3.1 INTRODUCTION

Influences of CAG repeat length on the age of onset has been a well-studied feature in most of the triplet repeats disorders such as Huntington’s disease (HD) and SCAs. Almost immediately after the discovery of the HD gene in the year 1993, studies were conducted associating the length of the trinucleotide expansion with the mean age of clinical onset (Duyao et al. 1993). A recent study on 4,078 HD patients showed that the HD pathogenesis is determined by a completely dominant action of the longest expanded allele and also by as yet unidentified genetic or environment factors (J.-M. Lee et al. 2012).

Although, inverse correlation of the age of clinical onset (AO) and the length of the CAG repeat in SCAs such as SCA1 (Ranum et al. 1994; Orr et al. 1993; Jodice et al. 1994; Dubourg et al. 1995; Chung et al. 1993), SCA2 (Hayes et al. 2000), SCA3 (Schöls et al. 1995) and SCA6 (Matsuyama et al. 1997) is known, the length of the CAG repeat explains only 50-80% of the age of onset variance (Matsuyama et al. 1997) suggesting that there are genetic or non-genetic factors other than trinucleotide repeat length that influence phenotypic variation within each SCA mutation (Bürk et al. 1996). The most striking non-genetic component hypothesized to contribute to the phenotypic variation is the environment. Therefore, in this study conducted in a homogenous community with shared environmental features, it was pertinent to understand the relationship of age of onset (AO) of the symptoms with the CAG repeat.

Germ-line instabilities in the repeat tract are known to further enhance the expansions in the successive generations and subsequently result in earlier disease onset with increased severity, a phenomenon called anticipation (Zoghbi et al. 1988; SCHUT JW 1950). Anticipation is common in other neurological disease such as HD (Ranen et al. 1995) , fragile X (Fu et al. 1991) and myotonic dystrophy (Mahadevan et al. 1992); SCA2 (Orozco Diaz et al. 1990) and SCA3 (Coutinho and Andrade 1978). In SCA1, anticipation is associated with expansions inherited from the paternal mutant allele (Chung et al. 1993; Jodice et al. 1994). This was understood to be due to strand slippage during spermatogenesis resulting in further expansion of the mutant allele and anticipation (Pearson 2003). In view of contradicting studies showing no influence of parental sex on the mutation (YOhsiro SUZUKI, 1995), we explored the influence of parental gender on the age of onset in the Adukkamparai cohort.
To fully understand the spectrum of the disease, longitudinal studies have been conducted in SCA1. The first of its kind was carried out in South Africa (Bryer et al. 2003) over a 10 year period in 47 SCA1 patients clearly charting out the evolution of clinical phenotypes observed during progression of mild to severe form of the disease. Subsequently, a 2 year study was conducted among the European population (Jacobi et al. 2011) with data collected at year 1 and 2 focusing on the influence of gender, repeat length – both normal and expanded, AO and disease duration on the disease progression. It was concluded that the progression is faster in earlier AO and patients with larger expansions. A recent two-year study conducted with data collected every 6 months in patients from the 12 participating centers of the US estimated the rate of progression of ataxia (SCA1, 2, 3 and 6) and have shown that among all the SCAs, SCA1 progressed the fastest similar to the European population (Jacobi et al) and hence progression rates are comparable between the US and Europe. Although significant inverse correlation was seen with the AO and CAG repeat, no correlation was seen with regard to progression of the disease indicating other confounding factors to affect the pathogenesis after its onset (Ashizawa et al. 2013).

SCA1 occurs in diverse ethnic groups worldwide (van de Warrenburg et al. 2005; Zhou et al. 2001; Wakisaka et al. 1995; Bryer et al. 2003; Jodice et al. 1993; Goldfarb et al. 1996; Illarioshkin et al. 1996). Varying clinical symptoms have been observed among the SCA1 patients worldwide (Goldfarb et al. 1996; Bryer et al. 2003; Rüb et al. 2013; Zhou et al. 2001). This variability has been attributed to ethnic diversity and environmental effects (van de Warrenburg et al. 2005). In order to understand the origin of the mutation and examine the possible existence of a common ancestral gene, studies have looked at haplotype markers along the SCA1 region in different populations (Ramesar et al. 1997; Wakisaka et al. 1995). In the South African cohort of mixed origins, 2 different founders were identified (Ramesar et al. 1997). Studies exploring common founders in Indian population have been carried out in SCA1 (Mittal et al. 2005), SCA2 (Sinha et al. 2004) and SCA12 (Bahl et al. 2005). In SCA1, among the 12 genetic markers studied three genetic markers - 2 SNPs (rs1476464 and rs2075974) and a microsatellite region D6S288, showed an association with expanded alleles by haplotype analysis. Patients from both Southern and Northern India showed a haplotype of C-4-C corresponding to the presence of allele C/G at rs1476464, a dinucleotide repeat of 25 at D6S288, and allele C/G at rs2075974, suggesting a founder effect for SCA1 in the Indian population (Mittal,
Sharma et al. 2005). Studies confirming the published data from independent isolated ethnic communities have not been conducted in India.

Therefore two identified SNP’s were studied in this cohort - rs2075974 (SNP1) and rs1476464 (SNP2). The position of the two SNPs is shown in Figure 3.1

![Figure 3.1 A Schematic diagram showing the position of the SNPs within the ATXN1 gene.](image)

rs2075974 is a single nucleotide variant mutation of A/G in the reverse strand at position 16327099 on the chromosome 6, resulting in a nucleotide change of GA”A” to GA”G“. It constitutes a synonymous mutation of glutamine in the protein. rs1476464 is located in the intronic region at chromosome position 16392057 and is a G/T transition.

3.2 MATERIALS AND METHOD

3.2.1 Patients

Data from longitudinal study of Adukkamparai cohort during the years 2009-2013 is shown in this chapter. The rate of progression of the disease was determined as follows. Rate 1 denotes progression from presymptomatic to mild, mild to moderate or moderate to severe; and Rate 2 denotes pre-symptomatic to moderate or mild to severe.

3.2.2 SNP analysis

Analysis of two SNPs, rs2075974 and rs1476464 in the ATXN1 gene was performed by the Taqman allelic discrimination assay obtained from Applied Biosystems, Life technologies. Approximately 20-50 ng of genomic DNA was mixed with 12.5 µL of SNP genotyping master mix and 1.25 µl of 20X primer probe mix (Life technologies pre-designed SNP genotyping assay, part number C__16167072_10 for rs2075974 and part number C__7615001_10 for rs1476464) and run on the Applied Biosystems 7500 fast real time PCR system. The cycling temperatures were as follows: denaturation at 95°C for 10 minutes followed by 40 cycles of 95°C for 15 sec and 60°C for 1 minute. A particular SNP is identified by the corresponding amplification peak and is confirmed by the allelic calls generated by
the 7500-software v2.0.5. A figure showing an overview of the process is shown below (Figure 3.2).

**Figure 3.2** Representation of the TAQMAN assay (A) homozygous for allele 1 would give a VIC fluorescence (B) homozygous for allele 2 would give FAM fluorescence. Heterozygous will give both the fluorescence as seen by the green dots in the Allelic discrimination plot (C).

### 3.2.3 Statistical Analysis

Pearson’s correlation coefficient was used to test linear association between repeat length and age of onset. The relationship between age of onset and inheritance with CAG repeats was calculated by Kruskal-Wallis ANOVA test with post-hoc Dunnett’s test. Chi-square test and one-way ANOVA was used to test significance of data. Significance was calculated using the GraphPad Prism 6.0 software and results were considered significant at P=0.05 level.
3.3 RESULTS

3.3.1 Correlation between age of onset and CAG repeat

Due to apparent heterogeneity in the age of onset among the genetically identified SCA1 cases in the cohort, a comparison of the age of onset of ataxia (AO) with the CAG repeat number was carried out to understand their correlation if any as shown in previous other studies (Ranum et al. 1994). It was found that the AO in the symptomatic individuals (n=16) correlated negatively with greater repeat numbers (r=-0.67, P=0.002), indicating that 45% (R²=0.45) of the variation in age of onset is accounted by the size of the repeat (Figure 3.3). When the distribution of repeat numbers with inheritance was analyzed, it was observed that the maternal and paternally transmitted alleles harbored repeat sizes in the range of 40-45 (shown as circles for maternally transmitted and squares for paternally transmitted alleles in Figure 3.3). Individuals with both parents affected (triangles in Figure 3.3) had longer repeats with the exception of one.

3.3.2 Correlation of inheritance and age of onset

To compare the AO with inheritance more closely, patients were classified based on inheritance of the mutant gene, into three groups as follows: Group 1 (n=6), with a maternally inherited mutated allele; Group 2 (n=5), with a paternally inherited mutated allele and Group 3 (n=5), in which both the parents were affected. The mean age of onset of disease in group 1 was 52.5±4.18 years, whereas in group 2 it was 40.2±4.81 years and in group 3 it was 38.6±5.17 (Figure 3.4). A significant difference was observed in the ages of onset between the groups 1, 2 and 3 by Kruskal-Wallis ANOVA test.
(P=0.0004). As significant results were observed group 1 was compared with group 2 and 3 by ordinary ANOVA with post hoc Dunnett’s test. Group 1 was found to be significantly different from group 2 (P=0.0015) and group 3 (P=0.005), whereas no significant difference was observed between group 2 and 3 (P=0.8). Thus, patients with maternally inherited alleles in the cohort show a later age of onset when compared to those with paternally inherited alleles. When both parents were affected, it appears that the age of onset was determined by the paternal allele. Similar difference in the age of onset between paternal and maternal inheritance of the CAG expanded chromosome has been observed in Huntington’s disease (Myers et al. 1983; Krawczak et al. 1991). To the best of my knowledge a similar finding has not been described for SCA1 to date. This is an interesting finding and could serve as valuable information for clinicians towards early detection of SCA1 in patients with affected fathers.

3.3.3 Longitudinal study of SCA1 patients

Now that we have seen that the AO of the disease is dependent on inheritance of the mutant allele and to some extent (45%) on the size of CAG repeats, I proceeded to understand if CAG repeat size played any role in progression of the disease. This was an interesting journey as many valuable and demarcating features of the disease emerged in a family specific manner.

3.3.3.1 Characteristic clinical signs among the SCA1 postive families:

Analysis of the clinical details showed that specific characteristics were present in individual affected families for which the details are given below and tabulated in Table 3.1.

Family O: This was the largest family with 3 generations and a total of 21 individuals. There were 2 severely affected individuals (O19, O8) and among them 1 (O8) expired due to end stage disease after our first visit. One mildly affected (O3) and 1 moderately affected individual (O11) showed progression to severe disease. A younger sibling (O3) progressed faster than his other siblings (O11 and O5). Both their parents mother O1 and father (expired) were affected. O3 developed symptoms at the age of 30 years and rapidly worsened to severe disease within 5 years. In this family 3 members had an unrelated disorder. They had bleeding manifestations and on evaluation were found to have von-Willebrand’s disease type 1. Wide eyed stare in the family members appears to be an early occurrence in this family.
Family Q: This family had a total of 7 adults, and 4 were symptomatic (3 moderately affected and 1 severely affected). They developed dysarthria at an early stage of disease. The 3 moderately affected individuals (Q1, Q3 and Q5) had not progressed 4 years after the first visit. One individual had pale optical disc on examination, and optic nerve dysfunction was seen on pattern shift visual evoked potential study. Two patients had sensorineural hearing loss, bucco-oral and limb dystonia and sensory ataxia in addition to cerebellar signs. Thus this family had extensive non-cerebellar involvement, including brain-stem dysfunction as evidenced by dysarthria and cranial nerve involvement.

Family J: Of the 11 members examined, 1 was severely affected (J5) and had a history of night cramps at a much earlier stage of the disease (5-6 years before onset). On our second visit one of the elderly pre-symptomatic patients (J2) had developed mild disease, and she had night cramps preceding ataxia.

Family F: Among the 13 individuals examined, 1 was severely affected and had expired after our visit while one was mildly affected. During our second visit we found 2 previously presymptomatic individuals (F1 and F3) who had developed mild ataxia. Another mildly affected patient (F8) had progressed to moderate disease. All members of this family had temporalis wasting. Two members had night cramps (F1 and F3) as well and two members had bucco-lingual dystonia (F1 and F6).

Family R: One member was affected in this family among the 7 examined. He had progressed from moderate to severe illness by the second visit. In addition to severe ataxia he also had frontal release signs, impaired short and intermediate term memory, severe dysarthria, bucco-lingual dystonia, tongue wasting, peripheral neuropathy, pyramidal signs and impaired cough and gag. Despite the severity of the disease he did not have any evidence of muscle wasting.

Family N: In this family one member was affected out of a total 6. He was presymptomatic at the outset (38 years old), and developed mild symptoms by the second visit.

Family G: This family had one moderately affected member, out of 5 examined and disease in this individual (G1) has not progressed after 4 years of examination. He developed significant peripheral neuropathy and wasting at an early stage of disease.

Family I: In this family 8 individuals were examined and although 4 of them were genetically positive for SCA1 none were symptomatic in the first visit. By the time of the second visit, 1 patient (I2) had developed mild symptoms.
### Section 3.3.3.2 Clinico-genetic analysis of the cohort during the second visit:

Attempts were made to understand the trend of disease progression in both pre-symptomatic and symptomatic individuals after a span of 4 years with the CAG repeat size and the result is summarized in Figure 3.5.

**Pre-symptomatics:**

On the second visit, 12 of the 21 pre-symptomatic individuals (ages 17 to 52) years did not show disease progression and did not manifest ataxia. However, 5 of them (N4, F1, J2, F3 and I2) had progressed to develop mild disease (ages 27 to 50) and four were lost to follow-up (unfilled symbols in Fig. 3.5). Amongst the 5 individuals who had shown progression of the disease, two individuals (N4, 43-year old and F1, 44-year old) had paternal inheritance, two (J2, 50-year old and F3, 27-year old) had maternal inheritance and one (I2, 39-year old) with both parents were affected. One of the patients (F3) with maternally inherited disease showed an early age of onset at 27 years. To understand the influence of CAG repeat number in disease progression in F3, we compared F3 with another individual J8 of similar age and genotype (indicated by red arrows in Fig. 3.5). J8 has shown no progression of
the disease in the span of four years suggesting that factors other than inheritance
pattern and CAG repeat size have influenced disease progression in F3.

**Symptomatics:**

Among the symptomatic individuals 5 had shown disease progression, 4 (F8, 
O1, O11, R1) had rate 1 progression and 1 (O3) had rate 2 progression (described in 
methods section). The genotypes of the individuals with rate 1 progression are 28/40 
(F8), 32/40 (O1), 32/49 (O11), 28/50 (R1) and that of the rate 2 progressed individual, 
O3 is 32/51. Despite sharing the higher range of repeat sizes (>49), individuals O11 
(45-year old), R1 (40-year old) and O3 (34-year old) differed in their rates of 
progression.

Heterogeneity in progression of disease was seen both within (intra) and 
between (inter) the families. F8 and O5 are 45-year-old males from two different 
families with repeat number 28/40 and 29/41. F8 shows rate 1 progression and O5 
remained in the same stage when evaluated after 4 years (yellow arrows in Fig. 2.5). 
Severely affected F6 with repeat numbers 28/43 expired in the span of four years and 
O19 with a higher repeat number of 28/47 remained in the same stage (orange arrows 
in Fig 3.5). Like wise, age dependent progression was not observed among certain 
older individuals (purple symbols in Fig. 3.5). And none had progressed in the four 
years of observation.

**Figure 3.5: Correlation between age, genotype, disease severity and repeat length of the individuals with SCA1 mutation.** Triangle (△) represents the phenotypes corresponding to a genotype in visit 1 and square (□) represents those in the visit 2. The results of the individuals between 17 and 35 years of age are shown in pink color, those between 36 and 45 years in blue, 46 and 55 years in green and above 55 in purple. Unfilled triangles represent individuals who have been lost on follow up, asterisk represent individuals who have expired during the study period. Highlighted in arrows are the specific cases described in the results.
3.3.4 Founder mutation study: SNP analysis

To understand the founder mutation in this population, we conducted SNP analysis on two regions known to be associated with the disease. Amplification plots and the Sanger sequencing results obtained for the SNP1 are shown in Figure 3.6.

SNP1 (rs2075974) and SNP2 (rs1476464) for 35 patients including the individuals J8, F8, O19, G1, Q1, and Q5 mentioned above showed clear association to the G allele (Figure 3.7). The association of the two SNPs with normal and SCA1 patient’s were significantly different (p=0.0001; Table 3.2). The allele G showed association with the expanded chromosome at SNP1 and SNP2 (Table 3.2). Presence of the G allele at both SNP1 and 2 in a patient with homozygous mutant alleles (O20) further supports our conclusion that the alleles G-G are linked to the disease locus in this cohort, as previously established for other Indian SCA1 families (Mittal et al. 2005).

![Figure 3.6: (A-C) Sanger sequencing using reverse primer for SNP1 region showing 3 the different allele combinations observed in the cohort; (D-F) representative images of the amplification plots obtained for the respective alleles using the TAQMAN assay.](image)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Normal (frequency)</th>
<th>SCA1 (frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-symptomatic</td>
</tr>
<tr>
<td>SNP1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>12/59 (0.20)</td>
<td>0/17 (0)</td>
</tr>
<tr>
<td>GG</td>
<td>12/59 (0.20)</td>
<td>11/17 (0.64)</td>
</tr>
<tr>
<td>AG</td>
<td>35/59 (0.59)</td>
<td>6/17 (0.35)</td>
</tr>
<tr>
<td>SNP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>0/59 (0)</td>
<td>1/17 (0.05)</td>
</tr>
<tr>
<td>TT</td>
<td>59/59 (1)</td>
<td>0/17 (0)</td>
</tr>
<tr>
<td>TG</td>
<td>0/59 (0)</td>
<td>16/17 (0.94)</td>
</tr>
</tbody>
</table>

*Table 3.2: Association of the SNPs between normal and SCA1 individuals of Adukkamparai*
3.4 DISCUSSION

As compared with other studies, where 66% of variation in age of onset is accounted by CAG repeat size (Ranum et al. 1994), our data indicate that the age of onset in Adukkamparai patients is only accounted to 45% by the repeat size. Moreover, the age of onset in individuals with paternally inherited alleles is earlier than in individuals with maternally inherited alleles and this appears to be independent of the size of CAG repeats. It is known from previous studies that paternal alleles undergo expansion and this results in anticipation (McMurray 2010). In our cohort, both maternal and paternal alleles harbored repeat sizes in the range 40-45 (Figure 3.3) and were not drastically different to cause a significant change in age of onset (Figure 3.4). No marked anticipation was observed. CAG repeat size could only partially determine the rate of progression/severity observed in the patients studied here. I thus speculate factors other than polyQ expansions such as other genetic or epigenetic modifiers to affect disease progression.

When the chromosomal background of the expanded ATXN1 was tested for SNPs known to be associated with CAG expansions, the results correlated with those from previous studies carried out in SCA1 patients from Southern and Northern India (Mittal et al. 2005). An association of the genotypes G and G at loci (rs1476464 and rs2075974) was found in these patients. These G-G alleles can thus provide additional markers for the diagnosis of SCA1. This is the first study highlighting this association of two SNP’s (G-G) within an ethnic community in Tamil Nadu. Therefore, although prevalence of SCA1 in southern India seems to be more, there does not seem to be a different founder from North India.
Very recent implementation of personalized gene silencing by SNP-directed approaches is being considered as a potential therapeutic strategy in Huntington’s disease, another triplet repeat disorder (Kay et al. 2014). SNP-directed gene silencing approaches are likely to be tried in future in the context of personalized medicine. By identifying SNPs associated with SCA1 in the cohort this study paves the way for selective silencing of disease alleles using SNP-targeted approaches in the future.

Genome wide linkage scans (DeStefano et al. 2002; J. H. Lee et al. 2008), exome sequencing (Guerreiro et al. 2012) and targeted gene sequencing (Jin et al. 2012) by Next Generation Sequencing could be employed for identifying the genetic modifiers proposed from the findings of this study and previous studies. Further molecular and genetic studies in individuals from the Adukkamparai cohort and the other cohorts with greater numbers of pre-symptomatic individuals carried out over time will aid in providing a mechanistic perspective and natural history of disease progression. These studies would eventually help in determining and targeting therapeutic interventions based on various stages of the disease.