison of mean CAG values revealed a significant difference ($p < 0.001$) between the black (95% CI: 16.72-17.05) and white (95% CI: 18.21-18.75) groups.

Using Stata, two-way histograms were constructed for each subpopulation. CAG repeat size is graphed on the x-axis and probability density on the y-axis (Figure 7). A normal density curve is overlaid on each of the histograms to approximate normal distribution. The highest point of the normal density curve is the average of the distribution; this point is observed to shift across the three subpopulations and is a clear indication of the differences in the calculated mean CAG values (Figure 7).

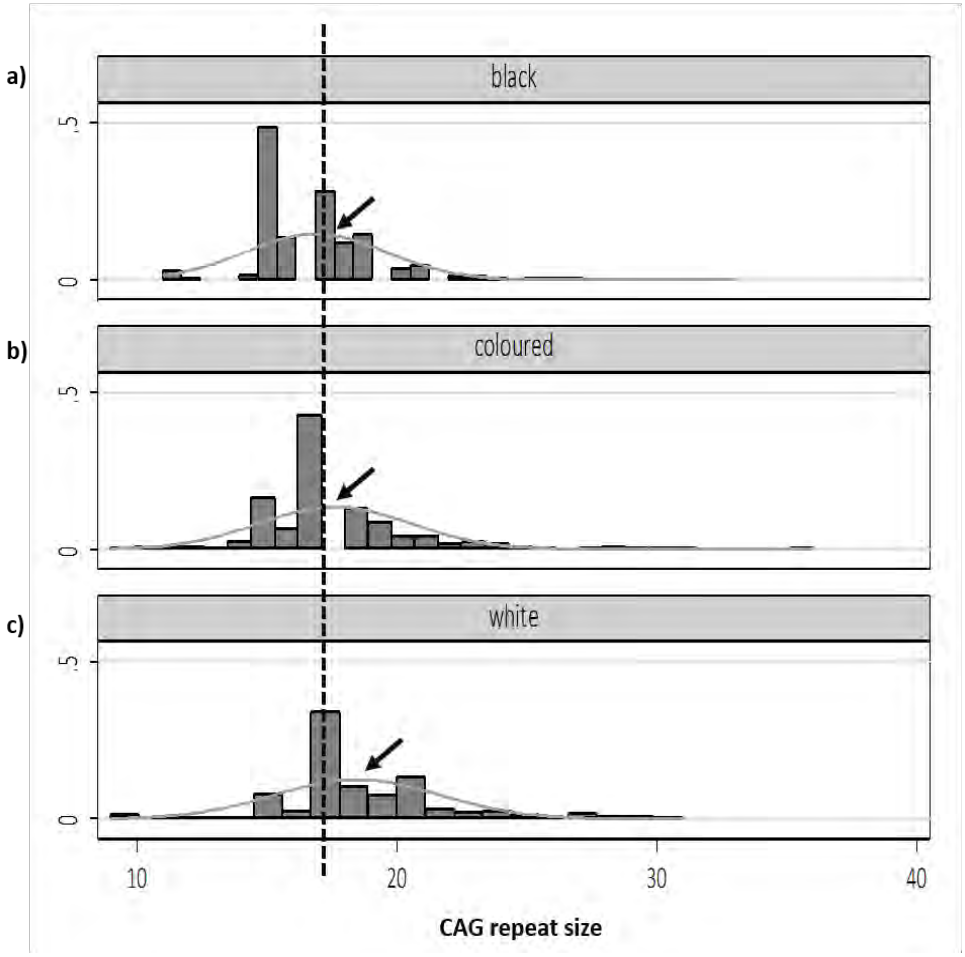


Figure 7. Histogram plots for the three subpopulations. Normal density curves are overlaid on each histogram. The dotted line approximates the median point (average) in the black subpopulation and arrows indicate this point in each group. There is an observable shift of this median in the graphs from the black subpopulation (a) to the coloured subpopulation (b); and the furthest shift is evident in the white subpopulation (c).

2.3.2. CCG alleles in the general population

The distribution of CCG alleles also showed marked variation across the three subpopulations.

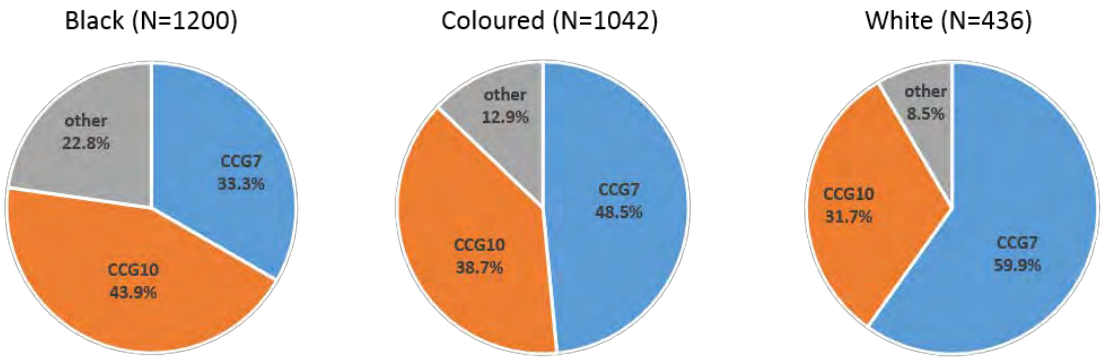


Figure 8. The distribution of CCG alleles across the three subpopulations. The largest proportion of CCG alleles in the white subpopulation are CCG7, compared to the black subpopulation where CCG10 is the most commonly observed. The black subpopulation also has a significant proportion of CCG alleles that are not the common CCG7 and CCG10.

The difference in the proportional variation of CCG alleles is evident from Figure 8above. A chi-square test revealed a significant difference ($p < 0.001$, Pearson chi-square test) between the proportions of CCG alleles in the different categories.

In the black subpopulation, the majority of alleles are CCG10 (43.9%), compared to CCG7 (33.3%). This is in contrast to both the white and coloured subpopulations where the CCG7 allele is more frequent than the CCG10.

A total of 22.8% of the CCG alleles in the black subpopulation are 'other' than CCG7/CCG10, compared to only 12.9% and 8.5% in the coloured and white subpopulations respectively (Figure 8). These results are highlighted (**in bold text**) in the detailed CCG sizing table below (Table 4).

Table 4. CCG sizing results for the general population cohort investigated in this project.

CCG allele	Black African		Coloured		White	
	N	proportion	N	proportion	N	proportion
4	0	0.0%	1	0.1%	1	0.2%
5	5	0.4%	4	0.4%	0	0.0%
6	37	3.1%	11	1.1%	0	0.0%
7	400	**33.3%**	505	**48.5%**	261	**59.9%**
8	95	7.9%	55	5.3%	12	2.8%
9	84	7.0%	52	5.0%	23	5.3%
10	527	**43.9%**	403	**38.7%**	138	**31.7%**
11	50	4.2%	10	1.0%	1	0.2%
12	2	0.2%	1	0.1%	0	0.0%
Total	1200	100.0%	1042	100.0%	436	100.0%

N - number

The majority of alleles from the white subpopulation lie within a narrow range (7 to 10 repeats), with only two alleles falling outside of this (one CCG4 and one CCG11). The range of CCG alleles is widest in the coloured subpopulation, with every allele identified in the South African population present in this group.

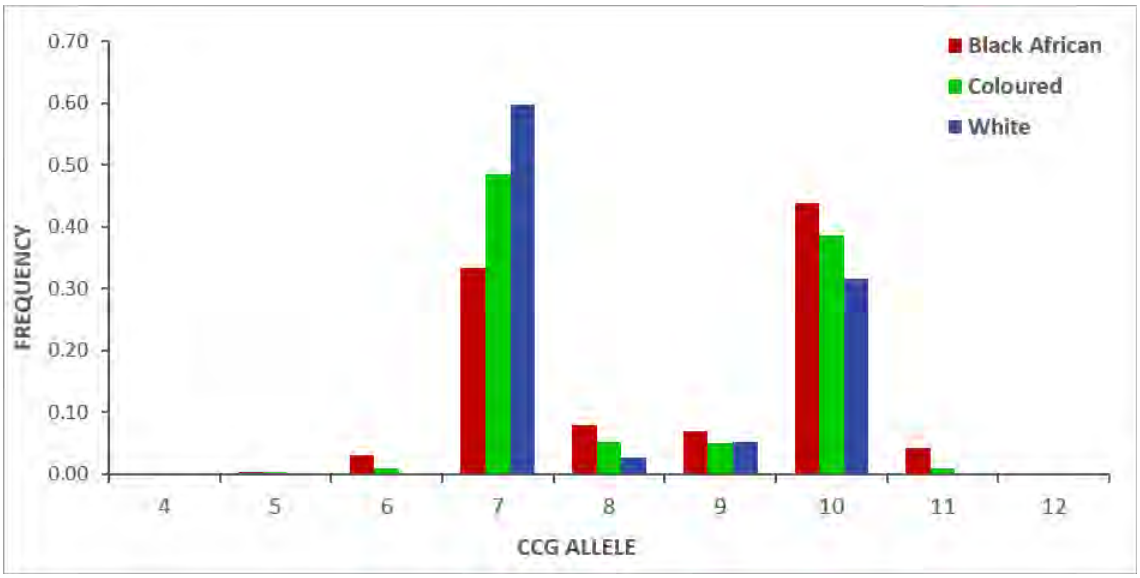


Figure 9. Frequency distribution of CCG alleles in the general population. The most common allele in the white and coloured subpopulations is CCG7, compared to CCG10 in the black subpopulation. The three subpopulations show distinctive patterns of CCG allele distribution.

The figure highlights the CCG sizing results in a bar graph. Differences in the CCG alleles present across the three subpopulation are clearly evident in the distinctive patterns shown in the frequency distribution.

2.4 Discussion

2.4.1. The mean CAG

Specific mean (average) CAG values are presented for the cohorts representing the different ethnic groups in SA. Importantly, the mean CAG in the black African subpopulation (16.88) is significantly lower than that calculated for each of the other groups. Likewise, the modal CAG (most frequent allele) observed in the black subpopulation (15) is lower than in the other two subpopulations (17). The lower modal CAG is consistent with the observed lower mean CAG in the black subpopulation.

Although the estimates are outdated (Hayden et al. 1980a), the pattern of similar HD prevalence figures for the white and coloured groups and a much lower occurrence in the black subpopulation, is consistent with the mean CAG values in this study. The lower mean CAG in the black subpopulation suggests that there are proportionately fewer HD patients in this group compared to the white and coloured groups.

The mean CAG in the black subpopulation (16.88) is comparable to values published for Asian populations; 16.9 for Chinese and Japanese (Warby et al. 2011) and 16.49 for Thais (Pulkes et al. 2014). This appears to reflect the reportedly low prevalence of HD in Asia (Harper 1992b; Pringsheim et al. 2012). However, the absence of current estimates makes it difficult to draw meaningful conclusions regarding the link between the mean CAG and the frequency of HD in these Asian populations.

The white South African population, as discussed in Chapter 1 of this thesis, had its origins in European settlers (Jenkins 1990). It was thus anticipated that the distribution of CAG repeat sizes and any statistical measures would be similar to other studies performed on European cohorts. Considering the mean and modal CAG values in Table 3 (page 51), the results show that the white South African population is indeed similar to other European cohorts (Squitieri et al. 1994; Warby et al. 2011), with comparable values (mean CAG ~ 18 and modal CAG = 17).

In the current study, the coloured subpopulation has a similar distribution of CAG repeats to the white subpopulation, with the most frequent allele observed as 17 in both groups. Other than indigenous African, the largest contribution to the makeup of the coloured population is European (de Wit et al. 2010; Quintana-Murci et al. 2010), which may explain the similarities seen between in the two groups.

The mean CAG recorded for the coloured subpopulation (17.62) is intermediate between the values for the black and white groups (16.88 and 18.48 respectively), and is possibly an indication of the admixture present in this group. Admixture is also evident in the wide range of alleles, presumably the result of genetic contributions from the different ethnic groups.

2.4.2. Proportion of intermediate alleles

Intermediate alleles hold primary significance due to their potential to expand into the pathogenic range in a single transmission (Goldberg et al. 1993b; Myers et al. 1993). The frequency of IAs has been estimated at ~1% in European populations (Goldberg et al. 1995; Sequeiros et al. 2010), however little is known about their occurrence in other populations.

The rate of new mutation has been estimated to be at least 10% of currently known cases in each generation, based on measures of mutational flow in disease alleles (Falush et al. 2001). This analysis highlights the likely under-ascertainment of individuals with less than 44 repeats and suggests that HD may not be so rare outside of known pedigrees. Another assessment of new mutations from IAs reported a possible rate of 8%, much closer to the estimate made using the mutational flow model (Almqvist et al. 2001).

It is conceivable, that given the presence of IAs in the general population and their contribution to *de novo* cases of HD, a larger proportion of IAs provides a bigger pool of alleles from which new mutations could arise. Thus, a comparatively higher proportion of IAs may correlate with a higher prevalence of HD in a given population.

The cohort investigated here shows consistency with the reported lower prevalence of HD in black South Africans (Hayden et al. 1980a), since the smallest proportion of IAs was identified in this group. Given the similar prevalence estimates in the coloured and white subpopulations, one would expect similar proportions of IAs in these two groups. Remarkably, the proportion of IAs in the white subpopulation is more than double that observed in the coloured subpopulation (Table 3, page 51).

One possibility is that new mutations (resulting from expansion of IAs in the general population) simply occur at a much lower rate in the coloured subpopulation, compared to the white subpopulation. This suggestion is consistent with anecdotal evidence (clinic records) which shows that a large number of patients in this group can trace ancestral links to well-known HD families.

However, it is also feasible that different populations exhibit specific mutation rates influenced by other factors, such as genetic background.

Subsequent to the observation of differences in the proportion of IAs, the category of large alleles (23 to 26 CAG repeats) was evaluated in the different subpopulations. The step-wise model (Warby et al. 2011) suggests that the pool of IAs is in turn supplied by the expansion of larger normal alleles. Thus, their distribution may be associated with HD prevalence since the incremental expansions that result in pathogenic repeat lengths, originate in this cluster. A population with a comparatively larger proportion of large alleles may therefore exhibit a correspondingly higher prevalence of HD.

Interestingly, the pattern exhibited by these proportions is more consistent with the reported estimates of prevalence (Hayden et al. 1980a); that is to say, similar in the white and coloured groups and smaller in the black subpopulation (Table 3, page 51). It may be worth noting, that the assessment of general population alleles included unaffected HD family members (considered as control individuals), and the unexpanded alleles from affected individuals. HD families may be more likely to have alleles with larger CAG repeats, thus contributing to the discrepancy observed in the proportions of IAs across the subpopulations.

In order to assess the frequency of new mutations in the South African population, an investigation of all individuals diagnosed with HD and determined to carry an expanded allele is required. New mutations (sporadic HD) are defined by the absence of a family history and CAG sizing of the individual's parents to confirm repeat sizes of less than CAG36 (Semaka et al. 2010). This would enable the calculation of a new mutation rate estimate for each of the subpopulations in SA.

In addition, an assessment of the occurrence of IAs within HD families may shed light on whether there is a significant contribution of these alleles to disease pathogenesis in the South African context.

2.4.3. CCG alleles and haplotypes

The most frequent alleles in the South African population are the CCG7 and CCG10 (Figure 8 and Figure 9, pages 56 and 58). The results are consistent with a previous study on a smaller cohort of South African individuals (September 1999). In this significantly larger sample, it is notable that there are numerous other alleles present, particularly in the black subpopulation. The CCG7 and CCG10 alleles are also the most commonly identified alleles in studies across various ethnic groups (Andrew et al. 1994a; Masuda et al. 1995; Pramanik et al. 2000; Ma et al. 2010).

Initial methods of PCR amplification for the CAG repeat in the *HTT* gene used primers flanking both the CAG-tract and the adjacent CCG repeat, under the assumption that the size of the CCG repeat was constant. However, subsequent analysis revealed that the repeat was polymorphic, ranging between 7 and 12 repeats (Andrew et al. 1994a). The CCG repeat has therefore been used as an intragenic marker for the purpose of haplotype construction.

Haplotype analyses showed that for most European populations, the CCG7 allele was predominantly associated (>80%) with expanded CAG repeats in HD patients, compared to the general population (approximately 60%) (Andrew et al. 1994a; Barron et al. 1994; Pecheux et al. 1995). Furthermore, CCG7 has also been associated with a higher mean CAG on unexpanded chromosomes; whereas CCG10 was linked to a reduced mean CAG length (Squitieri et al. 1994).

A similar association of the CCG7 allele with expansion was reported for the Indian population (Pramanik et al. 2000). The study reported a single 'unique' CCG4 allele from 380 control chromosomes (0.3%). From the results of the current study in SA, the coloured and white subpopulations each have a single CCG4 allele which equates to 0.1% and 0.2% respectively (Table 4, page 57).

In contrast to the association of a CCG7 with expansion, haplotype analysis in East Asian populations indicates that the CCG10 allele is associated with expansion (Masuda et al. 1995; Morovvati et al. 2008; Ma et al. 2010; Jiang et al. 2014). Thus, HD in both China and Japan appears to have one or more distinct origins, not shared with European populations.

In a Japanese cohort, only a single CCG6 was identified in 185 control alleles; the others being either CCG7 or CCG10 (Masuda et al. 1995). That population therefore exhibits a low level of heterogeneity in the *HTT* gene, a strong contrast with the results for the South African population (Figure 9, page 58).

The distribution of CCG alleles in the present study highlights, most importantly, the diversity contained in the South African population. Furthermore, the presence of multiple other alleles is a strong indicator of distinct origins of the HD mutation in the different population groups.

CHAPTER 3: THE HUNTINGTON DISEASE GENE SHOWS DISTINCT POPULATION-SPECIFIC GENETIC ORIGINS IN SOUTH AFRICA

3.1 Background

The presence of the HD mutation in many parts of the world was originally thought to be the result of European migration, including disease clusters such as had been identified in Tasmania and Venezuela. Genealogy that traced the origins of HD families in North America back to the 'old country' supported this notion (Shaw and Caro. 1982; Conneally 1984; Harper 1992b). Moreover, as sporadic cases were few and difficult to ascertain, the mutation rate for HD was estimated to be very low (although estimates varied depending on the sample population and the method of assessment).

These observations together with the high prevalence in populations with European ancestry, and comparatively low frequency in other population groups supported the hypothesis of a single or at least a small number of mutations for the majority of HD cases (Hayden 1981; Harper 1992b).

In the early 1990s, studies that attempted to isolate the HD gene using linkage analysis were confounded by seemingly contradictory recombination events (MacDonald et al. 1989; MacDonald et al. 1991; Snell et al. 1992). One suggestion was that this was the result of multiple origins of the mutation in HD chromosomes. Using multi-allelic polymorphisms in combination with the two-allele systems previously used for linkage analysis, MacDonald et al. showed that approximately one third of chromosomes shared a primordial haplotype (MacDonald et al. 1992).

Based on the chromosomes with a shared haplotype, the candidate region was narrowed to a 700 kb fragment between markers D4S180 and D4S182 (MacDonald et al. 1992; The Huntington's Disease Collaborative Research Group 1993). Several other haplotypes were identified in the study, indicating the presence of several different mutation origins.

Subsequent studies identified distinct disease-associated haplotypes in different population groups, using several intragenic and extragenic markers linked to the HD gene (Squitieri et al. 1994; Almqvist et al. 1995). The specific polymorphisms, CCG7 and Δ2642 found to be associated with expanded chromosomes, were shown to be similarly linked to normal chromosomes with a higher than average mean CAG size.

These observations led to the proposition that pre-mutation chromosomes existed in unaffected individuals in the general population and served to maintain the frequency of HD (Goldberg et al. 1993b; Snell et al. 1993). Intermediate alleles (IAs) with between 30 and 35 CAG repeats, were subsequently shown to have the potential to expand to pathogenic size on transmission.

Further investigation of new mutation families provided someevidence that particular chromosomes,exhibiting a few specific haplotypes, may be predisposed to expansion (Squitieri et al. 1994; Rubinsztein et al. 1995). This indicated the existence of a significantly higher HD mutation rate than previously estimated.

The Δ2642 deletion polymorphism was initially thought to contribute to repeat instability and/or disease pathogenesis (Rubinsztein et al. 1995). However, large normal alleles were also found to carry this deletion in similar proportions to the HD chromosomes. It was thus determined not to be causative of expansion but

merely a possible marker of haplotypes predisposed to expansion, having arisen on an ancestral chromosome with a high CAG repeat length.

Based on these findings, a step-wise model of expansion was proposed for the evolution of HD alleles: large normal chromosomes on a particular haplotype(s) predisposed to expansion, undergo small expansion events resulting eventually in an IA which is more prone to instability and has the potential to expand into the pathogenic range on transmission (Andrew and Hayden. 1995).

HD prevalence in a given population may therefore be correlated with the distribution of specific predisposing haplotypes; which haplotypes are themselves associated with a higher mean CAG length. This step-wise model of expansion has been extensively investigated by Warby and colleagues (Warby et al. 2009; Warby et al. 2011).

In SA, genetic haplotype differences have been studied in small cohorts representing the major population groups that comprise the South African population. Linkage analysis on eight HD families revealed the expansion to be associated with a CCG7 allele in six of the families (September 1999). One of the two black African families analysed had a CCG10 allele in phase with the expansion, and a family from the coloured subpopulation had a rare CCG5 allele. In both black families, a unique non-disease-associated CCG11 allele was identified.

One allele in the IA range (CAG30) was identified in 11 individuals from the same coloured family. The allele segregated on non-disease-associated chromosomes with a CCG10 allele, and was stably transmitted in all cases. A second IA (CAG31) was observed in a single individual from a different coloured family, and was also linked to a CCG10 allele (September 1999).

The second aim of this current study was therefore to perform a comprehensive SNP-based analysis of haplotypes across the *HTT* gene, in order to delineate specific association with disease and mutation origins in the different subpopulations. The genetic background of IAs identified in the general population was also investigated.

3.2 Methods

3.2.1. The Huntington disease databases

Division of Human Genetics, NHLS/UCT

In 1977, a registry was established at the Department of Human Genetics (currently the Division of Human Genetics) as part of a nation-wide study of HD families. The registry was initiated by Dr Hayden with support from Professor Peter Beighton (then head of department); and a DNA banking centre followed in 1986 (Greenberg 1990). This formed the basis of what is at present, an extensive electronic database with details of affected individuals and their families; including demographic information where available. In several cases this comprises large pedigrees, each with hundreds of individuals.

Biological material is stored under a unique identifier. All samples referred for HD testing have been coded 'H'; families are numbered sequentially, as are members within the family. For example, H 200.15KER is a fictional code that denotes a sample as the 15th member recorded from the 200th family referred for HD testing. The 'KER' is a further distinguishing feature taken from the first three letters of the individual's first name. It must be noted that external referrals typically have very limited information included on the request form.

A number of different methods have been used to isolate DNA from blood cells (lymphoblasts), resulting in varying purity and concentration. The current technique uses the PureGene DNA Isolation kit (Biozym, The Netherlands) with minor adjustments to the standard protocol.

Following isolation of genomic DNA from whole blood, the molecular test for the HD expansion is performed using standard published primers (Goldberg et al. 1993a; Andrew et al. 1994a) and results are recorded electronically. DNA aliquots are then stored at -20°C (short term storage) or -80°C (long term storage).

Division of Human Genetics, NHLS/Wits

The Division of Human Genetics at Wits was established in the latter part of the 1970s and has offered testing for HD since 2001 under the NHLS. Prior to this, patients were seen at affiliated hospitals and blood was drawn for DNA extraction. An aliquot of DNA was then transferred to the UCT molecular laboratory for testing, with additional aliquots stored at Wits.

As previously mentioned, this current project was performed in collaboration with Professor Amanda Krause, head of the clinical section (Division of Human Genetics,NHLS/Wits). It is noteworthy that the specialist neurogenetics clinic at GSH in Cape Town has to-date, never examined and diagnosed a black African patient. However, the genetics clinics in Johannesburg regularly evaluate and diagnose black African patients (personal communication). This collaboration was thus critical to ensure the evaluation of a comprehensive cohort of patients from across SA.

A similar electronic database records patient and family information at NHLS/Wits. Samples are coded 'HC' and families are numbered sequentially, as are individuals within a particular family. Genomic DNA is isolated from whole blood and the molecular test for the HD expansion performed. Thereafter, aliquots are stored for future research.

3.2.2. The patient cohort

The two databases (NHLS/UCT and NHLS/Wits) were scrutinised for all individuals diagnosed with HD and confirmed to carry an expansion in the *HTT* gene. Clinical files at both facilities were interrogated for demographic information to corroborate electronic records. Due to the long-standing nature of the HD registry in Cape Town, it was particularly challenging to ascertain complete demographics and in some cases recorded DNA samples could not be found and had likely been depleted. In addition, there was an overlap of records between the two databases prior to 2001.

As far as possible, DNA samples recorded as mutation positive were retrieved from storage and evaluated at either facility. Quality was assessed by gel electrophoresis and samples with marked degradation (i.e. smearing) were excluded. Concentration was determined using a Nanodrop spectrophotometer and working dilutions of 100 ng/µL prepared.

The final patient cohort comprised 128 individuals (37 of whom were black, 21 coloured and 70 white). DNA samples were transported to the Hayden laboratory at the CMMT (Vancouver, Canada) for quantification and genotyping as previously described for the control samples (Section 2.2.1. Sample collection and preparation, page 48).

3.2.3. CAG and CCG sizing

Repeat sizing was performed for the CAG-tract and the adjacent CCG repeat as has been described previously (Section 2.2.2. CAG and CCG sizing, page 48).

3.2.4. Statistical analysis

Statistical analyses were performed to compare the results obtained for the HD cohort from each subpopulation. The tests were described in a previous section (Section 2.2.3. Statistical analysis, page 50).

3.2.5. SNP genotyping

In order to construct haplotypes around the *HTT* gene, genotyping was performed at 96 SNP loci across the gene. The Illumina Bead Array platform was used to perform high throughput genotyping with a customised GoldenGate® Assay (Warby et al. 2011). The manufacturer's protocol outlining the workflow is provided in the Appendices.

Raw data were analysed using GenomeStudio®Data Analysis Software (Illumina) and following auto-calling, each locus was assessed manually to ensure correct genotyping. A list of all genotyped SNPs is given in the Appendices. Three of the 96 SNPs were excluded from further analyses as the results exhibited no defined genotype clusters.

3.2.6. Haplotype construction

Haplotypes were constructed by inference based on the genotypes of individuals at the remaining 93 loci across the *HTT* gene. SNP genotypes were imported into PHASE (v2.1) (Stephens et al. 2001), which uses a Bayesian algorithm to reconstruct haplotypes from population genotype information. By combining previous information with the probability of the haplotype in the observed data, this method treats unknown haplotypes as random quantities and calculates their conditional distribution and frequencies (Stephens and Donnelly. 2003).

Where available, family information was included in the analysis in order to assign haplotypes to known CAG-tract sizes. For the majority of patients (and for all controls), haplotypes were simply associated with CAG size. The assemblage of haplotypes into haplogroups was done manually based on previously determined criteria (Warby et al. 2009). Haplogroup assignment was based on 22 tSNPs across the *HTT* gene region, with variants defined by specific SNPs within the haplogroups (Figure 10, page 74).

Due to the diversity and distinctiveness of haplotypes in the South African population, it was necessary to define several novel haplogroup variants. The variants and tSNPs are shown in the figure below. Details of tSNPs (rs number and chromosomal position) are given in the Appendices.

tSNP	1	2	CAG	CCG	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	T/C	A/G			A/G	A/G	A/G	A/G	A/G	T/C	A/G	A/G	A/G	T/C	T/C	T/C	T/C	A/G	A/G	A/G	T/C	T/C	T/C	A/G
A1	C	A		7	G	A	G	G	G	C	A	A	G	C	C	T	T	A	G	G	C	T	C	A
A2	C	A		7	G	A	A	G	A	C	A	G	G	C	C	C	T	A	G	G	C	C	C	A
A3	T	A		7	G	A	G	G	G	C	A	A	G	C	C	T	T	A	G	G	C	C	C	A
A4	C	A		7	G	A	G	A	G	C	A	A	G	C	C	T	T	A	G	G	C	C	C	G
A5	C	A		7	G	A	G	G	G	C	A	A	G	C	C	T	T	A	A	G	C	C	C	A
A6	C	A		7	G	A	A	G	G	C	A	A	G	C	C	T	T	A	G	G	C	C	C	A
A7	C	A		7	G	A	A	G	G	C	A	G	G	C	T	C	T	A	G	G	C	C	C	A
B1	T	G		7	G	A	A	G	G	C	A	A	G	C	C	T	T	A	G	G	C	C	C	A
B2	C	G		7	G	A	A	G	G	C	A	A	G	C	C	T	T	A	G	G	C	C	C	A
C	*	G		8,9,10	A	*	A	G	G	*	*	*	A	C	*	C	*	A	G	*	*	C	*	*
C-SA	*	G		6,10	*	G	A	G	G	C	A	*	A	*	C	C	T	*	G	G	*	C	*	*

Figure 10. Defined haplogroups and variants in the South African population based on 22 tSNPs across the HTT gene. Haplogroup variants in REDtext are novel South African variants. Red, yellow and blue shaded SNP genotypes close to the CAG and CCG repeats define haplogroups A, B and C respectively. Green shaded SNP genotypes define variants within the specific haplogroups. Asterisks (*) are used where either SNP allele may be present in variants C and C-SA. The two repeats (CAG and CCG) are shaded grey. SNP details (rs number and chromosomal position) are given in the Appendices.

3.2.7. Phylogenetic analysis of haplogroup variants

To investigate relationships between haplogroup variants within each subpopulation, phylogenetic analyses were performed using the Molecular Evolutionary Genetics Analysis (MEGA) software, (v6.0) (Kumar et al. 1994; Tamura et al. 2013). This package infers patterns of evolution based on comparative analyses of DNA or protein sequences.

The main addition to MEGA6 is the capability for building molecular evolutionary trees scaled to time (TimeTrees) using the RelTime method which estimates relative times of divergence for all branching points (Tamura et al. 2013). This method was exploited to show relative divergence times for haplogroup variants identified in the South African population.

As outlined in the previous section, tSNPs were used to categorise the haplotypes constructed by PHASE into haplogroups; these were then further sub-divided (haplogroup variants). Representative haplotypesfrom each subpopulation were selected to analyse the relationships between these variants.

For each set of sequences, a Neighbour-Joining (NJ) tree was constructed; the NJ method does not assume a constant rate of evolution for all lineages. Evolutionary distances were computed using the Maximum Composite Likelihood method and the interior branch test (1000 bootstrap replicates) used to examine the reliability of the inferred phylogenies (Kumar et al. 2004). Subsequently, divergence times were calculated using the RelTime method (Tamura et al. 2013).

3.3 Results

In the diagnostic environment, a previous report showed that HD had been confirmed in over 200 families of coloured and white ethnicity, and in at least 50 black families (Krause and Greenberg. 2008). Approximately 600 records from the NHLS/Wits database and over 1000 records from the NHLS/UCT database were reviewed by the candidate; all individuals with a confirmed mutation status across both databases were identified.

High throughput SNP genotyping and CAG/CCG repeat sizing was performed for both expanded and unexpanded alleles. Haplotype analysis and assignment was successfully performed for 96 unrelated HD alleles and 368 unrelated general population alleles. Phase was determined and haplogroups assigned as described (Section 3.2.6. Haplotype construction, page 72).

3.3.1. CAG sizes in the patient cohort

An overview of the CAG repeats associated with disease in a cohort of South African patients is shown below (Table 5). The distribution is subsequently represented as a bar graph for the total patient cohort comprising 128 individuals (Figure 11).

Table 5. CAG-tract sizing results for the patient cohort investigated in this project.

	Black African	Coloured	White
N	37	21	70
Range	37-57	42-58	38-61
Median	44	44	44
Mean (SD)	44.51 (±4.02)	46.19 (±4.75)	44.77 (±4.09)
Mode	46	43	43

N – number, SD – standard deviation

The range of repeat sizes in this cohort is widest in the white subpopulation. Median values are the same while mean CAG varies across the three groups; although the differences do not reach significance. Modal CAG in the black subpopulation (46) is higher than in the other two groups.

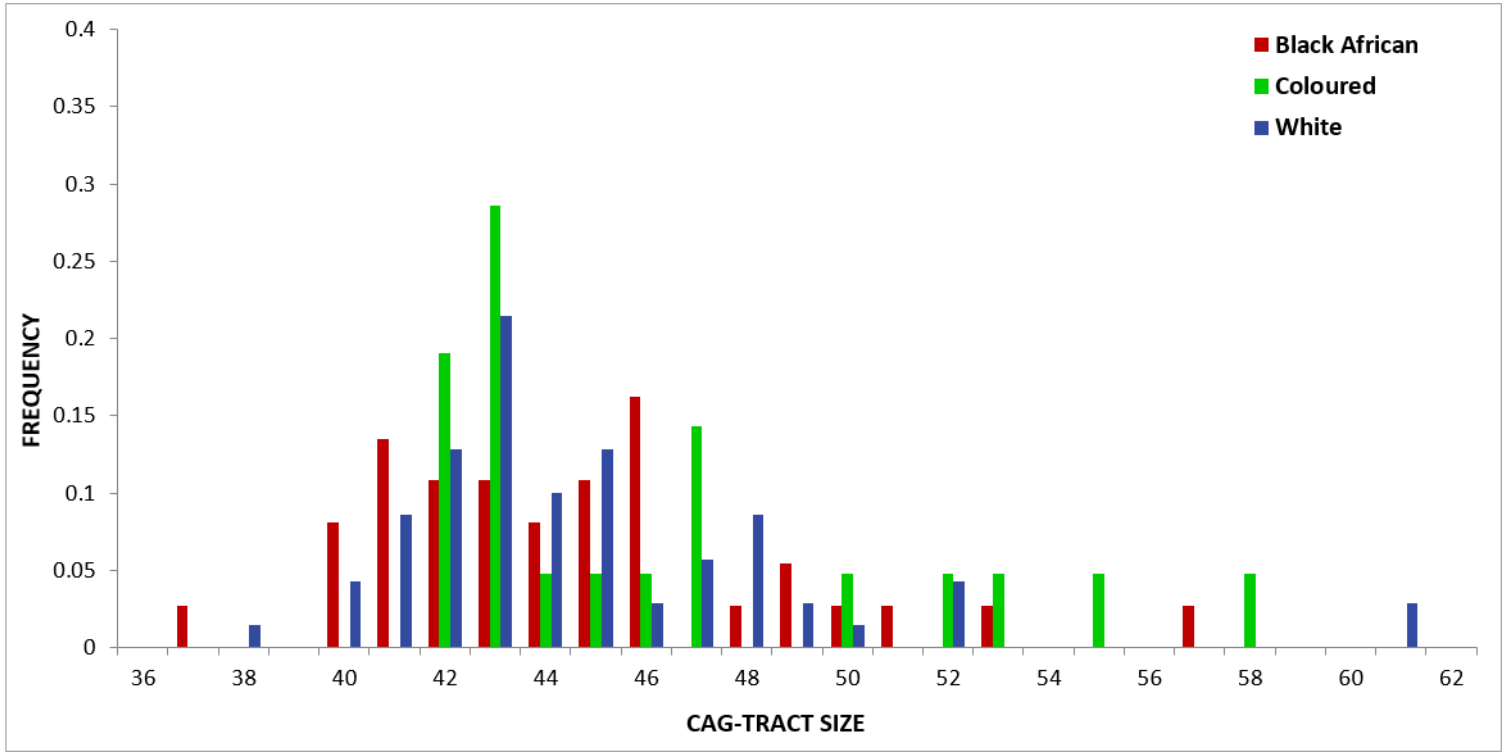


Figure 11. Frequency distribution of expanded CAG repeats in the patient cohort. There are subtle differences observable in the range of sizes and the modal CAG values (peaks) across the three subpopulations. Two individuals from the white subpopulation had a CAG repeat of 61, associated with JHD. There was one affected individual from each of the black and white subpopulations, with an allele in the zone of reduced penetrance (36-38 CAG).

3.3.2. Population-specific haplogroups and variants

The figures below show the haplogroups and variants identified in the South African population. The differences in distribution across the three subpopulations and between HD and general population alleles are evident. The haplotypes were defined using 22 tSNPs and each haplogroup is a collection of similar haplotype variants.

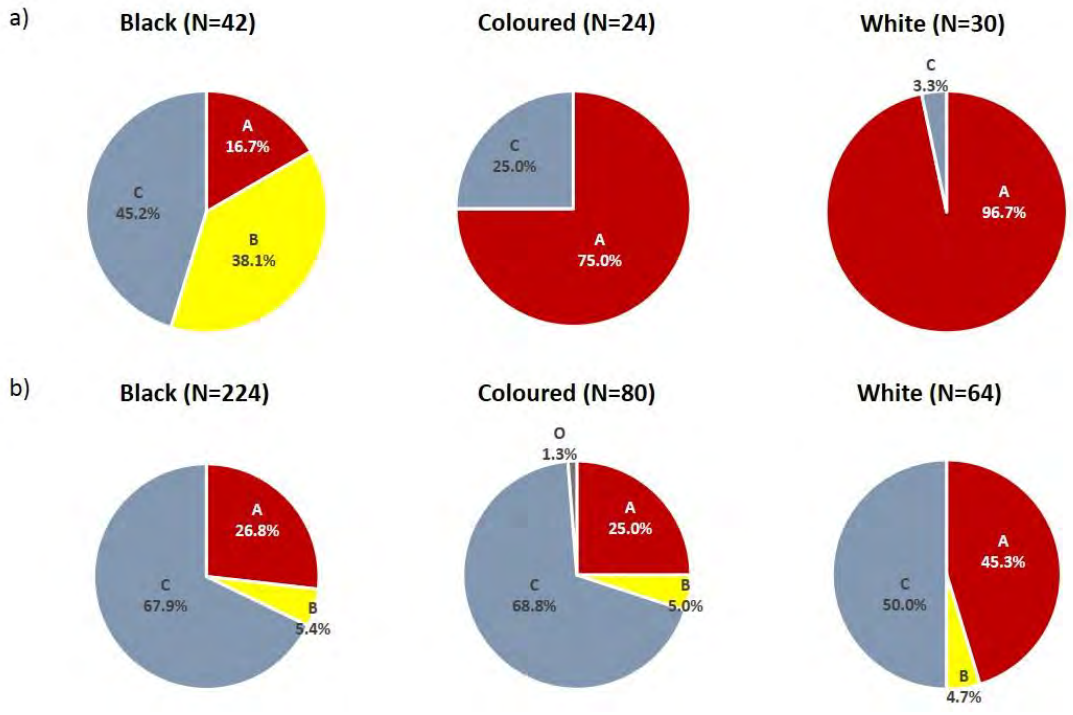


Figure 12. *HTT* haplogroups identified in the South African population. a) Haplogroups associated with expanded alleles (N=96) in the three subpopulations. A markedly different distribution is evident, with nearly all alleles (96.7%) from the white cohort assigned to haplogroup A. The majority of alleles from the black cohort however, are assigned to haplogroups B and C (83.3%). b) Haplogroups on general population alleles (N=386) in the three subpopulations. The largest proportion of alleles was assigned to haplogroup C; however, a comparatively large proportion of alleles in the white subpopulation was assigned to haplogroup A. A very small proportion in the coloured subpopulation was designated O for 'other'.

Haplotypes that could not be definitively assigned to haplogroups were clustered together in the category 'other' (O). A small proportion of general population alleles from the coloured subpopulation were classified as such in Figure 12 above. The haplogroups were further categorised into variants.

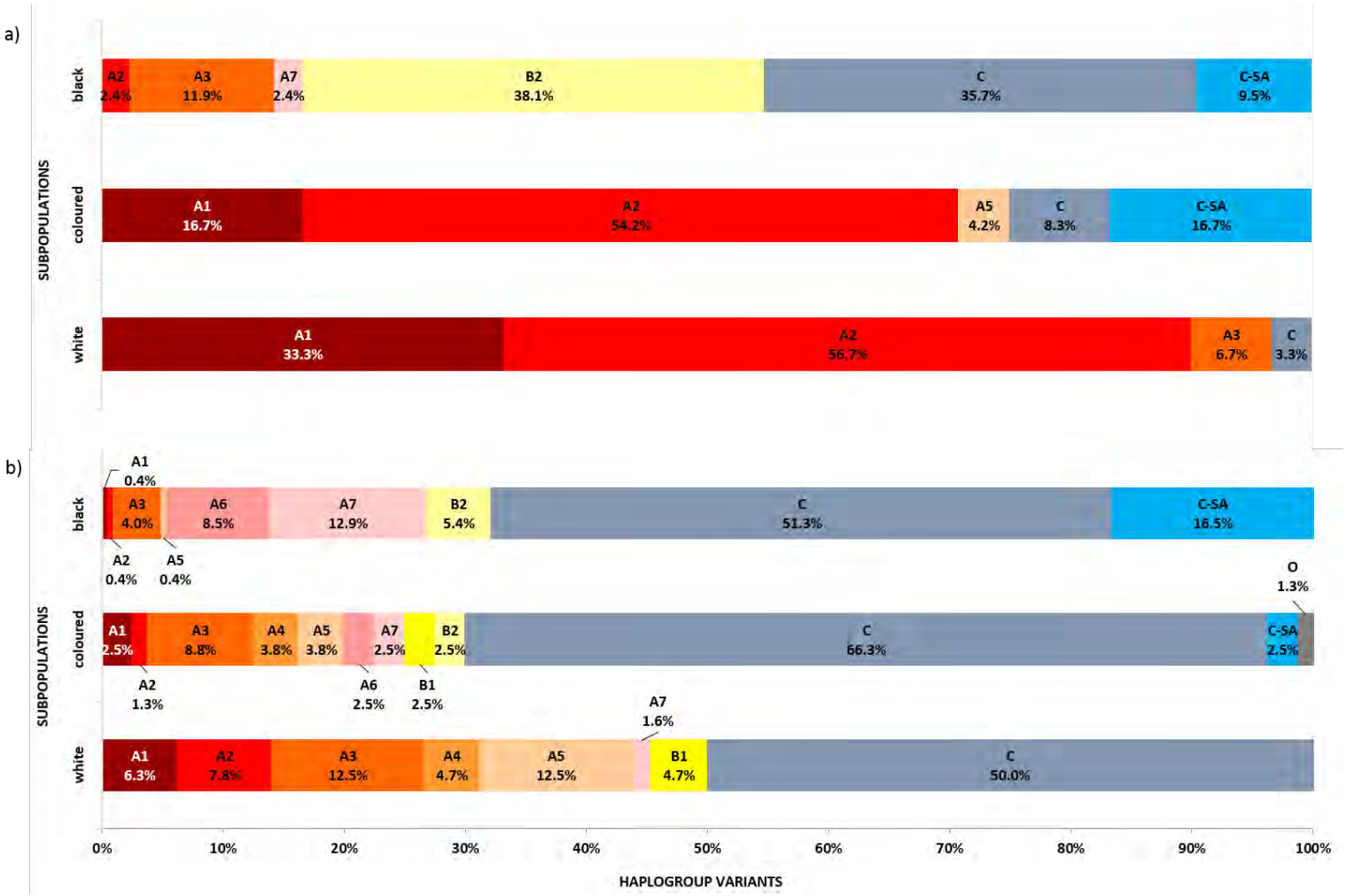


Figure 13. Haplogroup variants on phased unrelated alleles in the South African population. a) Haplogroup variants associated with expanded alleles (N=96). The largest proportion in the white and coloured cohorts consists of variants A1 and A2, while in the black cohort the largest proportion is the B2 variant, followed by C variants. b) Haplogroup variants on general population alleles (N=386). The distribution of haplotypes is distinct to each subpopulation.

In order to determine which variants are associated with specific categories of CAG-tract sizes, general population and HD alleles were compared. The variants determined to show significant differences in distribution, in each of the three subpopulations are represented in the figure below (published in Baine et al. 2013).

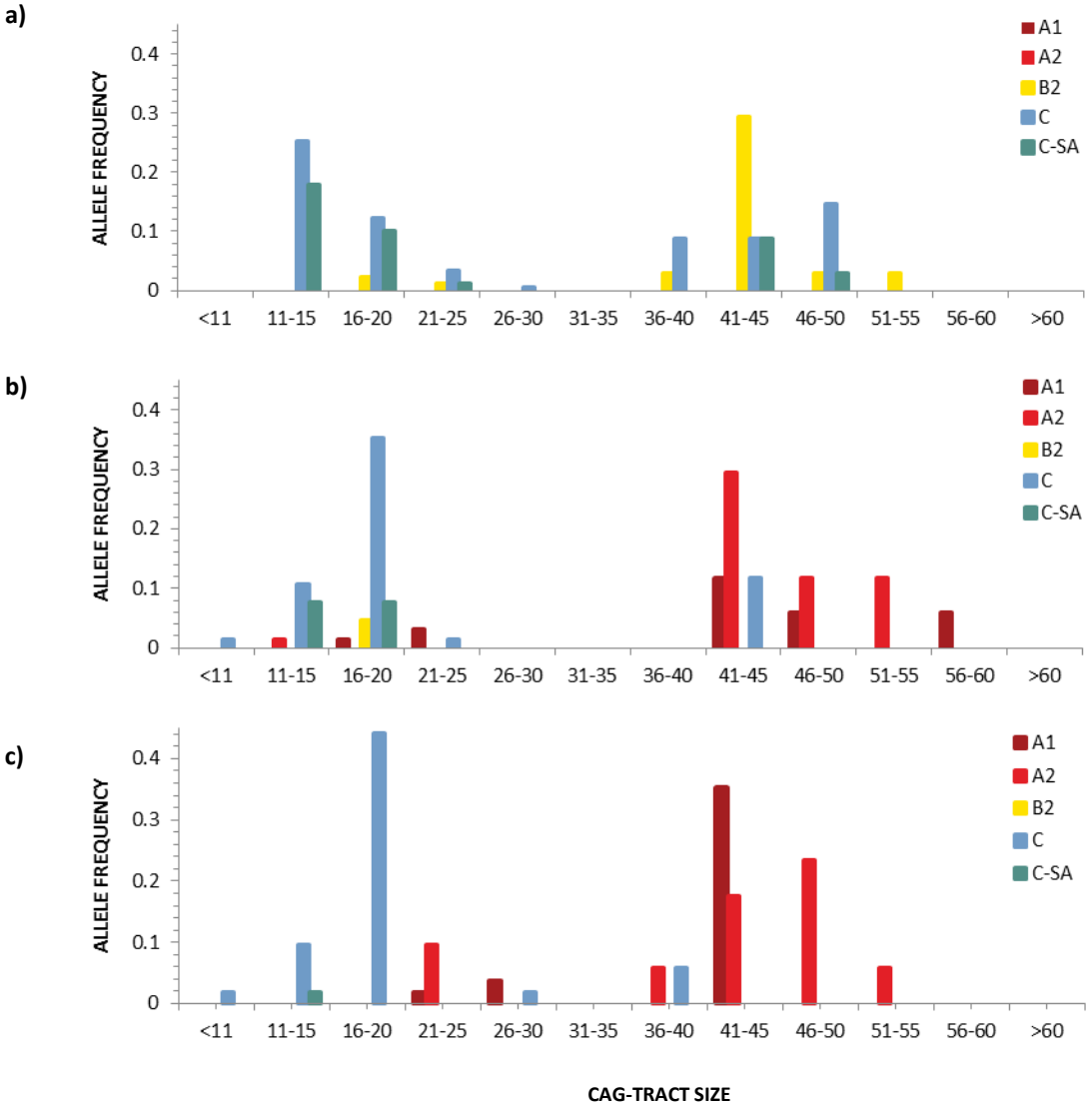


Figure 14. Haplogroup variants significantly associated with expanded alleles in the South African population.a) Black subpopulation b) Coloured subpopulation, c) White subpopulation.

In the white subpopulation, variants A1 and A2 are associated with HD alleles (CAG>35) (OR 7.76, p=0.0028 and OR 10.0, p=0.0003 respectively, Fisher's exact test). Haplogroup C is largely absent from HD alleles (OR 0.07, p=0.00093 Fisher's exact test). Similarly in the coloured subpopulation, variants A1 and A2 are associated with HD alleles (OR 5.87 p=0.0366 and OR 75.56 p=2.202E-07). Haplogroup C is uncommon on HD alleles versus those in the mixed general population (OR 0.12, p=0.0030, Fisher's exact test).

In stark contrast with the white and coloured groups, HD alleles from the black subpopulation HD alleles are predominantly associated with the novel variant B2 (OR 12.27, p=2.06E-07, Fisher's exact test). In addition the unique South African C group (C-SA) tends to be uncommon on HD alleles compared to the general population (OR 0.34, p=0.0552, Fisher's exact test). Notably, other haplogroup C variants do not differ between HD and general population alleles in the black subpopulation (OR 0.73, p=0.4543) (Baine et al. 2013).

3.3.3. Phylogeny of haplogroup variants

Phylogenetic analyses were performed using MEGA6 and the results are presented in TimeTrees in the figures below. TimeTrees were drawn as outlined in the methods section of this chapter (Section 3.2.7. Phylogenetic analysis of haplogroup variants, page 75).

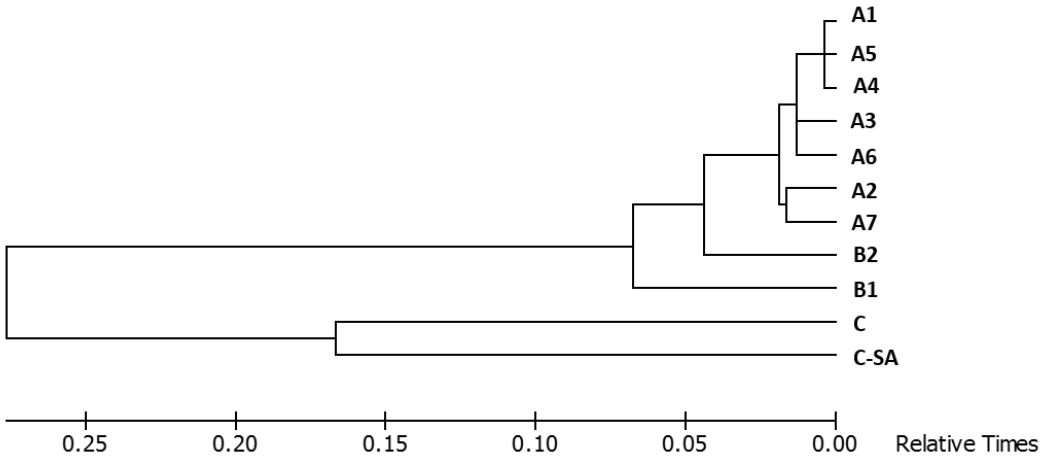


Figure 15. A time tree estimating phylogeny between selected haplogroup variants in the South African population. Evolutionary history was inferred using the Neighbour-Joining method in MEGA6. Divergence time points were calculated with the RelTime method for all branching points, using the branch lengths from the inferred tree. The analysis involved 11 nucleotide sequences with 93 SNP positions. The divergence time scale is indicated beneath the tree.

Analysis of a subset of representative haplotypes (11) from the whole population was performed in order to provide an overview of the relationships between the different haplogroup variants. Within haplogroup A, two main branches cluster the different variants. A1, A5 and A4 variants are most closely related; while A2 and A7 share a branch point and have evolved independently of the other variants. B1 and B2 also appear to have evolved fairly separately. The TimeTree indicates that haplogroup C encompasses variants that are comparatively older in origin and following the split, C and C-SA evolved along separate branches.

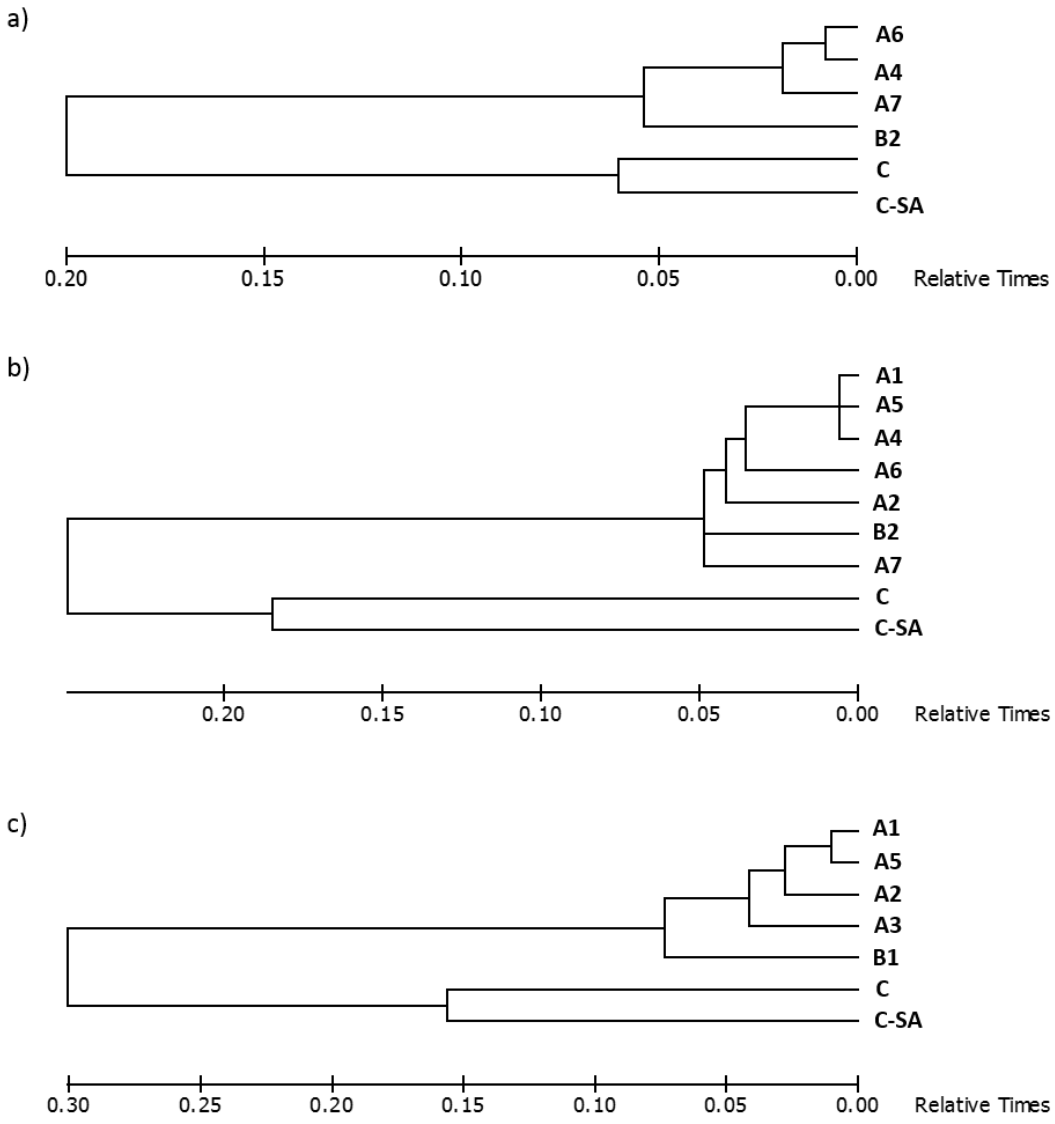


Figure 16. TimeTrees estimating phylogeny between selected haplogroup variants from each subpopulation. Evolutionary history was inferred using the Neighbour-Joining method in MEGA6 for: a) the black subpopulation (six variants), b) the coloured subpopulation (nine variants) and c) the white subpopulation (seven variants); each with 93 SNP positions. Divergence time points were calculated with the RelTime method for all branching points, using the branch lengths from the inferred trees. The divergence time scale is indicated beneath each tree and is specific to subpopulation.

Although the TimeTrees produced for each subpopulation (Figure 15, page 84) use different representative haplotypes, they show similar overall topology as the TimeTree for the entire population. Haplogroup A variants cluster together and are the most recently evolved, relative to the other haplogroups. Interestingly, the variants representing the black subpopulation appear to have comparable timelines of evolution.

A similar time frame is estimated for the evolution of A and B variant clusters, compared to the variants in haplogroup C. This is noticeably different in the coloured and white subpopulations where haplogroup C variants are significantly older than the A and B variants.

3.3.4. Haplogroup variants associated with intermediate alleles

Haplotype analysis was performed for general population alleles categorised as IAs following CAG-sizing (Section 2.3.1. CAG sizes in the general population, page 51). A total of 28 IAs from two of the subpopulations were successfully genotyped, and haplotype assignment was performed. The results are presented below for the coloured (N=11) and black (N=17) IAs.

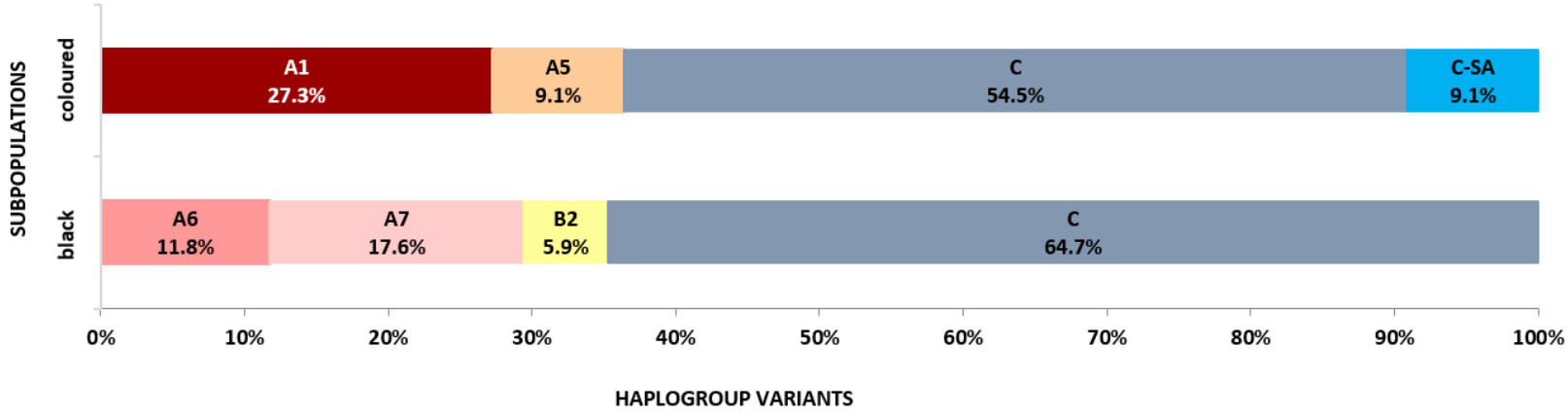


Figure 17. Haplogroup variants associated with IAs (N=28) from the coloured and black subpopulations. In the coloured subpopulation, the predominant variant is C followed by the A1 variant. The predominant variants in the black subpopulation are C and novel variants A6 and A7.

3.3.5. A haplogroup variant in a family with juvenile Huntington disease

The HD cohort analysed here included a family well-known to the specialist neurogenetics clinic at GSH in Cape Town. The family pedigree is a large one and since the 1980s, at least 20 individuals have received a diagnosis of HD with several more suspected. Fifteen family members have been confirmed molecularly to carry an expanded repeat in the HD gene.

The haplotypes identified in this family were of particular interest, due to the relatively early age of onset exhibited by most of the affected individuals (diagnoses generally made in the early- to mid-twenties). In addition, the presence of two cases of juvenile HD, including one with childhood onset (diagnosis at five years of age), suggested a role for genetic factors on repeat instability.

The haplogroup variant A2 was determined to be associated with the *HTT* expansion in this, and 12 other HD families from the coloured subpopulation. The non-disease associated variants identified within the family, were categorised under haplogroup C. The genetic background of the *HTT* gene in this family was therefore explored further and the molecular investigation is the subject of Chapter 4 of this thesis.

3.4 Discussion

3.4.1. Origins of the mutation

This study is the largest in-depth analysis of the *HTT* gene performed on an African population, and is thus a source of novel data regarding mutation origins. Early suggestions that the HD expansion arose from one or a few mutation events in Europe and was distributed around the world by migration (Hayden et al. 1980a; Harper 1992b), have been refuted by a number of studies in different populations (Morovvati et al. 2008; Paradisi et al. 2008; Ma et al. 2010; Pulkes et al. 2014).

As described in the introductory chapter, the population of SA is unique in its diversity, encompassing a range of ethnic origins and genetic backgrounds. The considerable admixture presented by the coloured subpopulation provides a rare opportunity for comparative analyses of the HD gene and mutation. This extensive analysis of the South African population contributes evidence of population-specific haplotypes, and confirms the presence of unique origins for the mutation in the different groups.

Haplotypes defined by 22 tSNPs closely associated with the disease-causing expansion, show distinct variation across the three subpopulations of SA (Figure 12, page 79; Figure 13, page 81; Figure 14, page 82). Given the country's history and evidence from previous genealogical and molecular studies (Hayden et al. 1980b; Scholefield and Greenberg. 2007), it is not surprising that expanded *HTT* alleles in the white and coloured subpopulations predominantly associate with the variants A1 and A2 (Figure 13, page 81).

These particular variants (A1 and A2) were found to be significantly associated with expansion in populations with European ancestry (Warby et al. 2009; Warby et al. 2011) and were labelled 'high-risk' haplotypes, due to the relatively higher prevalence of HDin European populations compared to other populations. In the cohort of black South Africans with expanded alleles, A1 was absent and A2 was identified in a single individual (equates to 2.4%, Figure 13, page 81).

This difference in the distribution of A1 and A2 variants across the subpopulations is significant and confirms unique origins for the HD mutation in the black population. Expanded *HTT* alleles in the black subpopulation are predominantly (~38%) associated with the haplogroup variant B2 (Figure 13, page 81). Furthermore, this observation indicates that a specific genetic background confers a proportionally higher probability of *HTT* expansion respective of ethnicity.

The *HTT* alleles from the coloured general population exhibit the largest number of haplotype variants, likely due to the admixture that resulted in this uniquely South African population. A small proportion of general population alleles could not be assigned a specific haplogroup and were designated 'other'; another likely indicator of admixture within the group. Similar to the white subpopulation, the majority of HD alleles in the coloured group are either A1 or A2 variants (total 70.9%) (Figure 13, page 81).

As discussed previously (Section 1.4.1. The people of South Africa, page 26), the Western Cape province is to date predominantly populated by the coloured

subpopulation; given that the Cape was the main point of entry for Europeans into Southern Africa.

In addition, a founder haplotype has been previously described, linking the coloured and white groups (Scholefield and Greenberg. 2007). The results presented here conclusively demonstrate that for the majority of coloured HD families, the mutation is indeed European in origin; with possible minor contributions from other ethnic groups.

3.4.2. Evolution of haplogroup variants

Phylogeny is usually defined as the evolutionary history of a group of organisms. In the current study, phylogenetic analyses were performed in order to establish relationships between sequence variants and estimate their relative ages.

A subset of sequences selected to represent the variants present in the overall population revealed the anticipated clustering of variants according to haplogroup (Figure 15, page 84). Haplogroup A variants are the most recently evolved, while haplogroup C variants, C and C-SA, are older and have evolved along very distinct branches.

Interesting to note is the definition of haplogroup C variants by the presence of a CCG allele other than CCG7 (Figure 10, page 74).The CCG7 allele has been associated with expanded CAG repeats and larger CAG-tract size in European populations (Squitieri et al. 1994). In the HD cohort investigated here, the majority of expanded alleles from the white and coloured subpopulations (96.7% and 75.1%, respectively) are categorised by the presence of a CCG7. In

the black subpopulation, only 54.8% of the expanded alleles carry a CCG7 (i.e. non-haplogroup C variants).

However, the variants predominantly associated with expansion across all three subpopulations are defined by the presence of a CCG7 allele; A1 and A2 in the white and coloured groups and B2 in the black group.

This observation corroborates the association of the CCG7 allele with expansion and corresponds with the results of the phylogenetic analysis performed here. In the TimeTrees for the three subpopulations (Figure 16, page 85), the separation of haplogroup C variants is again clearly evident.

Remarkably, the variants selected to represent the black subpopulation are estimated to be similar in age and therefore appear to have evolved quite independently (Figure 16, page 85). In the coloured subpopulation, the haplogroup C variants are considerably older than the other variants; while in the white subpopulation, haplogroup C variants are only marginally older compared to the other variants.

It is therefore reasonable to propose that the HD mutation arose independently, on several different backgrounds and at different time-points in the South African population.

3.4.3. Haplogroup variants on intermediate alleles

The hypothesis that *de novo* mutations originate from the pool of IAs in the general population was investigated by evaluating the genetic background of a cohort of IAs. It was anticipated that the distribution of haplogroup variants

associated with IAs would reflect the presence of those variants predominantly associated with expanded HD alleles.

The results from this cohort of IAs that were successfully assigned haplotypes indicate that this is true, at least for the black and coloured subpopulations for whom this analysis was possible. In the black subpopulation, the variants B2 and C associated with expanded alleles (Figure 13, page 81) were identified on 59% and 54.7% of IAs, respectively (Figure 17, page 87).

The coloured subpopulation reveals similar results, with the disease-associated variant A1 (Figure 13, page 81) identified on 16.7% of IAs (Figure 17, page 87). For both subpopulations, the majority of IAs occur on haplogroup C which is also significantly represented on expanded alleles.

Due to the limited number of IAs investigated, it is difficult to draw definitive correlations regarding the propensity of expansion exhibited by specific genetic backgrounds. Additional analyses of a larger cohort of IAs and the genetic background of new mutations is therefore recommended. Nevertheless, the results suggest that expansions do occur on a genetic background predisposed to expansion and that these alleles can be found in the general population.

CHAPTER 4: AN INVESTIGATION INTO THE PRESENCE AND POSSIBLE ACTION OF *CIS*-ACTING FACTORS ON CAG-TRACT INSTABILITY

4.1 Background

4.1.1. Repeat expansion disorders and instability

The human genome contains numerous polymorphic repeat sequences (microsatellites or short tandem repeats), that for many years only served as markers for linkage analysis (Paulson and Fischbeck. 1996). In the early 1990s however, a number of human diseases were shown to result from the expansion of trinucleotide repeat sequences: FXS (Verkerk et al. 1991), X-linked SBMA (La Spada et al. 1991) and HD (The Huntington's Disease Collaborative Research Group 1993).

Subsequently, other repeat expansions have been associated with disease; including a tetranucleotide repeat in patients with Myotonic Dystrophy type 2 (DM2) (Liquori et al. 2001) and massively expanded pentanucleotide repeats in SCA10 patients (Matsuura et al. 2000). Close to 30 different disorders, effecting neurological, neurodegenerative or developmental symptoms are now known (Castel et al. 2010; La Spada and Taylor. 2010; Kim and Mirkin. 2013).

The pathogenic mechanisms underlying these disorders can be categorised into three main groups: loss of gene function (e.g. in FXS), toxic gain of function by the protein (e.g. HD) and toxic gain of function by the RNA (e.g. DM2) (Nelson et al. 2013). The trinucleotide CAG repeat underlying HD, encodes the amino acid glutamine (Q) and is the basis of a toxic gain of function for at least eight other disorders (reviewed by (Orr and Zoghbi. 2007).

In spite of the shared mutational mechanism, the polyglutamine disease-causing genes share little structural or functional similarity. However, the pathogenic entity in all cases is believed to be the polyQ-containing protein or peptide, which induces selective neurodegeneration (Orr and Zoghbi. 2007).

Repeat instability is a key feature of these dynamic mutations with evidence for both somatic and germline changes in repeat length (Pearson et al. 2005). Tissue-specific somatic mosaicism has been shown for HD, SCA1 and a number of others (Telenius et al. 1994; Chong et al. 1995; Tanaka et al. 1996). However, the proposition that disease pathology and severity were correlated with longer repeats in the affected tissues was not be substantiated. Mosaicism was shown to be present also in tissues with a high cell turnover rate such as the liver and colon (Tanaka et al. 1996).

In HD, family studies revealed germ line instability with expansion occurring on both maternal and paternal transmissions (Kremer et al. 1995; Ranen et al. 1995; Telenius et al. 1995). Interestingly, maternal transmission occasionally resulted in contraction and large expansions were almost exclusively paternally transmitted.

Anticipation (earlier age at onset of disease in the child versus parent) was shown to be linked to paternal transmission, and the majority of juvenile onset cases (> CAG60) (Duyao et al. 1993; Kremer et al. 1995; Ranen et al. 1995). The sex of the transmitting parent is thus recognised as a significant factor in intergenerational differences in CAG repeat lengths. The repeat length itself is central to instability, with expanded repeats in the disease-associated range

showing a higher propensity for expansion on transmission no matter the sex of the parent (Kremer et al. 1995; Ranen et al. 1995).

DNA metabolic processes (repair, replication and recombination)have been shown to contribute to instability, either independently or in combination with other tissue-specific mechanisms (Pearson et al. 2005; Wells et al. 2005). The tendency of repeat sequences to form unusual DNA structures largely underpins this instability and an understanding of contributing factors is vital to the development of novel therapeutic approaches for these disorders (Castel et al. 2010).

Cis-elements, such as the repeat sequence and purity of the tract have been suggested to affect instability, with support for this hypothesis shown in animal models (Pearson et al. 2005; Castel et al. 2010). Thus far, only one such element has been conclusively associated with instability; the CCCTC-binding factor (CTCF) adjacent to the CAG repeat associated with SCA7 (Libby et al. 2008; Castel et al. 2010).

Furthermore, the suggestion has been made that the association of specific haplotypes with the HD expansion may be linked to *cis*-elements in and around the sequence (Warby et al. 2009; Castel et al. 2010; Warby et al. 2011). Research efforts to elucidate these *cis*-elements and epigenetic mechanisms are ongoing.

Evidence of repeat instability is nowhere more apparent than in families with the juvenile form of HD. In the survey of South African HD families undertaken by Hayden and colleagues in the 1970s, a notable finding was a high frequency of juvenile cases in the coloured population (Hayden et al. 1982).

4.1.2. Juvenile Huntington disease

The onset of HD symptoms is typically in middle age with the majority of patients being diagnosed between 35 and 45 years (Margolis and Ross. 2003; Roos 2010). In a subset of patients (5-7%), onset of symptoms occurs before the age of 21 years and this has been classified as early-onset or juvenile HD (JHD) (Nance and Myers. 2001; Quarrell et al. 2012b). Although George Huntington did not encounter juvenile cases and therefore made no mention of JHD in his classic account, medical records indicate that this rare form was recognised prior to 1872 (Roos 2009).

The phenotypic spectrum of JHD is distinctly different from the adult presentation, and additional variation is evident between patients affected before the age of 10 years (childhood onset) and after the age of 10 years (adolescent onset) (Nance and Myers. 2001; Gonzalez-Alegre and Afifi. 2006). Unlike the adult choreic form, the major motor feature is prominent Parkinson-like rigidity, also known as the Westphal variant of HD. Dystonia and bradykinesia may be present and a steady decline in IQ is a marked feature as the disease progresses (Nance and Myers. 2001; Gonzalez-Alegre and Afifi. 2006).

Genetic testing generally reveals a repeat length of over 60 CAG repeats, with a preponderance of male transmissions; an estimated 80% is inherited from affected fathers (Duyao et al. 1993; Siesling et al. 1997; Nance and Myers. 2001). A family history is a requirement prior to testing a child for the HD mutation and a complete clinical picture is of utmost importance. Premature diagnosis of disease is of concern within a known HD family, where behavioural problems

may be wrongly attributed to the onset of clinical symptoms (Siesling et al. 1997; Nance and Myers. 2001).

In all cases, a definitive age of disease onset is difficult to ascertain for HD patients (Walker 2007; Sturrock and Leavitt. 2010). Diagnosis is generally made at a time when symptoms are clinically distinct and does not account for subtle changes and psychiatric manifestations that are known to occur during the pro-dromal or pre-manifest stage. Furthermore, the HD phenotypic spectrum varies widely and disease presentation may be very different between patients, even within the same family.

The clinical distinction between juvenile and adult forms is thus largely arbitrary and is based on onset before the age of 21 years (Siesling et al. 1997).

4.1.3. Juvenile Huntington disease in South Africa

In 1974, the first official report was made of a family with JHD (from the coloured subpopulation) originating in Natal (Saffer et al. 1974). The family comprised four siblings, all of whom had juvenile onset of disease symptoms; and a paternal aunt with the adult form. The father (transmitting parent), was deceased at the time of investigation but was reported to have been affected by a movement disorder and progressive dementia, compatible with a diagnosis of HD.

On clinical examination, symptoms varied amongst the patients, ranging from chorea to rigidity and cerebellar signs, however, all five showed a degree of mental impairment (Saffer et al. 1974).

Subsequently, a national survey of individuals affected by HD in SA highlighted the unexpectedly high frequency of JHD in the coloured subpopulation (referred to in the study as 'mixed ancestry') (Hayden et al. 1982). A total of 11 individuals presented with juvenile onset HD, comprising 15.7% of the cohort of mixed ancestry patients. The overall frequency of JHD in SA was reported to be 7.7% with all recorded HD patients being from the white or coloured subpopulations.

Consistent with the studies discussed above, paternal transmission was more common than maternal. The unique genetic constitution of the coloured subpopulation was proposed as an explanation for the high prevalence of JHD, and it was suggested that genetic factors may play a modifying role in disease phenotype (Hayden et al. 1982).

The family with two cases of JHD highlighted in the previous chapter, was selected for further investigation. Furthermore, having identified the haplogroup variant (A2) associated with the mutation in this family, a unique opportunity was presented to explore the hypothesis that *cis*-elements may contribute to the propensity of specific variants to expansion (Warby et al. 2009; Castel et al. 2010; Warby et al. 2011).

4.2 Methods

4.2.1. Family details and sample selection

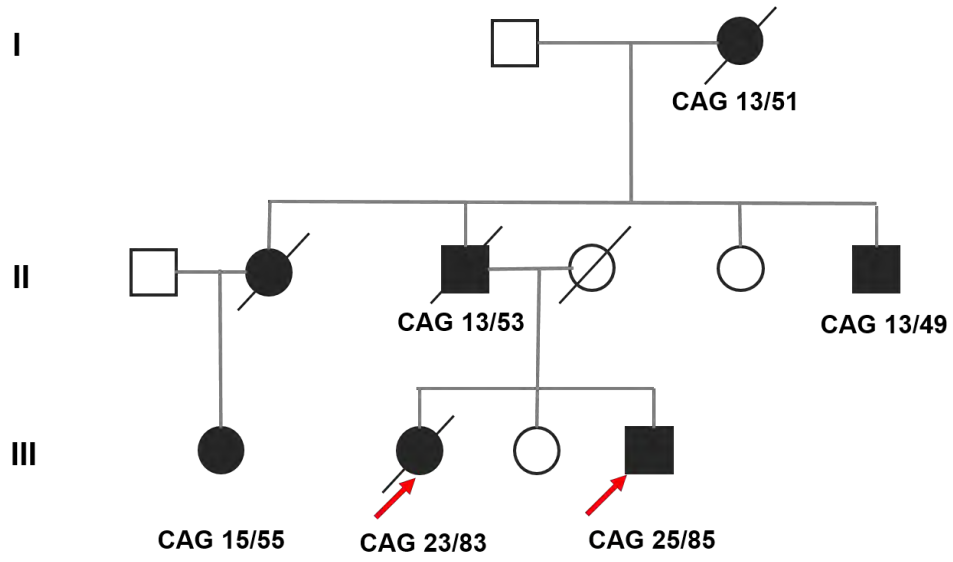


Figure 18. Pedigree of a JHD family from the coloured subpopulation. Standard symbols: circles represent females, squares represent males, filled in symbols represent affected individuals and a line through a symbol designates those deceased. Red arrows designate the two cases of JHD. The six individuals selected for sequencing were those for whom DNA samples available and their CAG repeat sizes are indicated.

A total of six samples were selected for sequencing to investigate the transmission of the *HTT* gene across two generations in this family: I:2, II:3,6 and III:1,2,4. CAG repeat sizes for these individuals were available from diagnostic test results recorded on the NHLS/UCT database. The significantly expanded alleles transmitted to the two individuals presenting with JHD are shown in the figure.

4.2.2. Sequencing of the CAG-tract

A set of primers was designed (HDCAGF and HDCAGR) to specifically amplify the CAG-tract and left unlabelled (i.e. no fluorescent tags) in order to minimise the background signal. PCR amplification was successfully performed in the molecular laboratory at UCT, following lengthy optimisation procedures (primer sequences and optimised reaction conditions are given in the Appendices). Each reaction included a negative control (no DNA) to ensure no reagents were contaminated.

Following amplification, the maximum volume of reaction product was slowly electrophoresed, through a 2% agarose gel (SeaKem® LE Agarose, *LONZA*) to allow the separation of expanded and unexpanded alleles. A molecular weight marker (GeneRuler™ 100bp Plus DNA Ladder, *ThermoScientific*) was electrophoresed alongside the PCR products for the purpose of estimating product size.

The gel was stained with SYBR® Safe DNA Gel Stain (*Invitrogen*) which binds to nucleic acids enabling visualisation. The UVIpro gel documentation system (*UVItec*) was used to visualise the PCR products under ultraviolet (UV) light and photographs taken using UVIpro software (*UVItec*). Exposure to UV light was minimised and the expanded allele quickly excised to reduce DNA degradation.

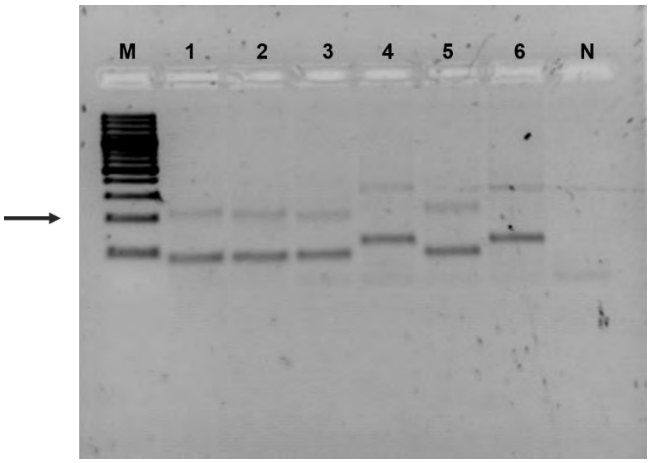


Figure 19. Gel image showing PCR amplified products (CAG repeat). A molecular weight marker (M) was electrophoresed alongside the products from individuals labelled 1 - 6. The arrow indicates the expansion, which varies depending on the length of the CAG repeat. Each amplification reaction included a negative control (N) to ensure no contamination. PCR products were electrophoresed through a 2% agarose gel at 100 volts for 2 hours to enable allele separation. The gel was stained with SYBR® and products visualised under UV light.

Gel slices were transferred to the Central Analytics Facility (CAF) at Stellenbosch University, for extraction and sequencing. Sequence results were visualised as electropherogram trace files (example in Figure 20 below); aligned to a reference sequence and edited where necessary using BioEdit software (v7.2.0) (Hall 1999).

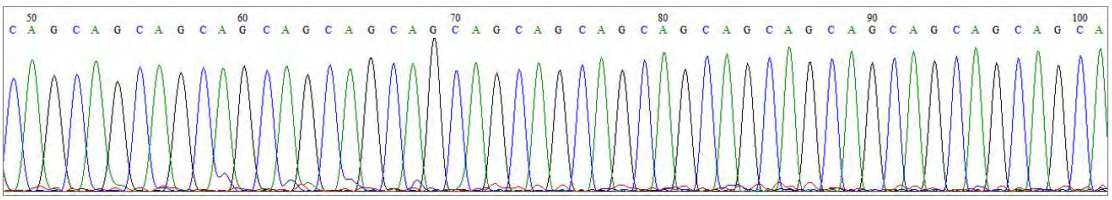


Figure 20. An example of the trace obtained from sequencing of the CAG-tract. The selection shows 18 repeats and is visualised using BioEdit.

Sequencing of the CAG-tract was successfully achieved for all six individuals, including those with very large repeats (>CAG80). There was no evidence of variation within the tract itself in any sample, therefore a larger region was selected for further investigation.

4.2.3. Sequencing of the *HTT* promoter and exon 1

The *HTT* promoter is reported to be 972 nucleotides long. Another set of primers (*HTT* promF and *HTT* promR) was designed to flank the gene promoter and exon 1, which contains the CAG-tract. The expected PCR product is approximately 1408 nucleotides in length but varies depending on the length of the CAG repeat. Direct Sanger sequencing is reliable for between 500 to 600 nt, therefore an internal primer (*HTT* promseq2) was required to enable complete sequencing of the region following amplification. Primer positions are shown in Figure 21 below.

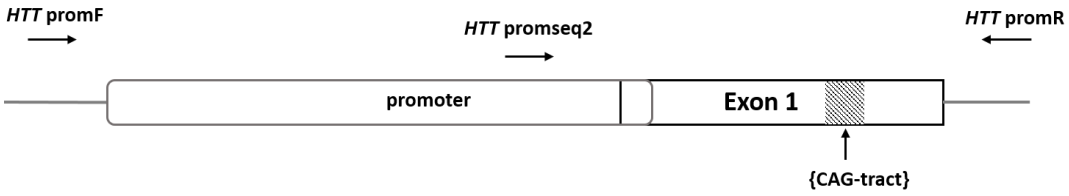


Figure 21. Schematic representation of the *HTT* gene promoter region. The promoter is indicated, overlapping with Exon 1 which contains the CAG-tract (shaded). *HTT* promF/R are the primers flanking the region of interest and used for PCR amplification and sequencing, *HTT* promseq2 is an internal primer designed to enable complete sequencing.

Amplification of this region was especially challenging due to its size and the presence of large repeats, and required extensive and lengthy optimisation. The primer sequences and final PCR conditions are given in the Appendices. Each reaction included a negative control (no DNA) to ensure no reagents were contaminated.

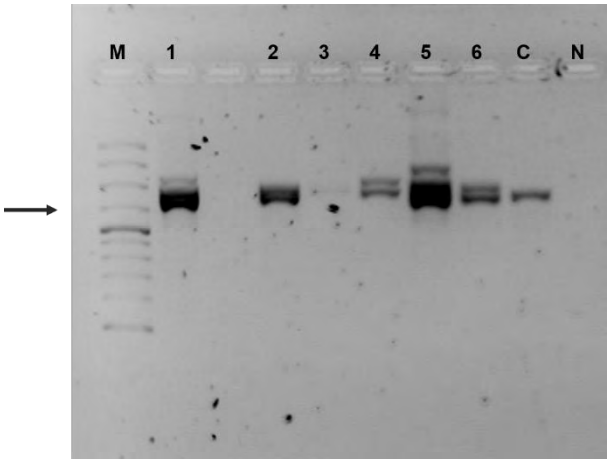


Figure 22. Gel image showing PCR amplified products (promoter and exon 1). A molecular weight marker (M) was electrophoresed alongside PCR products from individuals labelled 1 - 6. There is an empty well between individuals 1 and 2, C is an unaffected family member who does not carry the expansion. The arrow indicates the expansion, which varies depending on the length of the CAG repeat. Each amplification reaction included a negative control (N) to ensure no contamination. PCR products were electrophoresed through a 2% agarose gel at 100 volts for 2 hours to enable allele separation. The gel was stained with SYBR® and products visualised under UV light.

A similar procedure was followed for gel electrophoresis and excision as in the preceding section (Section 4.2.2. Sequencing of the CAG-tract, page 101), and gel slices transferred to the CAF for extraction and sequencing. The results were visualised and analysed using Bioedit software (v7.2.0) (Hall 1999).

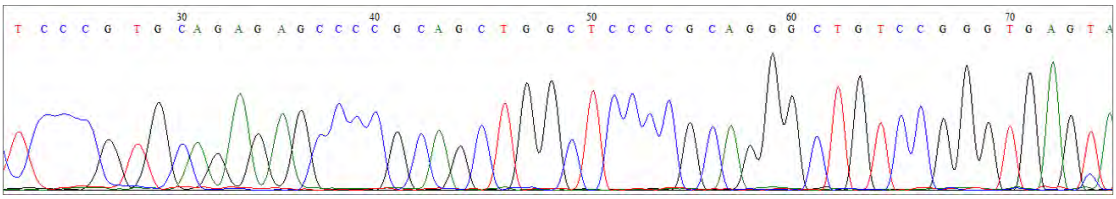


Figure 23. An example of the trace obtained from sequencing of the *HTT* promoter region. The selected sequence shows the first 75 nucleotides visualised using BioEdit.

Sequencing of the promoter region was successful for all but one sample, III:2 (Figure 18, page 100).

4.3 Results

Several variants were identified in the promoter region following sequencing with the primer *HTT* promF. Only one of these variants was shared between one JHD individual (III:3) and the transmitting parent (II:3). All are recorded SNPs.

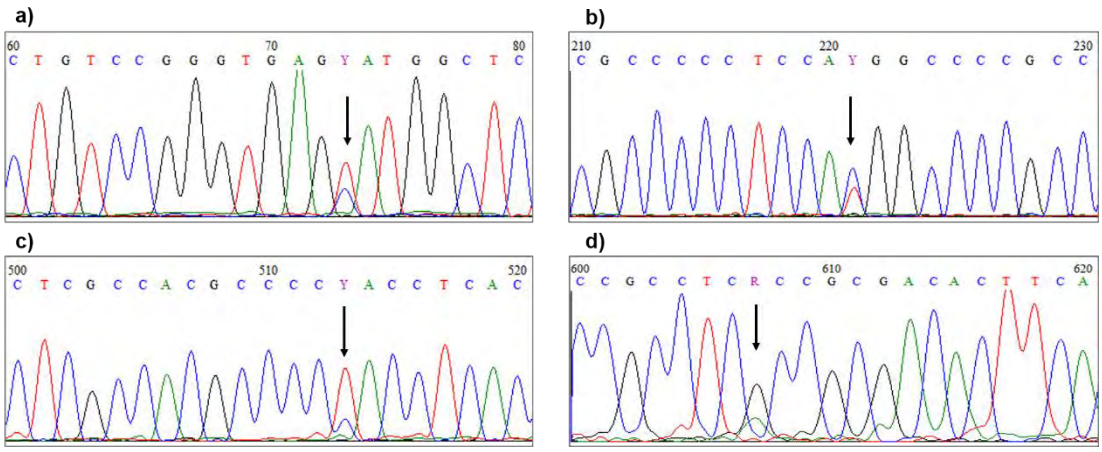


Figure 24. Variants identified in the *HTT* promoter region indicated by arrows: a) rs28431418 in one sample III:3, b) rs28616835 in two samples II:3 and III:3, c) rs149624523 in one sample III:3, d) rs116191541 in one sample II:3. The traces were visualised using Bioedit.

No additional variants were identified using the internal primer (*HTT* promseq2), however this could be due to the poor results obtained using this primer. Unfortunately time constraints on the project prevented subsequent follow-up.

4.4 Discussion

Although different publications have reported on the occurrence of JHD in the South African population (Saffer et al. 1974; Hayden et al. 1982), not much is recorded about the clinical presentation or the molecular background of this rare form of the disease. The prevalence study performed in the 1970s which reported a comparatively high frequency of JHD in the coloured subpopulation (Hayden et al. 1982), suggests that a detailed and systematic investigation is required in order to complete the profile of JHD in the South African context.

The two electronic databases (NHLS/UCT and NHLS/Wits) have on record at least 14 known cases of HD (Table 6, page 109) with an approximate age of onset before the age of 21 years, and a genotype result of more than 60 CAG repeats (Siesling et al. 1997; Nance and Myers. 2001). Of these 14 individuals, two are siblings who inherited a significantly expanded repeat from their affected father (Figure 18, page 100).

Due to the repeat instability observed in these transmissions, this family was selected for further investigation of the genetic sequence surrounding the CAG-tract. Furthermore, the expanded allele was found to be associated with haplogroup variant A2, the predominant variant in the coloured subpopulation (Figure 13, page 81). Thus, *cis*-elements that could potentially confer a predisposition to expansion were also sought.

Repeat instability as a mechanism underlying the dynamic nature of trinucleotide expansion disorders has been the subject of numerous investigations. Larger repeats are known to be associated with more severe disease and earlier onset of symptoms (Fortune et al. 2000). The earlier age of

onset in successive generations as a result of germline instability, a phenomenon known as anticipation, is of particular clinical significance (Harper et al. 1992).

Several hypotheses have been explored and studies of transgenic mouse models have identified both *cis*-elements and *trans*-factors that potentially modify repeat instability. *Cis*-elements such as DNA sequence, GC-content, CTCF binding sites, CpG-methylation, DNA replication processes have all been suggested to play a role (Cleary and Pearson. 2003).

Associated *trans*-factors include proteins involved in DNA replication, repair and recombination. In addition, interactions between *trans*-factors and *cis*-elements may lead to changes in the regulation of transcription, replication and DNA damage and repair processes (Castel et al. 2010).

This investigation undertaken on a family with JHD and a specific haplotype variant associated with expansion, aimed to elucidate sequence variants that may be linked to repeat instability. The gene promoter region and the CAG-tract were amplified and sequenced in separate reactions.

There were no novel variants identified using the approach undertaken here, however this family remains an enigma and offers potential for more complex and extensive analyses on a genome-wide level. Changes in gene expression and novel DNA variants may yet be revealed which could advance our understanding of the mechanisms involved in repeat instability.

The cohort of individuals recorded on our databases with a diagnosis of JHD is worth highlighting here. The code 'JHD' has been used for the sake of anonymity. Year of birth and age at which a clinical and molecular diagnosis was made are also indicated. The age at diagnosis is used here, rather than the age of onset, due to the challenge involved in accurately determining initial onset of symptoms.

CAG repeats are shown in the table where confirmed (lower CAG/upper CAG). The subpopulation is indicated, and the haplogroup variant associated with expansion in the family is also given where available from this current study.

Table 6. A cohort of South African patients diagnosed with JHD.

Sample code	Date of birth (year)	Age at diagnosis (years)	CAG repeats	Subpopulation	Haplogroup variant
JHD 1.1	1994	11	23/83	COLOURED	A2
JHD 1.2	2002	7	25/85	COLOURED	A2
JHD 1.3	1991	20	15/55	COLOURED	A2
JHD 1.4	1989	16	10/78	COLOURED	A2
JHD 2.1	1980	25	16/61	COLOURED	A2 (?)
JHD 3.1	1991	16	13/65	COLOURED	C-SA
JHD 4.1	1991	9	19/>80	COLOURED	A2
JHD 4.2	1976	22	-	COLOURED	A2
JHD 5.1	1978	22	17/61	COLOURED	UNKNOWN
JHD 6.1	1974	22	-	BLACK	A3
JHD7.1	1993	17	17/60	WHITE	UNKNOWN
JHD 8.1	1993	18	24/65	WHITE	UNKNOWN
JHD 9.1	1989	22	15/64	COLOURED	C
JHD 10.1	1991	19	16/59	COLOURED	UNKNOWN

The two individuals investigated in this study are designated JHD 1.1 and 1.2 in the table. JHD 1.3 is a known member of the same family, a cousin to the two probands (Figure 18, page 100). JHD 1.4 also belongs to this family, however questions surrounding paternity have been raised. This individual may be a sibling of JHD 1.1 and 1.2 OR a paternal cousin.

Interestingly, anecdotal information suggests that JHD 2.1 is related to the individuals above and belongs to a different branch of the family (designated here JHD 1). The haplogroup variant is therefore indicated as A2 (?) and requires confirmation of relationship.

JHD 4.1 and 4.2 are also members of a large pedigree. This family has been known to the division since the 1970s and currently comprises over 150 individuals, 40 of whom have received a diagnosis of HD and many more are suspected to be affected or likely to develop HD later in life.

This cohort provides a significant number of individuals for clinical and molecular investigations, to be undertaken in the future.

Section III

This section comprises Chapter 5 in which minimum estimates of HD prevalence have been determined, based on molecular genetic testing records. All data was evaluated and analysed by the candidate in consultation with a statistician (Mr Henri Carrara) and an epidemiologist (Dr Sumaya Mall) at the Faculty of Health Sciences, UCT.

CHAPTER 5: A MINIMUM ESTIMATE OF THE CURRENT PREVALENCE OF HUNTINGTON DISEASE IN SOUTH AFRICA

5.1 Background

As briefly outlined in Chapter 1, determining the prevalence and/or incidence of a monogenic disorder like HD can be a complex process requiring many different methods of ascertainment (Harper 1992b). At the very least, a careful and thorough interrogation of the following needs to be performed: hospital records, death certificates, reports from medical specialists and general physicians.

Subsequent to the identification of the gene in 1993 and direct mutation testing for the confirmation of an HD diagnosis, molecular genetic testing records have become vital for complete ascertainment. Several recent studies have utilised different methods, including examination ofdiagnostic laboratory records, in an effort to estimate current prevalence rates (Morrison et al. 2011; Evans et al. 2013; Fisher and Hayden. 2014). Results suggest that there has been a noticeable increase in the prevalence rates of HD for the populations assessed.

However, the reported rise in prevalence could conceivably be a direct result of improved ascertainment. In a number of first-world countries, a centralised health care system and electronic records, make it possible to assess many different forms of data. Evidence suggests that HD prevalence rates have indeed increased due to: a population that lives longer, the improved quality of life offered by available interventions and reduced stigmatisation resulting from increased societal awareness.

The lack of prevalence data for neurodegenerative disease affecting populations in sub-Saharan Africa has been noted (Lekoubou et al. 2014). Comparative studies of HD prevalence in African populations are therefore implausible.

In sub-Saharan Africa, genetic ascertainment is limited by the fact that a molecular test is generally requested by a clinician whose patient has sought medical attention. Thus, affected individuals who never seek medical care may be missed and ascertainment would be incomplete in populations for whom universal healthcare and specialist clinics are not available.

This limitation could possibly be addressed by door-to-door surveys, the methodology employed for common diseases and proposed by epidemiologists. However, the rarity of HD makes this a challenging technique, particularly in a resource-limited setting like SA and Africa in general.

In SA, genetic testing has been performed by the molecular laboratory at UCT since 1995. A second facility was established at Wits in 2001 and these two centres currently provide confirmatory molecular genetic tests for HD under the public NHLS system (Section 1.4.4. Clinical and testing services, page 35). Records from these two centres were extensively investigated by the candidate in order to identify individuals with a confirmed genetic diagnosis of HD for inclusion in the haplotyping study (Chapter 3 of this thesis).

Despite the limitations presented by the lack of a centralised healthcare system and different sources of ascertainment, the results of this study would be lacking in context if no attempt was made to estimate the current prevalence of HD in SA.

An estimate of HD prevalence in SA was therefore calculated based on the total number of individuals with a molecular confirmation of diagnosis (36 CAG repeats or more in the *HTT* gene), following referral for testing. This is by no means a complete ascertainment of HD cases in the country. The results presented here are minimum estimates, provided for the sole purpose of contextualising the results of the molecular investigations.

5.2 Methodology and Results

5.2.1. Review of molecular testing records

The testing records at the two public testing centres (NHLS/UCT and NHLS/Wits) were reviewed over a 16-year period, 1995 to 2010. HD has been reported to progress over 15-20 years from diagnosis (Sturrock and Leavitt. 2010; Walker 2007). The selected period (1995-2010), therefore provides a reasonable timeframe within which to estimate the total number of affected individuals. The cut-off dates were 1st January 1995 to 31st of December 2010.

The table below presents the results of a critical review of the available molecular testing records in the public sector. Stringent criteria were employed; only diagnostic tests results are reported, no predictive or pre-natal tests. A small proportion of records with no reported CAG repeat size, but merely 'E' for expanded were also excluded. Furthermore, every effort was made to ensure no overlap between the records at NHLS/UCT and NHLS/Wits prior to 2001.

Table 7. Summary of molecularHD testing performed in SA (1995-2010).

Year	Tests performed (N)	Positive (N)	Negative (N)	Equivocal (N)
1995	12	11	1	0
1996	31	20	9	2
1997	19	12	6	1
1998	28	20	6	2
1999	23	17	5	1
2000	30	15	15	0
2001	50	33	17	0
2002	37	26	10	1
2003	35	21	13	1
2004	55	27	21	7
2005	66	25	38	3
2006	42	23	16	3
2007	68	39	23	6
2008	46	22	20	4
2009	48	23	23	2
2010	48	19	27	2
Total	**638**	**353**	**250**	**35**

N - number

The total number of tests included for assessment was 638 with an average of 39.8 tests performed per year, between 1995 and 2010. Positive tests were recorded as 353, translating to a pick-up rate of approximately 55.3%. An average of 22 positive diagnoses are confirmed per year. A large number of negative tests (N=250) was observed which equates to 39.2% of the total number of tests performed. The sizing results for a representative sample of individuals tested is presented in (Figure 25, page 116).

An equivocal test result may be defined as one for which only a single allele can be identified using standard laboratory protocols. This result can be due to: 1) the presence of two homozygous alleles that cannot be differentiated, 2) allele dropout or 3) the presence of an expanded allele beyond the range of detection.

The objective of the molecular laboratory is to avoid, as far as possible, transmitting a result of this nature. In consultation with the clinical team, recommended techniques (Losekoot et al. 2013) may be employed where necessary, in order to resolve an equivocal result. A total of 35 tests were recorded as equivocal.

The positive molecular test records from the two centres are broken down by ethnicity in the table below.

Table 8. Summary of positive tests recorded, by subpopulation.

Subpopulation	**NHLS/UCT (1995-2010)**		**NHLS/Wits (2001-2010)**	
	N	**Proportion (%)**	**N**	**Proportion (%)**
Black	15	6.6	38	30.2
Coloured	58	25.6	14	11.1
White	151	66.5	70	55.6
Indian	3	1.3	4	3.2
	227	100.0	126	100.0

N - number

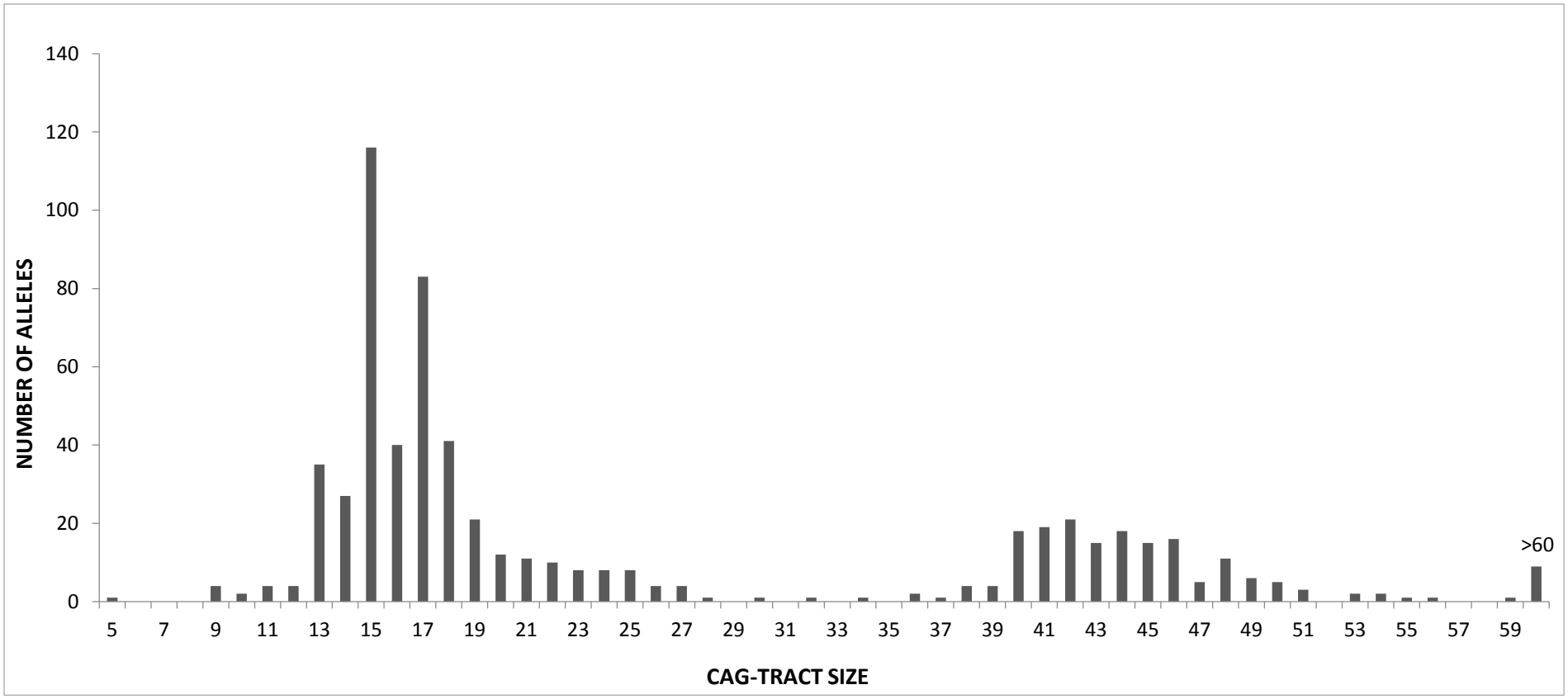


Figure 25. Frequency distribution of CAG sizing results obtained from molecular testing records. Normal alleles range from 5 to 26 CAG (N=439), with the two most frequently observed alleles being CAG15 and CAG17. Expanded alleles range from 36 to 85 CAG (N=168) and 9 alleles had more than 60 CAG repeats. Only 8 alleles were sized as IAs (27-35 CAG) and 11 were determined to be within the zone of reduced penetrance (36-39 CAG). This is only a representative sample (N=626 alleles) of CAG sizes available from the molecular diagnostic testing records, therefore does not correlate directly with the tabulated results.

5.2.2. Minimum estimates of prevalence

An important assumption was made for the estimation of prevalence: *the majority of referrals to the two public laboratories for HD testing originated within the provinces (the Western Cape and Gauteng) where these facilities are located*. This assumption was essential in order to limit the geographical area, a critical factor in the calculation of a prevalence estimate.

The most recent population census in SA was performed in 2011 and results are available from Statistics South Africa (www.statssa.gov.za). The population of the Western Cape was estimated at 5,822,734, while Gauteng was estimated to have a considerably larger population at 12,272,263 (Table 9, page 118). In spite of the significantly smaller land area covered by Gauteng (Figure 4, page 37), the population is more than double that of the Western Cape. This is predominantly the result of migration to the city of Johannesburg, the economic capital of SA, and its surrounding areas.

The table highlights the demographic differences between the two provinces. The overwhelming majority of the population of Gauteng is black African; in direct contrast to the Western Cape province where almost half the population identifies as coloured (Table 9).

Table 9. The distribution of subpopulations across the Western Cape and Gauteng provinces

Population group	Western Cape		Gauteng	
	N	Proportion (%)	N	Proportion (%)
Black African	1 912 547	32.8	9 493 684	77.4
Coloured	2 840 404	48.8	423 594	3.5
Indian/Asian	60 761	1.0	356 574	2.9
White	915 053	15.7	1 913 884	15.6
Other	93 969	1.6	84 527	0.7
Total	**5 822 734**		**12 272 263**	

N – number, Data from (www.statssa.gov.za).

Prevalence estimates were thus calculated using the following equation:

$$\frac{N\ (people\ diagnosed\ with\ HD)}{Total\ population\ (Western\ Cape + Gauteng)} * 100\ 000$$

The minimum current estimate for HD prevalence overall is approximately 1.95 per 100 000 individuals. Prevalence estimates for the three subpopulations are given in the table, with the total number of affected patients indicated. Minimum estimates of HD prevalence form the 1970s survey are also shown for comparison (Hayden et al. 1980a).

Table 10. Minimum estimates of prevalence for the South African population.

Subpopulation	HD affected (N)	Current estimate (per 100 000)	*Previous estimate (per 100 00)
Black African	53	0.5	0.01
Coloured	72	2.2	2.17
White	221	7.8	2.22

N –number, *(Hayden et al. 1980a)

Based on the above prevalence estimates of HD in the different subpopulations, the minimum number of individuals expected to be currently affected by HD across the country was calculated using the equation:

$$\frac{prevalence\ estimate}{100\ 000} * \ total\ subpopulation\ (N)$$

Population sizes for each subpopulation, were sourced from www.statssa.gov.za. The results are presented in the table below:

Table 11. Number of people potentially affected by HD in SA

Subpopulation	HD affected (N)	Current estimate (per 100 000)	Expected affected (N)
Black African	53	0.5	190
Coloured	72	2.2	101
White	221	7.8	358
Total	***346***		***649***

N - number

The approximate number of affected people possibly present in the South African population is almost double the number currently known. No statistical analyses were performed here due to the fact that these figures are very much approximated values.

5.3 Discussion

The prevalence of HD in SA has been minimally estimated here at 2 individuals per 100 000 (based on the population of two provinces). Population-specific minimum estimates are presented in Table 10 (page 118); these figures are markedly different and will be discussed further below.

Database records at the two public testing centres indicate that the total number number of tests performed in SA per year between 1995 and 2010, has risen steadily (Table 7, page 114). Between 2005 and 2010, an average of 53 referrals were made per year for a diagnostic HD test. Positive diagnoses average approximately 22 per year.

The pick-up rate (55.3%) is similar to that recorded elsewhere (52.4% in Spain and 55.8 in Slovenia) (Ramos-Arroyo et al. 2005; Peterlin et al. 2009). However, it is not as high as the rate reported (65.6%) for a large sample from Germany, Austria and Switzerland in the 1990s (Laccone et al. 1999). A recent report from Cuba highlights the ethnic variation present in that country and reports a pick-up rate of approximately 48% (Vázquez-Mojena et al. 2013).

Interestingly, there is a very distinct difference in the demographics of the individuals tested at the two facilities. The NHLS/UCT database reveals a total of 227 positive HD tests recorded since 1995. Of those, only 6.6% were black patients (Table 8, page 115). Since initiating testing in 2001, NHLS/Wits has recorded a total of 126 positive HD tests. In stark contrast to the NHLS/UCT records, 30.2% of these confirmatory diagnoses have been made for black individuals.

Another significant contrast is evident in the coloured subpopulation. NHLS/Wits records show only 11.1% of patients diagnosed molecularly are coloured, compared to 25.6% from the NHLS/UCT records (Table 8, page 115).

These observations are consistent with the distribution of the subpopulations of SA, between the two provinces of Gauteng and the Western Cape (Table 9, page 118). The observation that the majority of patients tested were from the white subpopulation at both centres (66.5% at NHLS/UCT and 55.6% at NHLS/Wits; Table 8, page 115) is a strong indicator that access to medical services is skewed.

It is worth noting again here, that the specialist neurogenetics clinic at GSH in Cape Town has to date not examined or diagnosed a single HD patient from the black subpopulation (personal communication, Professor Greenberg). Further scrutiny of the NHLS/UCT records reveals that the black patients (Table 8, page 115) were in fact referred from outside of the Western Cape.

Given the assumption made regarding geographical limits, there was the possibility that a proportion of positively diagnosed cases would originate from outside of the established provincial boundaries. For the purpose of this study, this was deemed an acceptable limitation for the calculation of prevalence estimates. A future study may attempt to address the occurrence of HD in each of South Africa's provinces, initially by following up referrals and subsequently investigating all medical records within the province.

The minimum estimates of prevalence presented in (Table 10, page 118) appear to be consistent with reports that HD in black African populations occurs at a significantly lower frequency compared to other population groups.

However, the prevalence estimate calculated in this study (0.5 per 100 000) is 50 times greater than that published previously (Hayden et al. 1980a). This is not surprising considering that since the 1970s two major changes have taken place, namely abolition of the apartheid system of governance and the availability of direct mutation testing.

Apartheid (an Afrikaans word meaning 'apartness') was a model of racial segregation implemented and enforced in SA by the National Party which won the national election in 1948 (Louw 2004). In essence the system was designed to protect the Afrikaner culture and establish sovereignty in response to Anglo (British) domination following the Anglo-Boer War (1899-1902). As a direct result, over several decades, black South Africans were forcibly removed to rural 'homelands' as urban areas were demarcated for the minority white population.

Designated 'second-class' citizens, black South Africans were subject to severe marginalisation, both economic and social (Louw 2004). This culminated in progressive poverty, poor education and a lack of access to services. Following the abolition of apartheid and the establishment of SA as a democratic state in 1994, radical changes have been observed.

Combined with the availability of direct mutation testing for the HD gene and increased access to previously denied services, the large increase in the frequency estimate of HD in the black subpopulation (Table 10, page 118) is comprehensible. The estimated prevalence is comparable to figures from Asian populations, where HD has been reported to occur at a similarly low frequency (0.4 per 100 000 individuals) (Pringsheim et al. 2012).

A significant difference is also observed in the estimated prevalence for the white subpopulation (Table 10, page 118). The current figure (7.8 per 100 000 individuals) is comparable to worldwide estimates for Caucasian populations (Pringsheim et al. 2012). Interestingly the minimum estimate of HD prevalence calculated in this study for the coloured subpopulation (2.2 per 100 000 individuals), is almost identical to that reported previously (Hayden et al. 1980a). It is thus conceivable that this figure is not far removed from the actual frequency of HD in this subpopulation.

Although no attempt was made in this study to investigate additional sources of ascertainment, it is almost certain that numerous cases of HD remain undiagnosed for various reasons. Extrapolating the minimum estimates of frequency to the South African population as a whole, provided an opportunity to calculate the number of people that may be affected by HD (Table 11, page 119). The results suggest that a significant proportion of individuals is unknown.

There are currently very few facilities available to provide specialist care, particularly for elderly individuals. Old age homes and hospices are not available in the public health care system, and only a small minority of the population has the resources to avail themselves of the private facilities. It is therefore a fairly reasonable assumption that only a few cases of HD would be missed by not investigating these facilities.

For the majority of the black African population, various suppositions can be made regarding individuals and families who manifest HD symptoms. Superstition still plays a significant role in this group and rather than seek

medical attention, patients and their families may be more likely to consult with traditional healers and/or witchdoctors (known as Sangomas in SA) (Sizer et al. 2012).

Outside of major hospitals and institutions, misdiagnosis is likely to contribute to fewer referrals for testing and a lower ascertainment rate overall. The diagnosis of HD requires specialised training in neurology and experience of movement disorders. Regrettably, medical resources in rural areas are limited to primary health care clinics and access to specialist services is not available.

The lack of medical services and specialist clinics is especially relevant for the majority black population across the country who reside in rural communities, removed from any urban development. In addition, a large proportion of the coloured subpopulation inhabits rural areas in the Western Cape.

African communities have largely maintained a traditional sense of family and community, particularly in rural areas. Individuals affected by HD may therefore be accepted and cared for by the extended family and the community as a whole, who view this simply as 'part of life'.

Although efforts have been made to redress the historical impact of the apartheid system, poverty and a lack of education combine to make this an ongoing challenge. Central to this transformation is the establishment of disease registries and advocacy groups, not only for HD but for the many other non-communicable diseases that are becoming more prevalent in SA.

Section IV

This section summarises the results of the molecular investigations reported on in Section II and the prevalence estimates calculated in Section III. Concluding remarks and future research prospects are also presented.

CHAPTER 6: CONCLUDING REMARKS AND FUTURE PROSPECTS

6.1 Population-specific differences

The differences between the various groups that comprise the South African population were extensively investigated during the course of this project. The results presented and discussed in Chapters 2 and 3, indicate clear distinctions between the subpopulations for the HD gene. Furthermore the diversity evident in these groups is particularly promising for future research.

The mean CAG values calculated for a large sample of alleles taken from the general population, show statistically significant differences between the black subpopulation and the white and coloured groups. There is also a difference in the mean CAG value observed for the white and coloured subpopulations, although this does not reach statistical significance.

Additional differences were evident in the range of repeat sizes and the modal CAG recorded for each subpopulation (Table 3, page 51). Results from CCG sizing of general population alleles (Figure 9, page 58) further highlight the variation in the *HTT* gene across the three subpopulations.

The detailed investigation of *HTT* haplotypes provides the greatest sense of the diversity present within the South African population. The disease-associated haplotypes identified (Figure 13, page 81) are specific to subpopulation and can be linked to the respective mean CAG values determined from the sizing study.

The lower mean CAG in the black subpopulation correlates with the distinctive distribution of *HTT* haplotypes in this group, compared to the white and coloured subpopulations.

6.2 Origin of the Huntington disease mutation

The identification of population-specific haplotypes in this study, definitively illustrates that the HD mutation in SA occurs on unique genetic backgrounds. For several haplotypes, clear assignment to the previously defined haplogroup variants (Warby et al. 2009) was not possible; necessitating the definition of novel variants. These variants included: A6, A7, B2 and C-SA (Figure 10, page 74).

In the coloured and white subpopulations the mutation is, as anticipated, predominantly associated with haplotypes identified in European populations (Warby et al. 2009). These so-called 'high risk' haplogroup variants (A1 and A2) are completely absent from the black subpopulation, where the mutation is predominantly associated with haplogroup B (specifically the novel variant B2) and haplogroup C (Figure 13, page 81). This suggests that any given population has its own 'high risk' haplotypes.

An analysis of phylogeny revealed specific relationships between the haplotypes identified in each subpopulation. Haplogroup A variants are generally the most recently evolved, while haplogroup C variants comprise 'older' haplotypes. C-SA in particular is a novel variant not previously identified; this variant may well be specific to African populations. The results suggest that the HD mutation arose at different time points on different genetic backgrounds, respective of the ethnic group.

The variation in the proportional distribution of IAs and large alleles warrants further investigation into the respective haplotype backgrounds. IAs have been shown to have the potential to expand on transmission into the pathogenic range, resulting in *de novo* HD cases. The stepwise model proposes that this propensity for expansion is linked to haplotype, with specific genetic backgrounds showing increased instability (Warby et al. 2009).

Although limited by the small size of the cohort, the results of the haplotype analysis on IAs identified in this cohort (Figure 17, page 87), indicate that this may well be true. Further investigation of a larger sample of IAs is therefore recommended. In addition, the occurrence of new mutations warrants exploration, in order to determine the occurrence of *de novo* HD cases in the South African population. The presence of *cis*-elements and *trans* acting factors in relation to the potential for expansion may also be investigated.

6.3 Repeat instability

An investigation into DNA sequence variations that could be associated with repeat instability was performed on a South African coloured family with JHD. The family investigated presented an opportunity to study the presence of *cis*-elements, both in the context of JHD and in relation to the haplogroup variant associated with the expansion. Due to limited available resources, the investigation was targeted to the CAG-tract and the immediate vicinity, a region encompassing the gene promoter and exon 1. No novel variants were identified using direct DNA sequencing.

Numerous factors have been suggested to play a role in repeat instability, a key component of disorders resulting from a repeat expansion. Instability at the HD gene locus has been shown to result in *de novo* pathogenic expansions, and effects anticipation within affected families. Understanding the related mechanisms is therefore of primary importance, particularly for the identification of novel therapeutic targets (Castel et al. 2010).

One of the indicated factors is the genetic background of the expanded repeat and both *cis*-elements and *trans* factors have been investigated with regard to instability. In HD, the association of specific haplotypes with expanded repeat sequences is a strong indication of the role played by genetic background (Warby et al. 2009). In addition, the presence of juvenile HD in a very small proportion of HD families, suggests that instability-related factors may well be family-specific and heritable.

In spite of the lack of success presented by the approach undertaken here, the results indicate the need for a broader strategy in an investigation of this nature. A future molecular study could examine epigenetic mechanisms, such as methylation status and the effect of micro RNAs in regulating gene expression. Whole exome sequencing may also be employed in order to identify novel genetic differences in comparison to a reference sequence.

6.4 Future prospects

6.4.1. RNA interference as a therapeutic

A number of options have been explored for the treatment of polyglutamine disorders, of which HD is but one. One of the most promising techniques involves harnessing the natural RNA interference (RNAi) pathway. This form of therapy targets mutant messenger RNA (mRNA), thereby suppressing the expression of the toxic protein product and its downstream effects. Various methods, both allele-specific and non-allele-specific, have been investigated (Scholefield and Wood. 2010; Watson and Wood. 2012).

Non-allele-specific methods involve the suppression of both the mutant and wild-type protein products by targeting the CAG repeat, and have shown promise in disease models (Boudreau et al. 2009; Hu et al. 2009; Evers et al. 2011). A key consideration is the function of the wild-type protein and whether or not its loss is tolerable (Watson and Wood. 2012). For instance, loss of the *HTT* gene product is embryonic lethal, whereas absence of the *Ataxin 1* (*ATXN1*) gene which causes SCA1, results in a fairly mild phenotype. Gene replacement therapy may thus be necessary in some cases to supplement function (Kubodera et al. 2005).

Even more important to consider is off-target effects resulting from targeting a repetitive sequence (CAG) that is widely present across the genome and significantly variable (Butland et al. 2007). However, significant advancements have been made in mitigating these effects using chemical modifications, alternative delivery systems and artificial microRNAs (Watson and Wood. 2012).

Despite the fact that non-allele-specific silencing is less technically challenging, wherever possible selective suppression of the expanded mutant mRNA is preferable. The ultimate in personalised medicine, antisense oligonucleotides (ASOs, or AONs), are designed to target SNPs associated with the HD expansion. The knowledge of population-specific haplotypes is thus central to the design and development of ASOs for a population group (Warby et al. 2009; Carroll et al. 2011).

In white (or Caucasian) populations where the HD mutation has been shown to be predominantly associated with one or two haplotypes, it has been estimated that ASOs against two SNP targets could treat up to 85% of HD patients. More recent analyses estimate this at 65-70% of patients depending on the SNP targets (Kay et al. 2014). This is, to a large extent, possible due to the homogeneity of the genetic background in these patients.

As evidenced by the results presented and discussed in this thesis, it would be exceptionally challenging to make a similar estimate for the South African population due to the diversity of haplotypes. Although the majority of known HD patients in SA share a European origin for the mutation, many individuals and families with unique genetic backgrounds would be neglected by this approach.

Apart from HD, a number of polyglutamine SCAs are present in SA (Smith et al. 2012) and the common pathogenic mechanism presents an opportunity to treat several different disorders using a similar technique. For the South African population as a whole, CAG-targeting therapeutics would likely provide the widest therapeutic reach. However, the challenges inherent to non-allele-

specific silencing make the alternative approach relatively encouraging (Scholefield and Wood. 2010; Watson and Wood. 2012; Wild and Tabrizi. 2014).

The molecular genetics laboratory at UCT has previously undertaken investigations of allele-specific silencing for SCA7 and SCA1 in cell models (Scholefield 2008; Baine 2010). Based on founder haplotypes identified in SCA1 patients from the coloured subpopulation, a single nucleotide difference was targeted to preferentially silence the mutant *ATXN1* transcript (Baine 2010). The results demonstrated proof-of-principle in a limited setting (cellular model); however, only a small subset of patients stand to benefit from this form of therapeutic.

On the other hand, the black subpopulation exhibits a strong founder SCA7 mutation; the majority of patients identified have a single SNP allele linked to the expanded CAG repeat (Greenberg et al. 2006; Scholefield et al. 2009). The results of a preliminary investigation on this cohort of patientshas contributed to the overall understanding of RNAi, and its application to selective silencing in polyglutamine disease. More recently, this preliminary investigation has been replicated in patient-derived cells and continues to show potential (Scholefield et al. 2014), supporting its development as a viable therapeutic approach for this group of patients.

The extensive haplotype investigation undertaken in this study, provides for the first time, an opportunity to assess the feasibility of allele-specific silencing for HD in an African population. The disease-associated haplotypes identified here will be further examined in order to identify potential SNP targets that may be exploited for therapeutic development.

Overall both gene silencing methods outlined here are being extensively researched. Pre-clinical studies for the safety and efficacy of effector molecules are nearing completion and clinical trials are imminent (Wild and Tabrizi. 2014). It is vital that the diversity of the HD community, including African populations, is represented in these efforts.

6.4.2. Prevalence study

In order to present a clear picture of the current state of HD in SA, a prevalence study of the disease is planned for 2015. Accurate estimates are vital for resource allocation and the delivery of services to the HD community in SA; a systematic survey of the population is therefore imperative and long-overdue.

The established collaboration between the academic institutions of UCT and Wits will form the foundation of this project, with the NHLS databases providing referral records. Cascade screening may be employed in order to identify additional affected individuals and those at risk of developing HD. This systematic method of approaching relatives of patients has proven beneficial, for example in familial hypercholesterolaemia (Ned and Sijbrands. 2011; Ademi et al. 2014).

Due to the known symptomatic overlap between neurodegenerative disorders (Wild et al. 2008; Martino et al. 2013), other available patient databases (e.g. the Spinocerebellar Ataxias, Parkinson disease), will be interrogated for cases of HD that may have been misdiagnosed.

Contact will be established with clinical specialists (neurologists and psychiatrists) around the country to explore their knowledge of and familiarity

with HD patients. Other medical personnel, such as general physicians who manage patients on a day-to-day basis, will also be contacted. Wherever possible, personal visits will be arranged with the clinicians around SA in order to interview them and review their records.

As mentioned previously, it is unlikely that old age homes would have undiagnosed affected individuals due to the specialised care required for these patients. However, psychiatric hospitals and similar facilities have been known to provide care for HD patients in the past. These may therefore provide an additional opportunity for the identification of HD cases and will be included in the search.

Lastly, a family advocacy group, the Huntington's Association of South Africa (HASA) established in 2004 has been recently revived (www.huntingtons.org.za). There is a concerted effort driven by HASA to raise awareness of this devastating condition in SA. Currently, facilitated support groups for families are available in Cape Town and Pretoria.

Apart from raising awareness, HASA's goal is to establish care facilities for HD patients. Determining the current burden of disease and how this is likely to change over time, is therefore essential to fundraising and lobbying in order to access the resources that will enable provision of services. The planned project outlined above, therefore has the full backing and support of this advocacy group.

6.4.3. Is there a place for South Africa in a global HD registry?

The international research HD community has for decades been closely associated with patient and family advocacy groups. These groups have been fundamental to the success that this field has experienced. The willingness of families to participate in research, combined with the dedication and passion poured into fundraising efforts has slowly improved our understanding ofHD and brought us ever closer to a cure.

A significant contribution to research has been through disease registries and biobanks that enable access to patient and family records and biological material. In the last decade, a number of large observational studies have been established for the purpose of collecting prospective data to map the course of disease. These types of studies also provide a foundation for the recruitment of patients to clinical trials.

As the most recently established of these prospective registry studies, Enroll-HD has a global outreach with recruitment sites in North and South America, Europe and Australia (www.enroll-hd.net). Enroll-HD aims to collate clinical data and biological material for further research into disease mechanisms and to facilitate the development and design of effective clinical trials for HD therapeutics.

Thus far, none of the international observational studies has included participants from the African continent. Pharmacogenomics studies of South African populations have highlighted the effect of genetic variation on drug metabolism and therefore efficacy, as reviewed by Warnich and colleagues

(Warnich et al. 2011). The development of therapeutic agents must therefore take into account the genetic context of various population groups.

The results of this South African study show irrefutable evidence of variation between the subpopulations investigated. It is thus of utmost importance that studies of these populations are included in the global effort to understand disease mechanisms. Of particular significance is the genetic diversity reported here, especially in the black African subpopulation which represents the most ancient and therefore most heterogeneous group of peoples.

The inclusion of African populations in a global longitudinal and observational study of HD may provide novel contributions to the understanding of disease, due to the genetic background. The candidate has therefore initiated a proposal for South Africa to be considered for inclusion in Enroll-HD.

This would be the very first site in Africa to recruit participants. Furthermore, as one of the more developed countries on the African continent, SA can act as a point of contact for the establishment of other African collaborations. Support from key players in the HD community in SA has already been established and the proposal is being developed as a collaborative effort between UCT and Wits, and with the support of the family advocacy group, HASA.

The existence of phenocopies (or genocopies) of HD in the South African population presents another exciting opportunity for research into previously unidentified mutations and mechanisms of disease that result in overlapping phenotypes. HD phenocopies have been estimated to account for between 1 and 7% of clinically diagnosed HD (Wild et al. 2008); however database records suggest that this figure is much higher in SA (unpublished data).

Subsequent to the establishment and roll-out of Enroll-HD, a secondary study is planned known as Enroll-HD-like to provide a similar foundation for the understanding of HD phenocopies (personal communication).

6.4.4. Do novel molecular causes underlie HD phenocopies?

HD phenocopies have been estimated to account for a proportion of diagnosed cases (Wild et al. 2008). The subsequent identification of a mutation in the *JPH3* gene as the underlying cause of HDL2 partly explained the presentation of HD phenocopies in SA. HDL2 is reported to be relatively common in the black patient population (Magazi et al. 2008). It has also been described in the coloured subpopulation with a possible difference in phenotype (Bardien et al. 2007). The presence of a founder event for this mutation is the subject of a submitted manuscript (personal communication, Professor Krause).

In the South African population, a significant proportion of individuals clinically diagnosed with HD have been found not to carry a repeat expansion in the HD gene. Database records reviewed during this study indicate that close to 40% of samples referred for HD testing do not carry mutations in either the HD or HDL2 genes (Section 5.2 Methodology and Results, page 113 and personal communication). It is therefore necessary to explore the spectrum of HD phenocopies in the available patient cohort which holds immense potential for the identification of novel mutations and disease mechanisms.

A complete clinical assessment is a required initial step. The differential diagnosis of HD and other neurological syndromes (Wild et al. 2008; Martino et al. 2013) has been previously highlighted and known mutations may be identified in this cohort.

However, a thorough clinical assessment may provide indications regarding potentially novel mutations in subsets of patients with similar phenotypes in the absence of known mutations. This is a subject for extensive molecular investigation and will be undertaken in the future.

6.5 Conclusion

This study presents a comprehensive analysis of the genetic background of the HD gene in the South African population. A large sample of alleles from the general population was genotyped and the normal distribution of CAG repeat sizes in the different subpopulations evaluated. The three main subpopulations were represented and differences between the various groups areextensively reviewed and discussed.

The distribution of expanded alleles was investigated in a cohort of South African HD patients. Disease-associated haplotypes were constructed based on extensive genotyping of SNP loci across the *HTT* gene. The South African population exhibited extraordinary genetic diversity and a number of unique haplotypes were identified. Haplotype analyses, including construction of phylogeny, confirmed distinct origins for the HD mutation in the different population groups.

In a family with juvenile onset HD, an investigation was undertaken of the genetic sequence surrounding the expanded CAG-tract in a number of affected members. There were no novel variants identified within the *HTT* promoter or exon 1 of the gene. More extensive analyses in the future may yield clues to the mechanisms involved in repeat instability, which could then be exploited for therapeutic benefit.

Minimum estimates of the current prevalence of HD in SA were calculated, based on available records of molecular tests for confirmation of diagnosis. The results suggest that a significant proportion of affected individuals remain unknown and undiagnosed. It is therefore essential that an epidemiological study be undertaken using varied methods for ascertainment in order to determine the current burden of disease and identify those individuals at-risk.

The inclusion of SA as a member of the global registry Enroll-HD, would stimulate the recognition of HD as an important non-communicable disorder and facilitate the much-needed allocation of resources. The HD community in SA stands to benefit significantly from the initiative, with the hope that this will subsequently extend into sub-Saharan Africa.

References

Ademi Z, Watts GF, Pang J, Sijbrands EJ, van Bockxmeer FM, O'Leary P, Geenhoed E, Liew D (2014). Cascade screening based on genetic testing is cost-effective: Evidence for the implementation of models of care for familial hypercholesterolemia. J Clin Lipidol 8(4): 390-400.

Almqvist E, Spence N, Nichol K, Andrew S, Vesa J, Peltonen L, Anvret M, Goto J, Kanazawa I, Goldberg Y (1995) Ancestral differences in the distribution of the Δ2642 glutamic acid polymorphism is associated with varying CAG repeat lengths on normal chromosomes: insights into the genetic evolution of Huntington disease. Hum Mol Genet 4(2):207-214.

Almqvist EW, Bloch M, Brinkman R, Craufurd D, Hayden MR (1999) A worldwide assessment of the frequency of suicide, suicide attempts, or psychiatric hospitalization after predictive testing for Huntington disease. Am J Hum Genet 64(5):1293-1304.

Almqvist E, Elterman D, MacLeod P, Hayden M (2001) High incidence rate and absent family histories in one quarter of patients newly diagnosed with Huntington disease in British Columbia. Clin Genet 60(3):198-205.

Anderson KE, Marder KS (2001) An overview of psychiatric symptoms in Huntington's disease. Curr Psychiatry Rep 3(5):379-388.

Anderson KE, Gehl CR, Marder KS, Beglinger LJ, Paulsen JS, Huntington's Study Group (2010) Comorbidities of obsessive and compulsive symptoms in Huntington's disease. J Nerv Ment Dis 198(5):334-338.

Andresen JM, Gayán J, Cherny SS, Brocklebank D, Alkorta-Aranburu G, Addis EA, Cardon LR, Housman DE, Wexler NS (2007a) Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds. J Med Genet 44(1):44-50.

Andresen JM, Gayán J, Djoussé L, Roberts S, Brocklebank D, Cherny SS, The US-Venezuela Collaborative Research Group, The HD MAPS Collaborative Research Group, Cardon LR, Gusella JF et al (2007b) The Relationship Between CAG Repeat Length and Age of Onset Differs for Huntington's Disease Patients with Juvenile Onset or Adult Onset. Ann Hum Genet 71(Pt 3):295-301.

Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, Squitieri F, Lin B, Kalchman MA (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet 4(4):398-403.

Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR (1994a) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 3(1):65-67.

Andrew SE, Goldberg YP, Kremer B, Squitieri F, Theilmann J, Zeisler J, Telenius H, Adam S, Almquist E, Anvret M et al (1994b) Huntington disease without CAG expansion: phenocopies or errors in assignment? Am J Hum Genet 54(5):852-863.

Andrew S, Hayden M (1995) Origins and evolution of Huntington disease chromosomes. Neurodegeneration 4(3):239-244.

Baine FK (2010) Identification of a suitable SNP for allele-specific silencing of the disease-causing gene in SCA1 patients in South Africa. Dissertation, University of Cape Town.

Baine FK, Kay C, Ketelaar ME, Collins JA, Semaka A, Doty CN, Krause A, Greenberg LJ, Hayden MR (2013)Huntington disease in the South African population occurs on diverse and ethnically distinct genetic haplotypes. Eur J Hum Genet 21(10):1120-1127.

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

Barron LH, Rae A, Holloway S, Brock DJ, Warner JP (1994) A single allele from the polymorphic CCG rich sequence immediately 3' to the unstable CAG trinucleotide in the IT15 cDNA shows almost complete disequilibrium with Huntington's disease chromosomes in the Scottish population. Hum Mol Genet 3(1):173-175.

Bates GP, MacDonald ME, Baxendale S, Youngman S, Lin C, Whaley WL, Wasmuth JJ, Gusella JF, Lehrach H (1991) Defined physical limits of the Huntington disease gene candidate region. Am J Hum Genet 49(1):7-16.

Bates GP (2005) History of genetic disease: the molecular genetics of Huntington disease—a history. Nat Rev Genet 6(10):766-773.

Beighton P (1976) Genetic disorders in Southern Africa. S Afr Med J 50(29):1125-1128.

Beighton P, Fieggen K, Wonkam A, Ramesar R, Greenberg J (2012) The University of Cape Town's contribution to medical genetics in Africa: from the past into the future. S Afr Med J102(6):446-448.

Boudreau RL, McBride JL, Martins I, Shen S, Xing Y, Carter BJ, Davidson BL (2009) Nonallele-specific silencing of mutant and wild-type huntingtin demonstrates therapeutic efficacy in Huntington's disease mice. Mol Ther 17(6):1053-1063.

Brocklebank D, Gayan J, Andresen J, Roberts S, Young A, Snodgrass S, Penney J, Ramos-Arroyo M, Cha J, Rosas H et al (2009) Repeat instability in the 27–39 CAG range of the HD gene in the Venezuelan kindreds: counseling implications. Am J Med Genet B Neuropsychiatr Genet 150B(3):425-429.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion J, Hudson T et al (1992) 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 68(4):799-808.

Butland SL, Devon RS, Huang Y, Mead C, Meynert AM, Neal SJ, Lee SS, Wilkinson A, Yang GS, Yuen MM et al (2007) CAG-encoded polyglutamine length polymorphism in the human genome. BMC Genomics 8:126.

Carroll JB, Warby SC, Southwell AL, Doty CN, Greenlee S, Skotte N, Hung G, Bennett CF, Freier SM, Hayden MR (2011) Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the Huntington disease gene/allele-specific silencing of mutant huntingtin. Mol Ther 19(12):2178-2185.

Castel AL, Cleary JD, Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol 11(3):165-170.

Chang CM, Yu YL, Fong KY, Wong MT, Chan YW, Ng TH, Leung CM, Chan V (1994) Huntington's disease in Hong Kong Chinese: epidemiology and clinical picture. Clin Exp Neurol 31:43-51.

Chimusa ER, Daya M, Möller M, Ramesar R, Henn BM, van Helden PD, Mulder NJ, Hoal EG (2013) Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. PloS One 8(9):e73971.

Chong SS, McCall AE, Cota J, Subramony S, Orr HT, Hughes MR, Zoghbi HY (1995) Gametic and somatic tissue–specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. Nat Genet 10(3):344-350.

Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, Goldberg YP, Haddad BR, Richards F, Sillence D et al (1997) Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. Hum Mol Genet 6(2):301-309.

Cleary JD, Pearson CE (2003) The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. Cytogenet Genome Res 100(1-4):25-55.

Coggon D, Barker D, Rose G (2009) Epidemiology for the Uninitiated. Hoboken, NJ: BMJ Books.

Conneally PM (1984) Huntington disease: genetics and epidemiology. Am J Hum Genet 36(3):506-526.

Daya M, van der Merwe L, Galal U, Möller M, Salie M, Chimusa ER, Galanter JM, van Helden PD, Henn BM, Gignoux CR et al (2013) A Panel of Ancestry Informative Markers for the Complex Five-Way Admixed South African Coloured Population. PloS One 8(12):e82224.

de Jager T, Corbett CH, Badenhorst JC, Brink PA, Corfield VA (1996) Evidence of a long QT founder gene with varying phenotypic expression in South African families. J Med Genet 33(7):567-573.

de Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, van Helden PD, Seoighe C, Hoal EG (2010) Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum Genet 128(2):145-153.

Di Maio L, Squitieri F, Napolitano G, Campanella G, Trofatter JA, Conneally PM (1993) Suicide risk in Huntington's disease. J Med Genet 30(4):293-295.

Dunglison R (1842) The practice of medicine. Philadelphia: Lea and Blanchard.

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, Ross C, Franz M, Abbott M et al (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 4(4):387-392.

Evans SJ, Douglas I, Rawlins MD, Wexler NS, Tabrizi SJ, Smeeth L (2013) Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. J Neurol Neurosurg Psychiatry 84(10):1156-1160.

Evers MM, Pepers BA, van Deutekom JC, Mulders SA, den Dunnen JT, Aartsma-Rus A, van Ommen GB, van Roon-Mom WM (2011) Targeting several CAG expansion diseases by a single antisense oligonucleotide. PloS One 6(9):e24308.

Falush D, Almqvist EW, Brinkmann RR, Iwasa Y, Hayden MR (2001) Measurement of mutational flow implies both a high new-mutation rate for Huntington disease and substantial underascertainment of late-onset cases. Am J Hum Genet 68(2):373-385.

Farrer LA, Opitz JM, Reynolds JF (1986) Suicide and attempted suicide in Huntington disease: implications for preclinical testing of persons at risk. Am J Med Genet 24(2):305-311.

Fisher ER, Hayden MR (2014) Multisource ascertainment of Huntington disease in Canada: Prevalence and population at risk. Mov Disord 29(1):105-114.

Folstein SE, Chase GA, Wahl W, McDonnell AM, Folstein MF (1987) Huntington disease in Maryland: clinical aspects of racial variation. Am J Hum Genet 41(2):168-179.

Fortune MT, Vassilopoulos C, Coolbaugh MI, Siciliano MJ, Monckton DG (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. Hum Mol Genet 9(3):439-445.

Futter M, Heckmann J, Greenberg L (2009) Predictive testing for Huntington disease in a developing country. Clin Genet 75(1):92-97.

Glass J, Saffer D (1979) Huntington's chorea in a black family: a report of 2 cases. S Afr Med J 56(17):685-688.

Goldberg YP, Andrew SE, Clarke LA, Hayden MR (1993a) A PCR method for accurate assessment of trinucleotide repeat expansion in Huntington disease. Hum Mol Genet 2(6):635-636.

Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, Adam S, Sajoo A, Starr E et al (1993b) Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. Nat Genet 5(2):174-179.

Goldberg YP, McMurray CT, Zeisler J, Almqvist E, Sillence D, Richards F, Gacy AM, Buchanan J, Telenius H, Hayden MR (1995) Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. Hum Mol Genet 4(10):1911-1918.

Gonzalez-Alegre P, Afifi AK (2006) Clinical characteristics of childhood-onset (juvenile) Huntington disease: report of 12 patients and review of the literature. J Child Neurol 21(3):223-229.

Greenberg LJ (1990). A Molecular Approach to Huntington Disease in Southern Africa. Dissertation, University of Cape Town.

Greenberg L, Martell R, Theilman J, Hayden M, Joubert J (1991) Genetic linkage between Huntington disease and the D4S10 locus in South African families: further evidence against non-allelic heterogeneity. Hum Genet 87(6):701-708.

Greenberg J, Beatty S, Soltau H, Bryer A (1996) A predictive testing service for Huntington disease (HD) and late onset spinocerebellar ataxia (SCA1) in Cape Town. CME Review 14:1364-1367.

Greenberg J, Solomon G, Vorster A, Heckmann J, Bryer A (2006) Origin of the SCA7 gene mutation in South Africa: implications for molecular diagnostics. Clin Genet 70(5):415-417.

Guida M, Fenwick RG, Papp AC, Snyder PJ, Sedra M, Prior TW (1996) Southern transfer protocol for confirmation of Huntington disease. Clin Chem 42(10):1711-1712.

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306(5940):234-238.

Gusella JF, MacDonald ME, Lee J (2014) Genetic modifiers of Huntington's disease. Mov Disord 29(11):1359-1365.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT 41:95-98.

Harper PS, Morris MJ, Tyler A (1990) Genetic testing for Huntington's disease. BMJ 300(6732):1089-1090.

Harper PS (1992a) Huntington disease and the abuse of genetics. Am J Hum Genet 50(3):460-464.

Harper PS (1992b) The epidemiology of Huntington's disease. Hum Genet 89(4):365-376.

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51(1):10-16.

Harper PS, Lim C, Craufurd D (2000) Ten years of presymptomatic testing for Huntington's disease: the experience of the UK Huntington's Disease Prediction Consortium. J Med Genet 37(8):567-571.

Hayden MR, Beighton P (1977) Huntington's chorea in the Cape coloured community of South Africa. S Afr Med J 52(22):886-888.

Hayden MR (1979). Huntington's chorea in South Africa. Dissertation, University of Cape Town.

Hayden MR, MacGregor J, Beighton P (1980a) The prevalence of Huntington's chorea in South Africa. S Afr Med J 58(5):193-196.

Hayden MR, Hopkins H, Macrae M, Beighton P (1980b) The origin of Huntington's chorea in the Afrikaner population of South Africa. S Afr Med J 58(5):197-200.

Hayden MR, Ehrlich R, Parker H, Ferera S (1980c) Social perspectives in Huntington's chorea. S Afr Med J 58(5):201-203.

Hayden MR (1981) Huntington's chorea. Springer-Verlag Berlin.

Hayden MR, MacGregor J, Saffer D, Beighton P (1982) The high frequency of juvenile Huntington's chorea in South Africa. J Med Genet 19(2):94-97.

Hayden MR, Goldblatt J, Wallis G, Winship IM, Beighton P (1987) Molecular genetics and Huntington's disease. The South African situation. S Afr Med J 71(11):683-686.

Hayden MR (2003) Predictive testing for Huntington's disease: a universal model? Lancet Neurol 2(3):141-142.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J et al (2012) Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet 8(1):e1002397.

Hu J, Matsui M, Gagnon KT, Schwartz JC, Gabillet S, Arar K, Wu J, Bezprozvanny I, Corey DR (2009) Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. Nat Biotechnol 27(5):478-484.

Hubers A, Reedeker N, Giltay E, Roos R, van Duijn E, van der Mast R (2012) Suicidality in Huntington's disease. J Affect Disord 136(3):550-557.

Hubers AA, van Duijn E, Roos RA, Craufurd D, Rickards H, Landwehrmeyer BG, van der Mast, Rose C, Giltay EJ, REGISTRY investigators of the European Huntington's Disease Network (2013) Suicidal ideation in a European Huntington's disease population. J Affect Disord 151(1):248-258.

Huntington Study Group PHAROS Investigators (2006) At risk for Huntington disease: The PHAROS (Prospective Huntington At Risk Observational Study) cohort enrolled. Arch Neurol 63(7):991-996.

Huntington G (1872) On Chorea. The Medical and Surgical Reporter: A Weekly Journal 26(15):317-321.

International Huntington Association (IHA) and the World Federation of Neurology (WFN)Research Group on Huntington's Chorea(1994) Guidelines for the molecular genetics predictive test in Huntington's disease. Neurology 44(8):1533-1536.

Jenkins T (1990) Medical genetics in South Africa. J Med Genet 27(12):760-779.

Jiang H, Sun Y, Hao Y, Yan Y, Chen K, Xin S, Tang Y, Li X, Jun T, Chen Y et al (2014) Huntingtin gene CAG repeat numbers in Chinese patients with Huntington's disease and controls. Eur J Neurol 21(4):637-642.

Kandil MR, Tohamy SA, Abdel Fattah M, Ahmed HN, Farwiez HM (1994) Prevalence of chorea, dystonia and athetosis in Assiut, Egypt: a clinical and epidemiological study. Neuroepidemiology 13(5):202-210.

Kay C, Skotte NH, Southwell AL, Hayden MR (2014) Personalized gene silencing therapeutics for Huntington disease. Clin Genet 86(1):29-36.

Killoran A, Biglan KM, Jankovic J, Eberly S, Kayson E, Oakes D, Young AB, Shoulson I (2013) Characterization of the Huntington intermediate CAG repeat expansion phenotype in PHAROS. Neurology 80(22):2022-2027.

Killoran A, Biglan KM (2014) Current therapeutic options for Huntington's disease: Good clinical practice versus evidence-based approaches? Mov Disord 29(11):1404-1413.

Kim JC, Mirkin SM (2013) The balancing act of DNA repeat expansions. Curr Opin Genet Dev 23(3):280-288.

Klintworth GK (1962) Huntington's chorea in South Africa. A preliminary communication drawing attention to its frequent occurrence. S Afr Med J 36:896-898.

Krause A, Greenberg LJ (2008) Genetic testing for Huntington's disease in South Africa. S Afr Med J 98(3):193-194.

Kremer B, Goldberg P, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, Lin B, Bassett A, Almqvist E et al (1994) A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. N Engl J Med 330(20):1401-1406.

Kremer B, Almqvist E, Theilmann J, Spence N, Telenius H, Goldberg YP, Hayden MR (1995) Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. Am J Hum Genet 57(2):343-350.

Kubodera T, Yokota T, Ishikawa K, Mizusawa H (2005) New RNAi strategy for selective suppression of a mutant allele in polyglutamine disease. Oligonucleotides 15(4):298-302.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5(2):150-163.

Kumar S, Tamura K, Nei M (1994) MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci 10(2):189-191.

La Spada AR, Wilson EM, Lubahn DB, Harding A, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352(6330):77-79.

La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet 11(4):247-258.

Laccone F, Engel U, Holinski-Feder E, Weigell-Weber M, Marczinek K, Nolte D, Morris-Rosendahl DJ, Zuhlke C, Fuchs K, Weirich-Schwaiger H et al (1999) DNA analysis of Huntington's disease: five years of experience in Germany, Austria, and Switzerland. Neurology 53(4):801-806.

Lane A, Soodyall H, Arndt S, Ratshikhopha M, Jonker E, Freeman C, Young L, Morar B, Toffie L (2002) Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies. Am J Phys Anthropol 119(2):175-185.

Langbehn D, Brinkman R, Falush D, Paulsen J, Hayden M (2004) A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 65(4):267-277.

Langbehn DR, Hayden MR, Paulsen JS,PREDICT-HD Investigators of the Huntington Study Group (2010) CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. Am J Med Genet B Neuropsychiatr Genet 153B(2):397-408.

Lanska DJ (2000) George Huntington (1850-1916) and hereditary chorea. J Hist Neurosci 9(1):76-89.

Lee JM, Ramos EM, Lee JH, Gillis T, Mysore JS, Hayden MR, Warby SC, Morrison P, Nance M, Ross CA et al (2012) CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. Neurology 78(10):690-695.

Lekoubou A, Echouffo-Tcheugui JB, Kengne AP (2014) Epidemiology of neurodegenerative diseases in sub-Saharan Africa: a systematic review. BMC Public Health 14:653.

Li J, Hayden MR, Almqvist EW, Brinkman RR, Durr A, Dodé C, Morrison PJ, Suchowersky O, Ross CA, Margolis RL et al (2003) A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS study. Am J Hum Genet 73(3):682-687.

Li J, Hayden MR, Warby SC, Durr A, Morrison PJ, Nance M, Ross CA, Margolis RL, Rosenblatt A, Squitieri F et al (2006) Genome-wide significance for a modifier of age at neurological onset in Huntington's disease at 6q23-24: the HD MAPS study. BMC Med Genet 7:71.

Libby RT, Hagerman KA, Pineda VV, Lau R, Cho DH, Baccam SL, Axford MM, Cleary JD, Moore JM, Sopher BL et al (2008) CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet 4(11):e1000257.

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science 293(5531):864-867.

Losekoot M, van Belzen MJ, Seneca S, Bauer P, Stenhouse SA, Barton DE, European Molecular Genetic Quality Network (EMQN) (2013) EMQN/CMGS best practice guidelines for the molecular genetic testing of Huntington disease. Eur J Hum Genet 21(5):480-486.

Louw PE (2004) The rise, fall, and legacy of apartheid. Greenwood Publishing Group.

Loy C, Lownie A, McCusker E (2010) Huntington's disease. Lancet 376(9751):1463.

Ma M, Yang Y, Shang H, Su D, Zhang H, Ma Y, Liu Y, Tao D, Zhang S (2010) Evidence for a predisposing background for CAG expansion leading to *HTT* mutation in a Chinese population. J Neurol Sci 298(1-2):57-60.

MacDonald ME, Haines JL, Zimmer M, Cheng SV, Youngman S, Whaley WL, Wexler N, Bucan M, Allitto BA, Smith B et al (1989) Recombination events suggest potential sites for the Huntington's disease gene. Neuron 3(2):183-190.

MacDonald ME, Lin C, Srinidhi L, Bates G, Altherr M, Whaley W, Lehrach H, Wasmuth J, Gusella J (1991) Complex patterns of linkage disequilibrium in the Huntington disease region. Am J Hum Genet 49(4):723.

MacDonald ME, Novelletto A, Lin C, Tagle D, Barnes G, Bates G, Taylor S, Allitto B, Altherr M, Myers R et al (1992) The Huntington's disease candidate region exhibits many different haplotypes. Nat Genet 1(2):99-103.

MacLeod R, Tibben A, Frontali M, Evers-Kiebooms G, Jones A, Martinez-Descales A, Roos R, Editorial Committee and Working Group 'Genetic Testing Counselling' of the European Huntington Disease Network (2013) Recommendations for the predictive genetic test in Huntington's disease. Clin Genet 83(3):221-231.

Magazi D, Krause A, Bonev V, Moagi M, Iqbal Z, Dludla M, Van der Meyden C (2008) Huntington's disease: Genetic heterogeneity in black African patients. S Afr Med J 98(3):200-203.

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

Margolis RL, Ross CA (2003) Diagnosis of Huntington disease. Clin Chem 49(10):1726-1732.

Martino D, Stamelou M, Bhatia KP (2013) The differential diagnosis of Huntington's disease-like syndromes: 'red flags' for the clinician. J Neurol Neurosurg Psychiatry 84(6):650-656.

Masuda N, Goto J, Murayama N, Watanabe M, Kondo I, Kanazawa I (1995) Analysis of triplet repeats in the huntingtin gene in Japanese families affected with Huntington's disease. J Med Genet 32(9):701-705.

Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L et al (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet 26(2):191-194.

Meissen GJ, Myers RH, Mastromauro CA, Koroshetz WJ, Klinger KW, Farrer LA, Watkins PA, Gusella JF, Bird ED, Martin JB (1988) Predictive testing for Huntington's disease with use of a linked DNA marker. N Engl J Med 318(9):535-542.

Montano V, Ferri G, Marcari V, Batini C, Anyaele O, Destro-Bisol G, Comas D (2011) The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. Mol Ecol 20(13):2693-2708.

Moolman-Smook JC, De Lange WJ, Bruwer EC, Brink PA, Corfield VA (1999) The origins of hypertrophic cardiomyopathy–causing mutations in two South African subpopulations: a unique profile of both independent and founder events. Am J Hum Genet 65(5):1308-1320.

Morovvati S, Nakagawa M, Osame M, Karami A (2008) Analysis of CCG repeats in Huntingtin gene among HD patients and normal populations in Japan. Arch Med Res 39(1):131-133.

Morrison PJ, Johnston WP, Nevin NC (1995) The epidemiology of Huntington's disease in Northern Ireland. J Med Genet 32(7):524-530.

Morrison PJ (2010) Accurate prevalence and uptake of testing for Huntington's disease. Lancet Neurol 9(12):1147.

Morrison P, Harding-Lester S, Bradley A (2011) Uptake of Huntington disease predictive testing in a complete population. Clin Genet 80(3):281-286.

Morrison PJ (2012) Prevalence estimates of Huntington disease in Caucasian populations are gross underestimates. Mov Disord 27(13):1707-1708.

Morrison PJ (2014) Huntington disease in 2013–genetic choices across the life cycle. Clin Genet 85(1):76-77.

Mountain A (2003) The First People of the Cape: A Look at Their History and the Impact of Colonialism on the Cape's Indigenous People. New Africa Books.

Murphy RM, Walters RH, Tobelmann MD, Bernacki JP (2012) When More Is Not Better: Expanded Polyglutamine Domains in Neurodegenerative Disease Non-fibrillar Amyloidogenic Protein Assemblies-Common Cytotoxins Underlying Degenerative Diseases Springer, pp 337-375.

Myers R, MacDonald M, Koroshetz W, Duyao M, Ambrose C, Taylor S, Barnes G, Srinidhi J, Lin C, Whaley W et al (1993) De novo expansion of a (CAG) n repeat in sporadic Huntington's disease. Nat Genet 5(2):168-173.

Nakashima K, Watanabe Y, Kusumi M, Nanba E, Maeoka Y, Nakagawa M, Igo M, Irie H, Ishino H, Fujimoto A et al (1996) Epidemiological and genetic studies of Huntington's disease in the San-in area of Japan. Neuroepidemiology 15(3):126-131.

Nance MA, Myers RH (2001) Juvenile onset Huntington's disease—clinical and research perspectives. Ment Retard Dev Disabil Res Rev 7(3):153-157.

Nance MA (2012) Therapy in Huntington's disease: where are we? Curr Neurol Neurosci Rep 12(4):359-366.

Ned RM, Sijbrands EJ (2011) Cascade screening for familial hypercholesterolemia (FH). PLoS Curr 3:RRN1238.

Nelson D, Orr H, Warren S (2013) The unstable repeats—three evolving faces of neurological disease. Neuron 77(5):825-843.

Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. Annu Rev Neurosci 30:575-621.

Osler W (1894) On chorea and choreiform affections. Philadelphia: Blakiston P.

Paradisi I, Hernández A, Arias S (2008) Huntington disease mutation in Venezuela: age of onset, haplotype analyses and geographic aggregation. J Hum Genet 53(2):127-135.

Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, Reich D, Hayes VM (2010) Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet 19(3):411-419.

Paulsen JS, Langbehn DR, Stout JC, Aylward E, Ross CA, Nance M, Guttman M, Johnson S, MacDonald M, Beglinger LJ et al (2008) Detection of Huntington's disease decades before diagnosis: the Predict-HD study. J Neurol Neurosurg Psychiatry 79(8):874-880.

Paulson HL, Fischbeck KH (1996) Trinucleotide repeats in neurogenetic disorders. Annu Rev Neurosci 19:79-107.

Pearson CE, Edamura KN, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6(10):729-742.

Pecheux C, Mouret JF, Durr A, Agid Y, Feingold J, Brice A, Dode C, Kaplan JC (1995) Sequence analysis of the CCG polymorphic region adjacent to the CAG triplet repeat of the HD gene in normal and HD chromosomes. J Med Genet 32(5):399-400.

Peterlin B, Kobal J, Teran N, Flisar D, Lovrečić L (2009) Epidemiology of Huntington's disease in Slovenia. Acta Neurol Scand 119(6):371-375.

Potter NT, Spector EB, Prior TW (2004) Technical standards and guidelines for Huntington disease testing. Genet Med 6(1):61-65.

Pramanik S, Basu P, Gangopadhaya PK, Sinha KK, Jha D, Sinha S, Das SK, Maity BK, Mukherjee SC, Roychoudhuri S et al (2000) Analysis of CAG and CCG repeats in Huntingtin gene among HD patients and normal populations of India. Eur J Hum Genet 8(9):678-682.

Pringsheim T, Jette N (2012) Reply to letter: Prevalence estimates of Huntington's disease in Caucasian populations are gross underestimates. Mov Disord 27(13):1708-1709.

Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N (2012) The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. Mov Disord 27(13):1083-1091.

Pulkes T, Papsing C, Wattanapokayakit S, Mahasirimongkol S (2014) CAG-Expansion Haplotype Analysis in a Population with a Low Prevalence of Huntington's Disease. J Clin Neurol 10(1):32-36.

Quarrell OW, Handley O, O'Donovan K, Dumoulin C, Ramos-Arroyo M, Biunno I, Bauer P, Kline M, Landwehrmeyer GB, European Huntington's Disease Network (2012a) Discrepancies in reporting the CAG repeat lengths for Huntington's disease. Eur J Hum Genet 20(1):20-26.

Quarrell O, O'Donovan KL, Bandmann O, Strong M (2012b) The prevalence of juvenile Huntington's disease: a review of the literature and meta-analysis. PLoS Curr 4:e4f8606b742ef3.

Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM (2010) Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. Am J Hum Genet 86(4):611-620.

Ramos EM, Latourelle JC, Lee J, Gillis T, Mysore JS, Squitieri F, Di Pardo A, Di Donato S, Hayden MR, Morrison PJ et al (2012) Population stratification may bias analysis of PGC-1α as a modifier of age at Huntington disease motor onset. Hum Genet 131(12):1833-1840.

Ramos-Arroyo MA, Moreno S, Valiente A (2005) Incidence and mutation rates of Huntington's disease in Spain: experience of 9 years of direct genetic testing. J Neurol Neurosurg Psychiatry 76(3):337-342.

Ranen NG, Stine OC, Abbott MH, Sherr M, Codori AM, Franz ML, Chao NI, Chung AS, Pleasant N, Callahan C et al (1995) Anticipation and instability of IT-15 (CAG)n repeats in parent-offspring pairs with Huntington disease. Am J Hum Genet 57(3):593-602.

Rawlins M (2010) Huntington's disease out of the closet? Lancet 376(9750):1372-1373.

Robins Wahlin T, Bäckman L, Lundin A, Haegermark A, Winblad B, Anvret M (2000) High suicidal ideation in persons testing for Huntington's disease. Acta Neurol Scand 102(3):150-161.

Roos RA (2009) The history of juvenile Huntington's disease. In: Juvenile Huntington's Disease: And Other Trinucleotide Repeat Disorders. Oxford University Press, Oxford pp 33-38.

Roos RA (2010) Huntington's disease: a clinical review. Orphanet J Rare Dis 5:40.

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Ramesar RS, Old J, Bontrop R, McMahon R, Barton DE, Ferguson-Smith MA (1994) Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. Nat Genet 7(4):525-530.

Rubinsztein DC, Leggo J, Goodburn S, Barton DE, Ferguson-Smith MA (1995) Haplotype analysis of the delta 2642 and (CAG)n polymorphisms in the Huntington's disease (HD) gene provides an explanation for an apparent 'founder' HD haplotype. Hum Mol Genet 4(2):203-206.

Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman J, Chotai K, Connarty M, Craufurd D, Curtis A et al (1996) Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am J Hum Genet 59(1):16-22.

Saffer D, Nathan D, Kahle P, Steingo B (1974) Huntington's Disease in a Coloured family. S Afr Med J 48(58):2399-2402.

Schneider SA, Walker RH, Bhatia KP (2007) The Huntington's disease-like syndromes: what to consider in patients with a negative Huntington's disease gene test. Nat Clin Pract Neurol 3(9):517-525.

Scholefield J (2005) Detailed investigation of the unstable (CAG) repeat and the immediate surrounding region of the IT15 gene in some South African families with Huntington disease. Dissertation, University of Cape Town.

Scholefield J, Greenberg J (2007) A common SNP haplotype provides molecular proof of a founder effect of Huntington disease linking two South African populations. Eur J Hum Genet 15(5):590-595.

Scholefield J (2008) RNAi based allele-specific silencing of the disease-causing gene in black South African patients with SCA7. Dissertation, University of Cape Town.

Scholefield J, Greenberg LJ, Weinberg MS, Arbuthnot PB, Abdelgany A, Wood MJ (2009) Design of RNAi hairpins for mutation-specific silencing of ataxin-7 and correction of a SCA7 phenotype. PloS One 4(9):e7232.

Scholefield J, Wood MJ (2010) Therapeutic gene silencing strategies for polyglutamine disorders. Trends Genet 26(1):29-38.

Scholefield J, Watson L, Smith D, Greenberg J, Wood MJ (2014) Allele-specific silencing of mutant Ataxin-7 in SCA7 patient-derived fibroblasts. Eur J Hum Genet 22(12):1369-1375.

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

Semaka A, Collins J, Hayden M (2010) Unstable familial transmissions of Huntington disease alleles with 27–35 CAG repeats (intermediate alleles). Am J Med Genet B Neuropsychiatr Genet153B(1):314-320.

September A (1999). Molecular investigation of the trinucleotide repeats within the Huntington disease gene in Southern Africa. Dissertation, University of Cape Town.

Sequeiros J, Ramos E, Cerqueira J, Costa MdC, Sousa A, Pinto-Basto J, Alonso I (2010) Large normal and reduced penetrance alleles in Huntington disease: instability in families and frequency at the laboratory, at the clinic and in the population. Clin Genet 78(4):381-387.

Shang H, Danek A, Landwehrmeyer B, Burgunder J (2012) Huntington's disease: new aspects on phenotype and genotype. Parkinsonism Relat Disord 18:S107-S109.

Shaw M, Caro A (1982) The mutation rate to Huntington's chorea. J Med Genet 19(3):161-167.

Siesling S, Vegter-van der Vlis M, Roos RA (1997) Juvenile Huntington disease in the Netherlands. Pediatr Neurol 17(1):37-43.

Sizer EB, Haw T, Wessels T, Kromberg JG, Krause A (2012) The utilization and outcome of diagnostic, predictive, and prenatal genetic testing for Huntington disease in Johannesburg, South Africa. Genet Test Mol Biomarkers 16(1):58-62.

Smith D, Bryer A, Watson L, Greenberg LJ (2012) Inherited polyglutamine spinocerebellar ataxias in South Africa. S Afr Med J 102(8):683-686.

Snell RG, Thompson LM, Tagle DA, Holloway TL, Barnes G, Harley HG, Sandkuijl LA, MacDonald ME, Collins FS, Gusella JF et al (1992) A recombination event that redefines the Huntington disease region. Am J Hum Genet 51(2):357-362.

Snell RG, MacMillan JC, Cheadle JP, Fenton I, Lazarou LP, Davies P, MacDonald ME, Gusella JF, Harper PS, Shaw DJ (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. Nat Genet 4(4):393-397.

Soodyall H, Makkan H, Haycock P, Naidoo T (2008) The genetic prehistory of the Khoe and San. Southern African Humanities 20:37-48.

Spinney L (2010) Uncovering the true prevalence of Huntington's disease. Lancet Neurol 9(8):760-761.

Squitieri F, Andrew S, Goldberg Y, Kremer B, Spence N, Zelsler J, Nichol K, Theilmann J, Greenberg J, Goto J et al (1994) DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum Mol Genet 3(12):2103-2114.

Starfield M, Hennies HC, Jung M, Jenkins T, Wienker T, Hull P, Spurdle A, Küster W, Ramsay M, Reis A (1997) Localization of the gene causing keratolytic winter erythema to chromosome 8p22-p23, and evidence for a founder effect in South African Afrikaans-speakers. Am J Hum Genet 61(2):370-378.

Stata Corporation (2011) Stata Statistical Software: Release 12.0. Stata Corporation.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68(4):978-989.

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73(5):1162-1169.

Sturrock A, Leavitt BR (2010) The clinical and genetic features of Huntington disease. J Geriatr Psychiatry Neurol 23(4):243-259.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30(12):2725-2729.

Tanaka F, Sobue G, Doyu M, Ito Y, Yamamoto M, Shimada N, Yamamoto K, Riku S, Hshizume Y, Mitsuma T (1996) Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeat in dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. J Neurol Sci 135(1):43-50.

Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, Adam S, Greenberg C, Ives EJ, Clarke LA et al (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nat Genet 6(4):409-414.

Telenius H, Almqvist E, Kremer B, Spence N, Squitieri F, Nichol K, Grandell U, Starr E, Benjamin C, Castaldo I et al (1995) Somatic mosaicism in sperm is associated with intergenerational (CAG)n changes in Huntington disease. Hum Mol Genet 4(2):189-195.

The American College of Medical Genetics/American Society of Human Genetics Huntington Disease Genetic Testing Working Group (1998) ACMG/ASHG Statement. Laboratory guidelines for Huntington disease genetic testing. Am J Hum Genet 62(5):1243-1247.

The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72(6):971-983.

The U.S.–Venezuela Collaborative Research Project, Wexler NS (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. Proc Natl Acad Sci USA 101(10):3498-3503.

Töpper R, Schwarz M, Lange HW, Hefter H, Noth J (1998) Neurophysiological abnormalities in the Westphal variant of Huntington's disease. Mov Disord 13(6):920-928.

van Duijn E, Craufurd D, Hubers AA, Giltay EJ, Bonelli R, Rickards H, Anderson KE, van Walsem MR, van der Mast RC, Orth M et al (2014) Neuropsychiatric symptoms in a European Huntington's disease cohort (REGISTRY). J Neurol Neurosurg Psychiatry 85(12):1411-1418.

Van Rij MC, De Rademaeker M, Moutou C, Dreesen JC, De Rycke M, Liebaers I, Geraedts JP, De Die-Smulders CE, Viville S, BruMaStra PGD working group (2011) Preimplantation genetic diagnosis (PGD) for Huntington's disease: the experience of three European centres. Eur J Hum Genet 20(4):368-375.

Vázquez-Mojena Y, Laguna-Salvia L, Laffita-Mesa JM, González-Zaldívar Y, Almaguer-Mederos LE, Rodríguez-Labrada R, Almaguer-Gotay D, Zayas-Feria P, Velázquez-Pérez L (2013) Genetic features of Huntington disease in Cuban population: implications for phenotype, epidemiology and predictive testing. J Neurol Sci 335(1-2):101-104.

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu Y, Kuhl D, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65(5):905-914.

Vonsattel JPG, Keller C, Pilar Amaya Md (2008) Neuropathology of Huntington's disease Handb Clin Neurol 89:599-618.

Walker FO (2007) Huntington's disease. Lancet 369(9557):218-228.

Warby SC, Montpetit A, Hayden AR, Carroll JB, Butland SL, Visscher H, Collins JA, Semaka A, Hudson TJ, Hayden MR (2009) CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am J Hum Genet 84(3):351-366.

Warby SC, Visscher H, Collins JA, Doty CN, Carter C, Butland SL, Hayden AR, Kanazawa I, Ross CJ, Hayden MR (2011) *HTT* haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. Eur J Hum Genet 19(5):561-566.

Warner JP, Barron LH, Goudie D, Kelly K, Dow D, Fitzpatrick DR, Brock DJ (1996) A general method for the detection of large CAG repeat expansions by fluorescent PCR. J Med Genet 33(12):1022-1026.

Warnich L, Drogemoller BI, Pepper MS, Dandara C, Wright GE (2011) Pharmacogenomic Research in South Africa: Lessons Learned and Future Opportunities in the Rainbow Nation. Curr Pharmacogenomics Person Med 9(3):191-207.

Wasmuth JJ, Hewitt J, Smith B, Allard D, Haines JL, Skarecky D, Partlow E, Hayden MR (1988) A highly polymorphic locus very tightly linked to the Huntington's disease gene. Nature 332(6166):734-736.

Watson LM, Wood MJ (2012) RNA therapy for polyglutamine neurodegenerative diseases. Exp Rev Mol Med 14:e3.

Wells RD, Dere R, Hebert ML, Napierala M, Son LS (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. Nucleic Acids Res 33(12):3785-3798.

Went L (1990) Ethical issues policy statement on Huntington's disease molecular genetics predictive test. International Huntington Association. World Federation of Neurology. J Med Genet 27(1):34-38.

Wetzel HH, Gehl CR, Dellefave–Castillo L, Schiffman JF, Shannon KM, Paulsen JS, Huntington Study Group (2011) Suicidal ideation in Huntington disease: the role of comorbidity. Psychiatry Res 188(3):372-376.

Wexler A (2010) Stigma, history, and Huntington's disease. Lancet 376(9734):18-19.

Whaley WL, Michiels F, MacDonald ME, Romano D, Zimmer M, Smith B, Leavitt J, Bucan M, Haines JL, Gilliam TC et al (1988) Mapping of D4S98/S114/S113 confines the Huntington's defect to a reduced physical region at the telomere of chromosome 4. Nucleic Acids Res 16(24):11769-11780.

Wild EJ, Tabrizi SJ (2007) Huntington's disease phenocopy syndromes. Curr Opin Neurol 20(6):681-687.

Wild EJ, Mudanohwo EE, Sweeney MG, Schneider SA, Beck J, Bhatia KP, Rossor MN, Davis MB, Tabrizi SJ (2008) Huntington's disease phenocopies are clinically and genetically heterogeneous. Mov Disord 23(5):716-720.

Wild EJ, Tabrizi SJ (2014) Targets for future clinical trials in Huntington's disease: What's in the pipeline? Mov Disord 29(11):1434-1445.

Winship W, Beighton P (2011) Genetic disorders in the Indian community of South Africa. S Afr Med J 101(7):481-484.

Young AB, Shoulson I, Penney JB, Starosta-Rubinstein S, Gomez F, Travers H, Ramos-Arroyo MA, Snodgrass SR, Bonilla E, Moreno H et al (1986) Huntington's disease in Venezuela: neurologic features and functional decline. Neurology 36(2):244-249.

Websites:

Enroll-HD: www.enroll-hd.net

Huntington's Associations of South Africa: www.huntingtons.org.za

Predictive Testing for HD: www.predictivetestingforHD.com

Statistics South Africa: www.statssa.gov.za

Appendices

Appendix 1: Standard consent form for molecular tests and project ethics approval

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory
Division of Human Genetics
IIDMM, LEVEL 3
UCTMedicalSchool, Observatory 7925

Tel: (021) 406 6425 Fax: (021) 406 6826

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel.
Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB
Keep blood in fridge at 4°C until able to send to laboratory

Please DO NOT send specimens on ice or frozen.

Please fill in all the information requested:

Surname:_____First Name(s):_____

New Family: Yes ☐ No ☐ (If no, please fill in family name) Family name: _____

Medical Aid: _____Medical Aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year:_____ Month:_____ Day: _____

Number of children: _____

Ethnic Origin: (please indicate ancestry of both your mother and father) _____

Contact Address: _____ Town: _____ Fax: Tel: _____

Referring Doctor/Sister: _____ Town: _____ Fax: Tel: _____

Hospital or Address: _____ Town: _____ Fax: Tel: _____

Reason for Referral (Clinical diagnosis):

Affected ☐ At Risk ☐ Carrier ☐ Spouse ☐ Query ☐ Unaffected ☐

Becker Muscular Dys. ☐	Duchenne Muscular Dys ☐	Colonic Carcinoma ☐
Fragile-X Syndrome ☐	Bipolar Disorder ☐	Huntington Disease ☐
Retinitis Pigmentosa ☐	Spinocerebellar Ataxia ☐	Waardenberg Syndrome ☐

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

Physical disability ☐ Mental retardation ☐ Deafness ☐ Impaired vision ☐ Night blindness ☐

Other: _____

Have samples from this patient been sent to a DNA lab before? (DELETE WHERE NOT APPLICABLE) YES / NO / Don't Know

If Yes, where: _____

For Laboratory use only:

DNA number: _____ *Vol.Blood:* _____*(ml)* *Other:* _____

Date Received: Year: _____ *Month:* _____ *Day:* _____ *Computer Index No:*_____

CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____ , request that an attempt be made using genetic material to assess the probability that: I / my child / my unborn child (DELETE WHERE NOT APPLICABLE) might have inherited a disease-causing mutation in the gene for:_____

2. I understand that the genetic material for analysis is to be obtained from: blood cells/skin sample/other (specify) (DELETE WHERE NOT APPLICABLE) :

3. I request that **no** portion of the sample be stored for later use. ☐ **(**MARK IF APPLICABLE)
   Or
   I request that a portion of the sample be stored indefinitely for (DELETE WHERE NOT APPLICABLE):
   (a) possible re-analysis
   (b) analysis for the benefit of members of my immediate family
   (c) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee, provided that any information from such research will remain confidential.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available.
   In addition, I authorise that they may be made known to: (DELETE WHERE NOT APPLICABLE) :
   other doctors involved in my care
   the following family members: _____
   other: _____

5. I authorise / do not authorise my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UCT.

6. I have been informed that:
   (a) there are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   (b) the analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   (c) the genetics laboratory is under an obligation to respect medical confidentiality .
   (d) genetic analysis may not be informative for some families or family members.
   (e) even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   (f) where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

_____ DATE: _____

Patient signature _____ **Witnessed consent** _____

NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM

UNIVERSITY OF CAPE TOWN

Health Sciences Faculty
Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6626 • **Facsimile** [021] 406 6411
e-mail: lamees.emjedi@uct.ac.za

29 September 2010

HREC REF: 450/2010 (linked to 229/2010)

Prof J Greenberg
Human Genetics,
Clinical Laboratory Sciences

Dear Prof Greenberg

PROJECT TITLE: HAPLOTYPE ANALYSIS OF DNA FROM SOUTH AFRICAN PATIENTS WITH HUNTINGTON DISEASE THAT IS STORED IN THE UCT NEURODEGENERATIVE DISORDERS REPOSITORY AND THE DETERMINATION OF THE FREQUENCY OF INTERMEDIATE AND EXPANDED HTT ALLELES IN THE GENERAL POPULATION IN SOUTH AFRICA

Thank you for submitting your new study to the Faculty of Health Sciences Human Research Ethics Committee.

It is a pleasure to inform you that the FHS HREC has **formally approved** the above-mentioned study.

Approval is granted for one year until 15 October 2011.

This approval includes permission to de-identify and irreversibly anonymise selected DNA samples from archived biological material referenced in the Master Registry for Neurodegenerative Disorders (HREC REF 229/2010) under the auspices of Professor Jacquie Greenberg and the Master Registry of Retinal Degenerative Disorders (HREC REF 226/2010) under the auspices of Professor Raj Ramesar and Professor Jacquie Greenberg

Please send us an annual progress report (website form FHS 016) if your research continues beyond the approval period. Alternatively, please send us a brief summary of your findings so that we can close the research file.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the REC. REF in all your correspondence.

lemjedi

UNIVERSITY OF CAPE TOWN

Faculty of Health Sciences
Human Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Ms S Ariefdien - Tel: [021]4066492 - Fax: [021]4066411
email: sumayah.ariefdien@uct.ac.za

07 May 2012

HREC REF. 215/2012

Prof J Greenberg,
Human Genetics
Suite n3
Wernher & Beit North Building

Dear Prof Greenberg,

PROJECT TITLE: THE MOLECULAR GENETIC BASIS OF JUVENILE HUNTINGTON DISEASE IN RELATION TO DISEASE-ASSOCIATED HAPLOGROUPS IN THE SOUTH AFRICAN POPULATION

Thank you for submitting your new study to the Faculty of Health Sciences Human Research Ethics Committee

It is a pleasure to inform you that the Ethics Committee has formally approved the above-mentioned study.

Approval is granted until 15 May 2013

Please submit an annual progress report (FHS016) if the research continues beyond the expiry date. Please submit a brief summary of findings if you complete the study within the approval period so that we can close our file (FHS010).

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC. REF in all your correspondence.

Yours sincerely

Signed by candidate

PROFESSOR MARC BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312

Appendix 2: Annotated *HTT* gene sequence

Key:

The *HTT* gene promoter sequence is highlighted in yellow

The start codon (ATG) is highlighted in red

Exon 1 is in UPPERCASE

CAG and CCG repeats are indicated in RED and GREEN respectively

Primer sequences for PCR amplification and sequencing of the promoter and exon 1 are highlighted in blue

. . .

8701 ccctgcccgc cacggcctgt gtcccaggcg tgaggggtg ccccacagacctctgctgag

8761 ctgctgctga atgacgcccc ttgggggtcc tgccggaagg tcagagcagg ggtgcactcc

8821 cataaagaaa cgcccccagg tcgggactca ttcctgtggg cggcatcttg tggccatagc

8881 tgcttctcgc tgcactaatc acagtgcctc tgtgggcagc aggcgctgac cacccaggcc

8941 tgccccagac cctctcctcc cttccggggc gctgcgctgg gaccgatggg gggcgccagg

9001 cctgtggaca ccgccctgca ggggcctctc cagctcactg ggggtggggt gggggtcaca

9061 cttggggtcc tcaggtcgtg ccgaccacgcgcattctctg cgctctgcgc aggagctcgc

9121 ccaccctctc cccgtgcaga gagccccgca gctggctccc cgcagggctg tccgggtgag

9181 tatggctctg gccacgggcc agtgtggcgg gagggcaaac cccaaggcca cctcggctca

9241 gagtccacgg ccggctgtcg ccccgctcca ggcgtcggcg ggggatcctt tccgcatggg

9301 cctgcgcccg cgctcggcgc cccctccacg gccccgcccc gtccatggcc ccgtccttca

9361 tgggcgagcc cctccatggc cctgcccctc cgcgccccac ccctccctcg ccccacctct

9421 caccttcctg ccccgccccc agcctcccca cccctcaccg gccagtcccc tcccctatcc

9481 cgctccgccc ctcagccgcc ccgcccctca gccggcctgc ctaatgtccc cgtccccagc

9541 atcgccccgc cccgcccccg tctcgccccg cccctcaggc ggcctccctg ctgtgccccg

9601 ccccggcctc gccacgcccc tacctcacca cgcccccccgc atcgccacgc cccccgcatc

9661 gccacgcctcccttaccatg cagtcccgcc ccgtcccttc ctcgtcccgc ctcgccgcga

9721 cacttcacac acagcttcgc ctcaccccat tacagtctca ccacgccccg tcccctctcc

9781 gttgagcccc gcgccttcgc ccgggtgggg cgctgcgctg tcagcggcct tgctgtgtga

9841 ggcagaacct gcgggggcag gggcgggctg gttccctggc cagccattgg cagagtccgc

9901 aggctagggc tgtcaatcat gctggccggc gtggccccgc ctccgccggc gcggccccgc

9961 ctccgccggc gcagcgtctg ggacgcaagg cgccgtgggg **GCTGCCGGGA CGGGTCCAAG**

10021 **ATGGACGGCC GCTCAGGTTC TGCTTTTACC TGCGGCCCAG AGCCCCATTC ATTGCCCCGG**

10081 **TGCTGAGCGG CGCCGCGAGT** CGGCCCGAGG CCTCCGGGGA CTGCCGTGCC GGGCGGGAGA

10141 **CCGCCATGGC GACCCTGGAA AAGCTGATGA AGGCCTTCGA GTCCCTCAAG TCCTTCCAGC**

10201 **AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAACAGC**

10261 CGCCA**CCGCC GCCGCCGCCG CCGCCGCCTC CTCAGCTTCC TCAGCCGCCG CCGCAGGCAC**

10321 **AGCCGCTGCT GCCTCAGCCG CAGCCGCCCC CGCCGCCGCC CCCGCCGCCA CCCGGCCCGG**

10381 **CTGTGGCTGA GGAGCCGCTG CACCGACC**gt gagtttgggc ccgctgcagc tccctgtccc

10441 ggcgggtccc aggctacggc gggatggcggtaaccctgcagcctgcggg ccggcgacac

10501 gaacccccgg ccccgcagag acagagtgac ccagcaaccc agagcccatg agggacaccc

10561 gccccctcct ggggcgaggc cttcccccac ttcagccccg ctccctcact tgggtcttcc

10621 cttgtcctct cgcgagggga ggcagagcct tgttggggcc tgtcctgaat tcaccgaggg

10681 gagtcacggc ctcagccctc tcgcccttcg caggatgcga agagttgggg cgagaacttg

10741 tttcttttta tttgcgagaa accagggcgg gggttctttt aactgcgttg tgaagagaac

Appendix 3: Primer sequences

Genotyping	**Primer**	**Sequence (5'-3')**
CAG repeat	HD344F_HEX*	CCT TCG AGT CCC TCA AGT CCT TC
	HD450R_PT#	GTT TGG CGG CGG TGG CGG CTG TTG
CAG and CCG	HD344F_6FAM*	CCT TCG AGT CCC TCA AGT CCT TC
	HD419F_NED*	AGC AGC AGC AGC AAC AGC C
	HD482R	GGC TGA GGA AGC TGA GGA G
DNA sequencing		
CAG repeat	HDCAGF	CTT CGA GTC CCT CAA GTC CTT C
	HDCAGR	CGG CGG TGG CGG CTG TTG C
HTT promoter and exon 1	*HTT* PromF	CGA CCA CGC GCA TTC TCT G
	HTT PromR	GCT GCA GGG TTA CCG CCA TC
	HTT Promseq2	ATC GCC ACG CCT CCC TTA C

*HEX (green), 6-FAM (blue) and NED (yellow) are fluorescent tags that enable the automatic separation and sizing of the PCR products based on the emitted light.

#PT: pig-tail, a modification that adds the sequence GTTT to the 5' end of the oligo to ensure that PCR products end in nucleotide A; this reduces stutter.

Appendix 4: Optimised PCR conditions

1. Genotyping (Pure CAG)

Reagent	Stock concentration	Final concentration (in 15 µL)
HD344F_HEX (Invitrogen)	10 µM	0.6 µM
HD450R_PT (Invitrogen)	10 µM	0.6 µM
PCR buffer	10X	1X
dNTPs	2.5 mM	0.2 mM
HiDi formamide	100%	3.5%
Glycerol	50%	15%
Taq polymerase (Roche)	5 U	0.05 U

1 µL DNA at 100 ng/ µL was added and made up to the final volume with sterile water.

Thermo cycling conditions:

95°C for 4 minutes, 35x (95°C for 45 seconds, 61°C for 45 seconds, 72°C for 45 seconds), 72°C for 5 minutes.

2. Genotyping (CAG and CCG)

Reagent	Stock concentration	Final concentration (in 15 µL)
HD344F_6FAM (IDT)	10 µM	0.4 µM
HD419F_NED (ABI)	10 µM	0.7 µM
HD482R (Invitrogen)	10 µM	1.2 µM
dNTPs	2.5 mM	0.2 mM
PCR buffer	10X	1X
HiDi formamide	100%	3.5%
Glycerol	50%	15%
Taq polymerase (Roche)	5 U	0.05 U

1 µL DNA at 100 ng/ µL was added and made up to the final volume with sterile water.

Thermo cycling conditions:

95°C for 4 minutes, 35x (95°C for 45 seconds, 61°C for 45 seconds, 72°C for 45 seconds), 72°C for 5 minutes.

3. Sequencing (CAG-tract PCR)

Reagent	Stock concentration	Final concentration (in 25 µL)
HDCAGF primer (IDT)	10 µM	0.32 µM
HDCAGR primer (IDT)	10 µM	0.32 µM
dNTPs	2.5 mM	0.1 mM
GoTaq Flexi buffer (Promega)	5X	1X
Magnesium chloride	25 mM	2.5 mM
Potassium chloride	250 mM	50 mM
Betaine (Sigma)	5 M	1 M
*DTT (Thermoscientific)	50 mM	1 mM
GoTaq polymerase (Promega)	5 U	0.04 U

1 µL DNA at 100 ng/ µL was added and made up to the final volume with sterile water.

*DL-Dithiothreitol.

Thermo cycling conditions:

95°C for 10 minutes, 35x (95°C for 1 minute, 66°C for 1 minute, 72°C for 30 seconds), 72°C for 5 minutes

4. Sequencing (*HTT* promoter and exon 1 PCR)

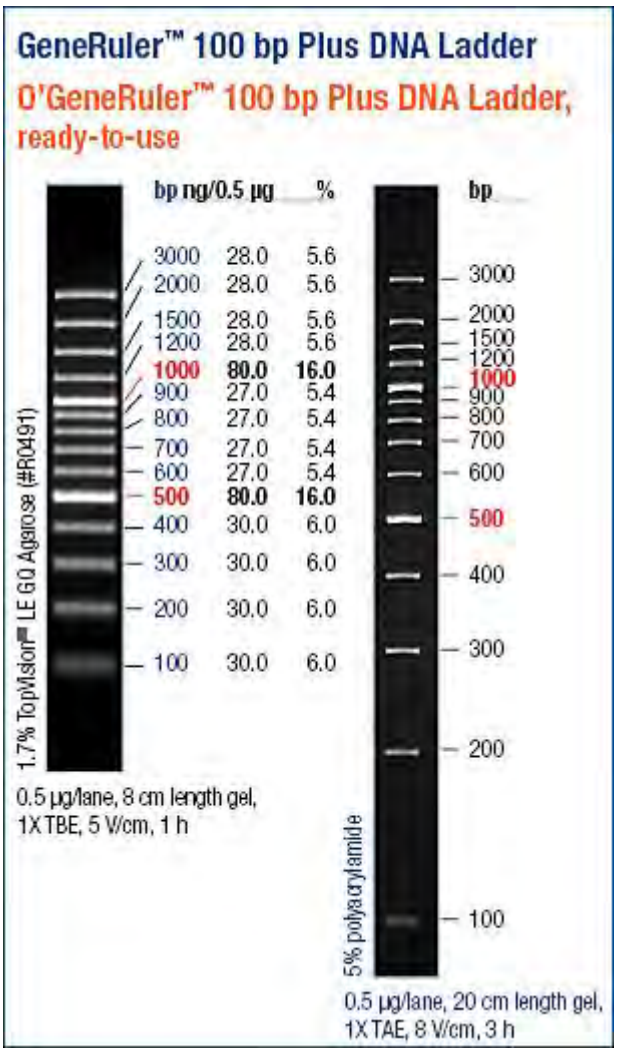
Reagent	Stock concentration	Final concentration (in 50 µL)
HTT PromF primer (IDT)	10 µM	0.4 µM
HTT PromR primer (IDT)	10 µM	0.4 µM
dNTPs	2.5 mM	0.15 mM
GoTaq buffer (Promega)	5X	1X
Magnesium chloride	25 mM	2.5 mM
Betaine (Sigma)	5 M	1 M
GoTaq polymerase (Promega)	5 U	0.04 U

1 µL DNA at 100 ng/ µL was added and made up to the final volume with sterile water.

Thermo cycling conditions:

95°C for 5 minutes, 35x (95°C for 1 minute, 62°C for 1 minute, 72°C for 1 minute and 30 seconds), 72°C for 10 minutes

Appendix 5: Molecular weight marker



www.thermoscientificbio.com

Appendix 6: SNP genotyping

The workflow for high throughput SNP genotyping using the Illumina custom-designed GoldenGate assay is clearly outlined by the figures below (www.illumina.com).

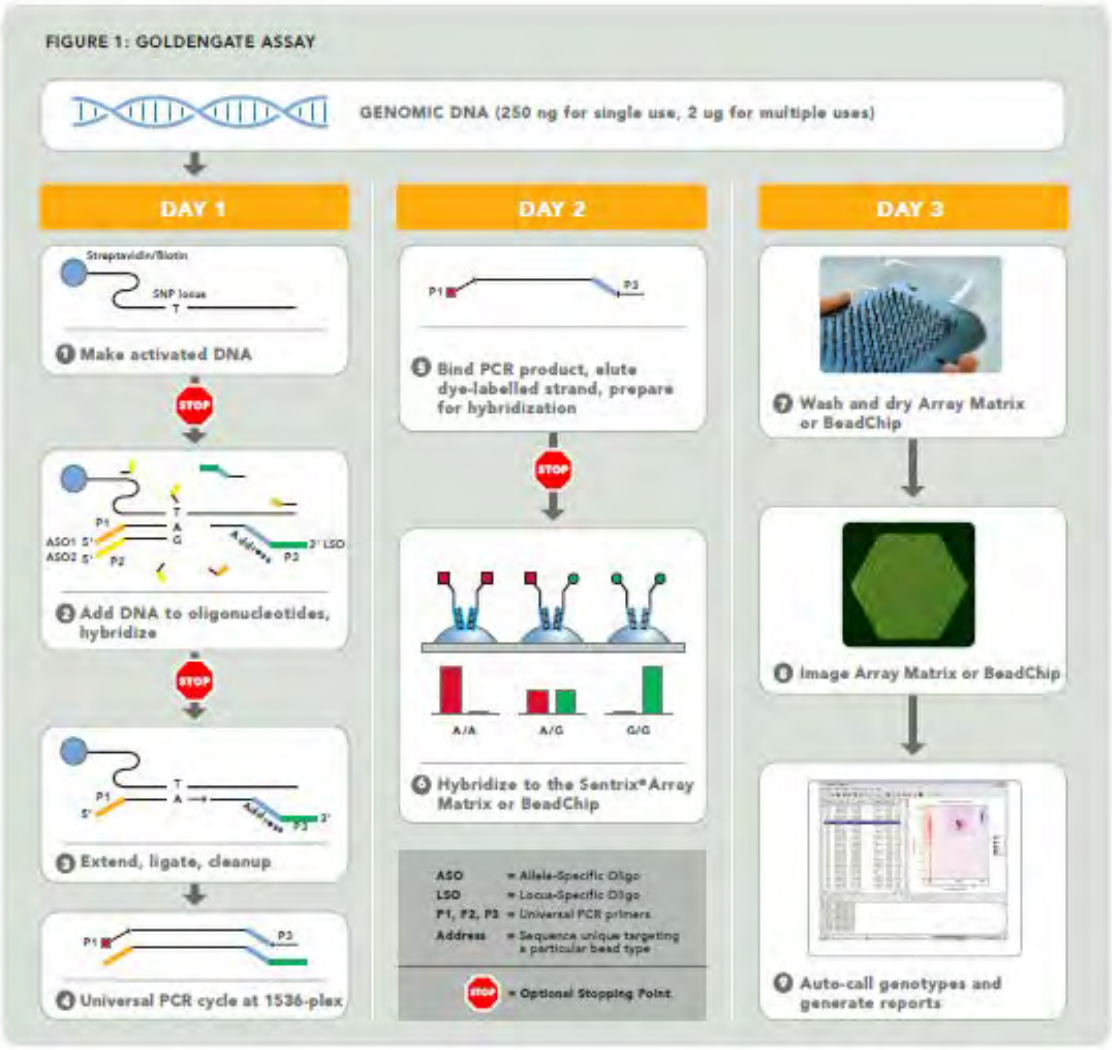


FIGURE 2: PROTOCOL TIME REQUIREMENTS

	PROTOCOL STEP	ONE PLATE Hands-on	ONE PLATE Total time	TWO PLATES Hands-on	TWO PLATES Total time	THREE PLATES[1] Hands-on	THREE PLATES[1] Total time	FOUR PLATES[1] Hands-on	FOUR PLATES[1] Total time
DAY 1	1 Make activated DNA	2:00	2:00	2:30	2:30	3:00	3:00	3:30	3:30
	2 Add DNA and hybridize to oligonucleotides	0:30	3:00	0:30	3:00	0:45	3:15	1:00	3:30
	3 Extend, ligate, and clean-up	1:30	2:00	2:00	2:30	2:45	3:15	3:30	4:00
	4 PCR cycle	–	3:00	–	3:30	–	3:00	–	3:00
	DAY 1 TOTALS	**4:00**	**10:00**	**5:00**	**11:00**	**6:30**	**12:30**	**8:00**	**14:00**
DAY 2	5 Bind PCR product, elute dye-labelled strand, prepare for hybridization	1:30	2:30	2:15	3:15	3:00	4:00	3:45	4:45
	6 Hybridize to Array Matrix[2]	0:30	0:30	0:30	0:30	0:30	0:30	0:45	0:45
	DAY 2 TOTALS	**2:00**	**3:00**	**2:45**	**3:45**	**3:30**	**4:30**	**4:30**	**5:30**
DAY 3	7 Wash and dry Array Matrix	0:10	0:30	0:10	0:30	0:15	0:45	0:15	0:45
	8 Image Array Matrix	–	2:00	–	4:00	–	6:00	–	8:00
	DAY 3 TOTALS	**0:10**	**2:30**	**0:10**	**4:30**	**0:15**	**6:45**	**0:15**	**8:45**
	TOTAL GENOTYPES (1536-PLEX)	**147,456**		**294,912**		**442,368**		**589,824**	

[1] Requires two technicians [2] Times listed are for processing on the Sentrix Array Matrix; analysis can also be carried out on the Sentrix BeadChip format.

Appendix 7: Genotyped SNPs

1. List of 96 SNPs

rs number	Chromosomal position		rs number	Chromosomal position		rs number	Chromosomal position
rs1263309	3051940		rs3025849	3143767		rs362331	3215835
rs13141939	3061282		rs12502045	3147268		rs916171	3216815
rs2798296	3062165		rs35892913	3148570		rs362325	3219326
rs2857936	3062310		rs1065746	3148624		rs3025818	3219499
rs7694687	3063817		rs1143646	3148653		rs2857790	3219613
rs9993542	3068545		rs363102	3149016		rs362322	3221365
rs762855	3074795		rs11731237	3151813		rs362321	3222028
rs13122415	3076181		rs3025843	3156038		rs3025816	3224176
rs13132932	3076266		rs4690073	3160150		rs362275	3224602
rs13102260	3076405		rs363144	3161295		rs3025814	3225190
rs10009935	3077037		rs3025838	3161446		rs362273	3227419
rs3856973	3080173		rs34315806	3162034		rs34600449	3230643
rs2285086	3089259		rs363099	3162056		rs2276881	3231661
rs7659144	3098321		rs3025837	3174845		rs3121419	3232257
rs16843804	3104390		rs363096	3180021		rs2269478	3234828
rs2024115	3104568		rs363129	3182293		rs362272	3234980
rs3733217	3107334		rs2298967	3185747		rs362271	3235518
rs10015979	3109442		rs2298969	3186244		rs3025807	3237923
rs7691627	3111410		rs10488840	3186993		rs3775061	3238754
rs2798235	3114832		rs34781439	3189430		rs362310	3239776
rs1936033	3117050		rs363125	3189547		rs362308	3241735
rs1936032	3117168		rs6844859	3190486		rs362307	3241845
rs4690072	3122507		rs363121	3195763		rs362306	3242100
rs6446723	3126813		rs363092	3196029		rs362303	3242307
rs2301367	3129124		rs363120	3196432		rs2237008	3244675
rs363070	3131573		rs363119	3196611		rs362299	3245718
rs34389685	3133113		rs2798232	3197205		rs362296	3247007
rs363081	3133627		rs3025830	3197648		rs3121417	3252678
rs363080	3133911		rs3025829	3198051		rs3129322	3252852
rs363075	3137674		rs3025827	3201590		rs108850	3258236
rs363064	3141410		rs7685686	3207142		rs1006798	3258373
rs363072	3142528		rs363088	3210330		rs3095074	3261152

Tag SNPs used to define the haplogroups and variants are shaded grey and listed in the table below. The blacked out SNPs were excluded from haplotype analyses due to poor allele-calling. SNP rs numbers and chromosomal positions are based on the Human Genome Assembly: GRCh37, and may have changed in the new assembly: GRCh38.

2. List of Tag SNPs

tSNP	rs number	Chromosomal position	Alleles
1	rs2857936	3062310	T/C
2	rs762855	3074795	A/G
3	rs3856973	3080173	A/G
4	rs2024115	3104568	A/G
5	rs10015979	3109442	A/G
6	rs363081	3133627	A/G
7	rs363075	3137674	A/G
8	rs363064	3141410	T/C
9	rs3025849	3143767	A/G
10	rs363102	3149016	A/G
11	rs4690073	3160150	A/G
12	rs3025838	3161446	T/C
13	rs363099	3162056	T/C
14	rs363096	3180021	T/C
15	rs2298967	3185747	T/C
16	rs362322	3221365	A/G
17	rs2276881	3231661	A/G
18	rs362272	3234980	A/G
19	rs362310	3239776	T/C
20	rs362307	3241845	T/C
21	rs362303	3242307	T/C
22	rs1006798	3258373	A/G

SNP rs numbers and chromosomal positions are based on the Human Genome Assembly: GRCh37, and may have changed in the current build: GRCh38 (www.ncbi.nlm.nih.gov).

Appendix 8: Published article