Development of a SCA7 patient-derived lymphoblast cell model for testing RNAi knock-down of the disease-causing gene

Danielle Claire Berkowitz

BRKDAN005

Supervisors: **Prof. Jacquie Greenberg**

Dr Janine Scholefield

Dr Marco Weinberg

Submitted to the University of Cape Town

In fulfilment of the requirements for the degree of MSc (Med) in Human Genetics

February 2011

Plagiarism Declaration

1. I know that plagiarism is wrong. Plagiarism is using another's work and to pretend that it is one's own.

2. I have used the American Journal of Medical Genetics as the convention for citation and referencing. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed and has been cited and referenced.

3. This dissertation is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

5. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

SIGNATURE: _____

DATE: _____

Acknowledgements

Firstly, thank you to my supervisors, Prof. Jacquie Greenberg, Dr Janine Scholefield and Dr Marco Weinberg. It has been a privilege to work with you over the last two years. I count myself blessed to have been mentored by not one, but three extraordinary scientists and individuals. Thank you for your unique contributions to this work and my development as a scientist.

Thank you to Prof. Raj Ramesar and the rest of the Division of Human Genetics. It has been a pleasure to call the HumGen lab my "home away from home".

To my "A-team" mates, Lauren and Fiona, thank you for your endless willingness to lend a sympathetic ear and give a helpful remark.

To Ms. Ingrid Baumgarten for her help with the transformations and tissue culture.

To the members of the Hatter lab for the use of their Rotor-Gene for qPCR experiments.

To Prof. Arieh Katz for the use of his tissue culture facilities.

To Prof. Matthew Wood for reading my draft.

The Marion Beatrice Waddel Scholarship, UCT Master's Scholarship, and DAAD for financial assistance.

The National Research Foundation and Ataxia UK for funding research costs.

To Mom and Graeme. You are my best friends, and your support has meant the world to me. Thank you.

Contents

List of Figures

List of Tables

Abbreviations

°C -	Degrees centigrade
µl -	Microlitre
µM -	MicroMolar
ADCA -	Autosomal dominant cerebellar ataxias
AGO2 -	Argonaute 2
APS -	Ammonium persulfate
bp -	Base pair
cDNA -	Complementary DNA
Ct -	Threshold cycle
DGCR8 -	DiGeorge syndrome critical region 8
DMEM -	Dulbecco's modified Eagle's medium
DNA-	Deoxyribonucleic acid
dNTP -	Deoxynucleotide triphosphate
dsRNA-	Double-stranded RNA
EBV -	Epstein-Barr virus
FCS -	Foetal Calf Serum
g -	Gram
G -	Gravitational acceleration
HD -	Huntington's disease
Hsp -	Heat shock protein
iPSC -	Induced pluripotent stem cell
JNK -	c-Jun N-terminal Kinase
LCL -	Lymphoblast cell line
LED -	Light-emitting diode
LNA -	Locked nucleic acid
miRNA -	MicroRNA
ml -	Millilitres
mM -	MilliMolar
mRNA -	Messenger RNA
ng -	Nanograms

NHLS - National Health Laboratory Services

NI - Nuclear inclusion

PACT - Activator of protein kinase PKR

pmol - PicoMolar

pre-miRNA - Precursor micro-RNA

pri-miRNA - Primary micro-RNA

RISC - RNA-induced silencing complex

RNA - Ribonucleic acid

rpm - Revolutions per minute

SA - South Africa

SBMA - Spinal bulbar muscular atrophy

SCA - Spinocerebellar ataxia

SCA7 - Spinocerebellar ataxia type 7

SEM - Standard error of the mean

shRNA - Short hairpin RNA

siRNA - Short interfering RNA

SNP - Single nucleotide polymorphism

STAGA - SPT3-TAF9-ADA-GCN5 acetyltransferase

TBE - Tris borate EDTA

TFTC - TATA-binding protein-free TAF containing complex

TRBP - HIV tar-RNA-binding protein

U - Unit

UCT - University of Cape Town

V - Volts

w/v - Weight per volume

XPO5 - Exportin 5

Abstract

Spinocerebellar ataxia type 7 (SCA7) is an inherited neurodegenerative disease caused by the expansion of a CAG repeat within the *ataxin-7* gene. The South African SCA7 population has been shown to have arisen due to a founder effect, and a single nucleotide polymorphism (SNP) within *ataxin-7* has been linked to the SCA7 mutation in all South African patients genotyped to date. Recently, this SNP has been exploited in a potential allele-specific RNA interference (RNAi) based therapy, in order to knock down the expression of the mutant transcript in heterozygous patients. Although this approach has been tested in an artificial cell-based model of SCA7, focus has shifted towards testing the therapy in SCA7 patient-derived transformed lymphoblast cell lines.

The CAG repeat length and SNP genotype of each patient within the South African cohort was determined in order to fully characterise the cohort and establish the percentage of patients who may be eligible for the therapy. Furthermore, a quantitative PCR (qPCR) assay was developed to specifically measure the expression of each allele of the SNP in patient- and control-derived transformed lymphoblast cell lines, in order to test the specificity of RNAi-based therapy. Finally, qPCR was employed to measure the mRNA expression levels of heat shock proteins (Hsps) 27 and 70 in patient- and control-derived cells, to identify phenotypic changes associated with the disease.

The CAG repeat lengths of each patient were calculated, and it was found that 43% of the current patient population was heterozygous for the SNP, and therefore potentially eligible for the therapy. Two allele-specific qPCR assays were designed and tested in patient- and control-derived cell lines. Finally, expression experiments showed a significant increase in *Hsp27* expression, and a significant decrease in *Hsp70* expression in SCA7 patient-derived cells when compared to unaffected controls. It is anticipated that these assays will be used in future studies to evaluate the efficacy of the RNAi-based therapy.

Chapter 1: Introduction

Chapter 1: Introduction

Spinocerebellar ataxia type 7 (SCA7), an inherited neurodegenerative disease, is present at an unusually high frequency within the South African population when compared to the other inherited ataxias (Bryer et al., 2003). Furthermore, there is evidence to suggest that the South African SCA7 population has arisen from a single founder population (Greenberg et al., 2006), providing researchers with a large, genetically homogenous population for study. As a result of this phenomenon, recent work within the Division of Human Genetics has focussed on the development of a novel therapy for these individuals. However, the need has arisen to determine the number of individuals who may be eligible for the therapy, and to develop an assay for testing the therapy *in vitro* in patient-derived cell lines.

1.1 Spinocerebellar ataxias

Spinocerebellar ataxias (SCAs), also known as autosomal dominant cerebellar ataxias (ADCAs), are a group of diseases characterised by cerebellar and spinal cord degeneration. There are currently 30 genes known to be causative of SCA (Matilla-Dueñas et al., 2010). The SCAs have a broad range of neurological symptoms, including ataxia of gait, stance, and limbs, cerebellar dysarthria, oculomotor disturbances of cerebellar and supranuclear genesis, retinopathy, optic atrophy, spasticity, extrapyramidal movement disorders, peripheral neuropathy, sphincter disturbances, cognitive impairment, and epilepsy (Schols et al., 2004) . A table of these symptoms, along with a brief description of each is presented below in Table 1. In the absence of comprehensive epidemiological data, the global prevalence of the SCAs has been estimated to be between 2 and 7 in 100 000 (Van de Warrenburg et al., 2002; Zhao et al., 2002; Duenas et al., 2006). The SCAs are highly heterogeneous, with the age of onset, severity and rate of disease progression differing significantly between individuals (Bryer et al., 2003). Clinical onset of the disease is usually between 30 and 50 years of age, although cases of onset in early childhood and after 60 years have been reported (Duenas et al., 2006). The diseases typically progress to death within 10 to 20 years (Orr and Zoghbi, 2007),

although the rate of disease progression is often influenced by the nature of the pathogenic mutation.

Table 1: Common neurological symptoms associated with the Spinocerebellar ataxias (adapted from Schöls et al., 2004)

Symptom	Description
Ataxia	Derived from the Greek word, "a taxis" meaning "without order" or "incoordination"
Cerebellar dysarthria	Caused by lesions in the cerebellum, resulting in jerky, uncoordinated movements of the speech musculature
Oculomotor disturbances	Dysfunction of the oculomotor nerve, which controls most of the movement of the eye, the constriction of the pupil, and maintaining an open eyelid.
Retinopathy	Non-inflammatory damage to the retina of the eye
Optic atrophy	Deterioration of the optic nerve, which carries nerve impulses from the eye to the brain
Spasticity	Continuous contraction of specific muscles, resulting in difficulty in moving and talking
Extrapyramidal movement disorders	Symptoms including extreme restlessness, involuntary movements and uncontrollable speech
Peripheral neuropathy	Damage to the peripheral nervous system
Sphincter disturbances	Difficulty in controlling bodily sphincters
Cognitive impairment	Impaired brain function
Epilepsy	Recurring seizures

A classification system based on the variable clinical presentations and neuropathologies of the ADCAs was first proposed by Harding in 1993, leading to the differentiation of the ataxias into three distinctive groups (Harding, 1993). ADCA type I contained the broadest range of clinical symptoms, and neuropathology associated with degeneration of the cerebellum and basal ganglia,

cerebral cortex, optic nerve, pontomedullary systems, spinal tracts or peripheral nerves. Due to the broad range of symptoms, the majority of the SCAs are classified as ADCA type I. ADCA type II is distinguishable by the presence of pigmented retinal degeneration, leading to pigmented retinopathy. SCA type 7 is the only SCA to be classified within this group. Finally, ADCA type III is a pure cerebellar syndrome resulting from cerebellar degeneration, leading to the classification of SCAs type 5, 6, 11, 14, 15, 16, 26 as ADCA type III disorders (Duenas et al., 2006).

1.1.1 Polyglutamine repeat disorders

Spinocerebellar ataxia type 7 (SCA7) is classified as a polyglutamine repeat disorder. In total, nine inherited neurodegenerative disorders are known to be caused by the expansion of an unstable trinucleotide CAG repeat within specific genes (Yamada et al., 2008). These include six of the SCAs, (SCA 1, 2, 3, 6, 7 and 17), as well as Huntington's disease (HD), dentatorubral-pallidoluysian atrophy, and spinal bulbar muscular atrophy (SBMA). The CAG repeat size is highly variable, particularly in patients carrying repeats within the pathogenic range. The translation of the CAG repeat leads to an expanded polyglutamine region within the corresponding protein. Unaffected individuals generally carry alleles of between 4 and 18 repeats, whereas mutant alleles in SCA7 patients can range between 38 and 460 repeats (Michalik et al., 2004). Individuals carrying alleles within the intermediate range (19-37) are typically unaffected, but have an increased risk of transmitting expanded alleles within the disease-causing range to offspring (Stevanin et al., 1998). It is generally accepted that the expanded polyglutamine tract changes the conformation of the protein, and leads to an altered function and interaction with other proteins (Orr and Zoghbi, 2007). Consequently, the consensus amongst researchers is that pathogenesis is a result of a toxic gain-of-function acquired by the mutant protein.

1.1.2 SCA7

The earliest reports of Spinocerebellar ataxia type 7 (SCA7) date back to the 1930s and 1950s. These describe patients with spinocerebellar degeneration and varying degrees of retinal degeneration, inherited in an autosomal dominant manner

(Froment et al., 1937; Bjork et al., 1956). Figure 1 shows comparative images demonstrating atrophy of the cerebellum and pons evident in SCA7 patients. The presence of pigmented retinal dystrophy in SCA7 patients has become a useful tool for distinguishing SCA7 from other cerebellar ataxias. In 1996 the gene was mapped to the locus on chromosome 3p12-p13 (David et al., 1996). Expanded polyglutamine repeats had already been identified as the pathogenic mutation underlying SCA1 and 3 (Orr et al., 1993; Kawaguchi et al., 1994). These disorders, along with SCA7, showed strong anticipation in age of onset, particularly through paternal transmission, therefore it was thought that an expanded CAG repeat may have been underlying SCA7. In 1997, David et al. identified the novel *ataxin-7* gene by positional cloning, and established that SCA7 was associated with the expansion of a trinucleotide CAG repeat within the coding region of the gene.

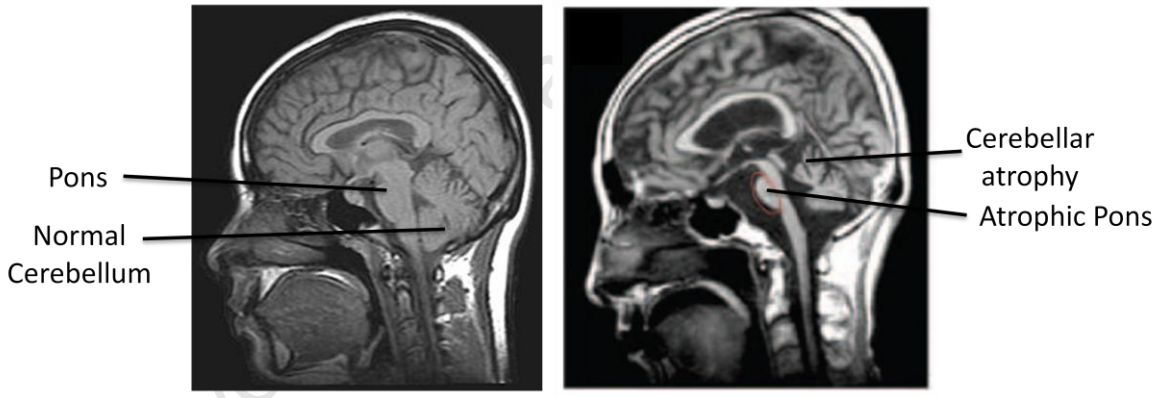


Figure 1: Areas of pathological neuronal loss in SCA7 patients. Image showing MRI scans of two individuals. An unaffected individual is shown on the left, and the right image showing a SCA7 patient. The image on the right demonstrates marked atrophy of the pons and cerebellum. Images obtained and adapted from http://hypothesisnow.files.wordpress.com/2010/03/mri_brain-nasa.jpg and Garden and La Spada (2007).

1.1.3 The CAG repeat region & genetic anticipation

An essential component of the diagnosis procedure for SCA7 is to determine the CAG repeat length of the individual. This information can be used by clinicians for more effective management of the patient, since there is a distinct inverse relationship between CAG repeat length and the age of onset of symptoms (patients with larger repeats develop symptoms at an earlier age) (David et al., 1997). Whilst the age of onset can not be predicted, clinicians may be able to

estimate the age of onset and the rate of progression of the disease in individuals who are pre-symptomatic. Furthermore, patients with SCA7 should be aware of the likelihood that any offspring inheriting the mutant allele will present with an earlier and more severe form of the disease if the repeat region expands further during transmission. Genetic anticipation describes a phenomenon in genetic disorders when the symptoms become apparent at an earlier age as the disorder is passed on to successive generations. Anticipation is common to a number of trinucleotide disorders, most likely due to the dynamic nature of the mutation. It is also common for the severity of symptoms to increase as the disorder is transmitted. A study by David et al. (1996) analysed 23 SCA7 parent-child couples, finding existence of noticeable anticipation, greater in paternal than maternal transmissions. Gonadal mosaicism has been reported in the sperm of SCA7 patients, and increases of up to 85 repeats have been observed in paternal transmissions (David et al., 1998).

Most hypotheses for the proposed mechanism of CAG repeat expansion have been linked to the tendency of the expanded repeat regions to form unusual DNA structures, or to promote DNA slippage during replication and repair (Cleary and Pearson, 2005). A recent study has shown that convergent bidirectional transcription can contribute to CAG repeat instability (Nakamori et al., 2011). However, many current reports propose that the genomic context of the repeat regulates instability. Libby et al. (2003) used transgenic mice to show that various genomic modifications, such as deletion of the 3' genomic region, could significantly stabilize intergenerational instability. In 2008 the same group found that mutations within the CTCF protein binding site, downstream from the CAG expansion in the *ataxin-7* gene, can promote repeat instability in germline and somatic tissues (Libby et al., 2008). These results support the hypothesis that *cis*-acting elements may predispose repeat regions to instability.

The most common method of sizing the SCA7 repeat region is through polymerase chain reaction-mediated amplification of the repeat region and subsequent separation and size estimation through capillary electrophoresis (Sequeiros et al., 2010). However, additional methods such as polyacrylamide gel electrophoresis or

cycle sequencing may be employed. Due to these different methods of size determination, it may be necessary in the future to develop a widely-accepted assay, along with published primer sequences, that is used by all laboratories to size the SCA7 repeat region (Sequeiros et al., 2010).

1.1.4 Ataxin-7 protein

When the gene underlying SCA7 was cloned, the normal function of the ataxin-7 protein was unknown. Northern blot analysis revealed a 7.5 kilobase transcript expressed in the heart, placenta, skeletal muscle, pancreas, brain, lung, liver, kidney and the central nervous system (David et al., 1997). The 2727 base pair (bp) open reading frame encodes a 892 amino acid protein with several identifiable regions, including nuclear localization signals and a nuclear export signal, suggesting a potential shuttling function.

Subsequent functional studies have indicated that ataxin-7 is a component of a mammalian transcription co-activator complex called STAGA (SPT3-TAF9-ADA-GCN5 acetyltransferase) (Palhan et al., 2005), which functions to mediate interactions between transcriptional activators and the RNA polymerase II transcription complex. STAGA and the closely related TATA-binding protein-free TAF containing complex (TFTC) promote histone acetylation, thereby mediating transcriptional activation (Garden and La Spada, 2007). Although the exact function of ataxin-7 within the STAGA/TFTC complex is still unknown, it has been shown that ataxin-7 mediates an interaction between STAGA and the cone-rod homeobox transactivator of photoreceptor genes, linking the function of ataxin-7 with the retinal phenotype observed in SCA7 patients (Palhan et al., 2005). Furthermore, polyglutamine-expanded ataxin-7 inhibits the histone acetyltransferase activity of STAGA, which is predicted to lead to retinal degeneration. A recent study using a yeast two-hybrid screen has identified further interacting partners of the ataxin-7 protein, highlighting the possibility that one or more unidentified partners may play a role in SCA7 pathogenesis (Kahle et al., 2011).

1.1.5 Pathogenesis

Numerous hypotheses regarding the pathogenesis of polyglutamine disorders have been suggested in the literature. These include (i) misfolding of the disease protein resulting in altered function; (ii) deleterious protein interactions engaged in by the mutant protein; (iii) formation of toxic oligomeric complexes; (iv) transcriptional dysregulation by the mutant protein; (v) mitochondrial dysfunction resulting in impaired bioenergetics and oxidative stress; (vi) impaired axonal transport; (vi) aberrant neuronal signalling; (vii) cellular protein homeostasis impairment; and (viii) RNA toxicity (summarized by Williams and Paulson, 2008). However, it is most likely that combinations of one or more of these mechanisms result in the eventual phenotypes evident in the polyglutamine disorders. There may be "non-specific" toxic effects of the mutant protein, or toxicity may be specific to the particular mutated protein, related to an alteration of its wild-type function.

The SCA7 phenotypic feature that has been most commonly linked with pathogenesis and subsequent neurodegeneration is the presence of protein aggregates, or neuronal inclusions (NIs). The NIs typically contain a range of proteins, including the mutant and wildtype polyglutamine protein, as well as components of the ubiquitin-proteosome system. The aggregation of the expanded polyglutamine proteins, and their inclusion into NIs is a phenomenon common to all of the known polyglutamine disorders, but the exact role of the NIs in disease progression is still unknown. Whilst initially it was suggested that the inclusions mediate neurodegeneration, recent groups have proposed that the inclusions are merely a consequence of the abnormal protein aggregation, or that NIs may even play a protective role by sequestering the mutant proteins (Orr and Zoghbi, 2007; Arrasate et al., 2004). For SCA7 in particular, it has been found that cerebellar Purkinje cells, a cell type severely affected by the disease, rarely display NIs. In contrast, areas spared of degeneration (such as the cerebral cortex) have been found to exhibit NIs. These findings suggest that the presence of NIs may not be causative of degeneration (Michalik et al., 2004). However, in the absence of detailed studies into the exact mechanisms of aggregation, inclusion formation and toxicity, no conclusions can be made on the exact relationship between NI

formation and pathogenesis. For example, it is possible that NIs form earlier on in affected neuronal subtypes, or current imaging methods may not detect smaller NIs which may contribute significantly to pathogenesis. Furthermore, there is some evidence in other polyglutamine diseases such as Huntington's disease that implicates toxic intermediates formed during the aggregation process, rather than the inclusions themselves (Ross and Poirier, 2004).

A characteristic of NI formation that is likely to contribute to degeneration is the recruitment of other cellular proteins into the nuclear inclusions. The absence or reduced occurrence of these proteins in the cellular environment may interfere with their standard functioning, and later contribute to toxicity. Whilst some recruited proteins are common to multiple polyglutamine diseases (such as particular molecular chaperones and ubiquitin), it is possible that future studies may reveal recruited proteins specific to the neuronal subtypes affected by the particular disease.

It is still unknown how expanded ataxin-7 causes degeneration only in specific neuronal populations, in spite of its expression throughout the central nervous system and other non-neuronal organs. Ataxin-7 has been shown to be expressed in numerous regions of the brain, and is not restricted to neuronal types affected by the disease (Lindenberg et al., 2000). Another noteworthy feature is that the localization of the ataxin-7 protein has been found to vary between nuclear and cytoplasmic loci. The protein localization is not specific to affected neuronal types. However, it has been suggested that the mutant protein may be more detrimental in cell types with high metabolic demands and protein turnover (Garden and La Spada, 2007). It is possible that future comparative expression studies may reveal new ataxin-7 protein targets and interactants (such as STAGA-dependent transcription factors) that may confer susceptibility or protection to specific cell types.

1.1.6 Treatment and therapy

Since there is currently no known treatment for SCA7, clinical management is usually focussed on the alleviation of specific symptoms and management of the cognitive and psychiatric issues that may be associated with the condition. However, there are numerous studies exploring therapeutic strategies underway. One of the most challenging aspects of treating the polyglutamine disorders in general, is the nature of the mutation. Since the mutation in polyglutamine repeat disorders is dynamic, and the repeat sizes of the expanded tract varies from patient to patient, it may be problematic to directly target the repeat region in order to specifically silence the mutant transcript. Since the specific function of the wild-type protein is not always known (as in the case of SCA7), many strategies have opted to target the mutant transcript specifically, rather than completely abolishing expression of the transcript altogether. However, in some cases, silencing of both the wild-type and mutant transcripts may be tolerated, as has been found in a rat model of SCA3 (Alves et al., 2010). Because the mutation is shared across the polyglutamine disorders, it is likely that a therapeutic strategy for a particular disorder may be applicable to many, if not all, the polyglutamine disorders.

The next problem that arises is how to specifically target the mutant transcript. In a strategy adopted by Hu et al. (2009), antisense oligomers have been employed to target expanded CAG repeats preferentially. It was found that this method can be used to specifically silence mutant *htt* (responsible for HD) and ataxin-3 protein expression in cultured cells. Whilst this method may be beneficial to a number of the polyglutamine disorders, it is unlikely that it could be used for patients with shorter expanded repeats that are in the pathogenic range.

Mouse studies can provide important clues towards the function of the wild-type protein, and can therefore direct therapeutic strategies. For example, *htt* knockout mice are embryonically lethal (Nasir et al., 1995), and therefore studies have focussed on targeting the mutant transcript specifically. In the absence of a well-

characterised SCA7 animal model, investigations into SCA7 therapies have adopted a similar approach.

Another popular therapeutic strategy employs the use of allele-specific silencing by targeting a polymorphism specific to the mutant transcript. Whilst this approach may be useful in smaller populations, it is unlikely that a single polymorphism will be present at a high enough frequency in a larger population to be beneficial to a large proportion of individuals. In addition, the individuals need to be heterozygous for the polymorphism in order for the therapy to be applicable. Therefore, it may be necessary to target multiple polymorphisms. Pfister et al. (2009) have developed a panel of five small interfering RNAs, targeting 3 single nucleotide polymorphisms (SNPs), that they propose can be used to treat three-quarters of the United States and European HD patients.

1.2 RNAi

RNA interference (RNAi) describes a conserved regulatory mechanism employed by most eukaryotic cells to control gene activity (Kim and Rossi, 2008). This phenomenon was first discovered in *Caenorhabditis elegans*, when it was noted that the introduction of double-stranded RNAs (dsRNAs) prompted the silencing of the complementary messenger RNA (mRNA) sequence (Fire et al., 1998). Three years later it was shown that the same mechanism could be triggered in mammalian cells (Elbashir et al., 2001). Another three years later, Phase I Clinical Trials on the first RNAi-based human therapeutics were launched. According the the US National Institute of Health, there are currently at least five clinical trials based on potential RNAi therapies (www.clinicaltrials.gov). Disorders targeted in these trials include cancer and AIDS.

The mechanism of cellular silencing using RNAi can be separated into two main pathways, illustrated in Figure 2. The first, transcriptional gene silencing, is mediated through mechanisms that are not well understood, but are likely to include chromatin modification (Grishok et al., 2005). The second, post-transcriptional gene silencing can be divided into two pathways, whereby the target mRNA is either cleaved in a sequence-specific manner or translation is repressed and the mRNA is degraded.

Long primary transcripts called primary-microRNAs (pri-miRNAs) act as the substrates for the endogenous RNAi pathway. The expressed transcripts are processed into 60-70bp short hairpins, called precursor-microRNAs (pre-miRNAs), by the Drosha-DGCR8 microprocessor complex (Lee et al., 2003). The pre-miRNAs are then transported from the nucleus to the cytoplasm by exportin 5 (XPO5) (Yi et al., 2003). In the nucleus, the pre-miRNAs are cleaved by an RNaseIII endonuclease called Dicer into shorter fragments (~21-25 nucleotides long), termed microRNAs (miRNAs) (Zhang et al., 2004). These miRNAs are in turn recruited by a protein complex known as the RNA-induced silencing complex, or RISC. Once loaded into RISC, the sense or “passenger” strand of the double-stranded miRNA is cleaved by the protein Argonaute 2 (AGO2) and released (Liu et al., 2004), whilst the anti-sense

or “guide” strand binds to its complementary mRNA transcript. Perfectly complementary mRNAs are cleaved by the catalytic domain of AGO2, whilst the translation of imperfectly complementary mRNAs is repressed. These mRNAs are then degraded in a non-sequence-specific manner by P bodies (Liu et al., 2005). Both mechanisms result in effective post-transcriptional gene silencing.

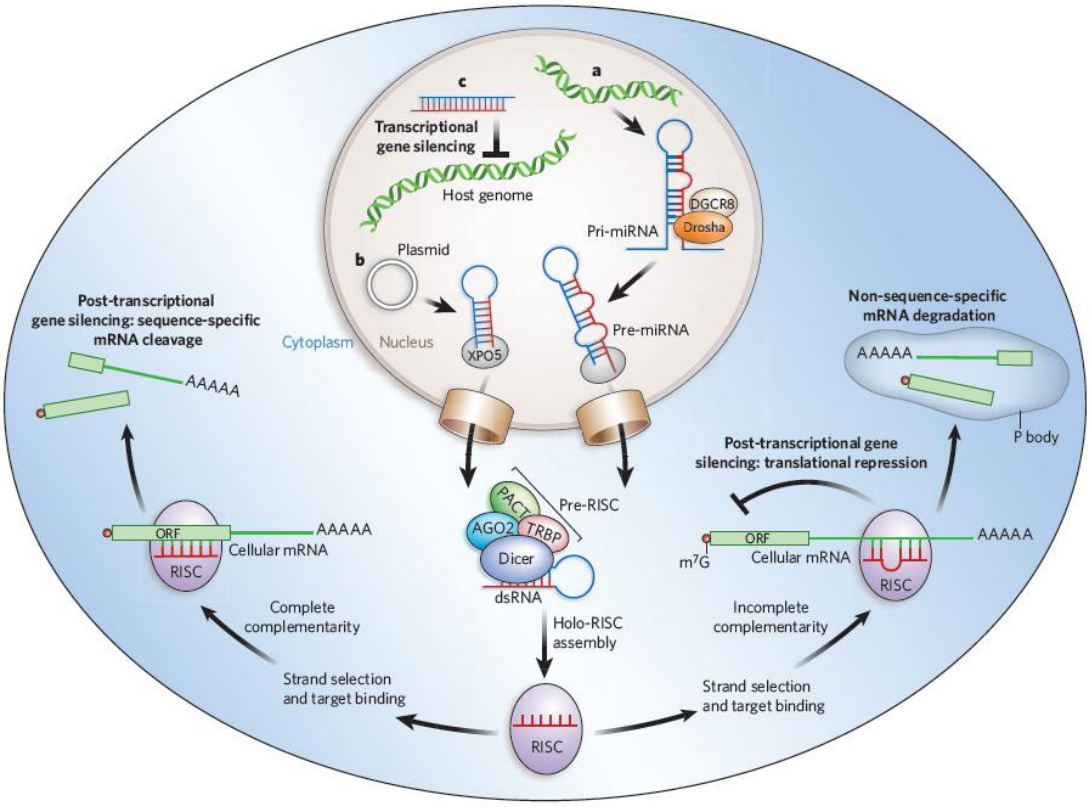


Figure 2: The mammalian RNAi pathway. (a) Primary microRNAs (Pri-miRNAs) are processed by Drosha and DiGeorge syndrome critical region 8 (DGCR8) into precursor miRNAs (Pre-miRNAs) and transported into the cytoplasm by exportin 5 (XPO5). In the cytoplasm the pre-miRNA is bound by pre-RNA induced silencing complex (pre-RISC) containing Dicer, the catalytic core Argonaute 2 (AGO2), and accessory proteins HIV tar-RNA-binding protein (TRBP) and activator of protein kinase PKR (PACT). The guide sequence binds to complementary sequences in cellular mRNAs. Perfect complementation leads to sequence-specific cleavage and mRNA degradation (left pathway), whilst incomplete complementation leads to translational inhibition accompanied by non sequence-specific degradation in P bodies. (b) Artificially transcribed short hairpin RNAs (shRNAs) follow a similar pathway. (c) siRNAs complementary to promoter regions can trigger transcriptional gene silencing via histone modifications and chromatin remodelling. Adapted from Castanotto and Rossi, 2009.

Pri-miRNAs do not act as the sole substrates for the RNAi pathway. Long dsRNA structures, from a variety of sources (including viral RNAs, transgenes or transposons), are also processed by Dicer to yield shorter fragments, called short interfering RNAs (siRNAs). These fragments progress through the pathway in the same fashion as miRNAs, resulting in post-transcriptional gene silencing.

1.2.1 RNAi therapeutics

In recent years, exploitation of the RNAi pathway has become a vital tool in molecular biology. Because the sequence of the dsRNA substrate specifies which mRNA is translationally repressed, it is possible to introduce synthetic substrates in order to potently and specifically silence a gene of choice. This may be for the purpose of silencing a disease-associated gene, or to elucidate the function of a gene by preventing its expression. Diseases and disorders including viral infections, dominant genetic disorders and cancers may benefit from RNAi-based therapies in the future.

Synthetic siRNAs can enter the RNAi pathway at a number of levels. The first method of introduction is by direct transfection of the siRNA molecule into the target cell, whereafter the siRNA is taken up by RISC (Martinez et al., 2002). In this manner siRNA introduction is able to produce transient transcriptional repression, since the intracellular concentrations of the siRNA molecules are diluted over time. The second method is by introduction of viral vectors containing Pol II or III promoters, which transcribe pri-miRNA mimics or short hairpin RNAs (shRNAs) (Amarzguioui et al., 2005). These dsRNA molecules enter the pathway at a higher level than siRNAs, and are processed by Drosha and/or Dicer. The obvious benefit is that the shRNAs can be expressed stably for a long period of time, and can be expressed in a tissue-specific manner if it is desired (Kim and Rossi, 2007).

1.2.2 Challenges of RNAi

Although the concept of using siRNAs to support genetic therapies is simple, some cautions and considerations must be taken into account. Because the synthetic siRNA or shRNA molecules introduced into mammalian cells are processed in the

same way as endogenous RNAi molecules, one of the main concerns when developing an RNAi-based therapy is that the molecules introduced may saturate the endogenous pathway. Since the role of miRNAs is to "fine-tune" gene expression, introduction of foreign molecules into the pathway may perturb the natural system, resulting in undesirable effects. A study where the gene encoding Dicer was disrupted resulted in embryonic lethality in mice, indicating that accurate performance of the RNAi pathway is essential for proper cell function (Bernstein et al., 2003). Therefore, during the siRNA/shRNA design process, it is essential that potent effectors are chosen that are efficient at low doses, and that moderate, controllable vector systems are used for the expression of shRNAs (Aagaard and Rossi, 2007).

Although eukaryotes use the RNAi system to modulate gene expression, the pathway is also used as a "molecular immune system", to protect the host from dsRNA viral infections, transposons, retrotransposons, and retroviruses (Bagasra and Prilliman, 2004). Therefore, the introduction of long dsRNA molecules results in an innate immune response. This response leads to inhibition of protein synthesis, expression of anti-viral genes, and ultimately cell death (Gantier and Williams, 2007). Therefore thorough testing using in vitro human primary cells and animal models should be completed before proceeding to clinical trials.

Microarray studies have shown that cells treated with siRNAs show off-target silencing of a large number of non-targeted genes (Jackson et al., 2003). Since a third of randomly-selected siRNAs have the ability to affect cell viability (suggesting that they can induce toxicity) (Fedorov et al., 2006), it is essential that siRNAs are thoroughly screened before being introduced into mammalian cells. However, it may be possible to eliminate the toxic effect produced by siRNAs by the introduction of chemical modifications (Fedorov et al., 2006).

One of the main challenges facing the design of effective RNAi-inducing molecules is that of how the effectors enter the cell. As discussed above, siRNA/shRNA effectors can enter either via viral or non-viral delivery methods. Non-viral delivery

usually requires the aid of packaging systems such as liposomes or nanoparticles, to allow the negatively-charged siRNA molecules to cross hydrophobic cellular membranes. Chemical modifications are also required to prevent degradation by RNases (Kim and Rossi, 2007). The siRNAs can either be delivered intravenously (so-called non-selective systemic delivery), or be directed towards their target cells via cell-specific surface receptors (selective systemic delivery). In most cases it is preferential to have cell-specific delivery, to reduce dosage amount, and to prevent potential off-target effects in non-targeted cells.

In order to treat chronic diseases it is necessary to stably express the shRNA effector targeting the gene of choice. For these purposes, viral vectors are used as a mode of delivery. Lentiviral vectors are able to stably tranduce both dividing and non-dividing cells. The virus integrates into the host genome, allowing for stable expression of the shRNA effector, however, this phenomenon may lead to endogenous gene disruptions. Vectors can also be derived from adenoviruses. These vectors do not integrate into the genome, and are therefore diluted over the course of successive cell divisions. Adenoviral vectors are used where long-term RNAi delivery is not required, since repeated doses may initiate strong immune responses.

1.3 Heat shock proteins

The heat shock proteins (Hsps) are a set of proteins present in both eukaryotes and prokaryotes. These proteins accumulate in the cell in response to external or intrinsic stimuli that induce stress within the cell, such as heat shock or oxidative stress. A key role of the Hsps (also referred to as molecular chaperones) is to prevent or correct the misfolding of cellular proteins, thus preventing aggregation (Jakob et al., 1993). The Hsps are classified into one of six families, based on their size (namely Hsp100, Hsp90, Hsp70, Hsp60, Hsp40, and small heat shock proteins) (Jolly and Morimoto, 2000). In each of the Hsp families, some members are constitutively expressed, whereas the expression of others is induced in response to stresses. One of the hallmarks of many inherited neurodegenerative disorders, in particular the polyglutamine disorders, is the aggregation of the disease-causing

proteins within inclusions in the cell. Furthermore, the inclusions often include various proteins from the molecular chaperone and ubiquitin-proteasome pathways. Therefore it is intuitive that dysregulation of the endogenous heat shock response within affected cells may contribute to pathogenesis.

1.3.1 Heat shock proteins & SCA7

Changes in Hsp expression have been found in a number of polyglutamine disorders. Katsuno et al. (2005) found that administration of Geranylgeranylacetone, a nontoxic antiulcer drug, increased expression of Hsp70, 90 and 105 in a cell model of SBMA. As a result, cell death was inhibited, and accumulation of pathogenic androgen receptor protein (containing the expanded polyglutamine tract) was prevented (Katsuno et al., 2005). In an HD mouse model, Hay et al. (2004) found that expression levels of Hsp70 were decreased to below 40% of the endogenous levels. Furthermore, aberrations in Hsp expression levels have been implicated in the pathogenesis of SCA7. In a study by Merienne et al. (2003), expression levels of Hsp70 were reduced in the presence of polyglutamine expansions. They went on to show that as a result, c-Jun N-terminal Kinase (JNK) phosphatases were inhibited, activating a pathway which may ultimately lead to cell death through JNK and c-Jun activation (Merienne et al., 2003).

A study by Tsai et al. (2005) explored the potential relationship between SCA7 pathogenesis and Hsps. Hsps have been shown to suppress cellular toxicity and aggregate formation associated with polyglutamine expansion diseases (Cummings et al., 1998; Warrick et al., 1999). Lymphoblasts isolated from SCA7 patients express the mutant ataxin-7 protein, providing researchers with a novel system to study the cellular effects of SCA7. Tsai et al. used these cells to investigate any potential changes in Hsp expression in SCA7 patients. Using lymphoblast cell lines from two SCA7 patients, the group investigated protein expression levels of Hsps 27, 60, 70 and 90. This group showed decreased endogenous expression of Hsp27 and Hsp70 protein in SCA7 patient cell lines, whilst the expression of Hsps 60 and 90 remained unaffected. Semi-quantitative RT-PCR assays suggested that the decreased level of Hsp70 expression was due to transcriptional repression. Levels

of *Hsp27* mRNA were similar between cells with and without mutant *ataxin-7*, suggesting that the decreased expression of Hsp27 protein is not due to transcriptional dysfunction. The authors hypothesized that the reduction of these heat shock proteins 27 and 70 in SCA7 cells may contribute to the accumulation of misfolded proteins and cellular stress associated with the condition (Tsai et al., 2005).

Additional reasons for reduced Hsp expression in polyglutamine disorders were proposed in a review by Huen et al. (2007). Firstly, the Hsps may be recruited to the nuclear inclusions. This has been demonstrated previously by groups such as Zander et al. (2001). Alternatively, the ubiquitin-proteasome system may degrade the Hsps, but these hypotheses can not account for changes in mRNA expression levels. Finally, the polyglutamine proteins may interfere with the transcription of the Hsps.

1.4 Previous work

1.4.1 SCA7 in South Africa

Until 2003, very little data was available on the frequency of the SCAs in SA. A ten-year population-based study by Bryer et al. (2003) undertook to investigate the spectrum and prevalence of the SCAs in SA. Of the total of 54 families investigated during this time, it was found that 12 of these (22.2%, all of Black African origin) exhibited the SCA7 genotype. Most countries report a frequency of SCA7 of between 4.5 and 11.6%, apart from a Scandinavian population in Finland and Sweden with a higher reported frequency, suggested to be a result of a founder premutation (Jonasson et al., 2000). Therefore the frequency of SCA7 reported in SA is believed to be one of the highest in the world. In order to investigate potential reasons for the skewed incidence in SA, a haplotype study was undertaken in order assess the evidence for a founder effect in South African SCA7 families. A founder effect may be used to explain a loss of genetic variation (or an increased prevalence of a mutation) in a population if the population was established by a small group of "founder" individuals. Furthermore, since the individuals share a common ancestor, it is more likely that they will share the same

alleles at specific genetic loci. Alleles at multiple loci which are transmitted together on the same chromosome are referred to as a haplotype. In the South African SCA7 population, a SNP and four microsatellite markers in a 5.21 centimorgan region flanking the *SCA7* gene were genotyped in five extended families. It was found that affected individuals shared alleles at three loci flanking the mutation, and the allele frequencies were higher in the SCA7 patient population than the unaffected population (Greenberg et al., 2006). These results are suggestive of a founder effect in the South African SCA7 Black African population. For the purposes of this study, it is important to note that the genotyped SNP was found to be linked to the SCA7 mutation, and more than 50% of the SCA7 patients had an "A" allele associated with the mutation, and a "G" allele associated with the wild-type *ataxin-7* transcript.

Previous work within the Division of Human Genetics at the University of Cape Town (UCT) has aimed at exploiting the endogenous RNAi pathway to develop a potential therapy for South African SCA7 patients. One of the central issues facing developers of potential therapies for the polyglutamine disorders is that of how to specifically target the mutant transcript. Although the function of the ataxin-7 protein has not been well characterised, it is likely that the wild-type allele is necessary for normal cellular function. It has been shown that antisense oligonucleotides are able to distinguish between mutant and wild-type alleles based on repeat length (Hu et al., 2009), however, other genes containing CAG repeat regions may be targeted in a non-specific manner. Therefore, rather than targeting both *ataxin-7* transcripts, or the expanded CAG repeat in the mutant allele, our group has exploited a SNP linked to the SCA7 mutation in our South African cohort to develop a potential RNAi-based therapy. Using a similar SNP-targeting approach, Alves et al. (2008) have demonstrated allele-specific silencing of the mutant *ataxin-3* transcript in a rat model of SCA3. Using this rat model, they were able to show that an shRNA effector designed to target a SNP within the mutant *ataxin-3* CAG tract was able to decrease mutant ataxin-3 protein production, and neuropathological symptoms associated with SCA3. However, the same group has recently found that non-allele-specific silencing of *ataxin-3* strongly reduces

pathology in a SCA3 rat model (Alves et al., 2010), suggesting that total silencing of the *ataxin-3* gene may be tolerated *in vivo*.

For a previous study on the South African SCA7 cohort, multiple shRNAs (incorporating the SNP of interest (rs377 4729) at different positions) were screened to test for effective silencing of the mutant *ataxin-7* transcript. In order to achieve this, a SCA7 cell-based disease model was created by transfecting expression plasmids containing full-length constructs of wild-type or mutant *ataxin-7* cDNA, fused to eGFP or dsRED reporter targets, into HEK293 cells. Quantification of fluorescence allowed for measuring of expression levels of the two transcripts after shRNA transfection. Using this method, an shRNA effector capable of silencing the mutant transcript by more than 90% was identified (Scholefield PhD thesis, 2008; Scholefield et al., 2009). It was also shown that transfection of the shRNA molecule resulted in decreased numbers of cells with intracellular aggregates, which are commonly associated with disease pathogenesis.

Figure 3 illustrates the method developed by Scholefield et al. (2009) for effective silencing of the *ataxin-7* mutant transcript. In the South African cohort, patients with an expanded CAG repeat have an A allele at the SNP locus rs 377 4729. Once the mutant transcript is transcribed, a mutant polyglutamine protein is produced, which leads to downstream pathogenic events, and the resultant SCA7 phenotype. The RNAi effector has been designed to target the A allele, thereby preventing the translation of the mutant transcript, resulting in no production of the mutant protein. If the patient is heterozygous at the SNP position, the wild-type transcript is not targeted by the effector, and the wild-type protein is produced.

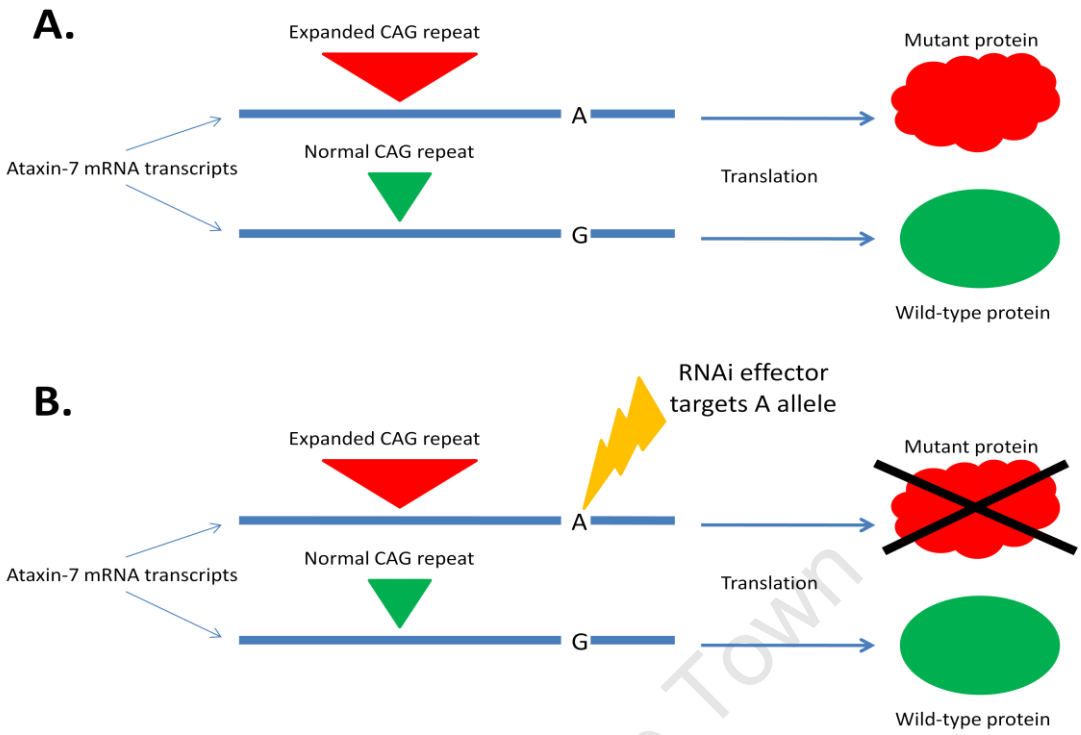


Figure 3: The method developed by Scholefield et al. (2009) for effective silencing of the *ataxin-7* mutant transcript. (A) In SCA7 patients the translation of the mutant *ataxin-7* transcript (containing an expanded CAG repeat) leads to the production of mutant protein, whilst the wild-type transcript is translated to produce normal ataxin-7 protein. In the South African cohort, the expanded CAG repeat is linked to the A allele of a downstream SNP (rs 377 4729). (B) In heterozygous patients, the designed RNAi effector selectively targes the A allele of the SNP (linked to the expanded repeat), enabling the production of wild-type ataxin-7 protein, but reducing the production of mutant protein.

Whilst the RNAi effectors have been shown to be effective in an artificial cell-based model, the next step in the development of the therapy would be to test the therapy in patient-derived cells. These cells express the mutant and wild-type *ataxin-7* transcripts endogenously, rather than from a vector. However, since the endogenous proteins are not fluorescently tagged (and therefore not detectable using microscopy), an alternative method must be developed to measure the expression of these transcripts and/or proteins in patient-derived cells.

1.5 Problem identification

The efficacy of the RNAi-based therapy described above was tested by measuring the levels of fluorescently-tagged mutant and wild-type *ataxin-7* transcripts. This was achieved by transfecting HEK293 cells with expression plasmids containing the target constructs with expression under the control of the strong viral

CytoMegalovirus promoter. Although these findings strongly suggested that the identified shRNA effector could successfully distinguish between alleles, and reduce expression of the mutant transcript, it has become necessary to measure the effect of the shRNA molecule on expression levels of endogenous wild-type and mutant *ataxin-7*. Since the SNP rs 377 4729 has been shown to be linked to the pathogenic SCA7 mutation within the South African patient cohort, the SNP can be used in allele-specific assays to measure the expression of each transcript in heterozygous samples. Allele-specific expression can be measured via conventional PCR and real-time quantitative PCR methods. The PCR primers can be designed to incorporate the SNP at the 3' end, to allow for efficient extension of matched primer-template duplexes, and little or no extension of mismatched duplexes (Kwok et al., 1994). Alternatively, Taqman probes could be designed to hybridize specifically to target sequences containing each allele of the SNP (Livak, 1999). Each probe is labelled with a different fluorogenic tag, so each allele can be measured quantitatively within a single assay.

Furthermore, it is necessary to identify and measure additional phenotypic markers (besides alleviation of aggregate formation) associated with the SCA7 phenotype in patient-derived cells, so that these markers may be measured post knock-down of the mutant transcript to give an indication of reversal of the SCA7 phenotype.

1.6 Goal of the current study

The goal of this study was to characterise the current South African SCA7 cohort, in terms of determining patient CAG repeat lengths and SNP (rs 377 4729) genotypes, in order to establish who may be eligible for the RNAi-based therapy. Transformed lymphoblast cell lines from a subset of these patients were then used to develop an assay for testing the efficacy of the proposed RNAi-based therapy. Furthermore, we aimed to develop a method for measuring *Hsp27* and 70 mRNA expression levels in patient lymphoblasts, in order to validate previous studies showing reduced expression levels of these proteins in SCA7 patient cells. This assay may give a further phenotypic measure of the efficacy of the therapy, if Hsp expression levels are restored after administration. The implications of this study may ultimately include two useful assays for pre-testing the efficacy of the therapy *in vitro* in cells isolated from patient whole blood samples, before offering the treatment *in vivo*.

1.6.1 Objectives

A. To identify all SCA7 patients in the Division of Human Genetics database (up until the end of 2010) and determine their CAG repeat size and SNP genotype.
B. To develop a quantitative assay to measure the levels of mutant and wild-type *ataxin-7* in SCA7 patient-derived transformed lymphoblasts.
C. To develop a quantitative assay to measure mRNA levels of heat shock proteins 27 and 70 in SCA7 patient-derived transformed lymphoblasts.

Chapter 2: Materials and Methods

Chapter 2: Materials & Methods

2.1 Preparation of samples

2.1.1 Database query and cohort selection

The individuals included in this study were selected from the neurodegenerative disorder biological material repository and database in the Division of Human Genetics, UCT (REC/REF 229/2010). The database contains information on South African families affected with or predisposed to the Spinocerebellar ataxias (SCAs). This information includes the results of clinical diagnostic tests and molecular genetic tests of potential SCA patients from throughout SA referred for genetic testing to the National Health Laboratory Services (NHLS) diagnostic laboratory at Groote Schuur Hospital. Patient information and samples were collected according to the World Medical Association Declaration of Helsinki (2008). The declaration outlines ethical principles for medical research involving human subjects. Ethical approval for the study was provided by the Research Ethics Committee of the University of Cape Town (REC REF 228/2009).

2.1.2 DNA isolation & integrity determination

The Puregene® Genomic DNA Purification Kit (Gentra Systems, USA) was used to isolate genomic DNA from peripheral blood lymphocytes in 300 microlitres (µl) of whole blood. The DNA isolation was routinely undertaken by a technologist within the Division of Human Genetics. Firstly, cells were lysed with an anionic detergent in the presence of a DNA stabilizer. The detergent facilitates lysis by disrupting cell membrane protein conformation, and the stabilizer prevents DNA degradation by limiting the activity of DNases. Contaminating proteins were removed by salt precipitation, and resulting DNA was recovered by ethanol precipitation and subsequent re-suspension in a hydration solution. The full protocol is listed in Appendix A. One sample was placed in short term storage at -20 degrees Centigrade (°C), while the other sample was placed in long term storage at -80°C. The details of the location of each sample are recorded in paper-based and electronic filing system used within the Division of Human Genetics.

Each of the DNA samples isolated from patient peripheral blood lymphocytes were retrieved from the appropriate storage locations. The quality of the DNA samples were determined by analysing the concentration of the stock samples using the NanoDrop ND-1000 Spectrophotometer and NanoDrop 1000 software (NanoDrop Technologies, USA).

The DNA stock solutions were diluted in sterile distilled water to working concentrations of 100ng/µl to be used in further investigations. The working stock solutions were stored at -20°C for the duration of the research project. The integrity of the diluted DNA solutions was determined by loading 200ng of the diluted DNA with 5µl loading dye (Appendix A) into the wells of a 1% weight per volume (w/v) agarose gel (Hispanger, Spain) (Appendix A). The DNA was electrophoresed for 30 minutes at 160 volts (V). SYBR gold nucleic acid staining was used (described in section 2.3.3.1), and DNA fragments were visualised under ultra-violet (UV) light with the UVIPro UVIGold transilluminator (UVItec Limited, UK).

2.2 Amplification of DNA by PCR

2.2.1 PCR

The polymerase chain reaction (PCR) is a method used for *in vitro* DNA amplification (Mullis and Faloona, 1987). Single copy genomic sequences can be amplified by a factor of up to ten million with high specificity. The process involves the addition of two oligonucleotide primers that flank the specific DNA sequence to be amplified. These hybridize to the template DNA strand to enable DNA synthesis by a heat stable DNA polymerase (called *Taq* polymerase, isolated from the thermo stable bacterium *Thermus aquaticus*). Amplification of the DNA fragment is achieved through repeated cycles of heat denaturation of the double strand, annealing of the primers to their complementary sequences on the target DNA strand, and extension of these primers by DNA polymerase in the presence of deoxynucleotide triphosphates (dNTPs) (Saiki et al., 1988).

All PCR reaction volumes were made up to a final volume of 25µl. A standard PCR reaction included 100ng DNA template, 0.4 microMolar (µM) of each primer

(forward and reverse), 200µM of each deoxyribonucleotide triphosphate (dNTP) (Bioline, UK), 0.1 unit (U) GoTaq DNA Polymerase (Promega, USA), 1x Colorless GoTaq Reaction Buffer (ph 8.5, 1.5mM $MgCl_2$) (Promega, USA) and distilled water made up to the final reaction volume.

The standard cycling conditions for each PCR reaction included an initial denaturation step at 95°C for five minutes. This was followed by 30 to 40 amplification cycles, consisting of cycles of 95°C for 30 seconds to denature the double-stranded DNA, 30 seconds at a variable annealing temperature (specific to each primer pair) and 40 seconds at 72°C. These cycles were followed by a final elongation step at 72°C for seven minutes. All PCR reactions were performed on the MultiGene Gradient Thermal Cycler (Labnet International, USA), unless stated otherwise.

Unless otherwise stated, all PCR and quantitative PCR primer sequences, reaction components and cycling conditions are listed in Appendices B and C.

2.3 Section A: Determination of CAG repeat size in SCA7 patients

Since 1997, for diagnostic and research purposes, the CAG repeat sizes of the SCA7 patients in the Division of Human Genetics database had been determined. However, these studies had been performed by more than one individual, and at least two different methods of repeat sizing had been used. Therefore, to obtain a consensus on the number of CAG repeats in each SCA7 patient, the following method was used on all samples for which genomic DNA was available.

2.3.1 Gene information

All genetic information (such as genomic DNA sequences and mRNA sequences) for the genes studied was obtained from the National Center for Biology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/) web-based gene browsers.

2.3.2 Primer design & PCR optimization

A PCR primer pair was previously designed for this research study by Ms F Baine (Division of Human Genetics, UCT) to flank the CAG repeat sequence within exon 3 of the *ataxin-7* transcript (NM_000333.3). The primer sequences were submitted to the Oligo Calculator programme (http://www.cnr .berkeley.edu/~zimmer/oligo TMcalc.html) to obtain the length, melting temperature (Tm), %GC content, molecular weight and optical density reading at one picoMolar (pM).

Primers were also submitted to the IDT OligoAnalyzer web-based tool (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx) to determine potential self-complementarity, homodimer formation and heterodimer formation. The designed primers were synthesized by the Department of Molecular and Cellular Biology, UCT. They were diluted in distilled water to working concentrations of 20µM in a final volume of 200µl and stored at 4°C. All primer sequences are given in Appendix B.

In order to determine the optimal conditions for the PCR reaction, a number of optimization assays were performed. Firstly, a temperature gradient PCR was

executed to establish the optimal annealing temperature for the reaction. This assay used standard cycling conditions, but varied the annealing temperature by approximately 1°C across twelve PCR reactions. Control DNA isolated from an individual with no known disease-causing mutations was used for all optimization assays.

Should a temperature gradient not produce optimal results, it may be necessary to change additional parameters in the PCR reaction. The following step in the optimization procedure for this particular PCR reaction was to utilize a PCR optimization kit (Roche, Switzerland). Firstly, 16 PCR reactions were prepared using buffers of varying pH values and magnesium chloride ($MgCl_2$) concentrations. Magnesium ions act as cofactors in the PCR reaction, and thus it is necessary to determine the optimal concentration in order to increase the specificity of the PCR reaction.

In cases where a magnesium gradient did not provide results with sufficient specificity, a dimethylsulfoxide (DMSO) PCR gradient was performed in the same manner. DMSO has been reported to reduce unspecific priming (Pomp and Medrano, 1991). Since the CAG region which was targeted for amplification had a high GC content, it was decided to utilize an annealing time shorter than the standard 30 seconds used in a PCR reaction. Mamedov et al. (2008) found that the optimal annealing times for GC-rich regions lie between three and six seconds.

The final step in the optimization process was to utilize FailSafe Premix buffer J (EPICENTRE Biotechnologies, USA) in the PCR reaction mix. The buffer contains reagents such as magnesium and betaine to increase the specificity of the PCR amplification. The PCR primer sequences, along with the PCR reaction composition and cycling conditions are presented in Appendix B and C.

2.3.3 PCR product detection

2.3.3.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used to detect the amplified PCR products. Agarose gels were prepared using Agarose D-1 LE (Hispanger, Spain) and 1x tris borate EDTA Buffer (TBE) (Appendix A). 1x TBE buffer was also used as the electrophoretic medium. All agarose gels prepared were 2 or 3% (w/v), unless otherwise stated.

The PCR fragments were visualised using SYBR Gold nucleic acid staining (Molecular Probes Invitogen, USA). SYBR Gold binds to double-stranded DNA and is detectable under UV light. SYBR Gold loading buffer was prepared by adding 2µl SYBR Gold stock solution (10 000X solution) to 1ml of 1x loading dye (Appendix A). Three microlitres of this solution was added to 5µl of the PCR product before electrophoresis. In all cases the GeneRuler 100bp DNA Ladder Plus (Fermentas, Canada) was electrophoresed alongside PCR products to allow for approximate size determination of the products (Appendix A). Electrophoresis was performed at 160 V for approximately 40 minutes. PCR products were visualised on the agarose gel under UV light using the UVIPro UVIGold transilluminator (UVItec Limited, UK). The UVIpro (version 12.3) software was used to capture and edit images.

2.3.3.2 Genotyping

The ABI Prism 3100 Genetic Analyzer (Applied Biosystems, USA) was used for accurate size detection of the PCR products. The automated fluorescence-based capillary electrophoresis system allowed for more sensitive sizing than conventional agarose gel electrophoresis methods. One microlitre of each PCR product was mixed with 0.4µl GeneScan 500 ROX Size standard (Applied Biosystems, UK) and 8µl Hi-Di Formamide (Applied Biosystems, UK). The resulting solution was added to the wells of a 96-well microtitre plate and denatured at 95°C for 5 minutes in a Hybaid Touchdown thermocycler (Hybaid Limited, UK), followed by snap freezing on ice. Following electrophoresis and detection, data collection was performed by the ABI Prism 3100 Genetic Analyzer Data Collection software (version 1.1, Applied

Biosystems). The resulting raw data files were analysed to provide PCR product sizing using the GeneMapper software (Applied Biosystems, UK).

2.3.4 Calculation of CAG repeat length

The size of each PCR product was determined using the GeneMapper software, as described above. The primer pair used in the PCR reaction was designed to flank the CAG repeat region, such that the size of the PCR product (excluding the repeat region) was 266 base pairs. Therefore, to provide an approximation of the number of repeats within a PCR product, 266 was subtracted from the PCR product size, and the resulting figure was divided by 3 (because the repeat is a trinucleotide), as follows:

$$Number\ of\ repeats = \frac{(PCR\ product\ size - 266)}{3} \quad [1]$$

A study by Dorschner et al. (2002) aimed to develop a multiplex PCR reaction to determine the CAG trinucleotide repeat lengths of SCA1, 2, 3, 6 and 7. During capillary electrophoresis of the SCA amplicons, they found that the fragments migrated faster than predicted, leading to an underestimation of repeat size compared to results obtained from polyacrylamide gel electrophoresis (PAGE). Therefore a size-correction formula was calculated to determine the accurate number of CAG repeats. The SCA7 formula [2] developed by this group was used to calculate the accurate repeat size of the individuals used in this project.

$$Number\ of\ repeats = \frac{(PCR\ product\ size - 266)}{3} \times 1.1211 + 5.0958 \quad [2]$$

2.4 Section B: Determination of SNP genotype (rs 377 4729) in SCA7 patients

Not all patients stored within the database had previously been genotyped for SNP rs 377 4729. Since this would be the SNP targeted by the RNAi-based therapy, this information could be utilized to determine whether patients could be eligible for the therapy, and to determine the scope of potential application of the therapy within the group of patients. Therefore the SNP was genotyped in all samples that had not previously been genotyped.

2.4.1 Primer design & PCR optimization

Primers previously designed to flank the region of the SNP of interest were obtained from the primer database within the Division of Human Genetics, UCT (Greenberg et al., 2006). A temperature gradient PCR was performed as described previously to determine the optimal annealing temperature for the reaction. The primer sequences, reaction components and cycling conditions are listed in Appendix B and C.

2.4.2 Restriction enzyme digestion

Restriction enzyme digestion was used to determine the SNP genotype for each individual. Restriction endonucleases are able to cleave double-stranded DNA at a nucleotide sequence specific to each enzyme, termed a restriction site. Because the SNP of interest involved a GA nucleotide change, the restriction enzyme Hsp92II (Promega, UK) was chosen due to its recognition sequence of CATG↑. Therefore it would preferentially cleave a DNA sequence containing an A allele at the position of interest, whilst leaving a sequence containing a G allele intact. The sequence of the PCR product, along with the cleavage sites are presented in Figure 4. The restriction digest reaction components are given in Table 2. PCR products were digested at 37°C for one hour, spiked with 10U Hsp92II, and left to digest overnight at 37°C.

AATGAACTGCCTGTCAACTCCCACGGCAGTTTTTCCCACTCACACACTCCTCTAGACAAACTCATAGGAAAGA
AAAGAAAGTGCTCACCCAGCTCGAGCAGCATCAACAACAGCAGCAGCAAACCCACAAAGGTTGCCAAAGTG
CCAGCCRTG↑AACAATGTCCACATG↑AAACACACAGGCACCATCCCAGGGGCACAAGGACTGATGAACAG
TTCCCTCCTTCATCAGGTAGGAAATGGACTGTGAGC

Figure 4: Sequence of the 248 base pair SCA7 SNP PCR fragment. The sequence is shown from 5' to 3'. The position of the A/G SNP is shown by R. Hsp92II cleavage sites are underlined.

Table 2: Hsp92II restriction digestion reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final concentration
dH_20	10	
Buffer K (Promega, UK) (10x)	2	1x
Hsp92II (Promega, UK) (10U/µl)	1	0.5U/µl
PCR product	7	
Total	**20**	

2.4.3 Product detection

The expected fragment sizes following restriction enzyme digestion for each of the three genotypes are presented in Table 3. Because some fragments differed in size by only 10 bases, it was necessary to resolve the fragments on a 6% polyacrylamide gel (Appendix A). PCR fragments were visualised using silver nitrate staining (Appendix A).

Table 3: SCA7 SNP genotypes & expected fragment sizes following Hsp92II digestion

Genotype	Fragment sizes (bp)
Uncut	248
GG	168, 80
GA	168, 153, 80, 15
AA	153, 80, 15

2.4.4 Validation of genotype by direct cycle sequencing

In some instances, direct cycle sequencing was employed to validate the genotype obtained from restriction enzyme digestion, or as an alternative means of genotyping. The chain-termination di-deoxynucleotide method for DNA sequencing

was developed by Sanger et al. (Sanger et al., 1977). Cycle sequencing follows the same principle as PCR, but with fluorescently-labelled dideoxynucleotides incorporated into the reaction mix. The incorporation of these nucleotides into the sequence results in sequence termination, since they lack a 3' hydroxyl group. Each of the four bases is labelled with a different coloured dye. The products are separated on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, USA), which separates the fragments according to size, and reads the dye incorporated at the 3' end of each fragment. Finally, the sequence of the PCR product is constructed from this data.

To carry out cycle sequencing, a PCR reaction designed to amplify the region of choice was completed as described previously (prior to restriction enzyme digestion), and PCR products were electrophoresed on an agarose gel. Thereafter, the cycle sequencing reaction was prepared as below in Table 4, using reagents from the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, UK). Cycling conditions are given in Table 5.

Table 4: Cycle sequencing reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
PCR product	3-8	
Forward or reverse primer (20µM)	0.5	0.5µM
Terminator mix (Applied Biosystems)	2	
Dilution buffer (Applied Biosystems)(5x)	4	1x
dH_2O	Up to 20µl	
Total	**20**	

Table 5: Cycling conditions for cycle sequencing

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	5 minutes	1
2	96	30 seconds	
	50	15 seconds	30
	60	4 minutes	

After the cycle sequencing reaction, samples were subjected to an ethanol precipitation clean-up reaction. Samples were transferred to 1.5ml tubes, whereafter 2µl of Sodium Acetate (NaAc) (3M, pH 4.5) and 50µl of 100% Ethanol was added. Samples were then incubated at -20°C overnight. After incubation, samples were centrifuged for 10 minutes at 10 000 revolutions per minute (rpm) in then Eppendorf MiniSpin microcentrifuge (rotor F45-12-11). The supernatant was discarded, and 30µl of 70% ethanol was added, followed by an additional centrifugation step at 10 000rpm for 10 minutes. The supernatant was discarded, and samples were air-dried. Thereafter, samples were re-suspended in 10µl dH_2O and subjected to capillary electrophoresis on the ABI Prism 3100 Genetic Analyzer (Applied Biosystems, USA). Results were analysed using the BioEdit Sequence Alignment Editor (Hall, 1999). Within BioEdit, the ClustalW multiple sequence alignment program was used to align sequences from test samples to the NCBI reference sequence (NG_008227.1), in order to identify the position and genotype of rs 377 4729.

2.5 Section C: Development of a quantitative assay to measure levels of wild-type and mutant *ataxin-7*

The aim of objective B of the study was to develop a quantitative assay to measure the levels of mutant and wild-type *ataxin-7* in SCA7 patient-derived transformed lymphoblasts, with the hope that this assay could be used to test the efficacy of the proposed RNAi-based therapy. Since the RNAi-effector was designed to specifically target the SNP linked to the expanded CAG repeat, it was necessary to develop an assay that was able to quantify the amount of each allele of the SNP in a given sample. Therefore, two allele-specific assays were designed, one to selectively measure the A allele (associated with the mutant transcript), and another to measure the G allele (associated with the wild-type transcript).

Because the aim of the assay was to quantify the levels of mRNA transcript, it was necessary to use cDNA, rather than genomic DNA, as the template for further PCR-based assays. Total RNA can be isolated from cells and used for downstream assays, however, the single-stranded nature of RNA and the high prevalence of RNases (enzymes which degrade RNA) means that RNA in its native form is highly unstable, and cannot be stored for extended periods of time. Therefore, reverse transcription enzymes (RNA-dependent DNA polymerases isolated from retroviruses) are used to synthesize complementary DNA (cDNA) from mRNA, which can be stored over extended periods of time and used in the same way as genomic DNA, but with the same nucleotide sequence as the mRNA transcript from which it was derived.

2.5.1 Tissue culture methods

Whole blood samples had previously been obtained from four SCA7 patients, hereafter referred to as CA65.3, CA65.4, CA65.5 (3 individuals from the same family), and CA165.2. In addition, numerous de-identified control samples were obtained from individuals unaffected with SCA7. Ten millilitres of whole blood was collected in a heparinised collection tube. B-lymphocytes within each sample were isolated and transformed using Epstein-Barr virus (EBV) in order to immortalise them. This provides an effective way for the lymphocytes to be grown continually

over a long period of time. Once the cells are transformed, they can be frozen and stored in liquid nitrogen for extended periods of time, providing a continuous source of DNA and RNA for future studies. Isolation and transformation of lymphocytes was performed by a member of the NHLS, using a method previously described (Freshney and Freshney, 1994).

2.5.1.1 EBV preparation

The EBV-producing marmoset B-lymphoblastoid cell line (B95-8) was cultured in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, USA) supplemented with 10% (v/v) Foetal Calf Serum (FCS) (Sigma-Aldrich, USA), and grown at 37°C supplemented with 10% CO_2 until medium was well conditioned (yellow in colour). The conditioned medium was collected in a 50ml conical tube and centrifuged at 1 000 rpm for 10 minutes (Sigma 302K centrifuge). The supernatant was removed and filtered through a 0.22µm filter to remove donor lymphoblastoid cells. An equal quantity of fresh culture medium was added and the solution was stored at 4°C for up to 6 months.

2.5.1.2 Lymphoblast isolation and transformation

The buffy coat and plasma was isolated from a 10ml whole blood sample (heparinized) by centrifuging for 10 minutes at 3 000 rpm at room temperature (Sigma 302K centrifuge). The buffy coat and plasma was collected using a 10ml pipette. The mixture was then layered on 5ml of Lymphoprep separation medium (Axis-Shield, Norway) in a 50ml conical tube, followed by centrifugation at 3 500 rpm for 20 minutes. The lymphocytes were removed via pipette from the interface between the Lymphoprep and plasma, and washed three times with culture medium at 1 000 rpm for 10 minutes. Thereafter cells were suspended in a conical tube containing 5ml of prepared EBV (see above) and transferred to a $25cm^2$ tissue culture flask with 100µl of Phytohaemagglutinin (PHA) (Invitrogen, USA) (final concentration 2mg/l) to inactivate regulatory T cells. The flask was incubated upright at 37°C (10% CO_2). After 3 days 2.5ml of the medium was removed and replaced with 2.5ml of fresh medium. Once fully transformed, one volume of cell

suspension was diluted in 2 volumes of fresh culture medium (DMEM with 10% FCS) and cultured at 37°C supplemented with 10% CO_2.

2.5.2 RNA extraction and cDNA synthesis

Prior to RNA extraction, 1 000 000 lymphoblast cells were seeded in 5ml of DMEM. After further culturing for three days, RNA was extracted following the Roche High Pure RNA Isolation kit (Roche, Switzerland) protocol. Following extraction, the RNA concentration was determined using the NanoDrop ND-1000 Spectrophotometer and NanoDrop 1000 software (NanoDrop Technologies, USA). Additionally, 8µl of the undiluted RNA sample (approximately 800ng) was loaded onto a 1% agarose gel and electrophoresed at 120V to determine the integrity of the sample. Total RNA should exhibit two high molecular weight bands, corresponding to 18S and 28S ribosomal RNA. Each sample of RNA was diluted in distilled water to give a concentration of 250ng/µl, and aliquoted into 2µg samples. Aliquots were subsequently stored at -80°C for later use, or used directly to synthesize cDNA following the Novagen First Strand cDNA synthesis kit (EMD Chemicals, USA) protocol, using oligo(dT) primers.

For initial experiments, the concentration of the cDNA was determined using the NanoDrop ND-1000. cDNA was then diluted to 100ng/µl to be used in further applications. However, cDNA cannot be reliably quantified spectrophotometrically without further purification. Therefore, for further experiments, cDNA was allocated a concentration based on the original quantity of RNA used to synthesize the cDNA. For example, if 2µg of RNA was used in a 20µl cDNA synthesis reaction, the resulting cDNA was allocated a concentration of 100ng/µl.

2.5.3 Design of allele-specific primers

PCR primers were designed to distinguish between the two alleles of the SNP of interest (rs 377 4729) using methods described previously. Various methods can be used to discriminate between two alleles, with the most common involving the incorporation of a mismatch to the SNP at the 3' end of the forward primer, or one or two bases upstream. Primers with and without mismatches to the DNA template

have different reaction efficiencies, therefore it is possible to design primers that are able to discriminate between alleles at a specific locus, allowing specific amplification of the allele of choice (illustrated in Figure 5).

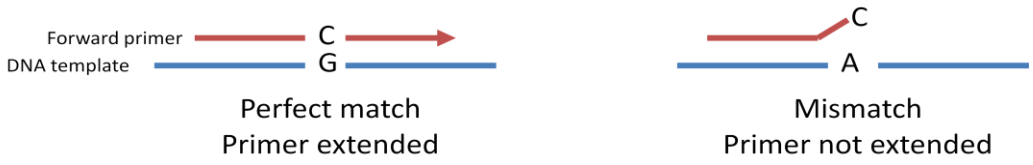


Figure 5: Allele-specific primer design. Primers matched to the DNA template at their 3' end are extended in the PCR reaction (left image), whilst primers with 3' end mismatches are not extended (right image).

Based on this rationale, two or three different forward primers were designed to specifically amplify each of the SNP alleles, using the same PCR design principles described previously. In addition, two different reverse primers were designed. The primer sequences are given in Table 6. The possible primer combinations, and the expected PCR fragment sizes, are listed in Table 7. Primers to amplify *ataxin-7* mRNA (regardless of genotype) were also designed and ordered from Primer Design (UK). Sequences, reaction components and cycling conditions are given in Appendix B and C.

Table 6: Allele-specific primer sequences. All sequences are given 5' to 3'. The position of SNP rs 377 4729 is underlined.

Primer	Sequence (5' to 3')
A forward 1 (F1)	CAAAGTGCCAGCC<u>A</u>
A forward 2 (F2)	CAAAGTGCCAGCC<u>A</u>T
A forward 3 (F3)	AAGTGCCAGCC<u>A</u>TG
G forward 1 (GF1)	CAAAGTGCCAGCC<u>G</u>
G forward 2 (GF2)	CAAAGTGCCAGCC<u>G</u>T
Reverse 1 (R1)	TAGCAAGCAGTTCAGACC
Reverse 2 (R2)	AGTGACCTACAGATGACAGTG

Table 7: Allele-specific primer combinations and expected fragment sizes

Allele to be amplified	Primer combination	Expected fragment size (bp)
A	F1 and R1	376
	F1 and R2	554
	F2 and R1	376
	F2 and R2	554
	F3 and R1	376
	F3 and R2	554
G	GF1 and R1	376
	GF1 and R2	554
	GF2 and R1	376
	GF2 and R2	554

2.5.4 Screening and optimization of allele-specific primers

2.5.4.1 A allele-specific primers

An initial screening was performed to determine which primer pair was most suitable for specific amplification of the A allele. A total of six primer pairs were screened (from three forward primers and two reverse primers). Since the primers would be used later in a quantitative PCR assay (qPCR), a kit designed for qPCR was used (SensiMix SYBR Kit, Bioline, UK), but in a standard thermocycler machine (MultiGene Gradient Thermal Cycler, Labnet International). Reaction components and cycling conditions are given in Appendix C. The PCR was performed on cDNA templates homozygous for each allele, that is, a sample with an AA genotype, and a sample with a GG genotype. PCR products were resolved on a 2% agarose gel (as described previously). The primer pair with the highest specificity for the A allele (with amplification of the AA sample, but not of the GG sample) was chosen for future experiments.

2.5.4.2 G allele-specific primers

Once the most specific primers to amplify the A allele had been identified, an additional two forward primers were designed, in order to amplify the G allele. A screening method was applied as described above, with the primer pair that was

able to specifically amplify the G allele (without amplification of the A allele) being chosen for further experiments.

2.5.5 Quantitative real-time PCR

Quantitative PCR (qPCR), also known as real-time PCR, was first developed in 1993 by Higuchi et al. (Higuchi et al., 1993). The basic concept involves monitoring the progress of a PCR reaction over time, by measuring the fluorescence of a dye that binds to double-stranded DNA. As the PCR reaction progresses, the amount of double-stranded DNA will increase exponentially and the fluorescence of the sample will increase accordingly. SYBR Green, the intercalator dye most commonly used for qPCR, fluoresces as it binds to newly-synthesized double-stranded DNA. If a certain threshold of detected fluorescence is set, the number of thermocycles a particular sample undergoes until it reaches the specified threshold is directly linked to the amount of starting template. For example, a sample with a high abundance of a given transcript will take fewer cycles to reach the threshold, than another sample with a lower level of the same transcript. The cycle at which the sample reaches the fluorescence threshold is referred to as the Ct (threshold cycle). Therefore qPCR is a method often employed to quantify the amount of template in a given sample. Since cDNA is double stranded, this method can also be used to quantify the amount of an RNA transcript in a sample.

Once primer screening had been completed on a conventional thermocycler, the qPCR reactions were performed on the Rotor-Gene 6000 Real-time rotary analyzer (Corbett Life Science, Australia). The Rotor-Gene differs from conventional thermocyclers by containing light-emitting diode (LED) and a photomuliplier detector. As each sample moves past the LED, light passes through the sample, and the fluorescence is measured by the detector. Therefore the amount of fluorescence in each sample is measured for each cycle.

Further optimization was required to obtain allele-specific amplification on the Rotor-Gene 6000. For the purposes of optimization, samples of each of the three genotypes (AA, GA and GG) were used. For example, for the A allele-specific

primers, the aim was to obtain ample amplification of the AA template, with less amplification of the GA template (since it would contain half the amount of the A allele transcript), and no amplification of the GG template. Likewise for the G allele-specific primers, where amplification of the GG template was required, with no amplification of the AA template. A number of parameters were changed in order to achieve adequate allele-specific amplification. These included: the concentration of primer added, the total volume of the reaction, the number of temperature steps in a cycle (two steps or three steps), the amount of cDNA added to the reaction, and the number of completed cycles. The optimized reaction components and cycling conditions for each primer pair are given in Appendix C.

2.5.6 Reference gene

The amount of RNA in a particular sample can be influenced by a number of factors. These can include the number of cells in the extraction, the RNA extraction efficiency, and the efficiency of the cDNA synthesis reaction. Therefore, in order to make comparisons between multiple samples, one must be able to control for the amount of RNA in each sample. This is most commonly achieved by quantifying the levels of an internal control gene (also called a reference gene). Ideally, the reference gene should be ubiquitously expressed over multiple cell types, and its expression should be unaffected by the experimental conditions. Once a reference gene is chosen, the abundance of the transcript of interest can be measured relative to the abundance of the reference gene.

For the purposes of this study, the *β-actin* gene was chosen as the reference gene. Our collaborators at the University of Oxford (UK) performed a screening experiment using the geNorm Reference Gene Selection kit (Primer Design, UK). This kit can be used to identify the best candidate reference gene for any particular experimental scenario. The expression of six candidate reference genes was measured over a series of experimental conditions (such as pre- and post-*ataxin 7* knockdown), and the most stable gene (identified as *β-actin*) was chosen as the reference gene for future experiments. Primers designed to amplify the *β-actin* gene in a qPCR reaction were ordered from Primer Design (UK). The reaction

components and cycling conditions used to amplify the *β-actin* transcript are given in Appendix C.

2.5.7 Standard curves

Two categories of mRNA quantification can be achieved through qPCR. Absolute quantification can allow the user to measure the input copy number of a particular mRNA transcript in a sample. Relative quantification is used to describe the change in expression of a particular transcript relative to another sample (usually an untreated control, or a sample at time zero in a time course experiment). For the purposes of this study, the relative quantification method of analysis was used, since the change in expression of each transcript (A or G allele) would be measured before and after administration of the RNAi effector.

In order to use the relative standard curve method of analysis, a dilution series (containing control cDNA of known concentrations) was prepared (40, 20, 10, 5, 2.5, 1.25 ng/µl), and used in qPCR reactions with each primer pair. These samples were used to prepare standard curves for each test gene (each primer pair), by plotting the log of the concentration of each sample (x-axis) versus the corresponding Ct value (y-axis). The equations of the resulting straight line graphs were obtained, to be used for analysis of the test samples.

2.5.8 Data analysis

Ct values for each sample were attained using the Rotor-Gene 6000 series software (Corbett Research, Australia). In order to calculate the concentration of cDNA in each unknown sample (test samples), the experimentally obtained Ct value was used in the equation acquired from the standard curve previously prepared for the gene of interest. This value could then be normalised to the *β-actin* value obtained for the same sample (as the endogenous control), to account for different amounts of starting material. These values could then be used for downstream analyses. A schematic diagram of the analytical process is given in Figure 6.

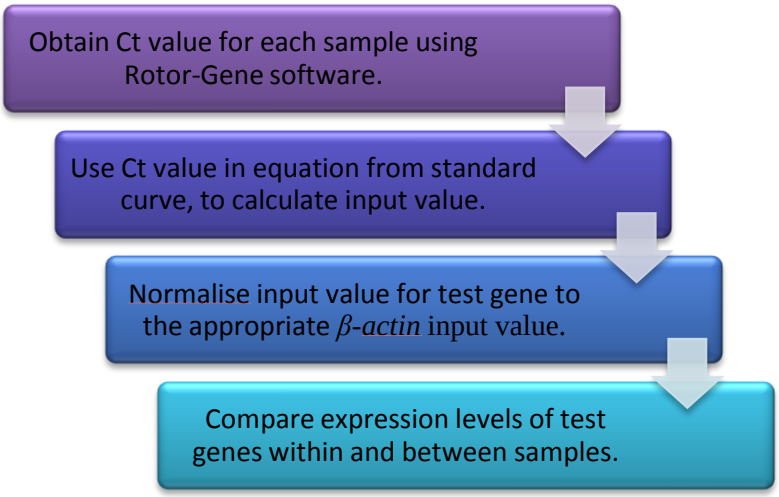


Figure 6: Schematic diagram illustrating the analysis of qPCR data

2.5.9 Experiment 1

2.5.9.1 Aim

To confirm specificity and sensitivity of the optimized allele-specific qPCR assay in patient- and control-derived lymphoblast cell lines of all three genotypes (AA, GA and GG).

2.5.9.2 Samples

Since all three genotypes (AA, GA and GG) were required for the experiment, the following patient- and control-derived cell lines were used:

Table 8: Samples used in Experiment 1

Patient	Genotype	Disease status
CA 65.3	AA	Affected
CA65.4	AA	Affected
CA 65.5	AA	Affected
CA 165.2	GA	Affected
Con 6	GA	Unaffected
Con 9	GG	Unaffected

2.5.9.3 Method

One million cells from each cell line were seeded in 5ml DMEM. RNA was extracted and cDNA synthesized as described previously. Three separate extractions and syntheses were completed for each cell line, resulting in a, b and c samples for each line (e.g. CA 65.3a, CA 65.3b, CA 65.3c). qPCR was completed on each sample as described previously, using the *ataxin-7*, *β-actin*, A-allele specific and G-allele specific primer pairs. Each sample was measured in technical triplicate (the same sample measured three times), and a no template control was included, where distilled water is used in the place of the cDNA template.

2.5.9.4 Analysis

Resulting Ct values were used in equations obtained from the standard curves to acquire the calculated input value for each sample. Input values were normalised

to the appropriate *β-actin* input value to account for differences in starting material. The resulting A-allele and G-allele transcript expression values were plotted on bar graphs as a percentage of total *ataxin-7* expression to indicate relative levels of each transcript in each sample. The expression values for the biological triplicates were averaged and plotted, to indicate general trends in expression between individuals. Thereafter the samples were grouped and averaged by genotype (AA, GA or GG). This gave an indication of sensitivity of the assay. For example, if the A allele-specific primers were used in qPCR reactions with samples of AA, GA and GG genotype, one would expect the AA sample to show a similar expression level as the total *ataxin-7* transcript (which was measured regardless of phenotype). The heterozygous samples should show approximately half the amount of transcript compared to total *ataxin-7*, and the GG samples should show no expression of the A allele transcript.

2.6 Section D: Development of a quantitative assay to measure levels of heat shock proteins 27 and 70 in SCA7 patient-derived lymphoblasts.

Previous reports have shown reduced levels of heat shock protein 27 and 70 (Hsp27 and 70) expression in patient-derived lymphoblasts, compared to cells derived from control individuals (Tsai et al., 2005). Our aim was to validate these results in our cohort of patients. Changes in expression of *Hsp27* and *Hsp70* between patient- and control-derived cells may serve as a measurable phenotype associated with the disease. This may provide a useful tool for measuring the effect of knock-down of the mutant *ataxin-7* transcript, since Hsp levels could be measured before and after administration of the therapy.

In order to quantify the amount of each transcript in our patient samples, qPCR primers were designed and ordered from Primer Design (UK). Since these primers had been designed and previously tested by the company, minimal optimization was required. Reaction components, cycling conditions and primer sequences are given in Appendix B and C.

2.6.1 Experiment 2

2.6.1.1 Aim

To determine the expression levels of *Hsp27* and 70 in patient- and control-derived lymphoblast cell lines, in order to confirm or refute previous studies reporting reduced expression of these transcripts in patient-derived cells.

2.6.1.2 Samples

The cell lines used for Experiment 2 (along with the individual's disease status) are listed in Table 9. A total of four patient-derived cell lines and three control-derived cell lines were used for analysis.

Table 9: Samples used in Experiment 2

Patient	Disease status
CA 65.3	Affected
CA 65.4	Affected
CA 65.5	Affected
CA 165.2	Affected
Con 6	Unaffected
Con 8	Unaffected
Con 9	Unaffected

2.6.1.3 Method

Preparation, RNA extraction and cDNA synthesis was performed as described for Experiment 1. qPCR was conducted on each sample, using the *β-actin*, *Hsp27* and *Hsp70* primer pairs. Each sample was measured in technical triplicate, and a no template control was included.

2.6.1.4 Analysis

Data was obtained, normalised and analyzed as described in Experiment 1. The average expression level of biological triplicates was determined and plotted on a bar graph to indicate trends in expression between individuals. Thereafter the expression levels between patient and control-derived lines were plotted to gauge

differences in expression between the two groups. The student's t-test was utilized to determine significant differences in expression between the two groups.

Chapter 3: Results and Discussion

Chapter 3: Results and Discussion

3.1 Database query and cohort selection

The SCA patient database of the Division of Human Genetics, UCT, was interrogated to select patients to be used in the study. In order to be included in the study, patients either had a confirmed diagnostic test for SCA7, or the results of previous SCA7 diagnostic testing had been inconclusive. Furthermore, patients were only included in the cohort if the Division was in possession of a genomic DNA sample. Thirty patients were selected to be included in the study. The samples were coded to maintain patient confidentiality.

3.2 DNA isolation & integrity

Once DNA stocks had been collected from the appropriate storage areas and diluted to a working stock of 100ng/µl, an aliquot of each sample was subjected to agarose gel electrophoresis in order to determine the integrity of the DNA sample. An example of a DNA integrity gel is given below in Figure 7.

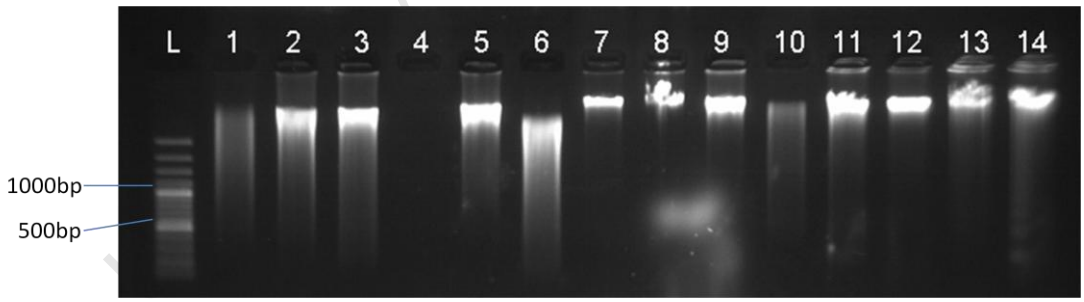


Figure 7: Determination of genomic DNA integrity by agarose gel electrophoresis. This is an image showing size-based separation of total genomic DNA on a 1% agarose gel. Genomic DNA was separated by electrophoresis at 160V for approximately 30 minutes. Visualisation was by SYBR Gold nucleic acid staining under UV light. Lanes 1 to 14 show genomic DNA isolated from peripheral blood lymphocytes (whole blood). Lane L shows a 100bp molecular weight marker (Fermentas).

In order for DNA to be used in downstream applications such as PCR, it is preferable that the DNA sample contains a high proportion of unfragmented DNA. This increases the likelihood that the desired DNA template is present in the sample. Unfragmented DNA samples show a high molecular weight band when

electrophoresed through an agarose gel, whilst DNA samples of poor integrity that are highly fragmented are often visible as smears or low molecular weight bands. The majority of DNA samples selected for the study exhibited bands of a high molecular weight (such as in lanes 7 and 12 in Figure 7). In cases where the sample showed poor integrity (lane 4), an alternative patient sample was sought from the storage facilities.

3.3 Section A: Determination of CAG repeat size in SCA7 patients

The first part of Objective A was to identify all SCA7 patients in the Division of Human Genetics database and determine their CAG repeat sizes. Since the discovery of the *ataxin-7* gene by David et al. in 1997, the standard method of diagnosis of SCA7 has been by sizing of the CAG repeat within exon 3 of the gene. Although original methods utilized estimations of repeat sizing by PCR and agarose gel electrophoresis, the development of capillary electrophoresis has allowed for more accurate measurements in recent years. However, discrepancies between apparatus, PCR primers and methods of analysis employed by different users may lead to inconsistencies in accurate repeat sizing. In the first instance, as previously mentioned, the accurate repeat size of an individual is required to determine whether either allele lies within the pathogenic range (i.e. more than 38 repeats in the mutant allele). Additionally, repeat sizing may have important implications for individuals who do not have alleles within the pathogenic range, but within the intermediate range (19-37 repeats). Whilst these individuals do not present clinically with SCA7, they carry a higher risk for transmitting alleles within the pathogenic range to their offspring, a reality that should be considered from a genetic counselling and family planning perspective.

3.3.1 PCR optimization

Following the design of PCR primers to amplify the region of choice (flanking the CAG repeat in exon 3 of *ataxin-7*), the first step in the PCR optimization procedure was to perform a temperature gradient assay in order to determine the optimal annealing temperature for the PCR reaction. In this assay, twelve PCR reactions were prepared and each subjected to a different annealing temperature during the PCR reaction. An example of the results of such a temperature gradient assay is given in Figure 8.

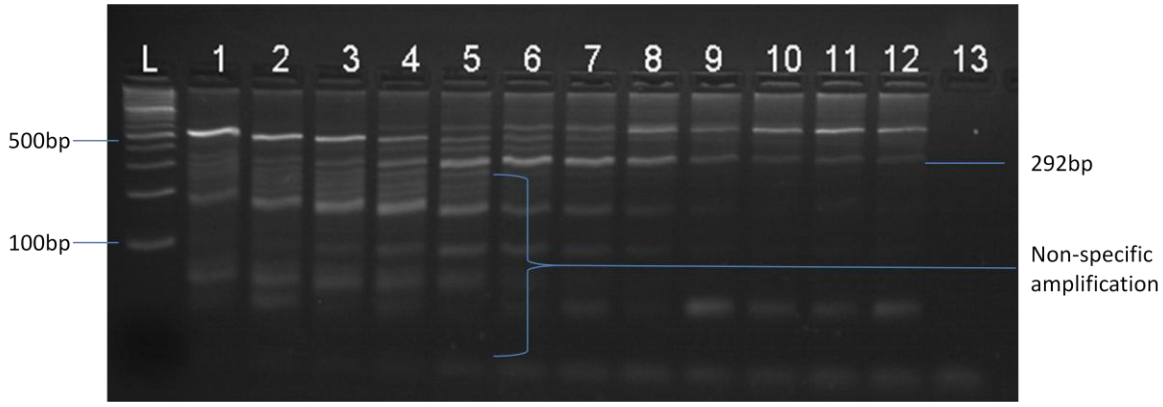


Figure 8: PCR optimization - temperature gradient. This is an image of size-based separation of the CAG repeat PCR fragments on a 3% agarose gel. PCR products were run at 160V for 40 minutes. PCR fragments were visualised by SYBR Gold nucleic acid staining under UV light. The expected size of the PCR fragment was 292bp. Lane L shows 100bp molecular weight marker (Fermentas). Lanes 1 to 12 show PCR fragments obtained from reactions using annealing temperatures ranging from 47°C (lane 1) to 58°C (lane 12). Lane 13 contained a water control sample (no DNA).

From the agarose gel represented in Figure 8 it was seen that the expected PCR product with a size of 292bp was present at all 12 temperatures, however, a large number of non-specific products accompanied the fragment in all cases. Although the non-specific PCR products diminished with increasing annealing temperature (lanes 1 to 9), an increase in annealing temperature also resulted in a reduction of the desired PCR product (lanes 8 to 12). Therefore further optimization was necessary in order to increase the specificity of the reaction. An annealing temperature of 54°C was used for the rest of the optimization procedures (corresponding to lane 8 in Figure 8). The Roche optimization kit was utilized to determine the optimal pH value and $MgCl_2$ concentration for the reaction. For this assay, 16 PCR reactions were prepared using buffers of varying pH values and magnesium chloride ($MgCl_2$) concentrations. The results of this optimization procedure are presented in Figure 9.

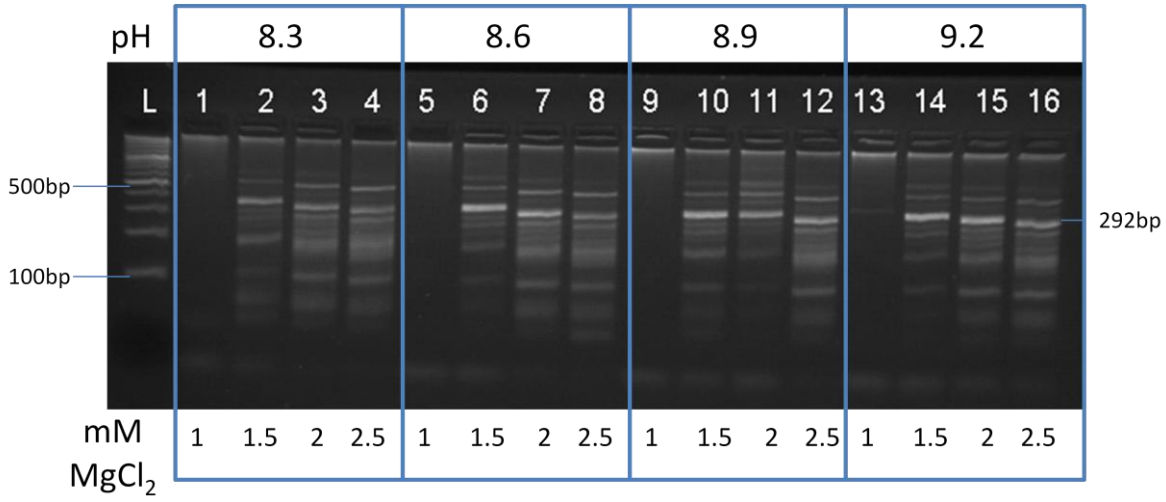


Figure 9: pH and $MgCl_2$ optimization. This is an image of the size-based separation of the CAG repeat PCR fragments on a 3% agarose gel. DNA fragments were visualised using SYBR Gold nucleic acid staining under UV light. Lane L shows a 100bp molecular weight marker (Fermentas). Lanes 1 to 16 show 16 reactions using buffers of varying pH values (pH 8.3, 8.6, 8.9 and 9.2) and $MgCl_2$ concentrations (1, 1.5, 2 and 2.5 mM).

It was evident from Figure 9 that none of the 16 reactions provided the optimal conditions for specific amplification. No amplification resulted from reactions containing buffers of low $MgCl_2$ concentrations (1mM), and increasing the $MgCl_2$ concentration resulted in more non-specific amplification. An increase in pH similarly had little effect. Therefore a DMSO gradient was performed to reduce non-specific priming. The results of the DMSO gradient PCR are presented in Figure 10.

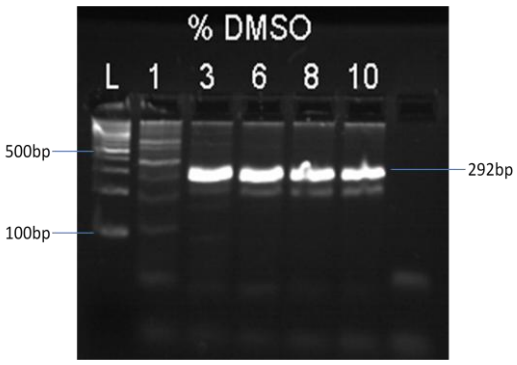


Figure 10: DMSO PCR gradient. An image showing the size-based separation of the CAG repeat PCR fragments on a 3% agarose gel. Visualisation was by SYBR Gold nucleic acid staining under UV light. Lane L shows a 100bp molecular weight marker (Fermentas). Subsequent lanes show 5 PCR reactions, each containing a varied percentage of Dimethyl sulfoxide (DMSO), indicated by the lane number (1, 3, 6, 8 or 10%).

The results of the PCR optimization step using a DMSO gradient showed that the incorporation of 3 to 10% DMSO increased the specificity of the PCR reaction to a large extent. Lanes 3, 6, 8 and 10 each showed the desired PCR fragment with an estimated size of 292bp, along with a smaller non-specific band. In order to further increase the specificity of the reaction, FailSafe Buffer premix J (EPICENTRE Biotechnologies, USA) was used. In addition, the annealing temperature time during the PCR reaction was decreased from 30 to 6 seconds, due to the GC-rich nature of the fragment.

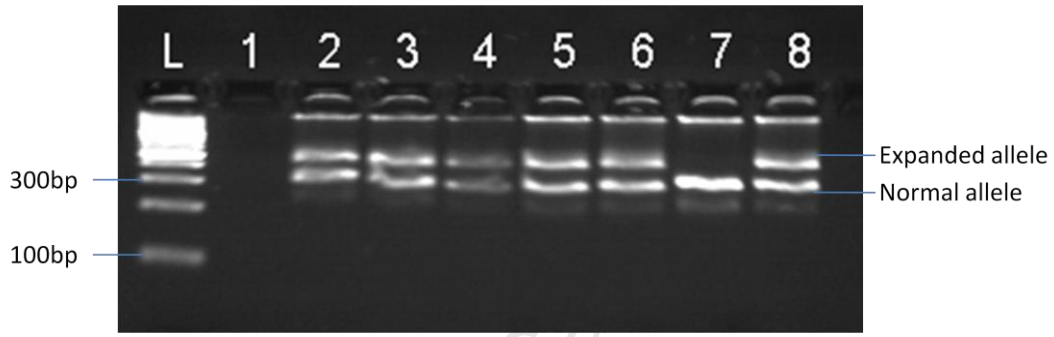


Figure 11: Optimized PCR reaction. This is an image showing the size-based separation of the CAG repeat PCR fragments on a 3% agarose gel. Visualisation was by SYBR Gold nucleic acid staining under UV light. Lane L shows a 100bp molecular weight marker (Fermentas). Lane 1 shows a water control. Lanes 2 to 8 show PCR fragments from test samples. Lane 7 contains an unaffected individual without an expanded allele.

Figure 11 shows the results of the optimized CAG repeat PCR reaction. The optimized CAG repeat PCR reaction was performed on seven different DNA samples obtained from SCA7 patients (lanes 2 to 6, and lane 8), and a single unaffected control individual (lane 7). Samples from affected individuals displayed two distinct PCR products, corresponding to the wild-type (smaller fragment) and mutant (larger fragment) alleles. However, the sample from the unaffected individual showed a single fragment corresponding to wild-type alleles in size. In the diagnostic setting, a single PCR product can indicate that the individual is not affected with SCA7, or that the expanded mutant allele is beyond the limit of detection of the PCR assay (too large). Single fragments visible on an agarose gel can often be resolved during capillary electrophoresis to display either two wild-type alleles of similar size, or a single allele (if the second allele is too large to be detected).

3.3.2 Genotyping

The CAG repeat PCR fragments obtained from patient DNA were subjected to capillary electrophoresis on the ABI Prism 3100 Genetic Analyzer (Applied Biosystems, USA) in order to determine the size of PCR fragments. An example of a resulting electropherogram is presented in Figure 12. The electropherogram showed green peaks corresponding to the PCR fragments. Since the DNA used for the assay was isolated from peripheral blood lymphocytes, the instability of the expanded CAG repeat was observed by the multiple green peaks around 360bp (since different individual cells differ in their repeat size).

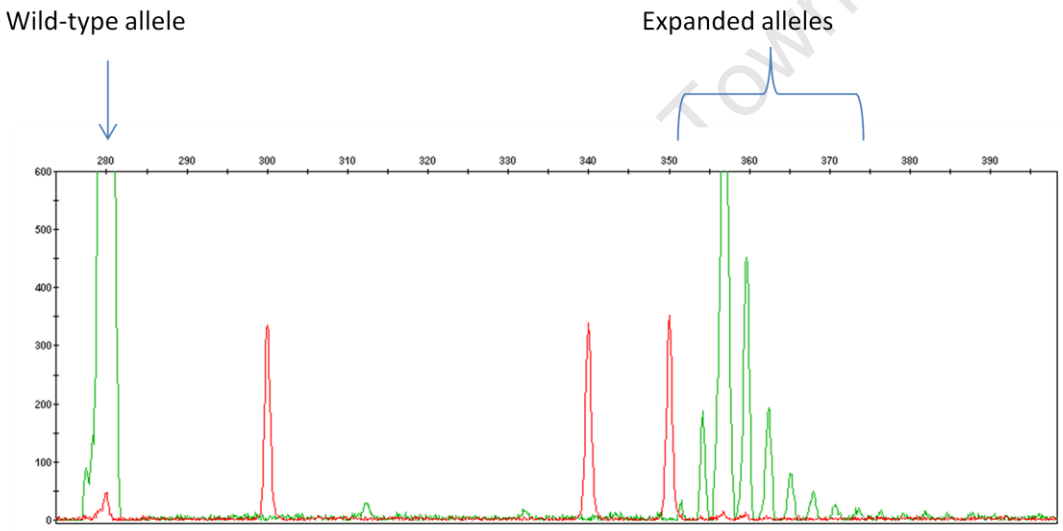


Figure 12: Electropherogram from capillary electrophoresis of a CAG repeat PCR fragment from a SCA7 patient's DNA. Relative fluorescence units are represented on the y-axis, and PCR fragment size in base pairs is represented on the x-axis. Red peaks show the GeneScan 500 ROX size standard. Green peaks show the migration of PCR fragments. The image shows an allele without a CAG expansion with a size of approximately 280bp, and an allele with a CAG expansion (approximately 357bp).

Since the PCR products were electrophoresed along with a size standard (Rox 500, shown by red peaks), the resulting raw data files were analysed to provide PCR product sizing using the GeneMapper software (Applied Biosystems, USA). Thereafter, the formula described previously by Dorschner et al. (2002) (Equation [2], page 44) was used to calculate the number of CAG repeats present in each PCR fragment. The CAG repeat sizing results for all genotyped SCA7 patients is presented in Table 10. Apart from individuals CA 104.3 and CA 637.5, all patient genotypes demonstrated alleles within the disease-causing range (more than 38

repeats). Individuals CA 104.3 and CA 637.5 were included in the study due to the lack of previous clinical and diagnostic information (their disease status was not known). However, from these results it is evident that these individuals are not affected with SCA7, due to the lack of an expanded allele.

Table 10: CAG repeat sizes of SCA7 patients

Patient	No. of CAG repeats (allele 1/allele 2)	Patient	No. of CAG repeats (allele 1/allele 2)
CA 104.1	11/127	CA 531.1	12/72
CA 104.2	12/43	CA 561.1	12/60
CA 104.3	11/12	CA 580.1	13/56
CA 15.1	12/63	CA 600.1	13/58
CA 15.2	11/62	CA 65.1	12/53
CA 15.3	13/50	CA 65.2	14/56
CA 165.1	12/56	CA 65.4	12/39
CA 165.2	7/51	CA 65.5	14/48
CA 213.1	12/54	CA 96.1	11/111
CA 240.1	14/68	CA 637.1	8/56
CA 309.1	12/59	CA 637.2	8/60
CA 355.1	14/83	CA 637.3	8/58
CA 406.1	12/55	CA 637.4	11/61
CA 417.1	14/51	CA 637.5	11/12
CA 521.1	14/62	CA 637.6	11/60

3.3.3 Previous sequencing experiments

As part of a previous study within the Division of Human Genetics, the CAG repeat lengths of a subset of SCA7 patients had been sequenced to accurately determine their repeat length (Ms Fiona Baine, unpublished data). A PCR of the CAG repeat region was completed on a subset of patients, and each allele (expanded and wildtype) was separately excised from an agarose gel, purified, and subjected to cycle sequencing. Although capillary or agarose gel electrophoresis can only allow for approximation of the PCR product sized compared to a size standard, cycle sequencing can allow for a more accurate determination of the number of repeats, since the actual DNA sequence is known. However, since cycle sequencing is more expensive than standard genotyping, and due to difficulties with sequencing of the unstable repeat region (particularly in large repeats), this method is not typically

used for diagnostic purposes. The results of the previous sequencing experiments, along with the genotyping results from this study, are presented below in Table 11.

Table 11: Comparison of sequencing and genotyping results for CAG repeat sizes in SCA7 patients.

Patient	Genotyping results (allele 1/ allele 2)	Sequencing results (allele 1/ allele 2)
CA 96.1	11/111	10/?
CA 165.1	12/56	10/46
CA 355.1	14/83	12/77
CA 417.1	14/51	12/56

From the results in Table 11 there was a clear discrepancy between estimated CAG repeat sizes obtained from the standard genotyping and cycle sequencing methods in each of the four SCA7 patients. There was an average difference of two repeats when comparing the wild-type allele sizes, and a difference of 5, 6 and 10 repeats when comparing the mutant allele sizes. The sequencing method failed to determine the mutant allele repeat number in patient CA 96.1, most likely due to the large size of the repeat.

In summary, a standard genotyping method employing PCR amplification and capillary electrophoresis has been used to estimate the repeat sizes of the SCA7 patients within the cohort for whom genomic DNA was available. This information can be used by clinicians for more effective management of the patients, since the rate of disease progression and severity of future symptoms can be approximated from the repeat sizes. Furthermore, patients can be counselled with regard to their risk of transmitting the disease-associated allele to future offspring. However, comparisons with previous sequencing experiments has highlighted that this method may not be the most accurate means of estimating repeat length, and alternative methods may need to be developed. In a recent publication by the European Molecular Quality Genetics Network, numerous "Best Practice Guidelines" for the molecular genetics testing of the SCAs have been proposed. For example, the group suggests that laboratories responsible for molecular genetic testing should annually participate in external quality assessment schemes, and validated standard operating procedures should be exchanged. Furthermore, there

should be a list of published primer sequences for general use, and sequenced controls should be used for accurate repeat sizing (Sequeiros et al., 2010). The adherence of diagnostic laboratories to these guidelines should allow for more accurate repeat sizing, and a higher level of uniformity between different laboratories.

3.4 Section B: Determination of SNP genotype (rs 377 4729) in SCA7 patients

A PhD graduate from the Division of Human Genetics successfully identified an RNAi-based effector to selectively target the mutant *ataxin-7* allele within the South African SCA7 cohort (Scholefield et al., 2009). This was achieved by targeting a SNP (rs 377 4729) linked to the mutant *ataxin-7* allele in 100% of cases tested to date. Due to the lack of a thorough understanding of the function of the ataxin-7 protein, the assumption was made that the wild-type protein should be maintained for proper cellular functioning, and therefore only patients who are heterozygous for the SNP may be eligible for the therapy. To obtain a current indication of the potential application scope of the therapy, all previously ungenotyped samples were subjected to PCR amplification and restriction endonuclease digestion.

3.4.1 Restriction enzyme digestion

PCR primers flanking the region of SNP rs 377 4729 were used to amplify the genomic region in all SCA7 patients who had not been genotyped previously. An example of gel electrophoresis of these PCR fragments is presented in Figure 13. All samples showed PCR products corresponding to the expected fragment size of 248 base pairs (lanes 2 to 8).

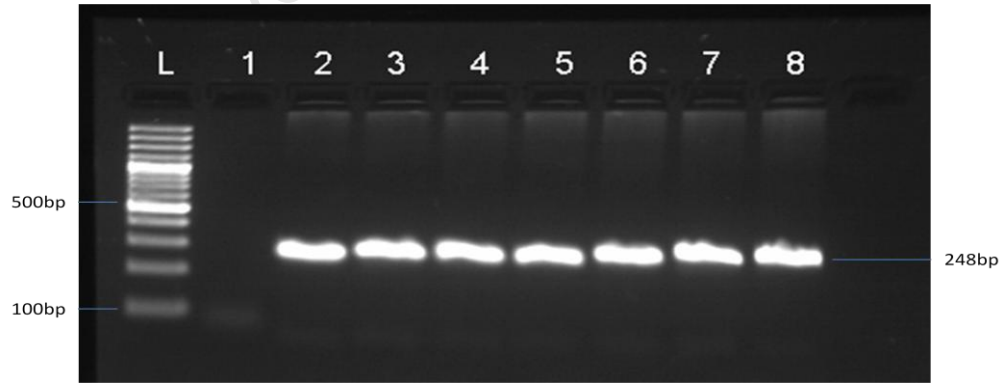


Figure 13: An image showing the size-based separation of the SNP region PCR fragments on a 2% agarose gel. Visualisation was by SYBR Gold nucleic acid staining under UV light. Lane L shows a 100bp molecular weight marker (Fermentas). Lane 1 shows a water control (no template control). Lanes 2 to 8 show PCR fragments obtained from SCA7 patient DNA with the expected size of 248bp.

Restriction enzyme digestion with Hsp92II was carried out on the resulting PCR fragments, and products were run on a 6% polyacrylamide gel visualised with silver nitrate staining as shown in Figure 14. Polyacrylamide gel electrophoresis was employed rather than agarose gel electrophoresis in order to resolve the two bands differing by 10 base pairs (153bp and 163bp). Samples of GG genotype displayed bands corresponding to 168 and 80bp in size (lane 6), whereas samples of AA genotype showed bands corresponding to 153, 80 and 15bp (although the 15bp fragments could not be visualised on the gel) (lane 2). Samples displaying bands corresponding to 168, 153, 80 and 15bp had a heterozygous genotype (lanes 3, 4, 5 and 7).

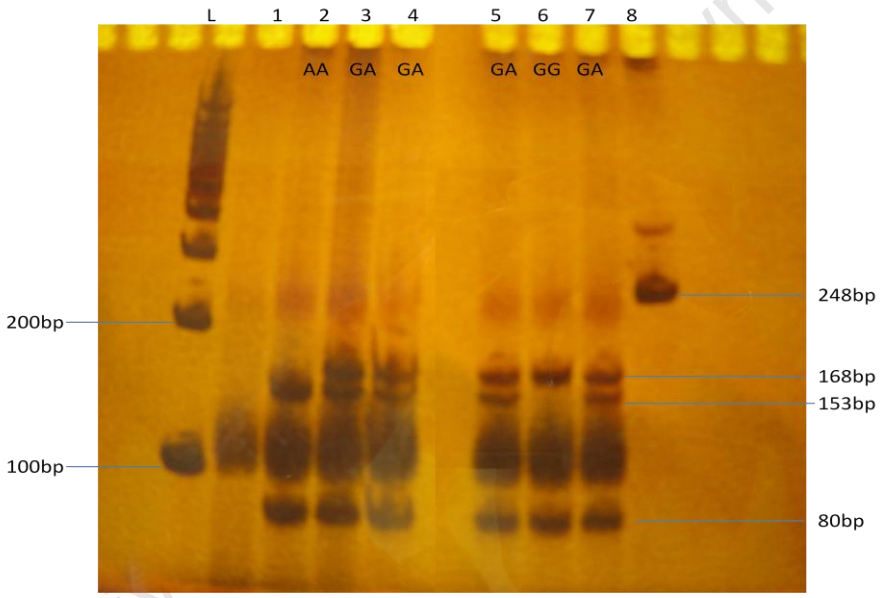


Figure 14: Polyacrylamide gel electrophoresis of PCR fragments subjected to restriction enzyme digestion by Hsp92II. Products were run on a 6% polyacrylamide gel and visualised by silver nitrate staining. Lane L shows a 100bp molecular weight marker (Fermentas). Lane 1 shows a water control (no template control). Lane 8 shows an undigested PCR fragment. The SNP genotype of each sample is indicated at the top of the lane (lanes 2 to 7).

The genotypes of the SCA7 patients genotyped for this study are presented in Table 12.

Table 12: Rs 377 4729 genotypes of SCA7 patients.

Patient	SNP 377 4729 genotype	Patient	SNP 377 4729 genotype
CA 213.1	AA	CA 65.2	AA
CA 309.1	AA	CA 65.4	AA
CA 355.1	AA	CA 65.5	AA
CA 406.1	GA	CA 96.1	GA
CA 417.1	AA	CA 637.1	AA
CA 521.1	AA	CA 637.2	AA
CA 531.1	GA	CA 637.3	AA
CA 561.1	GA	CA 637.4	GA
CA 580.1	AA	CA 637.6	GA
CA 600.1	AA		

From the total of 30 patients who had been genotyped (previously and for this study), 13 individuals had been found to be heterozygous for SNP 377 4729 (43%). The remaining 17 individuals had an AA genotype (57%).

3.4.2 Direct cycle sequencing

In some cases, direct cycle sequencing was used to genotype individual patients and controls, in order to validate results obtained from restriction enzyme digestion. Figure 15 shows an example of a chromatogram subsequent to cycle sequencing of the PCR product. The chromatogram showed coloured peaks, each colour corresponding to a different nucleotide base. The sequence from test samples was aligned to a reference sequence to determine the position and genotype of the SNP. The control sample in Figure 15 had a GG genotype, indicated by the single black peak at position 129. Samples of AA genotype showed a single green peak at the SNP position, and heterozygous samples showed overlayed green and black peaks. All samples that were validated by sequencing showed the same genotype that was determined by restriction enzyme digestion.

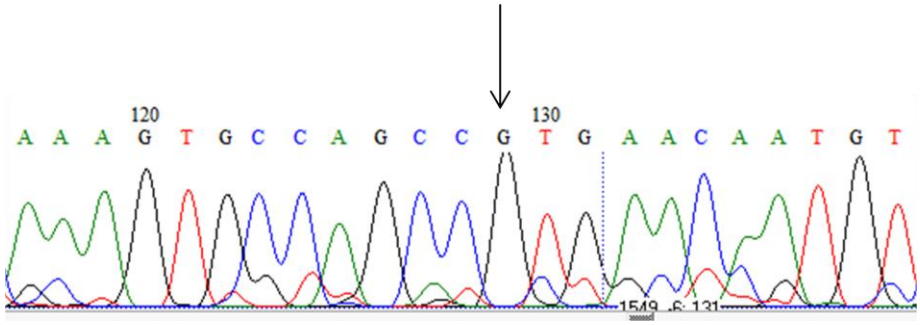


Figure 15: Chromatogram subsequent to cycle sequencing to genotype SNP rs 377 4729 in a single unaffected control individual. The position of the SNP is indicated by a black arrow (therefore this individual had a GG genotype).

Two separate approaches may be taken to enable further patients to be eligible for the therapy. Firstly, further SNPs can be included and targeted by the RNAi effectors. This approach was recently evaluated by Pfister et al. (2009), who found a single SNP to be heterozygous in 48% of their HD cohort. They subsequently developed five allele-specific siRNAs, targeting three SNPs, that could be used to treat three-quarters of the United States and European HD patient populations (Pfister et al., 2009). Along these lines, a further SNP (rs 373 3126) within the *ataxin-7* gene has been calculated to have a heterozygosity of 32% in the South African cohort (Ms Alice Foxon, Honour's project, 2010). Taken together, both SNPs could target approximately 50% of South African patients. Since there are approximately 20 known coding SNPs within the *ataxin-7* gene, it is likely that further SNPs of higher heterozygosities in the cohort may be identified in the future.

Another issue to note is that of the broader potential of application of the therapy in populations outside SA. Although no known studies have investigated whether the rs 377 4729 is linked to the SCA7 mutation in other populations and global regions, it should be noted that a family included within this study (Family CA 637) are of rural northern Namibian origin. Haplotyping studies have revealed that these individuals share the same common haplotype with the South African population at the SCA7 locus (unpublished data), and two of the five affected family members are heterozygous at the rs 377 4729 locus. Together these results suggest that the

scope of application of the RNAi-based therapy may be broader than previously anticipated, particularly on the African continent.

Another point for consideration is that it may not be necessary to knock down the mutant *ataxin-7* allele in a specific manner, and cells may be able to tolerate a complete aberration of expression. Work on a rat model of Machado-Joseph disease (SCA3) has shown that silencing of the wild-type ataxin-3 protein does not aggravate pathology, and that non-allele-specific silencing of the mutant and wild-type transcripts reduces neuropathology (Alves et al., 2010). Similarly, HD mice have been shown to tolerate silencing of the wild-type huntingtin protein for up to four months (Boudreau et al., 2009). However, in the absence of well characterised cellular models and knock-down mouse models for SCA7, the current assumption is that it would be beneficial to maintain expression of the wild-type ataxin-7 protein, if possible.

3.5 Section C: Development of a quantitative assay to measure levels of wild-type and mutant *ataxin-7*

The aim of objective B was to develop a method for measuring expression levels of mutant and wild-type *ataxin-7* in transformed lymphoblast cell lines, to be used in the future to test the efficacy and specificity of a recently developed RNAi-based therapy. In the event that this therapy reaches clinical testing stages, it is also anticipated that this assay could be used to test the therapy in patient-derived cells *in vitro*, before applying the therapy to the patient.

A quantitative PCR assay was developed to measure expression of each allele (G or A) of the SNP rs 377 4729, which is known to be linked to the SCA7 mutation in the South African patient population. Three forward PCR primers were designed for each allele, incorporating the position of the SNP at various positions of the 3' end of the forward primer. These were screened along with two possible reverse primers, to identify the most specific pair. The identified primer sets were then used in qPCR assays to measure expression of each transcript in mRNA isolated from cells of each of the three genotypes.

3.5.1 Screening and optimization of allele-specific primers

In order to identify a primer pair able to specifically amplify the A allele of rs 377 4729, a total of six primer pairs (from three forward primers and two reverse primers) were screened. cDNA templates homozygous for each allele (e.g. AA and GG genotypes) were used. An example of an agarose gel resulting from the screening assay is presented in Figure 16.

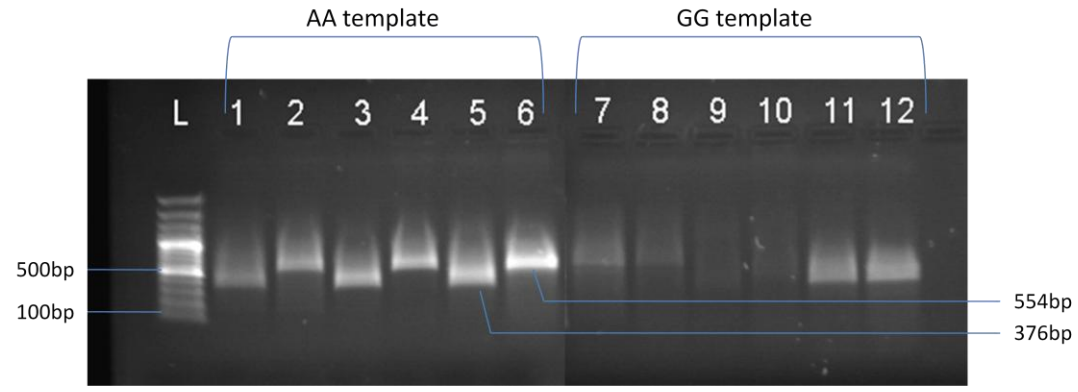


Figure 16: An image showing the size-based separation of PCR fragments on a 2% agarose gel for allele-specific primer screening. Visualisation was by SYBR Gold nucleic acid staining under UV light. Lane L shows a 100bp molecular weight marker (Fermentas). The PCR reactions represented by lanes 1 to 6 contained cDNA with an AA genotype, whereas the cDNA used in lanes 7 to 12 had a GG genotype. The primer pairs used in each lane was as follows: Lane 1 and 7 – primers F1 & R1, Lane 2 and 8 – F1 & R2, Lane 3 and 9 – F2 & R1, Lane 4 and 10 – F2 & R2, Lane 5 and 11 – F3 & R1, Lane 6 and 12 – F3 & R2. The expected fragment sizes were 376bp (if reverse primer R1 was used) or 554bp (if reverse primer R2 was used).

PCR fragments corresponding to the expected sizes (376bp if primer R1 was used, 554bp if primer R2 was used) were visible for all samples containing the DNA template of AA genotype. The primer combination of F3 and R2 showed the brightest PCR band (lane 6), whereas the primer combination of F1 and R1 resulted in the least amount of PCR product (lane 1). Conversely, when the GG DNA template was used, primer combinations using the F1 or F3 forward primers resulted in undesirable amplification of the DNA template (lanes 7, 8, 11 and 12). Samples containing the F2 forward primer resulted in the least amount of amplification of the GG DNA template (lanes 9 and 10). Since a primer pair was needed to specifically amplify the A allele was required, the pair that displayed the best amplification of the AA template with the least amount of amplification of the GG template was chosen for further experiments. Therefore primers F2 and R1 were chosen for the qPCR experiments (used in lanes 3 and 9 of Figure 16). A similar screening method was employed to identify a primer pair to specifically amplify the G allele of rs 377 4729 (data not shown). Primers GF1 and R1 were selected for further experiments.

3.5.2 Quantitative real-time PCR

Once appropriate allele-specific primers were identified using a standard thermocycler PCR machine and agarose gel electrophoresis, the experiments were performed on the Rotor-Gene 6000 real-time rotary analyzer (Corbett Life Science, Australia), in order to enable accurate quantification of each transcript. Since various parameters can vary between different thermocyclers, such as ramp speeds between cycling temperatures, it was necessary to further optimize the allele-specific PCR assays on the qPCR thermocycler. For these purposes, cDNA templates of all three genotypes were used. Cycling parameters were changed to obtain adequate amplification of samples homozygous for the allele for which the primers were designed, with less amplification of the heterozygous samples, and little or no amplification of the samples homozygous for the alternative allele.

3.5.3 Standard curves

Since the relative standard curve method of analysis was utilized, standard curves were prepared for each PCR primer pair by using cDNA samples of known concentration in each qPCR assay. These values were used to construct a standard curve for each primer pair, by plotting the log of the cDNA concentration versus the Ct value obtained for that sample. The equations of these curves were used to analyse further test samples, as described in section 2.5.8. An example of one of the standard curves is presented in Figure 17.

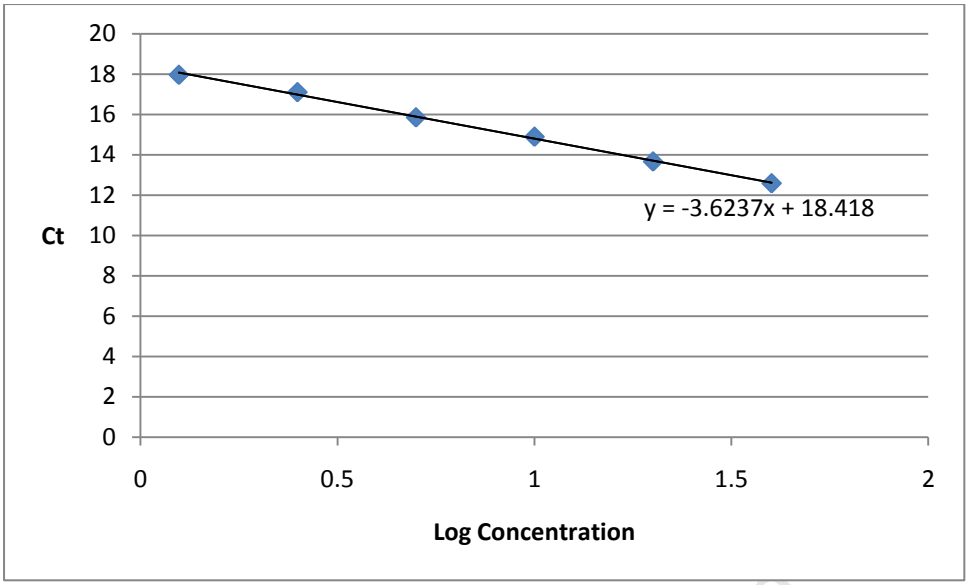


Figure 17: An example of a standard curve used for analysis of qPCR data. Ct values obtained for six serial cDNA dilutions (y-axis) were plotted against the log of each cDNA concentration (x-axis). The equation of the straight line through the data points is given on the graph.

The straight line of the standard curves displayed a negative slope, since dilutions with a higher concentration were detected earlier during the thermocycling process, and therefore had lower Ct values. The input amount for each of the test samples could then be calculated by applying the Ct value for each sample to the standard curve equation.

3.5.4 Experiment 1

The aim of experiment 1 was to confirm the specificity and sensitivity of the allele-specific qPCR assays in patient- and control-derived lymphoblast cell lines of all three genotypes. The assay was used to measure expression of each allele (A and G) in six lymphoblast cell lines (three samples of AA genotype, two samples of GA genotype, and a single sample of GG genotype). The expression of total *ataxin-7* and *β-actin* was also measured to serve as endogenous controls.

The following results show the relative expression of the A allele in each of the six samples, shown separately (by cell line). All values were normalized to *β-actin* expression and plotted as a percentage of total *ataxin-7* expression.

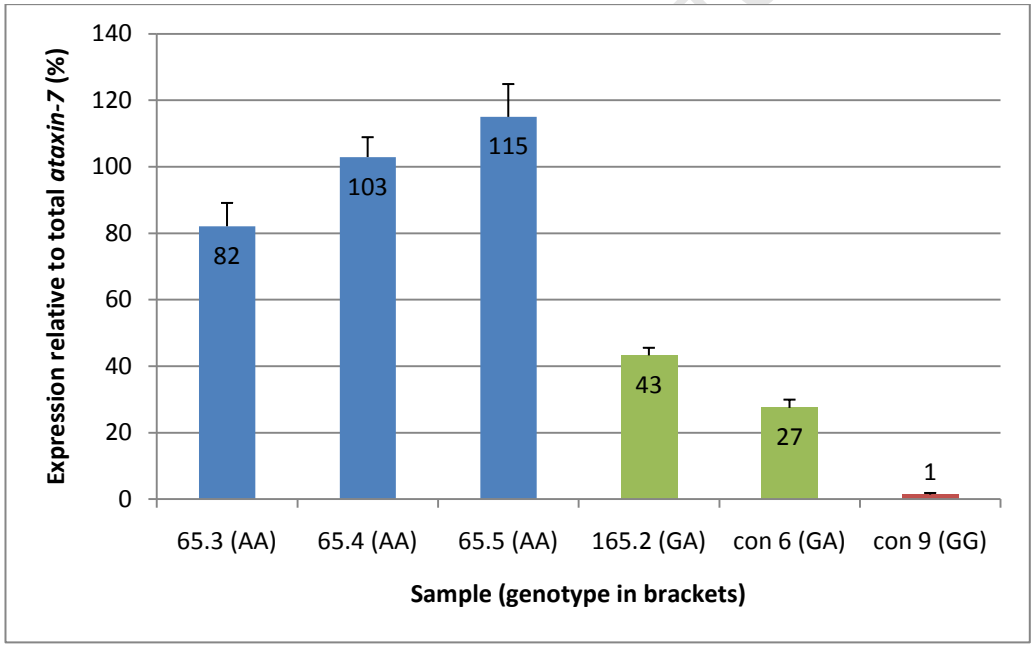


Figure 18: Results of allele-specific qPCR using primers designed for amplification of the A allele of SNP rs 377 4729. Samples show expression of the A allele transcript normalised to *B-actin* and expressed as a percentage of total *ataxin-7* expression (y-axis) in six cell lines (x-axis). Genotypes of cell lines are given in brackets. Individual values are shown within each column. Error bars show standard error of the mean (SEM).

Two distinct parameters were measured in this assay. Firstly, the specificity of the assay could be evaluated by the ability of the assay to distinguish between the different alleles, and preferentially amplify the allele for which it was designed.

Secondly, sensitivity could be assessed by the ability of the assay to accurately measure the transcript for which it was designed. Since the primers used in this assay were designed to preferentially amplify the A allele transcript, it was hypothesised that samples of AA genotype (shown with blue bars in Figure 18) would show 100% of total *ataxin-7* expression, since all *ataxin-7* transcripts present would contain the A allele. The results from Figure 18 showed a range of percentages from 82% (sample 65.3) to 115% (sample 65.5), with an average of 103%. Although these values are close to the expected value of 100%, the difference in values between the three cell lines suggest that the assay may not be particularly sensitive. Furthermore, heterozygous cell lines (green bars) showed expression values of 43% and 27% of total *ataxin-7* expression, once again indicating an insufficient sensitivity, since the expected value was 50%. However, the GG sample (red bar) showed minimal expression (1%), suggesting that the assay could successfully distinguish between the two alleles, and preferentially amplify the A allele but not the G allele.

Figure 19 shows the results obtained from a similar assay, where primers designed to specifically amplify the G allele of rs 377 4729 were used.

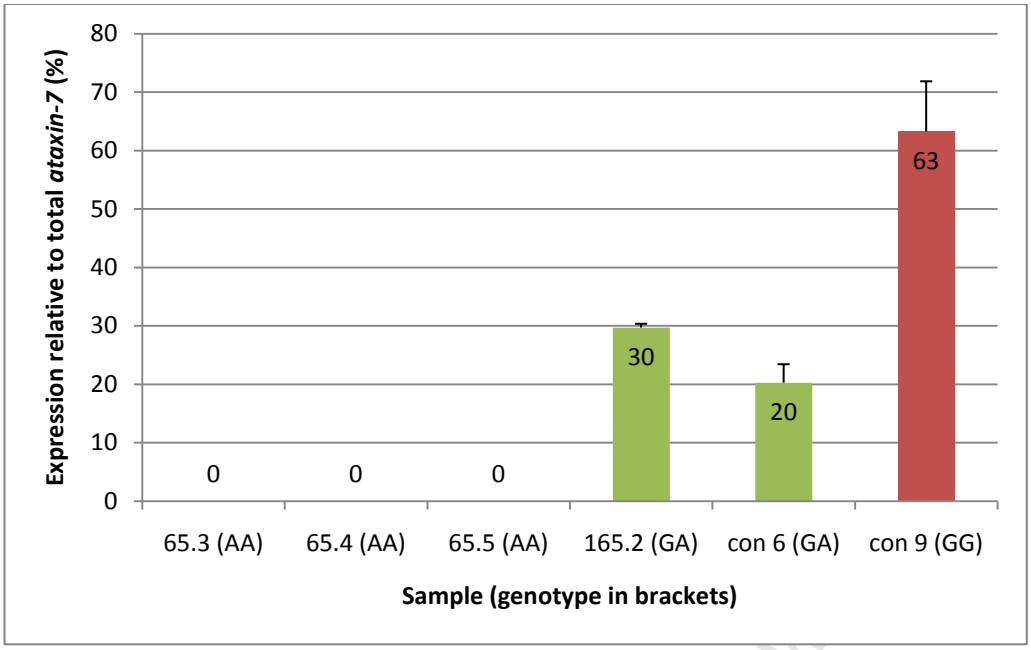


Figure 19: Results of allele-specific qPCR using primers designed for amplification of the G allele of SNP rs 377 4729. Samples show expression of the G allele transcript normalised to *B-actin* and expressed as a percentage of total *ataxin-7* expression (y-axis) in six cell lines (x-axis). Genotypes of cell lines are given in brackets. Individual values are shown within each column. Error bars show SEM.

Results of the G allele-specific assay showed detected expression of the G allele at 63% of the total *ataxin-7* expression in the cell line of GG genotype (red bar), as opposed to the expected value of 100%. Similarly, heterozygous cell lines (green bars) showed expression values of 30% and 20% (the expected value was 50%). Together these results suggested that the G allele-specific assay was not completely sensitive towards changes in levels of the transcript for which it was designed. However, there was no non-specific detection of expression in cell lines of AA genotype (65.3, 65.4 and 65.5), implying that the assay was specific towards the G allele transcript.

Altogether, the results represented in Figure 18 and Figure 19 show that two allele-specific assays have been developed to specifically measure either the A allele or G allele transcript of *ataxin-7* within cell lines of any genotype. However, these assays may not have the accurate sensitivities towards changes in expression levels of the transcripts for which they were designed. Nevertheless, they may still be used in future studies to determine the allele-specific effects of RNAi knock-down of the

mutant *ataxin-7* transcript (A allele transcript). For these experiments, the expression of each allele would be measured in the cell line of choice before RNAi administration. After application of the therapy, expression levels would be monitored to determine whether the A allele was selectively targeted by the RNAi effector (and would thus show reduced expression), with minimal effect on the G allele.

The incorporation of a single mismatch into a PCR primer is the simplest method of allele-specific PCR detection, due to the fact that mismatched 3' termini are extended less efficiently than matched termini during PCR amplification. Various other approaches may be taken to improve the specificities and sensitivities of allele-specific assays. Firstly, it is possible to incorporate additional mismatches one or two bases upstream from the 3' end of the primer, to further destabilize annealing to incorrect templates, and to increase discrimination between alleles (Kwok et al., 1994). An additional potential modification is the incorporation of a locked nucleic acid (LNA) base at the 3' end of the allele-specific primer (Latorra et al., 2003). LNAs are nucleic acid analogues with a 2'-O, 4'-C methylene bridge which locks the ribose ring in a C3'-endo conformation (see Figure 20), resulting in complementary duplexes with RNA or DNA being entropically favoured over mismatched duplexes (Koshkin et al., 1998).

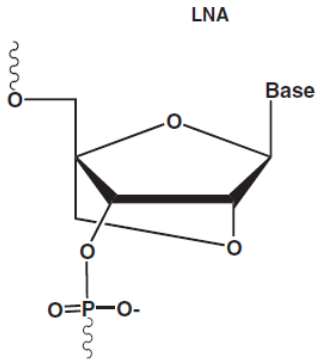


Figure 20: Molecular structure of a locked nucleic acid (LNA) residue. A 2'-O, 4'-C methylene bridge locks the ribose group into the C3'-endo conformation. Figure adapted from Latorra et al. (2003).

A further method of generating an allele-specific PCR assay includes the addition of "blocker" oligonucleotides, along with conventional allele-specific primers (Morlan et al., 2009). The blocker oligonucleotides aid in increasing the specificity of the assay, by binding to the alternative allele whilst the allele-specific primers bind to and amplify the allele of interest. The blockers are modified by phosphorylation at the 3' end, to prevent extension during the PCR process. The principle of this assay, termed Allele-Specific Blocker PCR, is illustrated below in Figure 21.

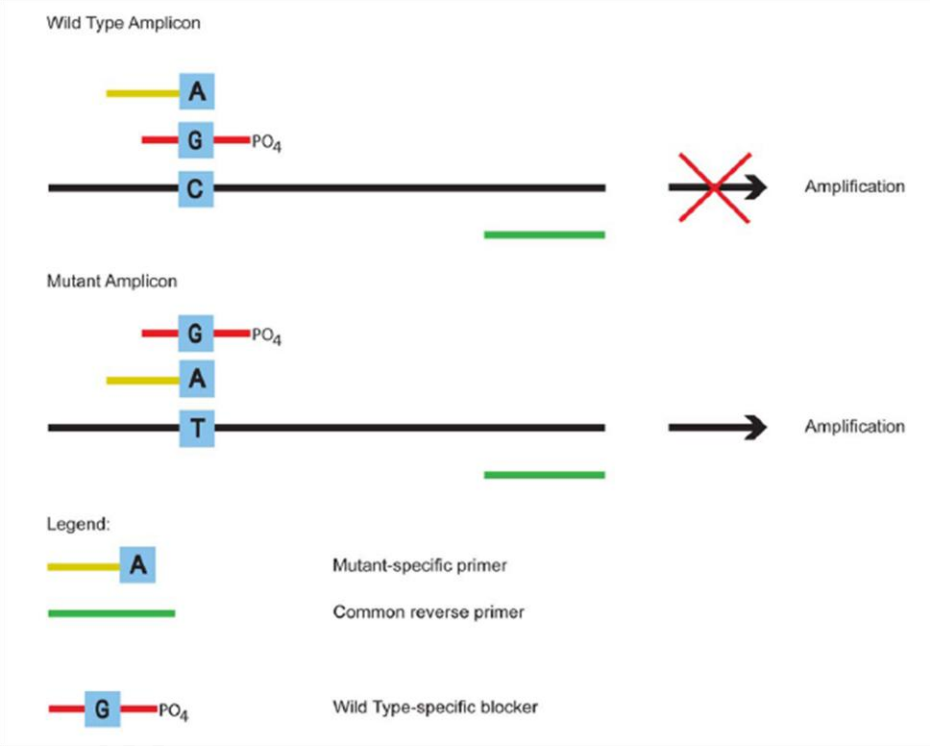


Figure 21: Principle of the Allele-Specific Blocker PCR method. A blocker oligonucleotides containing a phosphorylated 3' end is designed to bind to the alternative allele, preventing amplification (top image), whilst an unmodified oligonucleotides binds to the allele of choice, resulting in selective amplification of a single allele (bottom image). Figure adapted from Morlan et al. (2009).

Apart from modifications of primers, various parameters of the PCR reaction can be altered to improve specificity and sensitivity. Increased annealing temperatures, lower dNTP concentrations, lower primer concentrations and increased $MgCl_2$ concentrations have been said to enhance single nucleotide discrimination (Kwok et al., 1994). Finally, additions to the PCR reaction may aid in specificity. Apyrase, a nucleotide degrading enzyme, allows incorporation of nucleotides to the newly-

synthesized PCR product when the reaction kinetics are fast, but degrades nucleotides in the event of poor reaction kinetics (Ahmadian et al., 2001). Therefore, extension results from matched primers, but not from mismatched primers. Due to time constraints, the various potential improvements to the allele-specific assay could not be tested in this study, although it is likely that these may be beneficial in future projects.

3.6 Section D: Development of a quantitative assay to measure levels of heat shock proteins 27 and 70 in SCA7 patient-derived lymphoblasts.

Numerous studies have explored the relationship between the heat shock proteins and neurodegeneration, due to their endogenous role of protecting the cell during periods of stress. Therefore this group of proteins has been investigated with regard to their role in degeneration and protection, and as potential therapeutic targets. However, little focus has been placed on the role of heat shock proteins in SCA7.

A study by Tsai et al. (2005) found reduced expression of the Hsp27 and 70 proteins between two SCA7 patient lymphoblast cell lines, compared to a single control line. Semi-quantitative RT-PCR confirmed reduced expression of *Hsp70* at the mRNA level, but no differences in *Hsp27* expression. Together, these results suggested that the alteration of Hsp70 protein expression, but not Hsp27 expression, was due to transcriptional defects (Tsai et al., 2005).

3.6.1 Experiment 2

The third objective of this study was to measure the mRNA expression levels of heat shock proteins (Hsps) 27 and 70 in patient- and control-derived lymphoblast cell lines, to validate previous reports suggesting reduced levels of these transcripts in patient-derived lymphoblast cell lines (LCLs). Figure 22 shows the results of the qPCR assay to measure *Hsp27* expression in four patient-derived and three unaffected control-derived LCLs.

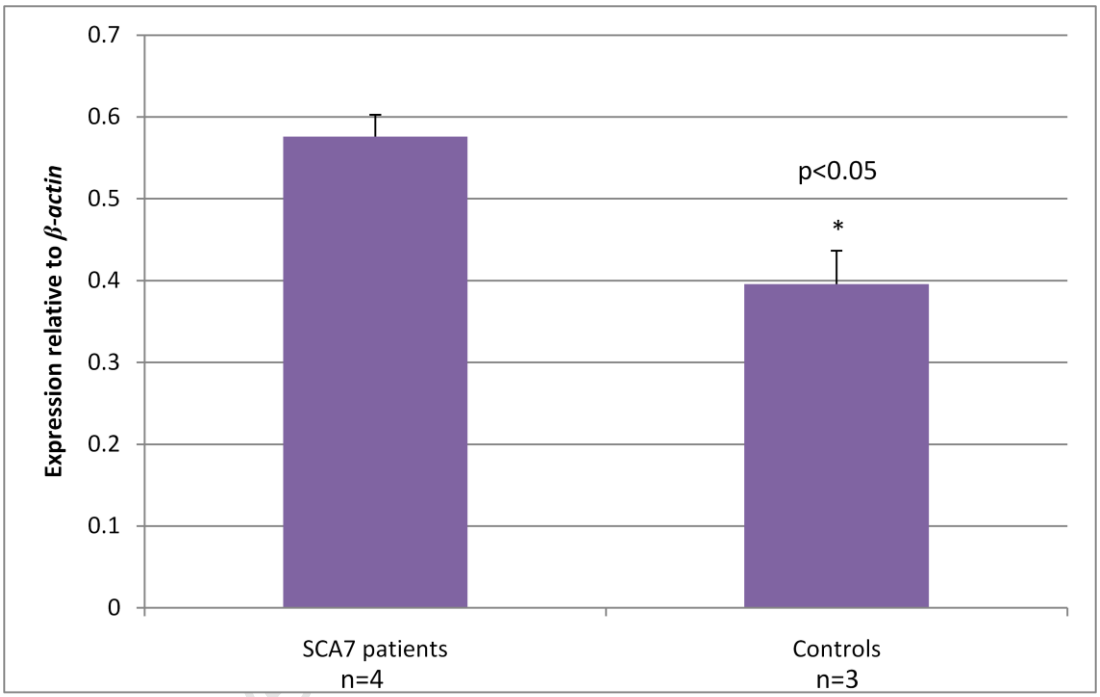


Figure 22: qPCR results showing relative expression of the *Hsp27* transcript in four patient-derived and three control-derived LCLs (x-axis). Expression is shown relative to *β-actin* (y-axis). Error bars show SEM. Statistically significant difference in expression is indicated by the p-value.

The comparison of *Hsp27* expression between patient- and control-derived lines showed a significant increase in expression in cells derived from SCA7 patients ($p<0.05$), contrary to results reported by Tsai et al., who found decreased expression of the Hsp27 protein in patient-derived cells, but no differences in mRNA levels between patients and controls when evaluated using semi-quantitative RT-PCR. It can be argued that due to the limited number of samples

used by Tsai et al. (two patients and a single control), these results may be a more accurate representation.

Hsp27 belongs to the group of small heat shock proteins, and is ubiquitously expressed at low levels in unstressed cells. Expression levels increase within several hours of exposure to stress. Studies have shown that Hsp27 increases the "chaperone capacity" of cells by acting as a substrate holder for Hsp70, and by keeping proteins competent for refolding (Kampinga et al., 2010). Therefore, the increased expression of *Hsp27* in SCA7 patient-derived cells may be due to an innate cellular response to mitigate pathogenesis. Studies have shown that overexpression of Hsp104 and Hsp27 reduces polyglutamine-mediated toxicity in cell-based and mouse models of HD (Perrin et al., 2007). Additionally, overexpression of Hsp27 rescues neuronal inflammation in SCA17 mice (Friedman et al., 2009). It should also be noted that mutations in *Hsp27* are causative in the development of the peripheral neuropathies in the diseases Charcot-Marie- Tooth disease type 2F and distal hereditary motor neuropathy type 2B (Friedman et al., 2009).

Before further conclusions with regard to the exact role of Hsp27 in SCA7 pathogenesis or protection can be made, it would be necessary to evaluate expression levels in a more appropriate model of the disease. Whilst SCA7 patient-derived lymphoblasts are known to express the mutant and wild-type ataxin-7 proteins, the cells are not pathologically affected. It would be beneficial to measure expression levels of *Hsp27* in neuronal cell lines expressing the mutant ataxin-7 protein, or in an appropriate SCA7 mouse model. It may be the case that the increase in *Hsp27* expression evident in lymphoblast lines is not seen in affected cell types, which would further implicate *Hsp27* in a protective role. Furthermore, it would be necessary to determine whether the changes in mRNA expression between SCA7 patients and unaffected controls are observed at the protein level.

The latter part of the objective was to measure expression levels of *Hsp70* between patient- and control-derived LCLs. The results from this experiment are represented by Figure 23.

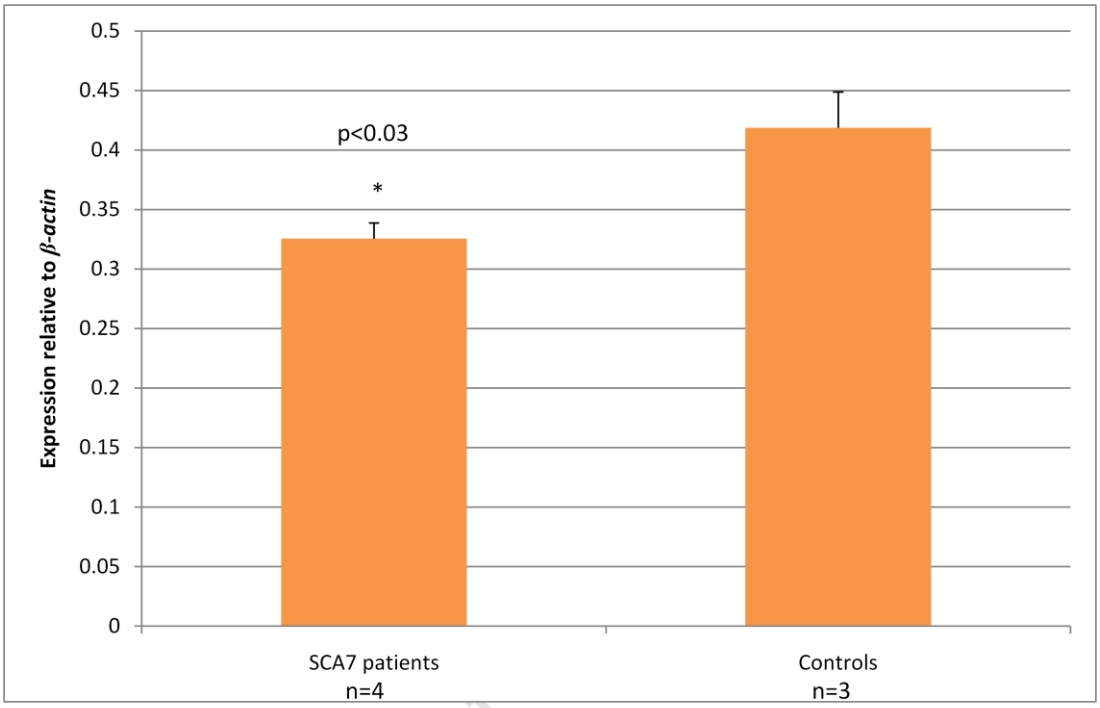


Figure 23: qPCR results showing relative expression of the *Hsp70* transcript in four patient-derived and three control-derived LCLs (y-axis). Expression is shown relative to β-actin (y-axis). Error bars show SEM. Statistically significant difference in expression is indicated by the p-value.

The results from the *Hsp70* expression assay showed a significantly decreased level of *Hsp70* in patient-derived LCLs when compared to control-derived cells ($p<0.03$), in accordance with results published by Tsai et al. Hsp70, a molecular chaperone protein, plays a pivotal role in the cell during periods of stress, and under normal cellular conditions. During normal growth, Hsp70 aids in folding newly synthesized proteins and degrading unwanted proteins. Under stressful conditions, Hsp70 prevents denaturation and degradation of protein substrates (Evans et al., 2010). Since a common phenomenon in the polyglutamine diseases is the accumulation and aggregation of proteins within the cell, Hsp70 and its co-chaperones have been linked to these diseases. Wacker et al. (2009) found that deletion of Hsp70 in a mouse model of HD lead to decreased survival, aggravated neurobehavioural and physical phenotypes, worsened motor deficit and an increased size of inclusion

bodies (Wacker et al., 2009). Furthermore, studies have shown the overexpression of Hsp70 reduces neurodegeneration in a mouse model of SCA1 (Cummings et al., 2001), and a *Drosophila* model of SCA3 (Warrick et al., 1999).

As mentioned previously, the lack of defined SCA7 animal models has hampered studies exploring the relationships between Hsps and disease pathogenesis. A single study by Helmlinger et al. (2004), utilized a transgenic mouse model of SCA7, which presents with severe retinopathy, to investigate the phenotypic effect of overexpressing Hsp70. Whilst overexpression of Hsp70 and its co-chaperone HDJ2 suppressed aggregation of mutant ataxin-7 protein in transfected mammalian cells in vitro, it failed to prevent neuropathy and aggregate formation in the transgenic mouse rod photoreceptors (Helmlinger et al., 2004). Therefore it remains to be shown whether reduced expression of *Hsp70* in SCA7 patient-derived cells could contribute to pathogenesis of the disease. Whilst our data show a significant decrease in *Hsp70* expression between patient- and control-derived cell lines, further studies should aim to evaluate the expression of the Hsp70 protein. Furthermore, it may be beneficial to perform functional studies, to determine whether cells with reduced expression levels are able to mount a stress response comparative to control cells.

In summary, a qPCR-based molecular assay has been developed and used to measure the expression levels of *Hsp27* and *Hsp70* in SCA7 patient- and unaffected control-derived lymphoblast cell lines. The results of these experiments showed significant differences in expression of the two genes between the two groups of cells, providing a potentially useful measurable phenotype associated with the SCA7 disease. In future experiments this assay may be used to measure expression of *Hsp27* and *Hsp70* in patient-derived cells before and after administration of the RNAi-based therapy. Any return of expression to "normal" levels after therapy administration may give an indication of amelioration of the disease-associated phenotype due to RNAi knock-down of the disease-causing *ataxin-7* transcript.

Chapter 4: Conclusions and Future work

Chapter 4: Conclusions and future work

In summary, this study has further characterised the South African SCA7 patient cohort by determining the CAG repeat length for patients for whom genomic DNA was available within the Division of Human Genetics during 2009 and 2010. The SNP rs 377 4729 was genotyped in these patients to establish the percentage of SCA7 patients who may be eligible for a potential RNAi-based therapy. Two quantitative PCR assays were developed and tested in patient- and control-derived transformed lymphoblast cell lines to selectively measure the expression of each allele of the *ataxin-7* SNP. Finally, qPCR was employed to measure differences in expression levels of the heat shock proteins 27 and 70 between patient- and control-derived cells. It is anticipated that these assays will be used in future studies to study the specificity and phenotypic effect of the RNAi effectors, to aid the development of the potential therapy for South African SCA7 patients.

As previously mentioned, current work into the mechanisms of SCA7 pathogenesis and the testing of potential therapies is hampered by the lack of appropriate patient material and a suitable mouse model. Unaffected cell types, such as fibroblasts and lymphoblasts, can be obtained relatively easily from SCA7 patients, however, these cells are likely to give a limited impression of disease pathogenesis compared to affected cell types such as neurons and photoreceptors. However, it may soon be possible to create patient-derived cells of any type, due to the recently developed induced pluripotent stem cell (iPSC) technology. This technique, developed by Takahashi and Yamanaka in 2006, employs the use of a set of defined transcription factors that are able to induce reprogramming of somatic cells (typically fibroblast cells) to a pluripotent stem cell state. These cells can then be subjected to specific conditions to "programme" them to a specific cell type of interest (Takahashi and Yamanaka, 2006). The development of this method has significant implications for diseases such as SCA7, where fibroblasts or lymphoblasts can be obtained from patients through relatively non-invasive methods, and affected cell types (such as neurons in the case of SCA7) can be developed for study. In some cases the iPSCs can be used for therapeutic purposes, in order to

replace diseased cells, or to test potential therapies. Work has already begun on developing iPSC lines from Friedreich Ataxia patient fibroblasts (Liu et al., 2010). Therefore future studies should focus on development and evaluation of this technology.

In the diagnostic setting, further attention should be placed on developing accurate assays for determining the CAG repeat lengths of potential SCA7 patients, to be published and used routinely in laboratories worldwide. In the South African context, future work should focus on determining the extent of the application of the potential RNAi-based therapy, by studying SCA7 patients from different populations, in southern Africa and worldwide. Furthermore, *ataxin-7* should be screened for additional coding SNPs with higher heterozygosity levels in the South African SCA7 population. Finally, the roles of heat shock proteins in disease pathogenesis should be further explored (ideally in cells that exhibit ataxin-7 mediated pathology) to determine whether these proteins may be exploited as therapeutic targets.

The method described in this dissertation has outlined a simple technique for testing allele-specific RNAi-based therapies in patient-derived lymphoblasts. Whilst it can be used to bring researchers a step further towards offering a therapy to patients with SCA7, it may also be applied to current and future studies adopting a similar allele-specific approach to treat other diseases. It is anticipated that this research, along with the continued development of novel technologies, will lead to a promising future for patients with polyglutamine diseases.

Appendices

Appendix A: Basic protocols

DNA purification from whole blood (Gentra Puregene Genomic DNA Purification kit)

900µl of red blood cell lysis solution was dispensed into a 1.5ml microcentrifuge tube, and 300µl of whole blood was added and mixed by inverting 10 times. The tube was incubated at room temperature for 1 minute, and inverted at least once during the incubation. Thereafter the sample was centrifuged for 20 seconds at 13 000 x G to pellet the white blood cells. The supernatant was discarded by pipetting, leaving approximately 10µl of the residual liquid with the pellet. The tube was then vortexed to resuspend the pellet in the residual liquid. 300µl of cell lysis solution was subsequently added, followed by vortexing to lyse the cells. 100µl of protein precipitation solution was added, followed by further vortexing for 20 seconds at high speed. Thereafter the sample was centrifuged for 1 minute at 13 000 x G, to form a pellet containing precipitated contaminating proteins. The supernatant was added to 300µl of isopropanol in a clean 1.5ml tube and mixed by inversion until DNA was visible as threads. The tube was then centrifuged to 1 minute at 13 000 x G and the supernatant discarded. Thereafter 300µl of 70% ethanol was added and the tube was inverted several times to wash the DNA pellet. This step was followed by centrifugation for 1 minute at 13 000 x G whereafter the supernatant was discarded. The pellet was air dried for 5 minutes and resuspended in 100µl of DNA Hydration Solution by vortexing for 5 seconds. Finally, the tube was incubated at 65°C for 5 minutes to dissolve the DNA.

1% (w/v) agarose gel

1g agarose D-1 LE (Hispanger, Spain) was added to 100ml 1x TBE (see below) in a glass flask and heated until agarose had dissolved. When cooled, the agarose solution was poured into a gel electrophoresis tray and allowed to set.

10x Tris borate buffer (TBE)

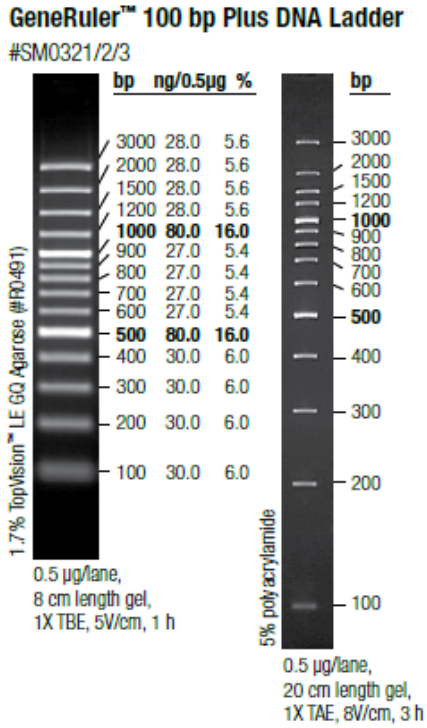
216g Tris (B&M Scientific, South Africa), 100g Boric acid (MP Biomedicals Inc, USA) and 14,8g EDTA (Merck, Germany) was dissolved in 2 litres of distilled water to give final 10x stock concentrations of 890mM Tris, 890mM Boric acid and 20mM EDTA.

Loading dye

A 40% sucrose solution was prepared and 0.25% (v/v) Bromophenol Blue (Merck, Germany) was added.

GeneRuler 100bp DNA Ladder Plus (Fermentas, Canada)

Image obtained from www.fermentas.com



6% Polyacrylamide gel

A 50ml total volume solution was prepared from 6.7 parts 40% acrylogel 5 solution (VWR International Ltd, England), 1x TBE and distilled water. 200ul of 10% APS (Merck, Germany) and 20ul Tetramethylethylenediamine (TEMED) (BDH Laboratory Supplies, England) was added to 16ml of the gel mix and poured between glass plates to set.

3µl of PCR product was added to 3µl of loading dye and subjected to electrophoresis of 155V for 45 minutes.

Silver nitrate staining:

A 0.1% Silver Nitrate solution (Merck, Germany) and 0.15% Formaldehyde & 1.5% (w/v) NaOH (Merck, Germany) solution were prepared. The polyacrylamide gel was rinsed in distilled water and submerged in the silver nitrate solution for 15 minutes. Thereafter the gel was rinsed in distilled water and submerged in the formaldehyde solution for 10 minutes. After further rinsing with distilled water the gel was placed on blotting paper and allowed to dry for visualisation.

Appendix B: Primer sequences

Gene	Primer Name	Length	Melting Temperature (°C)	Sequence (5' to 3')	Fragment size (base pairs)
Ataxin-7 (CAG repeat)	Atxn7 (CAG)n F	21	50	Hex-ATTGTAGGAGCGGAAAGAATG	262 excl. repeat region
	Atxn7 (CAG)n R	18	50	CCAGCATCACTTCAGGAC	
Ataxin-7 (SNP rs 377 4729)	SCA7 SNP F	20	59	AATGAACTGCCTGTCAACTC	248
	SCA7 SNP R	20	59	GCTCACAGTCCATTTCCTAC	
Ataxin-7	Atxn qPCR F	19	66.8	GCCAGCCGTGAACAATGTC	
	Atxn qPCR R	21	67.4	TTCCTCCCCGTGCTATTTTCA	
Hsp27	Hsp27 F	19	63.2	ACGAGCTGACGGTCAAGAC	
	Hsp27 R	19	68.6	GGGGGCAGCGTGTATTTCC	
Hsp70	Hsp70 F	24	60.4	ATGGAATCTATAAGCAGGATCTTC	
	Hsp70 R	24	59.4	CACATACAGAAACTTGATAAGCAG	

Appendix C: PCR and qPCR reaction components and cycling conditions

Table 13: CAG repeat PCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	4.08	
FailSafe buffer J (EPICENTRE) (2X)	5	1x
Forward primer (20µM)	0.2	0.4µM
Reverse primer (20µM)	0.2	0.4µM
DNA (100ng/µl)	0.4	4ng/µl
GoTaq DNA polymerase(Promega) (5U/µl)	0.12	0.06U/µl
Total	**10**	

Table 14: Cycling conditions for CAG repeat PCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	5 minutes	1
2	95	30 seconds	
	53	6 seconds	30
	72	40 seconds	
3	72	7 minutes	1

Table 15: SNP rs 377 4729 PCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	16.9	
GoTaq Buffer (Promega) (5x)	5	1x
dNTPs (Bioline, UK)	1	
Forward primer (20µM)	0.5	0.4µM
Reverse primer (20µM)	0.5	0.4µM
DNA (100ng/µl)	1	4ng/µl
GoTaq DNA polymerase(Promega) (5U/µl)	0.1	0.02U/µl
Total	**25**	

Table 16: Cycling conditions for SNP rs 377 4729 PCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	5 minutes	1
2	95	30 seconds	
	61	30 seconds	30
	72	40 seconds	
3	72	7 minutes	1

The 248bp PCR products were resolved on a 2% agarose gel as described previously.

Table 17: *Ataxin*-7 qPCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	3.6	
SensiMix SYBR (2x)	5	1x
Forward primer (10µM)	0.2	0.2µM
Reverse primer (10µM)	0.2	0.2µM
cDNA (15ng/µl)	1	2ng/µl
Total	**10**	

Table 18: Cycling conditions for *Ataxin-7* qPCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	10 minutes	1
2	95	15 seconds	
	60	30 seconds	35
	72	15 seconds	

Table 19: A allele-specific PCR reaction components (for primer screening). Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	9.5	
SensiMix SYBR (2x)	12.5	1x
Forward primer (10µM)	1	0.4µM
Reverse primer (10µM)	1	0.4µM
cDNA (100ng/µl)	1	4ng/µl
Total	**25**	

Table 20: Cycling conditions for A allele-specific PCR (primer screening)

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	10 minutes	1
2	95	15 seconds	
	61	30 seconds	35
	72	40 seconds	

Table 21: Optimized A allele-specific qPCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	3.4	
SensiMix SYBR (2x)	5	1x
Primer F2 (10µM)	0.3	0.3µM
Primer R1 (10µM)	0.3	0.3µM
cDNA (15ng/µl)	1	2ng/µl
Total	**10**	

Table 22: Optimized cycling conditions for A allele-specific qPCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	10 minutes	1
2	95	15 seconds	
	60	60 seconds	35

Table 23: Optimized G allele-specific qPCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	3.4	
SensiMix SYBR (2x)	5	1x
Primer GF1 (10µM)	0.3	0.3µM
Primer R1 (10µM)	0.3	0.3µM
cDNA (15ng/µl)	1	2ng/µl
Total	**10**	

Table 24: Optimized cycling conditions for G allele-specific qPCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	10 minutes	1
2	95	15 seconds	
	61	30 seconds	35
	72	40 seconds	

Table 25: *β-actin* qPCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	3.4	
SensiMix SYBR (2x)	5	1x
Primer mix (10µM)	0.2	0.2µM
cDNA (15ng/µl)	1	2ng/µl
Total	**10**	

Table 26: Cycling conditions for *β-actin* qPCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	10 minutes	1
2	95	15 seconds	
	61	30 seconds	35
	72	15 seconds	

Table 27: *Hsp27* and *Hsp70* qPCR reaction components. Stock solution concentrations are in brackets.

Reagent	Volume (µl)	Final Concentration
dH_20	3.6	
SensiMix SYBR (2x)	5	1x
Forward primer (10µM)	0.2	0.2µM
Reverse primer (10µM)	0.2	0.2µM
cDNA (15ng/µl)	1	2ng/µl
Total	**10**	

Table 28: Cycling conditions for *Hsp27* and *Hsp70* qPCR

Cycle number	Temperature (°C)	Time	Number of cycles
1	95	10 minutes	1
2	95	15 seconds	
	60	30 seconds	35
	72	15 seconds	

Electronic resources

- Ensembl (http://www.ensembl.org/)
- IDT OligoAnalyzer web-based tool (http://eu.idtdna.com/analyzer/ Applications/ OligoAnalyzer/Default.aspx)
- National Center for Biology Information (NCBI, http://www.ncbi.nlm.nih.gov/)
- Oligo Calculator programme (http://www.cnr .berkeley.edu/~zimmer/oligo TMcalc.html)
- US National Institute of Health registry of clinical trials (www.clinicaltrials.gov, accessed 07/02/2011)

References

Aagaard L, Rossi JJ. 2007. RNAi therapeutics: Principles, prospects and challenges. Adv Drug Deliv Rev 59:75-86.

Ahmadian A, Gharizadeh B, O'Meara D, Odeberg J, Lundeberg J. 2001. Genotyping by apyrase-mediated allele-specific extension. Nucleic Acids Res 29:e121.

Alves S, Nascimento-Ferreira I, Auregan G, Hassig R, Dufour N, Brouillet E, de Lima MCP, Hantraye P, de Almeida LP, Déglon N. 2008. Allele-specific RNA silencing of mutant ataxin-3 mediates neuroprotection in a rat model of Machado-Joseph disease. PLoS One 3:e3341.

Alves S, Nascimento-Ferreira I, Dufour N, Hassig R, Auregan G, Nobrega C, Brouillet E, Hantraye P, Pedroso de Lima MC, Deglon N. 2010. Silencing ataxin-3 mitigates degeneration in a rat model of Machado-Joseph disease: No role for wild-type ataxin-3? Hum Mol Genet 19:2380-2394.

Amarzguioui M, Rossi JJ, Kim D. 2005. Approaches for chemically synthesized siRNA and vector-mediated RNAi. FEBS Lett 579:5974-5981.

Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature 431:805-810.

Bagasra O, Prilliman KR. 2004. RNA interference: The molecular immune system. J Mol Histol 35:545-553.

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. 2003. Dicer is essential for mouse development. Nat Genet 35:215-217.

Bjork A, Lindblom U, Wadensten L. 1956. Retinal degeneration in hereditary ataxia. Br Med J 19:186-193.

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 17:1053-1063.

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

Castanotto D, Rossi JJ. 2009. The promises and pitfalls of RNA-interference-based therapeutics. Nature 457:426-433.

Cleary JD, Pearson CE. 2005. Replication fork dynamics and dynamic mutations: The fork-shift model of repeat instability. Trends Genet 21:272-280.

Cummings CJ, Mancini MA, Antalffy B, DeFranco DB, Orr HT, Zoghbi HY. 1998. Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. Nat Genet 19:148-154.

Cummings CJ, Sun Y, Opal P, Antalffy B, Mestril R, Orr HT, Dillmann WH, Zoghbi HY. 2001. Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice. Hum Mol Genet 10:1511-1518.

David G, Giunti P, Abbas N, Coullin P, Stevanin G, Horta W, Gemmill R, Weissenbach J, Wood N, Cunha S. 1996. The gene for autosomal dominant cerebellar ataxia type II is located in a 5-cM region in 3p12-p13: Genetic and physical mapping of the SCA7 locus. Am J Hum Genet 59:1328-1336.

David G, Abbas N, Stevanin G, Duerr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou E. 1997. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat Genet 17:65-70.

David G, Dürr A, Stevanin G, Cancel G, Abbas N, Benomar A, Belal S, Lebre AS, Abada-Bendib M, Grid D. 1998. Molecular and clinical correlations in autosomal

dominant cerebellar ataxia with progressive macular dystrophy (SCA7). Hum Mol Genet 7:165-170.

Dorschner MO, Barden D, Stephens K. 2002. Diagnosis of five spinocerebellar ataxia disorders by multiplex amplification and capillary electrophoresis. J Mol Diagn 4:108-113.

Duenas AM, Goold R, Giunti P. 2006. Molecular pathogenesis of spinocerebellar ataxias. Brain 129:1357-1370.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494-498.

Evans CG, Chang L, Gestwicki JE. 2010. Heat shock protein 70 (Hsp70) as an emerging drug target. J Med Chem 53:4585-4602.

Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, Leake D, Marshall WS, Khvorova A. 2006. Off-target effects by siRNA can induce toxic phenotype. RNA 12:1188-1196.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. Nature 391:806-810.

Foxon A. 2010. A haplotype study of South African SCA7 patients. Honour's project write-up, University of Cape Town.

Freshney RI, Freshney I. 1994. Culture of animal cells: A manual of basic technique. Wiley-Liss New York.

Friedman MJ, Li S, Li XJ. 2009. Activation of gene transcription by heat shock protein 27 may contribute to its neuronal protection. J Biol Chem 284:27944-27951.

Froment J, Bonnet P, Colrat A. 1937. Heredo-degenerations retiniennes et spino-cerebelleuses. Variantes ophtalmoscopiques et neurologiques par trois generations successives. Med Lyon 18:153-163. (Abstract in English)

Gantier MP, Williams BRG. 2007. The response of mammalian cells to double-stranded RNA. Cytokine Growth Factor Rev 18:363-371.

Garden GA, La Spada AR. 2007. Molecular pathogenesis and cellular pathology of spinocerebellar ataxia type 7 neurodegeneration. Cerebellum 99999:1-12.

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 70:415-417.

Grishok A, Sinskey JL, Sharp PA. 2005. Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. Genes Dev 19:683-696.

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser 41:95-98.

Harding AE. 1993. Clinical features and classification of inherited ataxias. Adv Neurol 61:1-14.

Hay DG, Sathasivam K, Tobaben S, Stahl B, Marber M, Mestril R, Mahal A, Smith DL, Woodman B, Bates GP. 2004. Progressive decrease in chaperone protein levels in a mouse model of Huntington's disease and induction of stress proteins as a therapeutic approach. Hum Mol Genet 13:1389-1405.

Helmlinger D, Bonnet J, Mandel JL, Trottier Y, Devys D. 2004. Hsp70 and Hsp40 chaperones do not modulate retinal phenotype in SCA7 mice. J Biol Chem 279:55969-55977.

Higuchi R, Fockler C, Dollinger G, Watson R. 1993. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. Nat Biotechnol 11:1026-1030.

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 27:478-484.

Huen NYM, Wong SLA, Chan HYE. 2007. Transcriptional malfunctioning of heat shock protein gene expression in spinocerebellar ataxias. Cerebellum 6:111-117.

Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. 2003. Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21:635-637.

Jakob U, Gaestel M, Engel K, Buchner J. 1993. Small heat shock proteins are molecular chaperones. J Biol Chem 268:1517-1520.

Jolly C, Morimoto RI. 2000. Role of the heat shock response and molecular chaperones in oncogenesis and cell death. J Natl Cancer Inst 92:1564-1572.

Jonasson J, Juvonen V, Sistonen P, Ignatius J, Johansson D, Bjorck EJ, Wahlstrom J, Melberg A, Holmgren G, Forsgren L, Holmberg M. 2000. Evidence for a common spinocerebellar ataxia type 7 (SCA7) founder mutation in Scandinavia. Eur J Hum Genet 8:918-922.

Kahle JJ, Gulbahce N, Shaw CA, Lim J, Hill DE, Barabási AL, Zoghbi HY. 2011. Comparison of an expanded ataxia interactome with patient medical records reveals a relationship between macular degeneration and ataxia. Hum Mol Genet 20:510-527.

Kampinga H, Bryantsev A, Kurchashova S, Golyshev S, Polyakov V, Wunderink H, Kanon B, Budagova K, Kabakov A. 2010. Regulation of stress-induced intracellular sorting and chaperone function of Hsp27 (HspB1) in mammalian cells. Biochem J 407:407-411.

Katsuno M, Sang C, Adachi H, Minamiyama M, Waza M, Tanaka F, Doyu M, Sobue G. 2005. Pharmacological induction of heat-shock proteins alleviates polyglutamine-mediated motor neuron disease. Proc Natl Acad Sci USA 102:16801-16806.

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I. 1994. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32. 1. Nat Genet 8:221-228.

Kim DH, Rossi JJ. 2007. Strategies for silencing human disease using RNA interference. Nat Rev Genet 8:173-184.

Kim D, Rossi J. 2008. RNAi mechanisms and applications. BioTechniques 44:613-616.

Koshkin AA, Singh SK, Nielsen P, Rajwanshi VK, Kumar R, Meldgaard M, Olsen CE, Wengel J. 1998. LNA (locked nucleic acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. Tetrahedron 54:3607-3630.

Kwok S, Chang SY, Sninsky JJ, Wang A. 1994. A guide to the design and use of mismatched and degenerate primers. Genome Res 3:S39-S47.

Latorra D, Campbell K, Wolter A, Hurley JM. 2003. Enhanced allele-specific PCR discrimination in SNP genotyping using 3'locked nucleic acid (LNA) primers. Hum Mutat 22:79-85.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S. 2003. The nuclear RNaseIII Drosha initiates microRNA processing. Nature 425:415-419.

Libby RT, Monckton DG, Fu YH, Martinez RA, McAbney JP, Lau R, Einum DD, Nichol K, Ware CB, Ptacek LJ. 2003. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. Hum Mol Genet 12:41-50.

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

Lindenberg KS, Yvert G, Müller K, Landwehrmeyer GB. 2000. Expression analysis of *ataxin-7* mRNA and protein in human brain: Evidence for a widespread distribution and focal protein accumulation. Brain Pathol 10:385-394.

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. Sci STKE 305:1437-1441.

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R. 2005. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol 7:1216-1266.

Liu J, Verma PJ, Evans-Galea MV, Delatycki MB, Michalska A, Leung J, Crombie D, Sarsero JP, Williamson R, Dottori M. 2010. Generation of induced pluripotent stem cell lines from Friedreich ataxia patients. Stem Cell Rev 1-11.

Livak KJ. 1999. Allelic discrimination using fluorogenic probes and the 5'nuclease assay. Genet Anal: Biomol Eng 14:143-149.

Mamedov T, Pienaar E, Whitney S, TerMaat J, Carvill G, Goliath R, Subramanian A, Viljoen H. 2008. A fundamental study of the PCR amplification of GC-rich DNA templates. Comput Biol Chem 32:452-457.

Martinez J, Patkaniowska A, Urlaub H, Lührmann R, Tuschl T. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110:563-574.

Matilla-Dueñas A, Sánchez I, Corral-Juan M, Dávalos A, Alvarez R, Latorre P. 2010. Cellular and molecular pathways triggering neurodegeneration in the spinocerebellar ataxias. Cerebellum 9:148-166.

Merienne K, Helmlinger D, Perkin GR, Devys D, Trottier Y. 2003. Polyglutamine expansion induces a protein-damaging stress connecting heat shock protein 70 to the JNK pathway. J Biol Chem 278:16957-16967.

Michalik A, Martin J, Van Broeckhoven C. 2004. Spinocerebellar ataxia type 7 associated with pigmentary retinal dystrophy. Eur J Hum Genet 12:2-15.

Morlan J, Baker J, Sinicropi D. 2009. Mutation detection by real-time PCR: A simple, robust and highly selective method. PLoS One 4:e4584.

Mullis, K. and Faloona, F. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155: 335-350.

Nakamori M, Pearson CE, Thornton CA. 2011. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats. Hum Mol Genet 20:580-588.

Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, Borowski A, Marth JD, Phillips AG, Hayden MR. 1995. Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. Cell 81:811-823.

Orr HT, Chung M, Banfi S, Kwiatkowski Jr TJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LPW, Zoghbi HY. 1993. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4:221-226.

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

Palhan VB, Chen S, Peng GH, Tjernberg A, Gamper AM, Fan Y, Chait BT, La Spada AR, Roeder RG. 2005. Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. Proc Natl Acad Sci USA 102:8472-8477.

Perrin V, Régulier E, Abbas-Terki T, Hassig R, Brouillet E, Aebischer P, Luthi-Carter R, Déglon N. 2007. Neuroprotection by Hsp104 and Hsp27 in lentiviral-based rat models of Huntington's disease. Mol Ther 15:903-911.

Pfister EL, Kennington L, Straubhaar J, Wagh S, Liu W, DiFiglia M, Landwehrmeyer B, Vonsattel JP, Zamore PD, Aronin N. 2009. Five siRNAs targeting three SNPs may provide therapy for three-quarters of Huntington's disease patients. Curr Biol 19:774-778.

Pomp D, Medrano J. 1991. Organic solvents as facilitators of polymerase chain reaction. BioTechniques 10:58-59.

Ross CA, Poirier MA. 2004. Protein aggregation and neurodegenerative disease. Nat Med 10:S10-S17.

Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K, Erlich H. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463-5467.

Scholefield JA. 2008. RNAi based allele-specific silencing of the disease-causing gene in Black South African patients with SCA7. PhD thesis, University of Cape Town.

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

Schöls L, Bauer P, Schmidt T, Schulte T, Riess O. 2004. Autosomal dominant cerebellar ataxias: Clinical features, genetics, and pathogenesis. Lancet Neurol 3:291-304.

Sequeiros J, Martindale J, Seneca S. 2010. EMQN best practice guidelines for molecular genetic testing of SCAs. Eur J Hum Genet 18:1188-1195.

Stevanin G, Giunti P, Belal G, Durr A, uberg M, Wood N, Brice A. 1998. De novo expansion of intermediate alleles in spinocerebellar ataxia 7. Hum Mol Genet 7:1809-1813.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663-676.

Tsai HF, Lin SJ, Li C, Hsieh M. 2005. Decreased expression of Hsp27 and Hsp70 in transformed lymphoblastoid cells from patients with spinocerebellar ataxia type 7. Biochem Biophys Res Commun 334:1279-1286.

Van de Warrenburg B, Sinke R, Verschuuren-Bemelmans C, Scheffer H, Brunt E, Ippel P, Maat-Kievit J, Dooijes D, Notermans N, Lindhout D. 2002. Spinocerebellar ataxias in the Netherlands prevalence and age at onset variance analysis. Neurology 58:702-708.

Wacker JL, Huang SY, Steele AD, Aron R, Lotz GP, Nguyen QV, Giorgini F, Roberson ED, Lindquist S, Masliah E. 2009. Loss of Hsp70 exacerbates pathogenesis but not levels of fibrillar aggregates in a mouse model of Huntington's disease. J Neurosci 29:9104-9144.

Warrick JM, Chan HYE, Gray-Board GL, Chai Y, Paulson HL, Bonini NM. 1999. Suppression of polyglutamine-mediated neurodegeneration in drosophila by the molecular chaperone HSP70. Nat Genet 23:425-428.

Williams AJ, Paulson HL. 2008. Polyglutamine neurodegeneration: Protein misfolding revisited. Trends Neurosci 31:521-528.

Yamada M, Sato T, Tsuji S, Takahashi H. 2008. CAG repeat disorder models and human neuropathology: Similarities and differences. Acta Neuropathol 115:71-86.

Yi R, Qin Y, Macara IG, Cullen BR. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 17:3011-3016.

Zander C, Takahashi J, El Hachimi K, Fujigasaki H, Albanese V, Lebre A, Stevanin G, Duyckaerts C, Brice A. 2001. Similarities between spinocerebellar ataxia type 7 (SCA7) cell models and human brain: Proteins recruited in inclusions and activation of caspase-3. Hum Mol Genet 10:2569-2579.

Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. 2004. Single processing center models for human dicer and bacterial RNase III. Cell 118:57-68.

Zhao Y, Tan E, Law H, Yoon C, Wong M, Ng I. 2002. Prevalence and ethnic differences of autosomal-dominant cerebellar ataxia in Singapore. Clin Genet 62:478-481.