Introduction

The word “ataxia” derived from Greek word, “*a taxis*” in the late 19th century and refers to “without order or incoordination”. This word is used to describe the symptom of incoordination that may be caused by trauma, infections, other diseases, Vitamin E deficiency or neurodegenerative changes. In addition to environmental and dietary factors, ataxia could also be caused inherently.

In 1863, Dr. Nikolaus Friedreich first described six patients with a recessively inherited ataxia in which the age of onset was below 20-year-old and it is known as Friedreich's ataxia. Thirty years later, Dr. Pierre Marie noticed another dominantly inherited form clinically distinct from that described by Friedreich and began between thirty and forty years of age, and patients described here is recently affirm as having spinocerebellar ataxia type 3 (SCA3)/Machado-Joseph disease (MJD) (Uchihara et al. 2004). Therefore, ataxias at that time were roughly divided into Friedreich's ataxia and Marie's ataxia, based on the age of onset.

These ataxias were not further clearly subclassified until in the 1980s, after Dr. Harding classifying autosomal dominant cerebellar ataxias (ADCAs) into three clinical subgroups types, including ADCA type I, type II, and type III. ADCA type I is characterized by a progressive cerebellar syndrome, with additional but variable associated features of supranuclear ophthalmoplegia, optic atrophy, mild dementia, peripheral neuropathy or extrapyramidal dysfunction. SCA1, SCA2, SCA3, SCA4, SCA8, SCA12, SCA17, dentatorubral pallidoluysian atrophy (DRPLA) and recently defined SCA27

and SCA28 fit in with this type. ADCA type II shows cerebellar ataxia accompanied with pigmentary macular dystrophy and only SCA7 is included. ADCA type III is a pure cerebellar syndrome and comprises SCA5, SCA6, SCA10, SCA11, SCA14, SCA15, SCA22, and SCA26. SCA13 does not fall within any categories described above.

Although each SCA is primarily categorized in terms of the clinical symptoms, different SCAs may sometimes have overlapping signs and it is difficult to distinguish simply on the basis of clinical grounds. However, with the advance of molecular biotechnology, mutation analyses have linked various genetic loci to the subtypes of SCA and afforded the genetic classification of clinical subtypes. Dissimilar to most inherited diseases caused by point mutation, deletion, insertion or translocation of causing gene, identified SCAs are largely caused by the unexpected prolongation of tandem nucleotide repeats. Among the mapped SCAs, SCA1, SCA2, SCA3/MJD, SCA6, SCA7, SCA17, and DRPLA have been shown to be associated with the expansions of coded CAG repeats translated into polyglutamine stretch that adds a toxic gain of function to the respective proteins (Orr et al. 1993; Kawaguchi et al. 1994; Koide et al. 1994; Imbert et al. 1996; Pulst et al. 1996; David et al. 1997; Zhuchenko et al. 1997; Koide et al. 1999), and SCA8, SCA10, and SCA12 are associated with expansions of 3' UTR CTG, intronic ATTCT, and promoter CAG, respectively (Holmes et al. 1999; Koob et al. 1999; Matsuura et al. 2000). Nevertheless, SCA5, SCA14 and SCA27 have been found caused by missense mutations (van de Warrenburg et al. 2003; Yabe et al. 2003; Stevanin et al. 2004; Alonso et al. 2005; Brusse et al. 2006; Ikeda et al. 2006) rather than tandem repeats

expansions. Some SCAs involved in this study will be discussed briefly in the following and the genetic summary of SCAs is listed in **Table 1**.

Spinocerebellar ataxia type 1 (SCA1)

SCA1 is the first mapped SCA in which the mutant gene *SCA1* localized on chromosome 6p23. It has been found that *SCA1* gene carries more than 40 CAG trinucleotide repeat units in affected individuals (Orr et al. 1993; Banfi et al. 1994) and it is now believed that expanded CAG-encoded polyglutamine (polyQ) tract within ataxin-1, the *SCA1* product, gains a novel function. SCA1 usually begins in the 3rd and 4th decade of life accompanied with symptoms such as ataxia, slow saccades, increased tendon reflexes at knee, and difficulty speaking and swallowing. In the view of molecular mechanisms, ataxin-1 interacts with glyceraldehyde-3-phosphotate dehydrogenase (GAPDH) (Koshy et al. 1996), cerebellar Purkinje-enriched leucine-rich acidic nuclear protein (LANP) (Matilla et al. 1997), calcium binding protein calbindin D28k (CaB) (Vig et al. 1998; Vig et al. 2000), proteasome and chaperon proteins (Cummings et al. 1998). Most associated proteins are recruited to nuclear aggregates formed under the expression of mutant ataxina-1 and this could be explained the fact that aberrant interactions involve in the pathogenesis.

Spinocerebellar ataxia type 2 (SCA2)

The typical clinical features of SCA2 include limb and gait ataxia, slow saccades, and hyporeflexia, while levodopa-respinsive parkinsonism is occasionally presented or dominant in the family (Gwinn-Hardy et al. 2000; Shan et al. 2001; Furtado et al. 2002; Ragothaman et al. 2004). SCA2 is caused by the expansion of unstable CAG repeats in *SCA2* gene encoding

ataxin-2. The most common normal ataxin-2 allele contains 22 glutamines. The range of glutamine repeats in mutant ataxin-2 varies from 33 to 77 with 37 glutamines being most common. Wild type ataxin-2 is primarily localized with the Golgi apparatus, whereas expression of mutant ataxin-2 disrupts the normal morphology of the Golgi complex (Huynh et al. 2003) and this dispersion can be reduced by the overexpression of parkin, an E3 ubiquitin ligase mutated in an autosomal recessive form of parkinsonism. In addition, ataxin-2 serves as a substrate of parkin for ubiquitination and degradation, and mutant ataxin-2 aggregates mask these effects of parkin by sequestrating it (Huynh et al. 2007). Thus, the correlation between ataxin-2 and parkin needs to be further investigated to uncover the pathogenesis of SCA2 and related parkinsonism.

Spinocerebellar ataxia type 3 (SCA3)

SCA3 is the most common dominantly inherited ataxia worldwide and also called Machado-Joseph disease (MJD), named for affected families of Azorean extraction, characterized principally by ataxia, spasticity, and ocular movement abnormalities. The mutation leading to SCA3/MJD is linked to chromosome 14q32.1, and the range of CAG repeats in mutant alleles is 45-86 repeats (Kawaguchi et al. 1994). The disease protein, ataxin-3, of 42kDa is smallest of the polyQ disease so far. Ataxin-3 consists of a structured N-terminus with deubiquitinating activity, two ubiquitin-interacting motifs (UIMs) and a C-terminal polyQ tract (Masino et al. 2003). *In vitro* and *in vivo* studies have shown that expression of expanded ataxin-3 induces cell death and results in accumulation of ubiquitinated intranuclear inclusions selectively in neurons of affected brain regions (Paulson et al.

1997) in which the recruitment of ubiquitin might act via UIM motifs (Donaldson et al. 2003).

Spinocerebellar ataxia type 6 (SCA6)

SCA6 is caused by a very mild CAG repeat expansion in *CACNA1A* gene encoding the α_{1A} subunit of voltage-gated calcium channels type P/Q (Zhuchenko et al. 1997). However, other mutations in this gene have been found to cause familial hemiplegic migraine (FMH) and episodic ataxia type 2 (EA2) (Ophoff et al. 1996), suggesting these diseases are allelic disorders. The clinical features of SCA6 include nystagmus, dysarthria, and ataxia, while in some cases signs of brainstem and pyramidal tract are also involved. It is believed that polyQ expansion in this calcium channel influences the trafficking of Ca^{2+} and a previous study has indicated that polyQ elongation caused a proportional negative shift of voltage-dependent inactivation, suggesting reduced Ca^{2+} influx may be associated with neuronal death in cerebellum (Toru et al. 2000).

Spinocerebellar ataxia type 7 (SCA7)

SCA7 is unique to SCAs for the phenotype of retinal degeneration in addition to typical cerebellar symptoms. SCA7 was mapped to Chromosome 3p14-p21.1 (Benomar et al. 1995). Normal SCA7 alleles contain 4-35 CAG repeats, while pathological alleles contain over 38 CAG repeats. Ataxin-7, encoded by *SCA7* gene, is nuclear localized and associates with the nuclear matrix and the nucleolus (Kaytor et al. 1999). In addition, ataxin-7 was considered as a component of the TATA-binding protein-free TAF-containing complex (TFTC) and the SPT3-TAF9-GCN5 acetyltransferase complex (STAGA) (Helmlinger et al. 2004), which both belong to histone

acetyltransferase (HAT) complexes, or related complexes. Although mutant ataxin-7 did not disturb the assembly of these complexes, it inhibited the histone acetylation function in a dominant-negative manner (Helmlinger et al. 2004; McMahon et al. 2005; Palhan et al. 2005). Taken together, the spatial and proteins interaction patterns help to explain some aspects of SCA7 pathogenesis.

Spinocerebellar ataxia type 8 (SCA8)

SCA8 was first described in 1999 as a novel autosomal SCA caused by CTG triplet repeats expansion in the nontranslated *SCA8* gene locating on chromosome 13q21 (Koob et al. 1999). Dysarthria, eye move disorder, and ataxia are commonly observed in patients. Incomplete penetrance is an interesting trait observed in SCA8 that expanded CTG repeats do not always cosegregate with clinical features. In addition, intergenerational transmission of CTG alleles is instable and shows a strong bias toward expansions in maternal transmission (Koob et al. 1999; Day et al. 2000; Moseley et al. 2000). Further description of SCA8 will be presented in the following.

Spinocerebellar ataxia type 10 (SCA10)

SCA10, characterized by ataxia, dysarthria and seizure not found in other SCAs (Grewal et al. 1998; Rasmussen et al. 2001; Grewal et al. 2002), is caused by expansion of ATTCT pentanucleotide repeat in intron 9 of a novel gene, *SCA10*, on chromosome 22q13.3. Although little is known about ataxin-10, the protein product of *SCA10*, has been localized to cytoplasmic and perinuclear compartments, and knocking it down induces sever neuronal loss (Marz et al. 2004). In addition, overexpression of ataxin-10 enhanced intracellular glycosylation activity via interaction with *O*-Linked β-N-

acetylglucosamine transferase (OGT) (Marz et al. 2006). Accordingly, ataxin-10 might play a vital role in cell.

Spinocerebellar ataxia type 12 (SCA12)

SCA12 was first described in pedigrees of German American and Indian descent (Holmes et al. 1999; Fujigasaki et al. 2001b). Patients usually begin with action tremor in their 3rd decade, followed clinical symptoms such as ataxia, hyperreflexia, dysarthria and bradykinesia (Holmes et al. 1999; O'Hearn et al. 2001; Srivastava et al. 2001). The causing mutation lies 5' untranslated region (UTR) of the gene *PPP2R2B*, which produces regulatory B subunit of serine/threonine phosphatase, PP2A. PP2A is implicated in various cell process including cell growth and differentiation, DNA replication, microtubule assembly, and apoptosis (Price and Mumby 1999; Virshup 2000). Down regulation of PP2A in brain has been shown to disrupt the balance of phosphorylation/dephosphorylaton state, resulting in accumulation of hyperphosphorylated tau, a microtubule-associated protein (Gong et al. 2000). It is speculated that expansion of CAG repeat misregulates the expression of downstream gene, which in turn affects the phosphatase activity of PP2A (Holmes et al. 2003).

Spinocerebellar ataxia type 17 (SCA17)

TATA-box binding protein (TBP) is well-known for its role in the initiation of RNA transcription, but it is now also considered associated with SCA17. SCA17 was first described in a Japanese family and the clinical features consist of ataxia, dementia, hyperreflexia, parkinsonian, and postural reflex disturbance (Koide et al. 1999; Fujigasaki et al. 2001a; Nakamura et al. 2001). Normal individuals carry 25-42 CAG repeats in *TBP* gene, while

those have 43-48 CAG repeats show an incomplete penetrance (Fujigasaki et al. 2001a; Nakamura et al. 2001; Silveira et al. 2002). Immunocytochemical analysis in SCA17 patients showed neuronal intranuclear inclusion bodies (Nakamura et al. 2001), and most neuronal nuclei were diffusely stained with 1C2 antibody, which recognizes expanded polyglutamine tracts (Nakamura et al. 2001). Expression of the full-length TBP protein resulted in repeat-length-dependent inclusion formation *in vitro*, and overexpression of expanded TBP increased Cre-dependent transcriptional activity, suggesting that mutant TBP might be involved in aberrant transcription activity (Reid et al. 2003). TBP is also found colocalize to neurofibrillary tangle structures in Alzheimer's disease (AD) (Reid et al. 2004), implying that ubiquitously expressed TBP might contribute to other diseases.

“Anticipation” refers to the phenomena that the disease begins at an earlier age and worsens while it is transmitted to the next generation. This phenomena is often found in diseases caused by trinucleotide repeat expansion, including Huntington's disease, myotonic dystrophy, SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA10, SCA12, SCA17, SCA22, and DRPLA, as the next generations often have more repeats and move severe symptoms.

The prevalence of SCAs varies among populations. Such differences were closely associated with the distributions of large normal alleles in Japanese and Caucasian populations (Takano et al. 1998). The frequency of SCA has been assessed in Chinese patients, with 5% for SCA1, 6% for SCA2, 48% for MJD/SCA3, and 0% for SCA6, SCA7, and DRPLA (Tang et al. 2000). A study of ethnic Chinese in Taiwan also revealed MJD/SCA3 as the most

common type (47%), followed by SCA6 (11%), SCA2 (11%), SCA1 (5%), SCA7 (3%), DRPLA (1%), and SCA8 (0%) (Soong et al. 2001). However, 22% of SCA patients in this study did not have mutations in the above 7 SCA genes, and there has been no report on SCA10, SCA12, and SCA17 among Chinese populations. In addition, and with the exception of the MJD/SCA3 and SCA6 genes, the distribution of SCA repeats in the Chinese population has not been well documented.

Among the growing numbers of the neurodegenerative disorder, expanded repetitive DNA sequences contribute some aspects to the pathogenesis. The expanded sequence might occur either within or outside of the coding region, and the latter includes the regions of introns, 3' untranslated, and 5' untranslated region (UTR). The autosomal inherited spinocerebellar ataxia type 8 (SCA8) is an example in which trinucleotide CTG repeat expansion occurs in the 3' UTR of the *SCA8* gene that lacks a significant open reading frame. SCA8 was first described in a large family characterized by the symptoms including limb ataxia, dysarthria, and horizontal nystagmus (Koob et al. 1999). The CTG repeat length is mitotic instably, and often interrupted by CTA repeats (CTG/CTA combined repeats) (Koob et al. 1999; Moseley et al. 2000; Silveira et al. 2000). In addition, the intergenerational changes in CTG repeat number are typically larger for SCA8 than for the other SCAs, and most cases are resulted form maternal expansive transmission (Koob et al. 1999).

Organization of SCA8 gene

Human *SCA8* gene spans around 32 kb in length on chromosome 13q21 and is composed of up to six exons that are alternatively spliced (**Figure 2A**).

The 5' end of *SCA8* transcript is variably spliced exon D, D', D'', D4, D5, which in turn joins to the variable exon C3, the invariable exons C2, C1, the alternative exon B, and terminates at either exon B or A (Nemes et al. 2000; Benzow and Koob 2002). Since the last exon is the region where CTG repeats reside, the expanded CTG tract might not be included in *SCA8* transcripts, accounting for the incomplete penetrance in some way. Another attractive feature of *SCA8* gene is that exon D is transcribed in the opposite direction through the first exon of *KLHL1* transcript (Nemes et al. 2000) (**Figure 2A**). The similar distribution patterns in the brain and the conserved sense/antisense organization of *SCA8* and *KLHL1* (Benzow and Koob 2002) indicate that endogenous *SCA8* transcripts might act as a natural, or so-called *cis*-encoded antisense transcripts of *KLHL1*, which further results in SCA8 neuropathogenesis.

The role of KLHL1 protein

KLHL1 protein is highly homologous to the *Drosophila* kelch protein, which is responsible for maintaining actin organization of ring canals connecting the oocyte to supporting nurse cells during oogenesis (Xue and Cooley 1993; Robinson and Cooley 1997). *KLHL1* spans a genomic size of over 400 kb and comprises of 11 exons encoding the amino-terminal region (NTR) following by the BTB/POZ (for Broad-Complex, Tramtrack and Bric a brac/Poxvirus and Zinc-finger domain) dimerization domain and the six antin-binding Kelch motif repeats (KREPs) separating by the intervening amino acid sequence (IVS) (Robinson and Cooley 1997). In addition, KLHL1 protein interacts with the α_{1A} subunit of voltage-gated calcium channels type P/Q, the product of SCA6 gene, by increasing its current

density and channel availability for opening (Aromolaran et al. 2007), indicating that KLHL1 might have a role for modulating calcium channel.

Natural antisense transcripts

Natural antisense transcripts (NATs) are endogenous RNAs complementary to sequences of already known function. Although some are transcribed from a locus that is different from the locus of the sense RNA, most antisense transcripts are issued from the same locus. They were first found in prokaryotes and have been considered regulating the expression of specific genes, which are often down regulated. In the last two decades, more and more NATs have been described in eukaryotes, including human. It is widely accepted that NATs might function as templates for translation or regulators of sense gene expression (Dolnick 1993; Li et al. 1996). Apart from being translation templates, NATs could exert its regulational effects via at least three mechanisms. First, the overlapping segments form double strand, which leads to the prevention of translation or the digestion to small fragments (Lai 2002). Second, NATs involve epigenetic regulations such as the methylation of promoters by some unknown mechanisms and the conversion of the chromosome structure (Wutz et al. 1997; Tufarelli et al. 2003). Third, sense and antisense RNAs form two voluminous RNA pol II complexes on opposite strands, leading to RNA polymerases clash in the overlapping region, which would ultimately interfere the activity of one or both protein complexes (Prescott and Proudfoot 2002).

Myotonic dystrophy type 1 (DM1)

Myotonic dystrophy type 1 (DM1) is also caused by a CTG trinucleotide repeat expansion in the 3' UTR of its disease-causing gene, *DMPK* (Buxton

et al. 1992; Harley et al. 1992; Kitsu et al. 1992; Mahadevan et al. 1992). The CTG repeat expansion within the *DMPK* gene primarily affects the neuromuscular system, while cardiac conduction defects, smooth muscle involvement, mental changes, hypersomnia, ocular cataracts, insulin dependent diabetes and testicular atrophy are also involved. At least three hypotheses have been proposed to explain the possible mechanisms of DM1. First, haploinsufficiency has been proposed to explain the dominant nature of DM. However, DMPK levels in adult and congenital patients were not significantly changed (Narang et al. 2000), and neither heterozygous nor homozygous *DMPK* knock out mice show the severely multisystemic DM phenotypes (Berul et al. 1999), suggesting that haploinsufficiency might not be directly involved in DM. Second, the CTG expansion is likely to disrupt the chromatin structure nearby the *DMPK* and further affect the transcription *DMPK* and its neighboring genes (Otten and Tapscott 1995). Third, the mutant *DMPK* transcripts tend to be retained within the nucleus in distinct foci, which in turn sequester the RNA-binding proteins, including CUG-binding proteins (CUGBPs), muscleblind (MBNL) protein family, and some other transcription factors, or affect the transport of other CAG repeats-containing RNA (Taneja et al. 1995; Fardaei et al. 2002; Ebralidze et al. 2004). In the view of being mediated by CTG expansion, the molecular similarity between SCA8 and DM1 should provide an opportunity to further define the molecular pathogenesis of SCA8.

Internal ribosome entry segment (IRES)

Initiation of translation of most eukaryotic mRNAs normally depends on the 5' m^7GpppN cap structure of mRNAs, which recruits 43S ribosome preinitiation complex via interaction with the cap binding protein eIF4E

(Sonenberg 1994). The translation machinery then migrates downstream until it meets the first AUG codon in the optimal context for initiation of translation (Kozak 1991). This scanning model predicts that any mRNA with long 5'-untranslated region (5'-UTR) and complex secondary structures may not be translated efficiently. In an alternative mechanism of translation initiation, the ribosome can be directly recruited to an internal site on the mRNA that can be some considerable distance from the cap structure (Hellen and Sarnow 2001). The latter mechanism requires the formation of a complex RNA structural element termed an internal ribosome entry segment (IRES). In the presence of *trans*-acting factors, IRES allows the internal ribosome entry (Stoneley and Willis 2004). Up to date, many eukaryotic cellular mRNAs have been suggested to contain such IRES activities (see http://www.rangueil.inserm.fr/IRESdatabase). Although reported non-coding (Koob et al. 1999), small ORFs in the *SCA8* transcripts were noted. Among them, a 102 amino acids containing-ORF1 and a 41 amino acids plus a polyleucine tract containing-ORF3 may be translated if *SCA8* RNA possesses a cap independent IRES activity. Our previous study had demonstrated the possible IRES activity within *SCA8* (Lin et al. submitted). Therefore, if small ORFs could be translated, the linkage between translatable ORFs and pathogenesis remains to be further investigated.

Animal models of SCA8

Animal models play a crucial role in studying pathological mechanisms of human diseases. Several models have been established to investigate SCA8. Ectopic expression of *SCA8* in the *Drosophila* retina induces late-onset and progressive neurodegeneration, and the neuronally expressed RNA binding protein, staufen is found to be recruited to the *SCA8* RNA with the presence

of 3' CUG-containing region, suggesting that SCA8 may be resulted from RNA-mediated molecular pathway (Mutsuddi et al. 2004). Recently targeted deletion of a single *Sca8* ataxia locus allele in mice leads to degeneration of Purkinje cell function, indicating partial loss of Klhl1 function with the pathogenesis of the disease (He et al. 2006). In addition, using a transgenic mouse model in which the human *SCA8* mutation is present on a BAC, a newly discovered gene, *ataxin 8* (*ATXN8*), which encodes a nearly pure polyglutamine expansion protein in the CAG direction, was reported (Moseley et al. 2006). The studies of BAC *SCA8* transgenic mice revealed 1C2-positive intranuclear inclusions in Purkinje and brainstem neurons, indicating polyglutamine expansion protein in the CAG direction with the pathogenesis of the disease (Moseley et al. 2006). However, BAC clones usually insert to the chromosome with relatively low copies, leading to lower expression of transgene. Hence, reinforcing transgene expression under the control of the specific promoter in replace of expressing low copies of artificial chromosome could sometimes get severe phenotypes and a earlier age of onset.

Aims

The first aim of this study was to set up a database of the trinucleotide- and pentanucleotide-repeat expansions leading to SCA, we have assessed the repeat size at the SCA1, SCA2, MJD/SCA3, SCA6, SCA8, SCA10, SCA12, SCA17, and DRPLA loci in 198 normal controls and 334 patients with ataxia and Parkinson's disease (PD) in Taiwanese. In addition, we also compared the frequencies of large normal alleles with the relative frequencies of SCA1, SCA2, MJD/SCA3, SCA6, and DRPLA among Taiwanese, Japanese, and Caucasians. Secondly, in order to uncover the molecular mechanisms underlying SCA8, we used *in vitro* cellular system to investigate the plausible pathogenesis, including the antisense effects between KLHL1 and SCA8, and the translatable open reading frame in SCA8. We also generated stable and inducible SCA8 cell lines carrying various CTA/CTG combined repeats to find out the repeats effects on stresses. In addition, to get insight the pathogenesis of SCA8 and for further potential therapy screening, we also generated transgenic mice harboring the human SCA8 cDNA carrying either normal or expanded CTA/CTG combined repeats under control of cerebellar Purkinje cell specific promoter, Pcp2/L7 promoter.