國立臺灣師範大學生命科學系博士論文

第八型脊髓小腦運動失調症分子致病機轉之研究

Molecular Genetic studies of spinocerebellar ataxia type 8

研 究 生：陳　怡　辰

I-Cheng Chen

指導教授：李　桂　楨 博士

Guey-Jen Lee-Chen

中華民國九十九年六月

國立臺灣師範大學生命科學系博士論文通過簽名表

系所別：生命科學系

姓名：陳怡辰　　學號： 894430063

博士論文題目：第八型脊髓小腦運動失調症分子致病機轉之研究

經審查合格，特予證明

論文口試委員

陳瓊美　博士
財團法人長庚紀念醫院神經肌肉疾病科 主治醫師

陳儀莊　博士
央研究院生物醫學科學研究所 研究員

謝秀梅　博士
國立臺灣師範大學生命科學系（研究所）教授

蘇銘燦 博士
國立臺灣師範大學生命科學系（研究所）教授

李桂楨　博士
國立臺灣師範大學生命科學系（研究所）教授
論文指導教授

系主任（所長）簽章：

中 華 民 國 九十九年 六月 一 日

國立臺灣師範大學學位論文授權書

本授權書所授權之論文為授權人在國立臺灣師範大學生命科學研究所九十八學年度第二學期取得博士學位之論文。

論文題目：第八型脊髓小腦運動失調症分子致病機轉之研究

指導教授：李桂楨教授

一、 授權人同意非專屬無償授權本校及國家圖書館將上列論文資料以微縮、數位化或其他方式進行重製，並可上載網路收錄於本校博碩士論文系統、國家圖書館全國博碩士論文資訊網及臺灣師範校院聯合博碩士論文系統，提供讀者基於個人非營利性質之線上檢索、瀏覽、下載、傳輸、列印或複印等利用。

二、 論文全文電子檔上載網路公開時間：【第一項勾選同意者，以下須擇一勾選】

□ 即時公開

■ 自 99 年 7 月 6 日始公開。

授權人姓名：陳怡辰　　（請親筆正楷簽名）

學　　號：894430063

註：1. 本授權書須列印並簽署兩份，一份裝訂於紙本論文書名頁，一份繳至圖書館辦理離校手續。

2. 授權事項未勾選者，分別視同「同意」與「即時公開」。

中　華　民　國　99　年　7　月　6　日

致謝

這本論文的完成，真的要感謝太多長輩家人朋友們在我的博士求學生涯中，一路上相助伴我學習成長，使我得以順利完成學業，千言萬語也無法表達我心中的感激！在此對所有曾幫助過我的人致上最真誠的謝意。

首先最要感謝的就是本博士論文指導教授李桂楨老師，李老師五年來的栽培與悉心指導讓我獲益良多，除了培養我獨立思考的能力，更給了我充分的機會與空間，引領我平穩的走上學術研究之路。而本論文的完成，也承蒙長庚醫院主治醫師陳瓊美醫師多來的提攜與關懷，沒有您當初的鼓勵就沒有這本論文的誕生！同時也感謝中研院生醫所陳儀莊老師以及本系謝秀梅老師、蘇銘燦老師經常提供寶貴的意見，並在百忙之中給予論文詳細的審閱與批示，使得論文更加嚴謹完整，在此致上最真誠的感謝。此外，也特別感謝長庚醫院吳逸如醫師提供得來不易的臨床檢體，使本研究更臻完備。特別感謝台師大分子影像核心實驗室以及台大貴重儀器中心協助本論文中免疫螢光染色之影像擷取。

再來要感謝的就是在實驗室一同打拼的夥伴們，特別要感謝一路上和我一起爲第八型小腦萎縮症致病機轉的研究努力的玄原學長、欣杰學長、雅今、葦苓、淑怡、士寰、金玨、郁芳、國修、修嘉，我深刻相信我們的成果總有一天一定會對疾病患者有所貢獻。也感謝李姐、秀觀學姐、政光學長、瑞宏學長、志信、雰茹、芷英、懿婷、玠璘、宜欣、承岳、襄銘、若芸、郡潔、慧茹、玄竺、昇翰、佩瑛、春嫻、品蓉、淑婷、詩涵、慈蓬、佩茹、文騰、允麟、明惠、婉玲…等 D311 成員，在這個大家庭中大家互相砥礪，分享生活中的點滴，這份同甘共苦的情誼實屬難能可貴，點點滴滴都是我在師大最珍貴的回憶。

此外，我很幸運的能擁有這一群大學死黨，雖然大學畢業後大家各奔東西，有的在台灣打拼，有的在國外奮鬥，但是彼此之間的默契以及情誼是永遠不變的，感謝最瞭我的王瑛，不用多說什麼就能接收我的訊息，時常鼓勵我給我意見安慰；還有溫琪，陪我一起跑遍各大影展，互相分享生活；淑真、雪其、承宗、

桂華、芷英、奕穎…時常捎來的關心訊息總是令我溫暖；還有遠在異鄉的孟穎、宏惠、駿逸是我旅途中最溫暖的港灣。感謝得民一直以來的鼓勵。

另外還要特別感謝行政院國家科學委員會提供我在英國一年的獎助學金，我非常幸運也很榮幸可以到英國倫敦帝國理工學院臨床研究中心從事幹細胞的研究，在英國一年承蒙李萌老師 (Meng Li)的不吝指導，使我獲益匪淺，有機會學習最新的幹細胞知識以及國外頂尖的科學技術，並將這些新技術帶回台灣。也特別感謝長庚醫院張國軒醫師在倫敦給予生活上極大的協助以及實驗上的指導；謝謝長庚李莉文醫師總是在我想念台灣味的時候剛好捎來一通電話，在下著雪的嚴冬走到隔壁宿舍我就有香噴噴的薑母鴨可以吃。也感謝在英國的好朋友 Lieven、Judit 及 Kailyn，雖然我們有著不同的文化，說著不同的語言，有時支支吾吾有時比手劃腳，但我會永遠記得每個徹夜談心的夜晚，每一次的旅行，每個笑聲每滴眼淚，以及在倫敦生活經歷的所有點點滴滴，這份跨越國界的友誼是我一生中最珍貴的寶藏。

最後，最感謝的當然是我的精神支柱親愛的爸爸、媽媽和弟弟，你們的包容和體諒是我走下去的動力，在我忙實驗忙報告時，桌上永遠都留有屬於我的那一份晚餐等我回家；在我猶豫不決到底該不該花一年的時間去英國時，也是二話不說要我放下台灣的一切大膽去追求夢想；有你們的支持當後盾，我才能有今天這一點點的成績，你們都辛苦了！我很幸運能夠擁有這麼溫暖可愛的家庭，雖然有遺憾在英國時沒能回台灣送奶奶走完最後一程，但我相信她在天上也會為我完成學業感到高興。

要感謝的人實在太多，如果未能提及也請見諒。在師大的求學生涯隨著論文的完成也即將劃下句點，但這不是結束而是開始，今後我將帶著大家的鼓勵繼續往前走，期待能發揮所學，貢獻一己之力。謝謝大家也祝福大家！

本論文感謝 財團法人罕見疾病基金會提供論文獎助金。

Index

中文摘要

脊髓小腦運動失調症 (Spinocerebellar ataxias)爲一群顯性遺傳的神經退化性疾病，患者小腦及腦幹區域發生漸進式的神經退化，其中第八型脊髓小腦運動失調症 (SCA8)與染色體13q21位置的*ATXN8OS*基因3'端非轉譯區CTG三核苷重複擴增相關；除此之外，目前根據前人研究顯示SCA8疾病除了與*ATXN8OS* CTG方向的擴增有關，其反向*ATXN8*基因之CAG擴增也可能扮演重要的角色。目前SCA8的致病機轉尚未十分明確，本論文主要核心爲研究SCA8疾病之遺傳與分子機轉。首先，我們利用人類胚胎腎細胞 (Human embryonic kidney 293 cell lines)建立可被誘導並穩定表現包含不同長度CTG擴增的ATXN8OS細胞株並進行*ATXN8OS* RNA表現之研究，實驗結果顯示帶有較長重複擴增之細胞，其*ATXN8OS* RNA被誘導表現的倍數較高，可能是由於其RNA比較穩定所導致。利用螢光原位雜交技術，我們也觀察到在帶有較長重複擴增之細胞中有RNA foci的形成。本論文的第二部份爲檢測*ATXN8OS*、*ATXN8*及*KLHL1* RNA在正常個體及*ATXN8OS* CTG擴增病人淋巴細胞株 (lymphoblastoid cell lines)中之表現，實驗結果發現在病人之淋巴細胞中*ATXN8* RNA表現量比正常人高並達顯著差異，推測可能與存在*ATXN8*啓動子中的-62 G/A多型性點有關。另一方面，雖然之前的研究認爲*ATXN8OS*基因不會進行轉譯，但我們實驗室的研究證實了*ATXN8OS*基因中的開放解讀架構(open reading frame, ORF)可以透過特殊IRES (internal ribosome entry segment)的轉譯活性製造出蛋白質，因此，*ATXN8OS* ORF蛋白質轉譯的調控機制以及ORF蛋白質在病理機制當中可能扮演的角色亦是本論文著重的議題。本論文中我們利用*ATXN8OS*融合*EGFP*基因證實了

ATXN8OS RNA具有轉譯的能力，除此之外我們利用*ATXN8OS* ORF蛋白質的抗體研究ORF蛋白質在不同細胞株中的表現，並更進一步利用質譜技術分析ORF蛋白質之胺基酸序列。我們也發現在病人的淋巴細胞株中，*ATXN8OS* ORF蛋白質的表現量高於正常人。預期本論文的實驗結果，將有助於了解SCA8的致病機轉，找尋適當的治療目標，並可以將結果應用到其他相關的神經退化性疾病中。

Abstract

The autosomal dominant spinocerebellar ataxias (SCAs) are a heterogeneous group of neurodegenerative diseases caused by a progressive degeneration of the cerebellum and brainstem. Among the SCAs, spinocerebellar ataxia type 8 (SCA8) involves the expression of a CTG/CAG expansion mutation from opposite strands producing CUG expansion transcripts (*ATXN8OS*) and a polyglutamine expansion protein (*ATXN8*). The pathogenic mechanism of *ATXN8OS* expansion is still unknown. The main purpose of this present proposal is to dissect the possible factors involved in SCA8 pathogenesis. Firstly, the stably induced cell lines expressing 0, 23, 88 and 157 CR exhibit low levels of *ATXN8OS* expression without doxycycline induction, and a repeat length-dependent repression of *ATXN8OS* expression was notable. Addition of doxycycline leads to 25~50 times more *ATXN8OS* RNA expression with a repeat length-dependent increase in fold of *ATXN8OS* RNA induction. The repeat length-dependent increase in induction fold is probably due to the increased RNA stability. RNA FISH experiments further revealed ribonuclear foci formation in cells carrying expanded 88 and 157 CR. Our results demonstrate that the expanded CUG-repeat tracts may affect *ATXN8OS* RNA expression and stability through epigenetic and post-transcriptional mechanisms. Secondly, *ATXN8* expression level is significantly higher in lymphoblastoid cells with SCA8 large alleles than that of the control cells. Our results suggest that *ATXN8* gene -62 G/A polymorphism may be functional in modulating *ATXN8* expression. Lastly, although reported non-coding, existence of IRES (internal ribosome entry segment) activity in the 5' UTR sequence of *ATXN8OS* has been demonstrated in our previous studies. Expression

of chimeric constructs with an EGFP gene fused in-frame to *ATXN8OS* ORF demonstrated *ATXN8OS* is translatable and the ORF protein formed aggregates and co-localized with mitochondria. Moreover, the ORF expression was validated in different human cells using ORF antiserum. ATXN8OS ORF was further confirmed by LC-MS/MS. The expression of ORF protein was significantly higher in lymphoblastoid cells carrying expanded *ATXN8OS*. The results of this study may suggest a broader hypothesis for further research in explaining the expanded CTG leading to neuronal dysfunction in SCA8.

List of tables and figures

Introduction

Spinocerebellar ataxias (SCAs)

The autosomal dominant spinocerebellar ataxias (SCAs) are a heterogeneous group of neurodegenerative diseases, characterized by progressive ataxia variably associated with other neurological signs and caused by a progressive degeneration of the cerebellum and brainstem (Harding, 1993). Previous studies in the genetic understanding of these diseases have identified over 30 genetically distinct subtypes of SCAs: SCA1 through SCA8, SCA10 through SCA31, and in addition, dentatorubral-pallidoluysian atrophy (DRPLA). A few SCA subtypes, including SCAs 1, 2, 3, 6, 7, 17 and DRPLA, are caused by the expansion of a CAG repeat sequence located within the coding region of specific genes, leading to an abnormally long polyglutamine tract in the encoded proteins (David et al., 1997; Kawaguchi et al., 1994; Koide et al., 1994; Koide et al., 1999; Orr et al., 1993; Pulst et al., 1996; Zhuchenko et al., 1997). A second group of SCAs, including SCAs 8, 10 and 12, are caused by a repeat expansion located outside of the coding region of the disease genes leading to dysregulation of gene expression (Holmes et al., 1999; Koob et al., 1999; Matsuura et al., 2000). Different mechanisms cause cerebellar ataxia and neurodegeneration in SCAs 5, 13, 14 and 27, where alterations in amino acid composition elicit disease symptoms in these four SCA subtypes (Chen et al., 2003; Ikeda et al., 2006; van Swieten et al., 2003; Waters et al., 2006; Yabe et al., 2003). In the rest of SCAs, the genes and, therefore, the mutations remain to be identified and characterized (Carlson et al., 2009).

Spinocerebellar ataxia type 8 (SCA8)

Among the SCAs, SCA8 was first described by Koob and co-workers in 1999 (Koob et al., 1999). Apart from the ataxias caused by polyglutamine expansions, an untranslated CTG expansion in 3' untranslated region (UTR) of *ATXN8OS* (ataxin 8 opposite strand, also known as *SCA8* gene) gene causes SCA8. Although the SCA8 mutation has molecular similarities to the myotonic dystrophy protein kinase (DMPK) mutation causing myotonic dystrophy type 1 (DM1) (Larkin and Fardaei, 2001), this second example of a pathogenic CTG expansion causes a central nervous system disease without the multisystemic features of DM1. Clinical and molecular analyses of different SCA8 ataxia families from the literature (Brusco et al., 2002; Day et al., 2000; Ikeda et al., 2000b; Juvonen et al., 2000; Koob et al., 1999; Silveira et al., 2000; Topisirovic et al., 2002) demonstrate that SCA8 consistently presents as a very slowly progressive ataxia that largely spares brainstem and cerebral function. Affected individuals have gait, limb, speech and oculomotor incoordination, spasticity, and sensory loss.

The sizing of SCA8 alleles has been clarified in various populations. In general, more than 99% of the normal SCA8 alleles had 16-37 repeats, while unrelated expanded alleles ranging from 68 to 800 repeats were found in familial and sporadic ataxia patients (Cellini et al., 2001; Ikeda et al., 2000b; Izumi et al., 2003; Jardim et al., 2001; Juvonen et al., 2000; Juvonen et al., 2002; Sobrido et al., 2001; Stevanin et al., 2000; Tazon et al., 2002; Topisirovic et al., 2002; Worth et al., 2000). Both expanded and normal SCA8 repeats are highly unstable (Silveira et al., 2000). The intergenerational changes in CTG repeat number are typically larger for SCA8 than for the other SCAs, and most cases are resulted from maternal expansive transmission

(Koob et al., 1999). In some affected families, expanded alleles do not always co-segregate with the disease phenotype (Cellini et al., 2001; Day et al., 2000; Ikeda et al., 2000a; Juvonen et al., 2000; Koob et al., 1999; Stevanin et al., 2000; Topisirovic et al., 2002; Vincent et al., 2000; Worth et al., 2000). The incomplete penetrance (Juvonen et al., 2000; Tazon et al., 2002; Worth et al., 2000) as well as the size of the expanded repeats that were not correlated among affected and healthy individuals (Schols et al., 2003; Sobrido et al., 2001; Stevanin et al., 2000; Sulek et al., 2004) have argued strongly that expansion may not be a susceptibility factor for the disorder. In addition, SCA8 alleles with expansion can be found in rare instances in the general population (Izumi et al., 2003; Juvonen et al., 2000; Koob et al., 1999; Stevanin et al., 2000; Tazon et al., 2002; Worth et al., 2000) as well as in patients with psychiatric disorder, Freidreich's ataxia, Parkinson's disease (PD), and Alzheimer's disease (AD) (Izumi et al., 2003; Pato et al., 2000; Sobrido et al., 2001; Vincent et al., 2000; Worth et al., 2000). Thus, the association between SCA8 CTG expansion and spinocerebellar ataxia is not straightforward.

Plausible pathogenesis of SCA8

The pathogenic role of *ATXN8OS* expansion remains uncertain. RNA-mediated neurotoxicity has been implicated in SCA8 pathogenesis. Human *ATXN8OS* is expressed in various brain tissues, including the cerebellum. It seems that the *ATXN8OS* gene is transcribed as a part of an untranslated RNA that overlaps with the transcription and translation start sites and the first splice junction of kelch-like 1 (*KLHL1*), a gene that encodes

a protein with structural similarities to a family of factors involved in the organization of the cytoskeleton (Koob et al., 1999). The *ATXN8OS* transcript may act as an endogenous antisense regulator of *KLHL1* expression in various brain tissues (Koob et al., 1999; Nemes et al., 2000). This antisense/sense transcriptional organization is evolutionarily conserved in both human and mouse (Benzow and Koob, 2002). Primate comparison shows human-specific features, with longer human alleles due to a novel variable trinucleotide repeat, not present in non-human primates, which increased the disease-causing expansion likelihood (Andres et al., 2003). Since the *ATXN8OS* gene functions as a gene regulator, it has been proposed that a RNA gain-of-function mechanism might underlie neurodegeneration in SCA8 as is the case for DM (Liquori et al., 2001; Philips et al., 1998). If the CTG expansion leads to an accumulation of the *ATXN8OS* transcript, it could prevent expression of the *KLHL1* gene. In addition to this possible antisense effect, the expanded CUG tract in the *ATXN8OS* transcript may impair nuclear cytoplasmic transport through the formation of extended hairpin loops resulting in nuclear retention and formation of nuclear foci and affecting the transport of other CAG repeats containing mRNAs (Amack et al., 1999; Davis et al., 1997; Sasagawa et al., 1999; Taneja et al., 1995). Furthermore, the RNA containing the CUG repeat expansion may increase its affinity to CUG-binding proteins, which regulates the alternative splicing of specific pre-mRNAs (Timchenko et al., 2001).

Recently, Moseley and colleagues reported a transgenic mouse model in which the full-length human *ATXN8OS* mutation is transcribed using its endogenous promoter. 1C2-positive intranuclear inclusions in cerebellar Purkinje and brainstem neurons in SCA8 expansion mice and human SCA8

autopsy tissue result from translation of a polyglutamine protein, encoded on a previously unidentified antiparallel transcript (ataxin 8, *ATXN8*) spanning the repeat in the CAG direction. The expression of noncoding *ATXN8OS* CUG expansion transcripts and the discovery of intranuclear polyglutamine inclusions caused by translation of *ATXN8* suggest that SCA8 pathogenesis may involve toxic gain-of-function mechanisms at both the protein and RNA levels (Moseley et al., 2006).

Kelch-like 1 protein is highly homologous to the *Drosophila* kelch protein, which is responsible for maintaining actin organization of ring canals connecting the oocyte to support nurse cells during oogenesis (Robinson and Cooley, 1997; Xue and Cooley, 1993). Previous study demonstrated that Kelch-like 1 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 Kelch-like 1 protein might have a role for modulating calcium channel. Loss of *KLHL1* expression from even a single allele causes abnormal gait, progressive loss of motor coordination, and Purkinje cell dendritic deficits. Mice with *KLHL1* specifically deleted in only Purkinje cells have the same phenotype, indicating that Kelch-like 1 protein is essential for normal motor coordination and for maintaining normal Purkinje cell functions (He et al., 2006).

The internal ribosome entry site (IRES) activity of *ATXN8OS* mRNA

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 implies that any mRNA with long 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 distance from the cap structure (Hellen and Sarnow, 2001). This cap-independent mechanism requires the formation of a complex RNA structural element termed an internal ribosome entry site (IRES) and the presence of *trans*-acting factors (Stoneley and Willis, 2004). IRESs were first identified in picornvirus RNAs (Jang et al., 1988), but many current examples of eukaryotic cellular mRNAs have been shown to contain IRES activities (http://www.rangueil.inserm.fr/IRESdatabase). Studies of factors required for IRES-dependent translation have revealed that RNA-binding proteins such as polypyrimidine tract binding protein (PTB), poly(rC) binding protein (PCBP), and upstream of N-ras protein (unr) specifically enhance the translation of IRES-containing mRNA by disrupting RNA-RNA interactions (Stoneley and Willis, 2004). As a result, the ribosome entry window attains an unstructured conformation and in doing so facilitates ribosome recruitment. In addition, the IRESs activity can be regulated by the cell cycle (Pyronnet et al., 2000), developmental stage (Creancier et al., 2000), apoptosis (Clemens and Bommer, 1999; Clemens et al., 2000), and cellular stress (Fernandez et al., 2002a; Fernandez et al., 2001; Fernandez et al., 2002b).

Although the *ATXN8OS* gene lacks a significant open reading frame, three small non-significant open reading frames (ORFs) in the *ATXN8OS* transcripts

were noted (Koob et al., 1999). Among them, a 102 amino acids containing-ORF1 and a 41 amino acids plus a polyleucine tract containing-ORF3 may be translated if *ATXN8OS* RNA possesses a cap independent IRES activity. Our previous study had demonstrated the possible IRES activity within *ATXN8OS* (Lin, 2007). Therefore, if small ORFs could be translated, the regulation mechanism of *ATXN8OS* IRES activity remains to be further investigated for the linkage between translatable ORFs and pathogenesis.

Cellular model approach for neurodegenerative diseases

Cell culture systems are generally a single cell type, presenting fewer confounding factors that might influence the experimental results. To investigate the molecular pathogenesis underlying the SCA8 cell dysfunction, *ATXN8OS* CR cell lines containing 0, 23, 88 and 157 CR were generated (Lin, 2007). These cell lines were originated from human embryonic kidney 293 cells, which express many neuron-specific mRNAs (Shaw et al., 2002). A large body of work on other repeat expansion diseases with similar neuronal pathology using this cell line has been reported (Anborgh et al., 2005; Handa et al., 2005). The derived *ATXN8OS* cell lines are isogenic except for the number of CTA/CTG combined repeats. The repeat number in these cell lines was stable (Lin, 2007).

For many neurological disorders, access to central nervous tissues for diagnostic or research purposes is not practically available. Although the pathology of SCA is mainly in cerebellum, most mutant proteins in SCAs are ubiquitously distributed including peripheral lymphocytes. Transformation of

peripheral B-lymphocytes by Epstein-Barr virus (EBV) is the method of choice for generating lymphoblastoid cell lines. Considering the similarity of certain gene expression and regulation between lymphoblastoid cell lines and neurons, lymphoblastoid cell lines have been considered as surrogate cells in the study of neurological disorders. Thus, lymphoblastoid cells are frequently used as a model to study SCA (Tsai et al., 2005; Wen et al., 2003) and can be a useful cell model to explore the disease pathogenesis.

Specific Aims

Many questions concerning association and pathogenic role between SCA8 expansion mutation and neurodegenerative diseases still remain unanswered. The consequences of many neurodegenerative diseases are similar, suggesting they may share a common pathogenic mechanism. Examination into the molecular processes might present novel possibilities for therapy. The main purpose of this present proposal is to dissect the possible factors involved in SCA8 pathogenesis. In view of the complex mechanisms leading to SCA8, stably induced HEK-293 cell lines carrying 0, 23, 88 and 157 CR were established to investigate the possible epigenetic and post-transcriptional regulations of the *ATXN8OS* expression. The stably induced HEK-293 cell lines were also used to perform gene expression through cDNA microarray to gain insights into the events of pathogenesis. Secondly, since SCA8 involves bidirectional expression of CUG (*ATXN8OS*) and CAG (*ATXN8*) expansion transcripts and is known to be an antisense regulator of *KLHL1*, the expression of *ATXN8OS*, *ATXN8* and *KLHL1* were evaluated using our lymphoblastoid cell models. Furthermore, the *trans*-acting factors and stress situations which might be involved in the regulation of *ATXN8OS* IRES activity were examined. Site-directed mutagenesis approach was applied to ascertain the *ATXN8OS* ribosome entry window. Additionally, antiserum against ATXN8OS-ORF1-GST (*S. japonicum* GST from pGEX plasmid) fusion protein was raised in rabbits to detect possible *ATXN8OS* ORF proteins. The application of these analyses could be very informative for identification of the mechanism involved in SCA8. Efforts were made in

anticipation of isolating potential therapeutic targets and ultimately facilitating drug discovery.

Materials and Methods

I. Analysis of *ATXN8OS* stably induced HEK-293 cell lines

Flp-In T-REx 293 cell lines stably expressing *ATXN8OS* cDNA

The inducible Flp-In T-REx 293 cell lines stably expressing *ATXN8OS* with different CTG repeats were generated by Hsuan-Yuan Lin (Lin, 2007) based on the Flp-In T-REx system (Invitrogen). Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplement with 10% fetal bovine serum (FBS), 1.0 mM sodium pyruvate, 1.5 g/L sodium bicarbonate, 100 U/ml penicillin, 100 U/ml streptomycin, and 15 μg/ml blasticidine at 37°C in an atmosphere containing 5% CO_2. To induce expression of *SCA8* carrying 0, 23, 88, and 157 CTG repeats, cells were treated with 1 μg/ml doxycycline for varying periods of time.

RNA isolation

Total RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's specifications. After PBS washing, 1ml of Trizol reagent was added to the culture dishes and cells were scraped from the culture dish. The cell suspension was incubated on ice for 5 minutes, mixed well with 1/5 volume of chloroform and incubated on ice for another 5 minutes. RNA was separated from DNA and proteins by centrifugation at 4˚C for 15 minutes. The

colorless, upper aqueous phase was carefully removed to a fresh tube avoiding the material that collected at the interface, and mixed with 0.8 volume of isopropanol. The mixture was sat at -20˚C for at least 1 hour and centrifuged at 4˚C for 15 minutes to precipitate RNA. The supernatant was discarded and RNA pellet was rinsed with 70% ethanol in DEPC-ddH_2O. RNA was air dried and dissolved in DEPC-ddH_2O.

Real-time RT-PCR (Quantitative RT-PCR)

Total RNA isolated from inducible HEK-293 *ATXN8OS* cell lines was treated with RNase-free DNase (Stratagene) and used as a template to synthesize the first-strand cDNA by High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with random primers. Real-time quantitative PCR experiments were performed in the ABI PRISM® 7000 Sequence Detection System (Applied Biosystems). Amplification was performed on a cDNA amount equivalent to 250 ng total RNA and with gene-specific TaqMan fluorogenic probes Hs01382089-m1 (exon C2 and C1 boundary) for *ATXN8OS*, Hs00153277-m1 for *CDKN1B*, Hs00391824-m1 for *SPON1*, Hs00168310-m1 for *GSTP1*, Hs01099687-m1 for *ISL1* and 4326321E for *HPRT1* (endogenous control) (Applied Biosystems). The amount of HaloTag mRNA was determined by customized Assays-by-Design probe (Forward primer: CCGACGTGGGACGAATGG, Reverse primer: CGGAAGGCCTGGAAGGT, TaqMan® probe: GAATTCGCCCGTGA) (4331348, Applied Biosystems). Fold change was calculated using the formula $2^{\Delta Ct}$, $\Delta C_T = C_T$ (control) – C_T (target), in which C_T indicates cycle threshold.

Western blot analysis

ATXN8OS CR cells were lysed in RIPA buffer (10 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA pH 8.0, 0.1% sodium dodecyl sulphate (SDS), 1% deoxycholate, 1% NP-40) containing the protease inhibitor mixture (Sigma). After sonication and sitting on ice for 30 minutes, the lysates were centrifuged at 13,000 g for 30 minutes at 4°C. Protein concentrations were determined with the Bio-Rad protein assay kit, using albumin as standards. Laemmli sample buffer was then added to 30 μg of protein and heated in a boiling water bath for 10 minutes. Equal amounts of protein from each sample were fractionated in a 12% SDS-polyacrylamide gel electrophoresis (PAGE). The fractionated protein samples were transferred onto a nitrocellulose membrane (Schleicher and Schuell), and non-specific binding was blocked in 5% non-fat dry milk for overnight at 4°C. After washing with Tris-buffered saline (TBS), the blots were probed with a 1:1000 dilution of HaloTag antibody (Promega) or a 1:10000 dilution of actin antibody (Chemicon) in TBS/1% bovine serum albumin/0.1% Tween 20. After extensive washing, the blots were probed with a 1:10000 dilution of goat anti-mouse or goat anti-rabbit IgG conjugated to horseradish peroxidase (Jackson ImmunoResearch). The blots were washed extensively and the protein detected using Immobilon™ western chemiluminescent HRP substrate (Millipore).

Immunocytochemical staining

ATXN8OS CR cells on coverslips were washed with PBS and fixed in 4% paraformaldehyde in PBS for 10 minutes, followed by 20 minutes incubation with 0.1% Triton X-100 in PBS to permeabilize cells, and overnight

incubation with 0.5% bovine serum albumin in PBS to block non-specific binding. The primary anti-HaloTag antibody, diluted 1:500 in 1% BSA in TBS, was used to stain cells at 4°C overnight. After washing, cells were incubated for 2 hours at room temperature in FITC-conjugated secondary antibody diluted to 1:500 in TBS containing 1% BSA, and washed in PBS. Nuclei were detected using 4'-6-diamidino-2-phenylindole (DAPI). The stained cells were examined after mounted in Vectashield (Vector Laboratories Inc.) using a Leica TCS confocal laser scanning microscope.

Fluorescent *in situ* hybridization (FISH)

To examine the ribonuclear foci, cells were grown on coverslips, washed, and fixed for 15 minutes at room temperature in 4% formaldehyde and 10% acetic acid. After 0.1% Triton X-100 treatment for 10 minutes, a Cy3-$(CAG)_{10}$ (Operon) or Cy5-CTGCGACTCCGCTGGAAACTCTTCAGCCA (unique to *ATXN8OS*) oligonucleotide probe was added at 37°C for 2 hours for fluorescent *in situ* hybridization (FISH) experiments (http://www.singerlab.org/protocols). Nuclei were detected using DAPI. Fluorescent signals are visualized using a Leica TCS confocal laser scanning microscope optimized for simultaneous dual fluorescent imaging.

RNA clean-up

ATXN8OS CR cells were induced by 1 μg/ml doxycycline for 7 days and treated with 50 nM Staurosporine (an external apoptotic stimulus) for 24 hours. Isolated RNA was cleaned up using RNeasy Mini Kit (Qiagen). As described

in the manufacture, the volume of RNA sample was first adjusted to 100 µl with DEPC-ddH_2O, mixed with 350 µl Buffer RLT containing 1% β-mercaptoethanol, followed with 250 µl of ethanol. The mixture was then applied to the RNeasy mini column and centrifuged for 15 seconds at ≥ 8,000 g. The column was transferred into a new collection tube, added 500 µl Buffer RPE, and centrifuged for 15 seconds at ≥ 8,000 g. Another 500 µl Buffer RPE was added to the column and centrifuged for 2 minutes at ≥ 8,000 g. And another centrifugation for 1 minute was carried out to eliminate residual ethanol. Finally, 20 µl of DEPC-ddH_2O was added the column which has been transferred to a new tube, stayed at room temperature for 3 minutes, and centrifuged at ≥ 8,000 g for 1 minute. The quality and quantity of cleaned-up RNA samples were determined by the agarose electrophoresis and the absorbance at 260 nm, respectively.

Microarray analysis

The cleaned-up RNA was committed to Phalanx Biotech Group, Inc. to identify differentially expressed genes between inducible HEK-293 *ATXN8OS* cell lines by microarray analysis. The array image was scanned and the acquired signal intensities were normalized. The cut-off values of up-regulated and down-regulated transcripts were set as 3 and 0.7 fold changes, respectively.

Protein sample preparation

The HEK-293 cells were transiently transfected with pEF vector,

pEF-ATXN8OS-23 CR and 157 CR plasmids (Lin, 2007). Cells were harvested and lysed in 100 µl Lysis solution (8 M urea, 4% CHAPS, 2% biolyte 3-10, 40 mM DTT). Protein extracts were centrifuged at maximum speed at 4°C for 30 minutes and the supernatant was transferred to a fresh tube. To separate proteins in the crude extracts from contaminating substances, the protein extracts were applied to the 2-D Clean-Up Kit (Amersham Biosciences). The cleaned-up protein pellet was resuspended in 100 µl of Lysis solution with sonication. The quantity of cleaned-up protein sample was determined using 2-D Quant Kit (Amersham Biosciences).

Proteomic analysis

For the first dimension, the protein sample was mixed with rehydation buffer (8 M Urea, 2% CHAPS, 0.5% IPG buffer, 0.002% bromophenol blue) to make up for the volume 250 µl (for 13-cm IPG strips). The sample mixture was loaded in the strip holders evenly and a pH 3-10, 13-cm Immobiline DryStrip (Amersham Biosciences) was faced down to absorb the protein. Strips were rehydrated and protein samples were separated according to each isoelectric point (pI) on IPGphor II (Amersham Biosciences). After IEF step, strips were equilibrated in SDS Equilibration buffer (2% SDS, 50 mM Tris-HCl pH8.8, 6 M Urea, 30% Glycerol, 0.002% bromophenol blue) containing 1% (w/v) dithiothreitol (DTT) for 15 minutes followed by the second equilibration in SDS Equilibration buffer containing 2.5% (w/v) iodoacetamide for 15 minutes. The strip then was transferred to the polyacrylamide gel and sealed with 0.5% agarose. Each gel was run at 15 mA for the first 15 minutes followed by 30 mA for 5 hours. The gels were fixed in

fix buffer (50% EtOH, 8% Phosphoric acid) for overnight. After briefly rinsed in ddH_2O, the gels were stained in Coomassie Blue G-250 for 24 hours. After staining, the gels were destained in 1% acetic acid for 30 minutes and scanned on scanner. ImageMaster 2D Platinum Software (Amersham Biosciences) was used for analyzing the images. Differentially expressed protein spots were picked from the stained gel, rinsed with ddH_2O, and then washed with 50 mM ammonium bicarbonate (NH_4HCO_3)/acetonitrile. The gel piece was incubated with freshly diluted trypsin (2.5 ng/μl in NH_4HCO_3) at 37˚C overnight and the resulting peptides were extracted by 50% acetonitrile containing 1% trifluoroacetic acid. Trypsinized protein fragments were identified using the MALDI-TOF MS (conduction of Genomic Medicine Research Core Laboratory in Chang Gung Memorial Hospital). The MALDI spectra used for protein identification from trypsinized fragments were searched against the National Center for Biotechnology Information (NCBI) protein databases using the MASCOT search engine (www.matrixscience.com).

II. Analysis of *ATXN8OS* lymphoblastoid cell models

Lymphoblastoid cell lines

Lymphoblastoid cell lines from two SCA patients [P1, 88 repeats; P3, 95~185 repeats] and four PD patients [P2, 92 repeats; P4, 87 repeats; P5, 82 repeats; P6, 84 repeats] as well as five normal controls (C1 ~ C5) were established (**Table 1**) (Food Industry Research and Development Institute, Taiwan) after obtaining informed consent. Patients P1 and P3 were recruited

from Taipei Veterans General Hospital; the rest patients and controls were recruited from Chang Gung Memorial Hospital. Cells were maintained in RPMI 1640 medium (GIBCO) containing 10% FBS.

Real-time RT-PCR (Quantitative RT-PCR)

Total RNA was extracted, treated, quantified, and reverse-transcribed to cDNA as pervious description. Using ABI PRISM® 7000 Sequence Detection System (Applied Biosystems), real-time quantitative PCR was performed on a cDNA amount equivalent to 250 ng total RNA with TaqMan fluorogenic probes Hs01382089-m1 for *ATXN8OS*, Hs00252991_m1 for *KLHL1* and 4326321E for *HPRT1* (endogenous control) (Applied Biosystems). The amount of *ATXN8* mRNA was determined by customized Assays-by-Design probe (Forward primer: ACTAGAATGACAGGTCAAGATGGTTAGT, Reverse primer: TCGAAGTATGAGAAGTACCTATCATATTTTGGTA, TaqMan® probe: CTGGAGCCAAGAACTAG) (4331348, Applied Biosystems). Fold change was calculated using the formula $2^{\Delta Ct}$, $\Delta C_T = C_T$ (control) – C_T (target), in which C_T indicates cycle threshold. Statistical analysis of differences between the groups was carried out using one-way analysis of variance (ANOVA).

Genotyping, sequencing and RFLP analysis of *ATXN8OS* exon A -62 G/A SNP

DNA was extracted from peripheral blood leukocytes using the DNA Extraction Kit (Stratagene). The alleles with CTG expansion were gel purified

and sequenced directly to accurately assess repeat size and the presence of interruptions. The DNA fragments containing the -62 G/A SNP were amplified using PCR with primers for RFLP analysis (Forward: 5'-TGATGTTATAATTGTTATATATTTATGCA; Reverse: 5'-AACTAACT CAACATCCAGATAATTT). The SNP was differentiated using the *Mph*1103I restriction enzyme.

III. Analysis of the IRES activity of *ATXN8OS* RNA and identification of *ATXN8OS* ORF protein

cDNA cloning of IRES initiation *trans*-acting factors

For PTB, the coding region was amplified by PTB-F (5'-GCGGTCTGCTCTGTGTGC) and PTB-R (5'-TGATGGAAGTTG TCGCCAG). After cloning into pGEM-T Easy and sequencing, the 660 bp of PTB coding sequence was then cloned into the *Eco*RI site of pcDNA3.1 vector (Invitrogen). For PCBP1, the 1.1 kb PCBP1 cDNA was amplified from cDNA of HEK-293 cells using PCBP1-U1 (5'-AAGCTTCATGGATGCC GGTGTGACTGAAAG) and PCBP1-D1 (5'-CTCGAGATCATGGGAGAAC AGCAGAAAGGG) primers. After cloning into pGEM-T Easy and sequencing, the 1.1 kb PCBP1 coding sequences with *Hin*dIII and *Xho*I sites at 5' and 3' ends respectively was subcloned into pcDNA3.1. The pcDNA3.1 constructs were verified by restriction mapping.

Transfection

The rapid transfection process was carried out according to the Lipofectamine™ 2000 reagent (LF2000) (Invitrogen) protocol. For transfection performed in the 24-well plate, 1.5 μg of DNA and 2 ~ 3 μl of LF2000 were first diluted into 50 μl Opti-MEM® I Reduced Serum Medium (Invitrogen), respectively, and incubated for 5 minutes at room temperature. Then DNA and LF2000 were combined and incubated for 20 minutes at room temperature. At the meanwhile of DNA-LF2000 complexes formation, the cell suspension of appropriate number per well was prepared in 10% FBS-DMEM without antibiotics. The cell suspension was added to the DNA-LF2000, mixed gently by rocking back and forth, and incubated at 37˚C for at least 24 hours for further studies.

Dual luciferase reporter assay

The dual luciferase reporter constructs were established by Hsin-Chieh Shiau. To evaluate the effects of stress-induced drugs, different concentrations of thapsigargen were added after transfection for 12 hours. Then cells were harvested after 12 hours and washed by PBS. The cell lysates were prepared following the Dual-Luciferase® Reporter Assay protocol (Promega). The IRES activity was directly measured by the ratio of the firefly luciferase level to the *Renilla* luciferase level by a luminometer using a dual luciferase assay system.

In vitro transcription

Plasmid DNA containing *ATXN8OS* 23 CR and *ATXN8OS* 157 CR were linearized by *Spe*I digestion. Transcripts were synthesised according to the manufacturer's recommendations (Ambion) in a reaction mixture containing 1× reaction buffer, 7.5 mM ATP, 7.5 mM CTP, 7.5 mM GTP, 7.5 mM UTP, 1 μg of linear DNA template and 2 μl T7 polymerase enzyme mix to a final volume of 20 μl. For biotinylated RNAs, Biotin-11-UTP and UTP were mixed in a ratio of 1:7. The reaction mixture was incubated at 37°C for at least 16 hours and DNA template was removed by TURBO DNase at 37°C for 15 minutes. The RNA samples were then recovered by lithium chloride (LiCl) precipitation and determined by the agarose electrophoresis and the absorbance at 260 nm.

RNA-binding assays and protein identification

To prepare protein lysates, HEK-293 cells were harvested by centrifugation at 1,000 g for 5 minutes followed with cold PBS washing twice. The cell pellet was resuspended in 100 μl RIPA buffer (10 mM Tris pH7.5, 150 mM NaCl, 5 mM EDTA pH8.0, 0.1% SDS, 1% DOS, 1% NP-40) containing the protease inhibitor mixture, sonicated 15 pulses three times, and sat on ice for 30 minutes. Protein extracts were centrifuged at maximum speed at 4°C for 30 minutes and the supernatant was transferred to a fresh tube. The protein concentration was determined using the Bio-Rad Protein Assay (Cat. No. 500-0006, Bio-Rad). 15 μg Biotin-11-UTP-labeled *ATXN8OS* RNAs were incubated with 1 mg protein lysates at 30°C for 15 minutes in RNA mobility

shift buffer (40 mM KCl, 5 mM HEPES pH7.1, 2 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, and 0.25 μg/μl tRNA). For the binding reaction with biotinylated RNAs, streptavidin magnetic particles and and a magnetic particle separator (Roche) were used to capture the RNA-protein complex. The captured proteins were purified by a 2-D Clean-Up Kit (Amersham Biosciences) and the first-dimension separation was performed on 7 cm IPG strips by cup loading. The second-dimension separation was performed by vertical PAGE (12.5%). After gel electrophoresis, proteins were stained with SYPRO Ruby (Cat. No. S12000, Molecular Probe) and the gels were imaged on the Molecular Dynamics Typhoon 9210 (Amersham Biosciences). ImageMaster 2D Platinum Software (Amersham Biosciences) was used for analyzing the images. The RNA binding proteins were identified by MALDI-TOF/TOF MS (Bruker Daltonics).

Site-directed mutagenesis

The introduction of each upstream AUG was performed by PCR mutagenesis. Site-directed mutagenesis was carried out using QuikChange™ XL Site-Directed Mutagenesis Kits (Cat. No. 200517, Stratagene). The reaction was set up in 25 μl of volume including 2.5 μl of 10 × reaction buffer, 10 ng of template DNA, 62.5 ng of both sense and antisense primers, 50 μM dNTPs, 1.5 μl of QuickSolution and 0.5 μl of PfuTurbo DNA polymerase. The reaction was temperature cycled with the condition: 1 cycle of denaturation at 95˚C for 1 minute, 18 cycles of reaction comprising denaturation at 95˚C for 50 seconds, annealing at 60˚C for 50 seconds, and elongation at 68˚C for 12 minutes, followed by elongation at 68˚C for 7 minutes. Following cycling, the

reaction was cooled and 0.5 μl of *Dpn*I (10 U/μl) was added to digest the parental DNA template at 37˚C for 1 hour. Then 5 μl of the reaction was taken to transform the competent cells and the site-directed mutant DNA was extracted from the cultured clones and sequenced.

Fluorescence activated cell sorting (FACS) analysis

HEK-293 cells were transfected with the various cDNA constructs using a lipofection procedure. pEGFP-N1 was used as a positive control. 48 hours later, cells were harvested for fluorescence activated cell sorting (FACS) analysis. The amounts of GFP expressed were analyzed in a FACSCalibur flow cytometer (Becton-Dickinson), equipped with an argon laser operating at 530 nm. A forward scatter gate was established to exclude dead cells and cell debris from the analysis. 10^4 cells were analyzed in each sample.

***ATXN8OS* ORF-EGFP constructs**

The 1.3-kb *ATXN8OS* cDNA containing exons D, C2, C1, B, and A (Nemes et al., 2000) was cloned as previously described (Lin, 2007). The ORF translation termination sequence in C1 exon was removed and a *Sma*I restricted site (underlined) added by PCR using primer 5'-GCGCCCGGGACACTTCAACTTCCTATACATACA. The *Eco*RI (in MCS of pGEM-T Easy vector)-*Sma*I fragment containing *ATXN8OS* ORF was in-frame fused to the EGFP gene in the pEGFP-N1 vector (between the *Eco*RI and *Bst*UI sites). Portion of the Kozak consensus translation initiation sequence (ACCATG) in the EGFP gene was further removed by site-directed

mutagenesis (primer 5'-CGGGCCCGGGATCCACCGGTCGCCΔGTGAGC AAGGGCGAGGAGCTG, Δ = ACCATG). The resulting pCMV/+801 construct (where +801 represents transcription start site of exon D) (**Fig. 9C**) was verified by DNA sequencing.

To construct pCMV/+1 (+1 representing transcription start site of exon D5), an *Xho*I site (underlined) was added to the 5' end of +1 sequence by PCR using primer 5'-CGCCTCGAGTTTATGTCTATGAAGG. The *Xho*I-*Sac*I fragment containing *ATXN8OS* +1 ~ +848 sequences was inserted between the *Xho*I (in MCS of pEGFP-N1) and *Sac*I (in exon D of *ATXN8OS*) sites of pCMV/+801 to generate pCMV/+1. To construct pATXN8OS/-481, a 2.1-kb *ATXN8OS* gene 5' fragment (AF252279 reversed complemented strand: 108333 ~ 110454) was cloned by PCR and sequenced. The *Afl*III-*Sac*I fragment containing *ATXN8OS* -481 ~ +848 sequences was used to replace the corresponding fragment containing CMV promoter in pCMV/+800 to generate pATXN8OS/-481. Then the *Afl*III-*Eco*RV fragment containing *ATXN8OS* -481 ~ -115 sequences in pATXN8OS/-481 was further removed to generate pATXN8OS/-114.

HEK-293 cell cultivation and transfection

Human embryonic kidney HEK-293 cells cultivated in DMEM containing 10% FBS were plated into 6-well (6×10^5/well) or 12-well (2×10^5/well) dishes, grown for 20 hours and transfected by the lipofection method (GibcoBRL) with the ORF-EGFP or pIRES-EGFP plasmid (4 μg/6-well or 2 μg/12-well). The cells were grown for 48 hours for the following EGFP fluorescence (12-well dishes), RNA and protein (6-well dishes) studies.

MitoTracker staining

HEK-293 cells were grown on coverslips and transfected with different cDNA constructs. After 48 hours the media were removed and the cells were incubated with the staining solution containing 400 nM MitoTracker® probe (Invitrogen) for 2 hours. After staining is complete, the cells were washed with PBS and fixed with 4% formaldehyde at 37°C for 15 minutes. Cells were examined after mounted in Vectashield by Leica TCS confocal laser scanning microscope.

LysoTracker staining

HEK-293 cells were grown on coverslips and transfected with different cDNA constructs. After 48 hours the media were removed and the cells were incubated with the staining solution containing 75 nM LysoTracker® probe (Invitrogen) for 2 hours. After staining is complete, the cells were washed with PBS and fixed with 4% formaldehyde at 37°C for 15 minutes. Cells were examined after mounted in Vectashield by Leica TCS confocal laser scanning microscope.

Confocal microscopy

For visualizing intracellular ORF-EGFP proteins, transfected cells on coverslips were fixed in 4% paraformaldehyde for 15 minutes. Nuclei were detected using 0.05% DAPI. Cells were examined after mounted in Vectashield for GFP and DAPI fluorescence using a Leica TCS confocal laser

scanning microscope optimized for simultaneous dual fluorescent imaging.

Real-time RT-PCR (Quantitative RT-PCR)

Total RNA was extracted, treated, quantified, and reverse-transcribed to cDNA as pervious description. Using ABI PRISM® 7000 Sequence Detection System (Applied Biosystems), real-time quantitative PCR was performed on a cDNA amount equivalent to 250 ng total RNA with TaqMan fluorogenic probes Hs01382089-m1 (exon C2 and C1 boundary) for *ATXN8OS* and 4326321E for *HPRT1* (endogenous control) (Applied Biosystems). Additional customized Assays-by-Design probe (Forward primer: ACTGCATTT CAGGAGCAAAAAGAGA, Reverse primer: GTCCCTGTGGTTTGAATCT ATTCCA, TaqMan® probe: CAGTGGCCTCATTTTG) (*ATXN8OS* exon D5/D4 region, Applied Biosystems) was used for *ATXN8OS* mRNA quantification. Fold change was calculated using the formula $2^{\Delta Ct}$, $\Delta C_T = C_T$ (control) – C_T (target), in which C_T indicates cycle threshold. Statistical analysis of differences between the groups was carried out using one-way analysis of variance (ANOVA).

Western blot analysis

Cells were lysed in RIPA buffer (10 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA pH 8.0, 1% sodium deoxycholate, 1% NP-40 and 0.1% SDS) containing the protease inhibitor mixture (Sigma). After sonication and sitting on ice for 20 minutes, the lysates were centrifuged at 14,000 rpm for 30 minutes at 4°C. Protein concentrations were determined with the Bio-Rad

protein assay kit, using albumin as standards. Proteins (25 μg) were electrophoresed on 10% SDS-polyacrylamide gel and transferred onto nitrocellulose membrane by reverse electrophoresis. After being blocked, the membrane was stained with anti-GFP (1:200 dilution, Santa Cruz Biotechnology) or anti-actin (1:10000 dilution, Chemicon International) antibody. The immune complexes were detected using horseradish peroxidase-conjugated goat anti-mouse (Jackson ImmunoResearch) or goat anti-rabbit (Rochland) IgG antibody (1:10000 dilution) and Immobilon™ Western Chemiluminescent HRP substrate (Millipore).

Lymphoblastoid and neuroblastoma cell lines

Lymphoblastoid cell lines were maintained as previous description. Human neuroblastoma SK-N-SH, SH-SY5Y and IMR-32 cells were cultivated in DMEM (IMR-32 and SK-N-SH) or 1:1 mixture of DMEM and F12 medium (SH-SY5Y) containing 10% FBS.

GST-ORF construct and antiserum

To construct GST-tagged ORF for antiserum production, *Bst*BI and *Eco*RI sites (underlined) were added to the 5' and 3' ends of *ATXN8OS* ORF by PCR using primers 5'-GCGCTTCGA**ATG**TGCTTCACATCGAAGTC and 5'-CC GGAATTC**TCA**ACACTTCAACTTCCTATAC (initiation and termination codons in boldface). The 317-bp *Bst*BI-*Eco*RI fragment containing *ATXN8OS* ORF sequences was then inserted between the *Acc*I (location 928) and *Eco*RI (location 944) sites of pGEX-5X-3 (GE Healthcare). The location 928 *Acc*I

site (underlined) used was added by site-directed mutagenesis using primer 5'-GATCTGATCGAAGGTCG**AC**GGATCCCC AGGAATTCC (mismatch nucleotides in boldface). The resulting pGST-ORF construct was verified by DNA sequencing and introduced into BL21(DE3)pLysS (Novagen). After IPTG induction, the 36-kDa antigen was purified using GST·Bind™ resin (Novagen) and used to raise antiserum in rabbit (LTK BioLaboratories).

ORF identification

Cells (10^7; lymphoblastoid, neuroblastoma or HEK-293) were lysed in 8 M urea lysis buffer (8 M urea, 4% CHAPS, 2% biolyte 3-10, 40 mM DTT). After ultrasonic homogenizing, protein extracts were centrifuged at 13,000 rpm at 4°C for 30 minutes and the supernatants transferred to new tubes. Insoluble pellets were then resuspended in 1:5 v/v SDS buffer (1.7% SDS, 20mM Tris) and 9.8 M urea lysis buffer by sonication. For Western blotting, proteins (30 μg) and aliquot of pellet suspension were separated on 12% SDS-PAGE and blotted. For 2D PAGE and 2D immunoblot, aliquots of pellet suspension were first separated using Immobiline DryStrip (7 cm, pH 3-10) (GE Healthcare) and further separated by a 12.5% SDS-PAGE. The blotting membranes were stained with ORF antiserum (1:200 dilution) or actin antibody (1:10000 dilution, Chemicon) and immune complexes detected as described. The 2D gel was stained with SYPRO Ruby (Molecular Probe) and scanned on a Typhoon 9400 imager (GE Healthcare). The map was compared to the 2D immunoblot. The ORF-specific spots were punched out and subjected to reduction and alkylation by DTT/iodoacetamide, followed by in-gel digestion with freshly prepared Trypsin Gold (2.5 ng/μl, Promega) at

37°C for overnight. The obtained peptides were extracted with 50% acetonitrile containing 1% trifluoroacetic acid and tandem mass spectra were generated by liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) at Proteomics and Protein Function Core Laboratory, Center of Genomic Medicine, National Taiwan University. MS/MS data were searched using the Mascot search engine (www.matrixscience.com) in a database containing theoretical trypsinized fragments of 23-kDa ORF protein initiated at +953 GUG codon.

Results

I. Analysis of *ATXN8OS* stably induced HEK-293 cell lines

Repeat length-related change in *ATXN8OS* expression

The expression of the *ATXN8OS* in 0, 23, 88 and 157 CR lines was driven by the same hybrid CMV/$TetO_2$ promoter (**Fig. 1A**). To examine *ATXN8OS* mRNA expression in *ATXN8OS* stably induced HEK-293 cell lines, *ATXN8OS* RNA levels were measured by real-time PCR quantification using *ATXN8OS*-specific probe and primers. The expression of the endogenous *ATXN8OS* RNA in vector only cell line was too low to be efficiently detected. In the absence of doxycycline, all *ATXN8OS* CR cell lines expressed low level of *ATXN8OS* RNA, ranging from 0.017 to 0.042 compared with endogenous *HPRT1* (**Fig. 1B**). A repeat length-dependent repression of *ATXN8OS* expression is notable, with $(CR)_n$/relative *ATXN8OS* RNA level being $(CR)_0$/1.0, $(CR)_{23}$/1.0, $(CR)_{88}$/0.6, and $(CR)_{157}$/0.4 (**Fig. 1C**).

The *ATXN8OS* RNA levels were further examined in these CR cells. After induction with doxycycline for 1 and 2 days, the amount of *ATXN8OS* RNA in 0, 23, 88 and 157 CR cells increased, respectively, to 29~32, 25~27, 24~25, and 19~20 times more that in 0 CR cells at the time of doxycycline addition (0 d) **(Fig. 1C**). When the amount of *ATXN8OS* RNA present at the time of doxycycline addition was set to 100% for each CR cell line, the fold of induction increase with repeat length, with $(CR)_n$/fold of induction being

$(CR)_0$/29~32, $(CR)_{23}$/25~27, $(CR)_{88}$/41~42, and $(CR)_{157}$/47~50 (**Fig. 1D**).

Repeat length-dependent repression of *HaloTag* gene located next to *ATXN8OS* cDNA gene

The expression of the *ATXN8OS* in 0, 23, 88 and 157 CR lines was driven by the same hybrid CMV/$TetO_2$ promoter (**Fig. 1A**). As CUG triplet repeat expansion in DM1 may alter the adjacent chromatin structure (Otten and Tapscott, 1995), the observed repeat length-dependent repression of *ATXN8OS* expression may be due to chromatin remodeling. Our previous results of methylation of H3-K9 and hypoacetylation of H3-K14 in 157 CR cells (Lee, 2008) also support the notion that the chromatin remodeling participated in the repeat length-dependent repression of *ATXN8OS* expression.

The cloning vector used to generate *ATXN8OS* CR lines was modified by placing a HaloTag gene downstream of the *ATXN8OS* cDNA gene (**Fig. 1A**). If chromatin structure was affected in expanded CR lines, reduced expression of HaloTag would be expected. To examine this, HaloTag RNA level relative to endogenous *HPRT1* RNA was first quantified by real-time PCR. As shown in **Fig. 2A**, HaloTag RNA expression in expanded 88 and 157 CR lines was significantly reduced (70~71%) when compared to the normal 23 CR line ($P < 0.05$). To confirm the expression change, proteins were collected and subjected to western blotting with HaloTag and β-actin antibodies. Consistent with the results of real-time PCR quantification, expression levels of HaloTag protein were significantly decreased in *ATXN8OS* 88 and 157 CR lines as compared to that of 23 CR line (**Fig. 2B**, 79~81%, $P < 0.05$). Fluorescence microscope examination after immunocytochemical staining using HaloTag

antibody also revealed the reduced expression of HaloTag protein (**Fig. 2C**).

Increased *ATXN8OS* transcript stability and ribonuclear foci formation with CUG repeat expansion

To examine the observed repeat length-dependent increase of fold of induction, the effect of repeat length on the stability of *ATXN8OS* RNA was investigated. The *ATXN8OS* cells were grown in the presence of doxycycline for 48 hours and actinomycin D (1 μg/ml) was added to block transcription of new RNA molecules. The stability of the *ATXN8OS* RNA was then determined by real-time PCR quantification of the amount of *ATXN8OS* RNA present in cells harvested at different time points after actinomycin D addition. The amount of *ATXN8OS* RNA present at the start of the experiment immediately before actinomycin D addition was set to 100%. As shown in **Fig. 3A**, using *HPRT1* mRNA as an internal control, the levels of *ATXN8OS* mRNA at 12 hours after addition of actinomycin D in 23, 88 and 157 CR cells were 7.2%, 12.1% and 22.0%, respectively. Therefore, expanded CR causes stabilization of *ATXN8OS* mRNA and subsequently reduces RNA decay.

Since the mutant *DMPK* transcripts accumulated in the nuclei of DM1 patient cells and aggregated to form distinct foci (Taneja et al., 1995), we investigated whether the expanded *ATXN8OS* CUG repeats form ribonuclear foci. The *ATXN8OS* 23, 88 and 157 CR cells were grown in the presence of doxycycline and FISH experiments using a Cy3-labeled $(CAG)_{10}$ oligonucleotide probe was performed two days later. As shown in **Fig. 3B**, no ribonuclear foci were seen in cells expressing *ATXN8OS* 23 CR. However, distinct ribonuclear foci, mostly perinuclear, were observed in cells expressing

expanded 88 and 157 CR. Similar results were obtained using an oligonucleotide probe specific to *ATXN8OS*. Since this probe can only bind to the *ATXN8OS* RNA in single copy, the inability to detect ribonuclear foci with 23 CR is not an artifact of the copy number of the repeats in the 88 or 157 CR cells.

Identification of targets affected by mutant *ATXN8OS* using microarray and proteomic approaches

To investigate whether human *ATXN8OS* gene with expanded CTG tract would affect global RNA expression, total RNAs from 50 nM staurosporine treated inducible HEK-293 *ATXN8OS* cell lines with 23 and 157 CR were genome-wide analyzed. After normalization, differentially expressed genes were selected using fold-change analysis. Genes that showed greater than 3-fold or smaller than 0.7-fold changes were considered to be significantly differentially expressed (**Table 2**). Among these candidate genes up-regulated in *ATXN8OS* expanded cell lines, *CDKN1B*, *SPON1*, *GSTB1* and *ISL1* were further validated by real-time PCR. CDKN1B (cyclin-dependent kinase inhibitor 1B, p27, Kip1) binds to and thereby prevents cyclin-CDK complexes from phosphorylating their protein substrates (Polyak et al., 1994). The biological consequence of this molecular interaction in cultured cells is cell cycle arrest, primarily in the G1 phase. SPON1 (spondin 1, extracellular matrix protein) is highly expressed in developing and damaged nerve tissue (Burstyn-Cohen et al., 1999). *GSTP1* (glutathione S-transferase pi) encodes a multifunctional detoxifying enzyme. GSTP1 may provide protection against oxidative stress (Holley et al., 2006). ISL1 (transcription factor,

LIM/homeodomain) was first reported to control various aspects of motoneuron identity in early embryogenesis (Tsuchida et al., 1994).

As shown in **Fig. 4**, no significant difference was observed in *CDKN1B* and *ISL1* expression level. Interestingly, the expression level of *SPON1* is significantly higher in *ATXN8OS* 157 CR cells compared to vector cells (P = 0.024) and *ATXN8OS* 23 CR cells (P = 0.034). Contrarily, *GSTP1* expression was highly reduced in *ATXN8OS* 157 CR cells.

Proteomic technology allows analysis of changes in protein level and is increasingly viewed as a potent tool for neurodegeneration disease research. To find out the protein targets affected by *ATXN8OS*, protein expressions of HEK-293 cells transiently transfected with *ATXN8OS* carrying normal 23 CTG repeats or disease-causing expanded 157 repeats were examined by proteomic approach. Several protein targets were identified by conventional 2D gels stained with Coomassie Blue G-250 (**Table 3**). The results showed that peptidylprolyl isomerase A isoform 2 (PPIA), triosephosphate isomerase (TPI), ACTB protein (ACTB), enolase 1 (ENO1), heterogeneous nuclear ribonucleoprotein (HNRNPL) and phosphoglycerate mutase 1 (PGAM1) were down-regulated by *ATXN8OS* expansion while proteasome 26S ATPase subunit 4 isoform 1 (PSMC4) and proline 4-hydroxylase β-subunit (P4HB) were up-regulated.

II. Analysis of *ATXN8OS* lymphoblastoid cell models

Analysis of *ATXN8OS*, *ATXN8* and *KLHL1* expression

Since SCA8 involves the expansion of CTG/CAG repeats from the overlapping *ATXN8OS* and *ATXN8* genes, *ATXN8* expression may be affected by *ATXN8OS* expansion and involved in SCA8 pathogenesis. Furthermore, the natural overlapping organization of the *ATXN8OS* and *KLHL1* coding regions suggested that *ATXN8OS* may function as a natural antisense regulator of *KLHL1* and affect *KLHL1* expression.

To evaluate the expression level of *ATXN8OS*, *ATXN8* and *KLHL1* genes, a total of six lymphoblastoid cell lines from patients with SCA8 large alleles and five age/gender-matched normal control lymphoblastoid cell lines were used (**Table 1**). RNA levels were measured by real-time PCR quantification using *ATXN8*, *ATXN8OS* and *KLHL1*-specific probes (**Fig. 5A**) and primers. Compared with endogenous *HPRT1*, the expression level of *ATXN8*, *ATXN8OS* and *KLHL1* were relatively low ($2^{-10} \sim 2^{-13}$). As shown in **Fig. 5B**, when the expressed level in control C1 was set as 1.0, *ATXN8* median (range) for patients and controls were 3.380 (1.670 ~ 5.754) and 1.284 (0.542 ~ 1.992), respectively. The difference in the *ATXN8* expression level between patients and controls was significant ($P = 0.012$). Contrarily, no significant difference was observed for *ATXN8OS* median (range) between patients [2.110 (0.993 ~ 3.667)] and controls [1.551 (0.929 ~ 2.512)] ($P = 0.364$) as well as *KLHL1* median (range) between patients [0.679 (0.321 ~ 1.205)] and controls [0.913 (0.767 ~ 1.231)] ($P = 0.137$). Our results suggest that *ATXN8* expression level is higher in *ATXN8OS* expanded cell lines. *KLHL1* expression may not be

affected by expanded CTG repeats.

ATXN8 -62 G/A promoter SNP

During DNA sequencing to assess the SCA8 repeat size, a novel C/T SNP 62 nucleotides downstream from the CTG repeats in *ATXN8* promoter was detected (**Fig. 6A**). The SNP was named -62 G/A, where translation initiation of polyglutamine protein +1 is used as temporary transcription start site of *ATXN8*. The SNP can be differentiated using mismatched PCR and *Mph*1103I restriction analysis (**Fig. 6B**). Previous results showed that *ATXN8* proximal promoter construct containing -62G displayed significantly higher luciferase activity compared with -62A in HEK-293, SK-N-SH and IMR-32 cells (Kao, 2008). According to real-time PCR analysis, the median (range) relative expression value of *ATXN8* mRNA for GG homozygote (three patients and one control) was higher (but not significantly) than GA heterozygote (three patients and four controls): 3.422 (1.847 to 5.754) and 1.859 (0.542 to 4.060), respectively ($P = 0.160$) (**Fig. 6C**).

III. Analysis of the IRES activity of *ATXN8OS* RNA and identification of *ATXN8OS* ORF protein

Possible factors involved in regulation of *ATXN8OS* RNA IRES activity

Dicistronic DNA test has been used in previous studies to show the existence of IRES activities in the 5'-UTR sequence of *ATXN8OS*. The 5'-UTR fragments 1~1045 directed firefly luciferase synthesis to a level of ~20% in 293 cells, as compared to the encephalomyocarditis virus (ECMV) (Gurtu et al., 1996) IRES2 positive control (**Fig. 7A**). A number of studies have demonstrated that cap-independent translation may need additional *trans*-acting factors (Stoneley and Willis, 2004). To further explore the factors required for regulation of *ATXN8OS* IRES activity, biotinylated *ATXN8OS* RNAs carrying 23 or 157 CTG repeats were synthesized and RNA binding assay was performed. Several plausible *ATXN8OS* RNA binding proteins were identified by MALDI-TOF MS (**Table 4**). However, the MOWSE scores of some identified proteins did not reach the criterion. A repetition of this experiment needs to be carried out to confirm this data.

In general, cellular IRESs work very inefficiently *in vitro*, but it is possible to stimulate certain cellular IRESs by the addition of known viral *trans*-acting factors. Studies of factors required for IRES-dependent translation have revealed that RNA-binding proteins such as PTB and PCBP1 enhance the translation of IRES-containing mRNA (Stoneley and Willis, 2004). Therefore, we cloned the PTB and PCBP1 cDNA and their effects on the *ATXN8OS* IRES in a dicistronic RNA were tested. Nevertheless increase of *ATXN8OS* IRES activity was not observed when PTB or PCBP1 was co-transfected into

HEK-293 cells (**Fig. 7A**).

Apart from *trans*-acting factor regulation, IRES-mediated translation may be activated under stress conditions when cap-dependent translation decreases (Hellen and Sarnow, 2001). To test the effect of stresses on translation from the *ATXN8OS* IRES, thapsigargin (induces ER stress by depleting ER Ca^{2+} stores) was used in different dosage. The results indicated that stress caused by thapsigargin could not activate *ATXN8OS* IRES activity (**Fig. 7B**).

Identification of the ribosome entry window within *ATXN8OS* RNA

The ribosome entry window attains an unstructured conformation and in doing so facilitates ribosome recruitment (Mitchell et al., 2003). Translation of a small upstream open reading frame appears to unfold inhibitory structures within the *Cat-1* IRES. As a consequence, the IRES undergoes major structural remodeling and assumes an active conformation (Yamin et al., 2003). To further identify the ribosome entry window within the *ATXN8OS* IRES, AUG codons were introduced into the 5' UTR sequence at suitably spaced intervals by using site-directed mutagenesis (**Fig. 8A**). When these mutant 5' UTR sequences were inserted into *ATXN8OS* fused EGFP construct, the AUG codons were both upstream of and out of frame with EGFP. If an out-of-frame AUG codon is introduced into the 5' UTR downstream of the ribosome entry window, ATXN8OS-EGFP synthesis can not be detected by FACS analysis. Conversely, an out-of-frame AUG codon placed upstream of the ribosome entry window will have no effect on ATXN8OS-EGFP synthesis. Comparing the expression of EGFP from each mutant construct to that from the wild-type construct reveals that AUG codons introduced into the 5' UTR at nucleotides

329, 350, 374, 391, 413, 425 and 434 had little or no effect (**Fig. 8B**). These data suggest that the ribosome entry window of *ATXN8OS* RNA has not been found.

ATXN8OS ORF-EGFP constructs

To investigate if indeed the *ATXN8OS* ORF could be translated, the *ATXN8OS* cDNA (NR_002717) with in-frame fused an EGFP tag to the C terminal of the *ATXN8OS* ORF was prepared (**Fig. 9C**, pCMV/+801). The transcripts made from this construct will be initiated from exon D (+801). As the promoter region upstream of exon D5 was identified by comparing human and mouse genomic DNA sequences flanking the 5' end of the transcripts (Benzow and Koob, 2002), *ATXN8OS* gene sequences +1 ~ +800 were included in construct pCMV/+1 so that transcripts made will be initiated from exon D5 (+1). In constructs pATXN8OS/-481 and pATXN8OS/-114, proximal *ATXN8OS* promoter fragments were used to drive *ATXN8OS* expression to mimic the *in vivo* situation.

ORF-EGFP expression

The constructs were transiently expressed into HEK-293 cells. After two days the levels of EGFP fluorescence were evaluated by FACS analysis. As shown in **Fig. 10A**, the 1.4~59.0% EGFP production was seen in cells transfected with ORF-EGFP constructs as compared to the pIRES2-EGFP (cap-independent EGFP expression). As transcripts initiated from +801 (pCMV/+801) expressed significantly more EGFP fluorescence than

transcripts initiated from +1 (pCMV/+1, pATXN8OS/-481 and pATXN8OS/-114) (59.0% vs. 1.4~3.5%, $P = 0.000$), *ORF-EGFP* RNA levels were measured by real-time PCR quantification using *ATXN8OS*-specific probe C2/C1 and primers. As shown in **Fig. 10B**, when the expressed level in pATXN8OS/-481 cells was set as 1.0, *ORF-EGFP* RNA levels for transcripts initiated from +801 (pCMV/+801) versus transcripts initiated from +1 (pCMV/+1, pATXN8OS/-481 and pATXN8OS/-114) were 3.7 and 1.0 ~ 2.4, respectively. Similar 1.0 ~ 2.6 range of *ORF-EGFP* RNA levels for transcripts initiated from +1 (pCMV/+1, pATXN8OS/-481 and pATXN8OS/-114) were also observed using *ATXN8OS* D5/D4 probe (**Fig. 10B**). On western blot immunostaining with GFP antibody, while no specific polypeptide was detected in mock-transfected HEK-293 cells, a 50 kDa protein was detected in pCMV/+801 transfected cells, as compared to a 26 kDa protein in pEGFP-N1 transfected cells (**Fig. 10C**). Nevertheless the observed 50 kDa protein is larger than the predicted ORF-EGFP protein (350 amino acids with MW of 39677).

ORF-EGFP aggregation

To further investigate the expression of ORF-EGFP protein, confocal microscopy examination of GFP fluorescence was carried out after transfection of pIRES-EGFP, pCMV/+801 and pATXN8OS/-481 constructs into HEK-293 cells. As shown in **Fig. 11**, strong GFP fluorescence was distributed diffusely in pIRES-EGFP-transfected cells. With pCMV/+801 construct, small and dispersed aggregates appeared mainly in the cytoplasm, in addition to showing diffuse cytoplasm expression. Cells transfected with

pATXN8OS/-481 or pATXN8OS/-114 (data not shown) gave sparse aggregates and very weak, diffuse GFP fluorescence. Furthermore, MitoTracker and LysoTracker were used to confirm the localization of ORF-EGFP protein. The results indicated that ORF-EGFP co-localized with mitochondria but not lysosome (**Fig. 12**).

ORF immunodetection

To validate if indeed *ATXN8OS* ORF is expressed in human cells, GST-ORF (*S. japonicum* GST from pGEX plasmid) fusion protein was prepared as antigen to raise antiserum in rabbit. As the average molecular weight of proteins that dissolve exclusively in urea buffer is up to 60% higher than in RIPA buffer (Ngoka, 2008), 8 M urea lysis buffer was used for lymphoblastoid protein extraction. On western blot staining with ORF antiserum, while no specific polypeptide was detected with pre-immune serum, a 23 kDa protein was detected in insoluble pellet fraction (**Fig. 13A**). The same 23 kDa protein was also observed in urea buffer-insoluble pellet fraction prepared from human embryonic kidney HEK-293 cells, neuroblastoma IMR-32, SK-N-SH and SH-SY5Y cells (**Fig. 13B**). Again the observed 23 kDa protein is larger than the predicted ORF: 102 amino acids with MW of 11220.

ORF identification

Utilization of alternative non-AUG translation initiation codons has been demonstrated with increasing frequency in mammalian species (Touriol et al.,

2003). Using Vector NTI software, translation from an upstream in-frame GUG codon (+953) predicted a 200 amino-acid ORF protein (MW 22654) or a 448 amino-acid ORF-EGFP fusion protein (MW 50528). To identify the translation start of *ATXN8OS* ORF, lymphoblastoid proteins from urea buffer-insoluble pellet fraction were subjected to 2D PAGE (**Fig. 14A**) and 2D immunoblot (**Fig. 14B**). The identity of the three ORF-specific spots was determined using LC-MS/MS and Mascot data search in a database set up for the predicted 200 amino-acid ORF. As shown in **Fig. 14C**, six matched peptide with sequence coverage of 47% was obtained, including the N-terminal peptide VPCPGAPCCS LVATGSR which can only be generated from translation start.

ORF expression and SCA8

To study if *ATXN8OS* expansion mutation altered ORF expression, the ORF protein level in lymphoblastoid cells expressing normal and expanded *ATXN8OS* transcripts was examined using western blot. When the average expressed level of the controls was set as 100%, ORF protein expression in lymphoblastoid cells carrying expanded *ATXN8OS* transcripts (P1, P3~P5) was significantly increased when compared to the lymphoblastoid cells with normal *ATXN8OS* transcripts (median: 158% vs. 100%, range: 113~204% vs. 86~115%, $P = 0.049$) (**Fig. 15**).

Discussion

ATXN8OS stably induced HEK-293 cell models

Clinical and genetic studies have shown that SCA8 is a slowly progressive inherited disorder with reduced penetrance, in addition to its rare occurrence of phenotype in individuals carrying a much larger repeat expansion (Koob et al., 1999; Mosemiller et al., 2003; Schols et al., 2003; Sobrido et al., 2001; Worth et al., 2000). Thus the pathogenic mechanisms underlying SCA8 are expected to be complicated and remain unclear. To better understand the possible molecular mechanisms of SCA8, we developed a number of otherwise isogenic human cell lines expressing *ATXN8OS* transcripts with 0~157 CTA/ CTG CR. The repeat number in these cell lines was stable and these cells were used to investigate the *ATXN8OS* expression. The implications of the findings in the pathogenesis of SCA8 are discussed as the following.

It is widely believed that amino-terminal extensions of histones are subject to a variety of posttranslational modification (Strahl and Allis, 2000). In chromatin domains that are transcriptionally repressed, high levels of histone H3-K9 methylation and H3-K14 hypoacetylation were observed. Therefore, it is possible to predict the transcriptional competence of a particular genomic region by examining the H3 methylation and acetylation patterns. The results showed that there is a repeat-dependent repression of *ATXN8OS* transcripts (**Fig. 1C**) and the expression of adjacent HaloTag was reduced in *ATXN8OS* cell line carrying 157 CR (**Fig. 2**). This observation is correlated with the

recent findings of H3-K9 dimethylation and H3-K14 hypoacetylation in the 157 CR cells by ChIP assay (Chen et al., 2009). In myotonic dystrophy, the DNA structure of expanded CTG repeats most likely produces changes in chromatin configuration which affect the expression of adjacent genes. Expansions of GAA repeat in Friedreich's ataxia also conferred variegation of expression on a linked transgene in mice (Saveliev et al., 2003). Silencing was correlated with a decrease in promoter accessibility and was enhanced by the classical position effect variegation modifier heterochromatin protein 1 (HP1), which is able to bind to methylated histone H3-K9 (Bannister et al., 2001). Elevated levels of histone H3 dimethylated on K9 were also seen in Friedreich's ataxia cells consistent with a repressive chromatin organization (Greene et al., 2007). As reduced expression of adjacent HaloTag gene was also seen in the 88 CR cells (**Fig. 2**) independent of H3-K9 dimethylation and H3-K14 hypoacetylation, DNA methylation or other histone modifications such as arginine methylation and serine/threonine phosphorylation (Kim et al., 2009) may be responsible for the observed repression of *ATXN8OS* RNA in the 88 CR cells. Nevertheless the epigenetic change induced by large SCA8 repeat expansions may be one of the genetic factors that suppress the disease symptoms in control individuals carrying large *ATXN8OS* repeat expansions.

Furthermore, increased *ATXN8OS* transcript stability was also observed in this study. It was reported that expanded CUG repeat transcripts form stable hairpins (Tian et al., 2000) and muscleblind and MBNL1 increase steady state levels of CUG repeat RNA (Houseley et al., 2005). Therefore, the possibility of stable hairpin formation and proteins binding may account for the observation that expanded CR causes stabilization of *ATXN8OS* mRNA (**Fig. 3A**), and in turn leads to repeat length-dependent increase in fold of *ATXN8OS*

induction (**Fig. 1D**). In DM1 and DM2, the expanded repeat RNA forms discrete ribonuclear foci to sequester CUG binding proteins and subsequently jeopardize the normal cellular functions of these proteins, which would then lead to abnormal RNA splicing of several genes (Ranum and Day, 2002). Although the induced expression levels of *ATXN8OS* RNA in these CR cells were low (ranging from 0.659~1.346 compared with endogenous *HPRT1*), ribonuclear foci were detected in our *ATXN8OS* 88 or 157 CR cells **(Fig. 3B**). Most of the RNA foci formed are located near nuclear membrane, which may be compatible with the observation by Koch and colleague that the hairpin structure formed by long CUG repeats (> 44) cannot pass through nucleic pores (Koch and Leffert, 1998). The ribonuclear foci observed in the nucleus may also result in transcriptional dysfunction to lead to the disease, as indicated by transcription factors leaching from chromatin by mutant RNA in myotonic dystrophy (Ebralidze et al., 2004).

In this study, gene expression profiles of *ATXN8OS* cell lines were examined by microarray and validated by real-time PCR. Interestingly, *SPON1* is up-regulated in 157 CR cells **(Fig. 4**). Previous studies have shown that *SPON1* is highly expressed in developing and damaged nerve tissue and involved in neuronal development and repair. It could promote neurite attachment and outgrowth in the cultured spinal cord and sensory neurons (Klar et al., 1992). Up-regulation of *SPON1* may promote sensory nerve regeneration (Burstyn-Cohen et al., 1999). It also had been found to be implicated in Alzheimer's disease, binding to the conserved central extracellular domain of amyloid-β precursor protein (APP) and inhibits β-secretase cleavage of APP (Ho and Sudhof, 2004).

Despite being a well-established technique for protein analysis, the lack of

reproducibility between traditional 2D gels leads to significant system variability making it difficult to distinguish between system variation and induced biological change. Therefore, accurate quantification to measure changes in protein expression remains one of the most important aspects of proteomics. With the recent advances in differential dye labeling and the introduction of the two-dimensional difference gel electrophoresis (2D-DIGE) technique, an attractive technology for the quantitative analysis of differences in protein profiles has become available (Marouga et al., 2005; Tonge et al., 2001; Yan et al., 2002). 2D-DIGE enables multiple protein extracts to be separated on the same 2D gel by labeling of each extract using spectrally resolvable dyes. 2D-DIGE involves use of a reference sample, known as an internal standard, which comprises equal amounts of all biological samples in the experiment, leading an accurate quantification between different samples. To find out the targets affected by expanded *ATXN8OS* tract, 2D-DIGE was applied to the inducible *ATXN8OS* cell model system (data not shown). In spite of advantages of 2D-DIGE, it does have some limitation in this study. First, few differentially expressed protein spots were found by using 2D-DIGE system. Moreover, the protein spots were too weak to be picked and analyzed by MALDI-TOF after SYPRO Ruby staining.

Although we could not find protein targets by 2D-DIGE, several targets were identified by conventional 2-D electrophoresis in this study (**Table 3**). Among the *ATXN8OS* expansion down-regulated proteins, some proteins have been involved in other neurodegenerative diseases. Peptidylprolyl isomerase A exhibits antioxidant properties and protects neurons against oxidative stress induced injury. It may prevent $A\beta_{25-35}$-induced neurotoxicity through attenuating oxidative stress induced by $A\beta_{25-35}$ (Ge et al., 2009). Inhibition of

triosephosphate isomerase has been reported to induce neuronal death in cultured murine cortical cells (Sheline and Choi, 1998). Interestingly, triosephosphate isomerase and enose 1 have been identified as targets of protein nitration in Alzheimer's disease brain (Castegna et al., 2003) and phosphoglycerate mutase 1 showed elevated levels of protein carbonyls in inferior parietal lobule from subjects with early stage-Alzheimer's disease (Sultana et al. 2010). ACTB is known to be correlated with KLHL1 which may play a role in SCA8 pathogenesis (Koob et al., 1999). HNRNPL may play roles in regulation on the level of mRNA splicing (Hui et al., 2003b), IRES-mediated translation (Hahm et al., 1998), and mRNA stability (Hui et al., 2003a; Shih and Claffey, 1999). However, MALDI-TOF analysis only could provide predicted candidate proteins and could not reflect the actual biological status of the cells. Thus, immunoblotting should be performed for further validation of these candidate proteins.

ATXN8OS lymphoblastoid cell models

Although the pathology of SCA is mainly in cerebellum, most SCA genes are ubiquitously distributed including peripheral lymphocytes. Thus lymphoblastoid cells are frequently used as a model to study SCA (Tsai et al., 2005; Wen et al., 2003) as well as Alzheimer's disease (Munoz et al., 2008; Sala et al., 2008). In our lymphoblastoid cell models, transcripts encoded by *ATXN8*, *ATXN8OS* and *KLHL1* genes are expressed at low steady-state. Real-time PCR analysis revealed that *ATXN8* expression level is significant higher in patients with SCA8 large alleles than controls (**Fig. 5B**). Within patient group, the *ATXN8* expression level was not related to disease onset.

However, the two SCA8 patients displayed higher (although not significantly) median *ATXN8* expression level than that of the four PD patients, indicating that higher *ATXN8* expression levels may be more likely to be associated with a SCA phenotype. While ataxin8 function is unclear, the *ATXN8OS* RNA transcript is thought to serve as an anti-sense RNA that inhibits the brain-specific transcript for the actin-binding protein KLHL1 (Benzow and Koob, 2002). Nevertheless, no significant difference was observed for *ATXN8OS* and *KLHL1* expression levels between patients and controls (**Fig. 5B**). As major haplotypes are found among patients with ataxia (Ikeda et al., 2004), genetic variation(s) in linkage disequilibrium with expanded repeats may affect *ATXN8* expression leading to the observed increased expression level in patients with SCA8 large repeats. During assessing the SCA8 repeat size, a novel -62 G/A SNP was identified with 90% (9/10) expanded alleles associated with G allele. Our recent results indicated that fragment containing -62G displayed significant higher promoter activity compared to fragment containing -62A in all three human cell lines tested (Wu et al., 2009). Thus the increased *ATXN8* expression level in patients with SCA8 large alleles was likely resulted from the -62 G/A variation. As the correlation of *ATXN8* mRNA expression with -62 G > A genotype is not absolute (**Fig. 6C**), possible hairpin formation and protein binding to increase the stability of expanded CUG repeat transcripts (Houseley et al., 2005; Tian et al., 2000) may also contribute to the observed increased *ATXN8* expression level in patients with SCA8 large alleles.

Previously intranuclear polyglutamine inclusions in cerebellar Purkinje and brainstem neurons were reported in SCA8 expansion mice (Moseley et al., 2006). Clinicopathologic investigation of a SCA8 patient also revealed

intracytoplasmic 1C2-positive granular structures (Ito et al., 2006). Using dot-blot filter retardation assay, we could detect aggregation from lymphoblastoid cells with large *ATXN8OS* expansion by anti-TFIID (N-12) antibody (Santa Cruz Biotechnology) (data not shown). However, we did not observed polyglutamine inclusions in the lymphoblastoid cells from patients with expanded SCA8 alleles by western blot (data not shown). Further experiments are therefore necessary to clarify this issue.

IRES activity of *ATXN8OS* RNA and identification of *ATXN8OS* ORF protein

SCA8 was first proposed to be caused by an RNA gain-of-function mechanism and analysis of *ATXN8OS* sequence did not reveal any possible spliced isoform possessing an ORF to extend through the expansion (Ranum and Day, 2004). However, previous studies have shown the possible IRES activity existing in the 5'-UTR sequence of *ATXN8OS*. Although considerable experiments have been done to find out the factors involved in regulation of *ATXN8OS* IRES activity, possible candidates have not been identified in this study. Other proteins or factors have also been implicated in the regulation of IRES activity and are worthy of further investigation.

To study the plausible pathogenesis of SCA8, we initially cloned the *ATXN8OS* cDNA containing spliced exons D, C2, C1, B, and A (**Fig. 9B**) (Nemes et al., 2000). Sequence analysis revealed the existence of a 102 amino-acid ORF 446 nucleotides from the 5' end of ATXN8OS RNA (**Fig. 9C**, construct pCMV/+801). As the conserved promoter region upstream of the *ATXN8OS* gene exon D5 was identified (Benzow and Koob, 2002), we also

cloned the *ATXN8OS* cDNA containing exon D5 and D4 thus the 102 amino-acid ORF will then be 1246 nucleotides from the 5' end of *ATXN8OS* RNA (**Fig. 9C**, constructs pCMV/+1, pATXN8OS/-114, pATXN8OS/-481). Interestingly, although there is no big difference of RNA expression level between pCMV/+801 and pCMV/+1 (**Fig. 10B**), the protein expression was significantly reduced when pCMV/+1 was transiently expressed into HEK-293 cells (**Fig. 10B**). It suggests that exon D5 and D4 region may alter RNA conformation and therefore affect protein translation. Western blot with anti-GFP antibody suggested alternative translation initiation from an upstream non-AUG and an in-frame downstream AUG codon, resulting in the production of 52 and 30 kDa isoforms, respectively (**Fig. 10C**). Moreover, the GFP-tagged ORF protein formed aggregates (**Fig. 11**) and co-localized with mitochondria (**Fig. 12**). These data implied important information that *ATXN8OS* is indeed translatable.

Using antiserum raised against ORF, our findings indicated that *ATXN8OS* ORF protein was present in 8 M urea lysis buffer-insoluble fraction of lymphoblastoid cells as well as HEK-293, IMR-32, SK-N-SH and SH-SY5Y cells (**Fig. 13**). When the ORF protein in lymphoblastoid cells carrying normal or expanded CTG repeats was examined, significant difference (100% vs. 158%, $P = 0.049$) was observed between cells carrying normal or expanded CTG repeats (**Fig. 15**). Previously it was reported that expanded CUG repeat transcripts form stable hairpins (Tian et al., 2000) and muscleblind and MBNL1 increase steady state levels of CUG repeat RNA (Houseley et al., 2005). In addition, the results of inducible cell model in this study also revealed increased RNA stability associated with CUG-expanded transcripts. Thus these may have contributed to the higher ORF expression observed in

CTG expanded *ATXN8OS* lymphoblastoid cells.

Recently, three yeast genes, CARP2A (acidic ribosomal protein P2A in *C. albicans*) (Abramczyk et al., 2003), GRS1 (one of the two glycyl-tRNA synthetase genes in *S. cerevisiae*) (Chang and Wang, 2004), and ALA1 (AlaRS gene in *S. cerevisiae*) (Tang et al., 2004) have been reported to use naturally occurring non-AUG triplets as translation initiators. In the case of CARP2A, a non-AUG codon GUG serves as the exclusive translation initiator, while in the case of ALA1 and GRS1, non-AUG codons ACG and UUG act as alternative translation initiators that are accompanies by a downstream in-frame AUG initiation codon. The molecular mechanism that enables the non-AUG condon to serve as translation initiation sites remains poorly understand. Further studies in GRS1 suggest that efficiency of UUG initiation codon is significantly affected by its sequence context (Chen et al., 2008).

Utilization of alternative non-AUG translation initiation codons has been reported with increasing frequency in mammalian species, in addition to initiating at a downstream in-frame AUG codon (Touriol et al., 2003). Translation initiation on such mRNAs results in the synthesis of proteins harboring different amino terminal domains potentially conferring on these isoforms distinct functions. In this work, we demonstrated that the expressed *ATXN8OS* ORF uses an upstream in-frame GUG triplet as a native translation initiation codon by using 2D PAGE, 2D immunoblot and tandem MS determination (**Fig. 14**). Whether the sequence context or the secondary structure surrounding the GUG codon governs the use of non-AUG codons for translation of *ATXN8OS* ORF protein remains to be determined. As alternative initiation sites are utilized for the synthesis of proteins that regulate biological processes in health and disease (Bruening and Pelletier, 1996; Hann et al.,

1994; Lock et al., 1991), the biological meaning of the *ATXN8OS* ORF protein and its role in the pathogenesis of SCA8 remains to be determined.

Previously bioinformatics analyses demonstrated that distinct consensus sequences (at -7 and -6 positions), upstream AUGs, 5'-UTR sequence length, G/C ratio and IRES secondary structure are important for categorizing mRNAs as those with and without alternative translation initiation sites (Wegrzyn et al., 2008). Among these properties, 5'-UTR of the alternative translation initiation sites showed conservation of G/C at the -6 position and C at the -7 position. In contrast, the AUG initiation sites showed consensus at position -3 for A/G and position +4 for G/A (Kozak, 1997; Wegrzyn et al., 2008). The *ATXN8OS* ORF GUG initiation codon has conserved C at the -7 position but less abundance U at the -6 position, the downstream in-frame AUG codon has conserved A at the -3 position but also less abundance U at the +4 position (**Fig. 14C**). Although not well conserved at the -6 position, the conserved C at the -7 position and other un-analyzed properties may support the use of the second most common alternative translation initiation GUG site (Wegrzyn et al., 2008) for the translation of *ATXN8OS* ORF protein.

In conclusion, as the findings in this study indicated, the expanded CTG tract may affect epigenetic and post-transcriptional regulations of the *ATXN8OS* expression, which may help to explain the reduced penetrance and RNA gain-of function of the disease. The results obtained from human lymphoblastoid cell model suggest that the novel *ATXN8* -62 G/A promoter polymorphism may function in modulating *ATXN8* expression in. Moreover, our study indicated that the *ATXN8OS* putative ORF protein could be expressed via a naturally occurring non-AUG start codon. The role of *ATXN8OS* ORF and its connection to SCA8 are deserving of further

investigation. The results of this study may lead to a better understanding of the disease nature.

References

Abramczyk, D., Tchorzewski, M., and Grankowski, N. (2003). Non-AUG translation initiation of mRNA encoding acidic ribosomal P2A protein in Candida albicans. Yeast (Chichester, England) 20, 1045-1052.

Amack, J.D., Paguio, A.P., and Mahadevan, M.S. (1999). Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. Human molecular genetics 8, 1975-1984.

Anborgh, P.H., Godin, C., Pampillo, M., Dhami, G.K., Dale, L.B., Cregan, S.P., Truant, R., and Ferguson, S.S. (2005). Inhibition of metabotropic glutamate receptor signaling by the huntingtin-binding protein optineurin. The Journal of biological chemistry 280, 34840-34848.

Andres, A.M., Soldevila, M., Saitou, N., Volpini, V., Calafell, F., and Bertranpetit, J. (2003). Understanding the dynamics of Spinocerebellar Ataxia 8 (SCA8) locus through a comparative genetic approach in humans and apes. Neurosci Lett 336, 143-146.

Aromolaran, K.A., Benzow, K.A., Koob, M.D., and Piedras-Renteria, E.S. (2007). The Kelch-like protein 1 modulates P/Q-type calcium current density. Neuroscience 145, 841-850.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120-124.

Benzow, K.A., and Koob, M.D. (2002). The KLHL1-antisense transcript (KLHL1AS) is evolutionarily conserved. Mamm Genome 13, 134-141.

Bruening, W., and Pelletier, J. (1996). A non-AUG translational initiation event generates novel WT1 isoforms. The Journal of biological chemistry 271, 8646-8654.

Brusco, A., Cagnoli, C., Franco, A., Dragone, E., Nardacchione, A., Grosso, E., Mortara, P., Mutani, R., Migone, N., and Orsi, L. (2002). Analysis of SCA8 and SCA12 loci in 134 Italian ataxic patients negative for SCA1-3, 6 and 7 CAG expansions. J Neurol 249, 923-929.

Burstyn-Cohen, T., Tzarfaty, V., Frumkin, A., Feinstein, Y., Stoeckli, E., and Klar, A. (1999). F-Spondin is required for accurate pathfinding of commissural axons at the floor plate. Neuron 23, 233-246.

Carlson, K.M., Andresen, J.M., and Orr, H.T. (2009). Emerging pathogenic pathways in the spinocerebellar ataxias. Current opinion in genetics & development 19, 247-253.

Castegna, A., Thongboonkerd, V., Klein, J.B., Lynn, B., Markesbery, W.R., and Butterfield, D.A. (2003). Proteomic identification of nitrated proteins in Alzheimer's disease brain. J Neurochem 85, 1394-1401.

Cellini, E., Nacmias, B., Forleo, P., Piacentini, S., Guarnieri, B.M., Serio, A., Calabro, A., Renzi, D., and Sorbi, S. (2001). Genetic and clinical analysis of spinocerebellar ataxia type 8 repeat expansion in Italy. Archives of neurology 58, 1856-1859.

Chang, K.J., and Wang, C.C. (2004). Translation initiation from a naturally

occurring non-AUG codon in Saccharomyces cerevisiae. The Journal of biological chemistry 279, 13778-13785.

Chen, D.H., Brkanac, Z., Verlinde, C.L., Tan, X.J., Bylenok, L., Nochlin, D., Matsushita, M., Lipe, H., Wolff, J., Fernandez, M., et al. (2003). Missense mutations in the regulatory domain of PKC gamma: a new mechanism for dominant nonepisodic cerebellar ataxia. Am J Hum Genet 72, 839-849.

Chen, I.C., Lin, H.Y., Lee, G.C., Kao, S.H., Chen, C.M., Wu, Y.R., Hsieh-Li, H.M., Su, M.T., and Lee-Chen, G.J. (2009). Spinocerebellar ataxia type 8 larger triplet expansion alters histone modification and induces RNA foci. BMC molecular biology 10, 9.

Chen, S.J., Lin, G., Chang, K.J., Yeh, L.S., and Wang, C.C. (2008). Translational efficiency of a non-AUG initiation codon is significantly affected by its sequence context in yeast. The Journal of biological chemistry 283, 3173-3180.

Clemens, M.J., and Bommer, U.A. (1999). Translational control: the cancer connection. The international journal of biochemistry & cell biology 31, 1-23.

Clemens, M.J., Bushell, M., Jeffrey, I.W., Pain, V.M., and Morley, S.J. (2000). Translation initiation factor modifications and the regulation of protein synthesis in apoptotic cells. Cell death and differentiation 7, 603-615.

Creancier, L., Morello, D., Mercier, P., and Prats, A.C. (2000). Fibroblast growth factor 2 internal ribosome entry site (IRES) activity ex vivo and in transgenic mice reveals a stringent tissue-specific regulation. The

Journal of cell biology 150, 275-281.

David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., et al. (1997). Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nature genetics 17, 65-70.

Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., and Housman, D.E. (1997). Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. Proceedings of the National Academy of Sciences of the United States of America 94, 7388-7393.

Day, J.W., Schut, L.J., Moseley, M.L., Durand, A.C., and Ranum, L.P. (2000). Spinocerebellar ataxia type 8: clinical features in a large family. Neurology 55, 649-657.

Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K., and Junghans, R.P. (2004). RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. Science (New York, NY 303, 383-387.

Fernandez, J., Bode, B., Koromilas, A., Diehl, J.A., Krukovets, I., Snider, M.D., and Hatzoglou, M. (2002a). Translation mediated by the internal ribosome entry site of the cat-1 mRNA is regulated by glucose availability in a PERK kinase-dependent manner. The Journal of biological chemistry 277, 11780-11787.

Fernandez, J., Yaman, I., Mishra, R., Merrick, W.C., Snider, M.D., Lamers, W.H., and Hatzoglou, M. (2001). Internal ribosome entry site-mediated translation of a mammalian mRNA is regulated by amino acid

availability. The Journal of biological chemistry 276, 12285-12291.

Fernandez, J., Yaman, I., Sarnow, P., Snider, M.D., and Hatzoglou, M. (2002b). Regulation of internal ribosomal entry site-mediated translation by phosphorylation of the translation initiation factor eIF2alpha. The Journal of biological chemistry 277, 19198-19205.

Ge, Y.S., Teng, W.Y., and Zhang, C.D. (2009). Protective effect of cyclophilin A against Alzheimer's amyloid beta-peptide (25-35)-induced oxidative stress in PC12 cells. Chinese medical journal 122, 716-724.

Greene, E., Mahishi, L., Entezam, A., Kumari, D., and Usdin, K. (2007). Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic acids research 35, 3383-3390.

Gurtu, V., Yan, G., and Zhang, G. (1996). IRES bicistronic expression vectors for efficient creation of stable mammalian cell lines. Biochemical and biophysical research communications 229, 295-298.

Hahm, B., Kim, Y.K., Kim, J.H., Kim, T.Y., and Jang, S.K. (1998). Heterogeneous nuclear ribonucleoprotein L interacts with the 3' border of the internal ribosomal entry site of hepatitis C virus. Journal of virology 72, 8782-8788.

Handa, V., Yeh, H.J., McPhie, P., and Usdin, K. (2005). The AUUCU repeats responsible for spinocerebellar ataxia type 10 form unusual RNA hairpins. The Journal of biological chemistry 280, 29340-29345.

Hann, S.R., Dixit, M., Sears, R.C., and Sealy, L. (1994). The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site. Genes & development 8, 2441-2452.

Harding, A.E. (1993). Clinical features and classification of inherited ataxias. Advances in neurology 61, 1-14.

He, Y., Zu, T., Benzow, K.A., Orr, H.T., Clark, H.B., and Koob, M.D. (2006). Targeted deletion of a single Sca8 ataxia locus allele in mice causes abnormal gait, progressive loss of motor coordination, and Purkinje cell dendritic deficits. J Neurosci 26, 9975-9982.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes & development 15, 1593-1612.

Ho, A., and Sudhof, T.C. (2004). Binding of F-spondin to amyloid-beta precursor protein: a candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage. Proceedings of the National Academy of Sciences of the United States of America 101, 2548-2553.

Holley, S.L., Rajagopal, R., Hoban, P.R., Deakin, M., Fawole, A.S., Elder, J.B., Elder, J., Smith, V., Strange, R.C., and Fryer, A.A. (2006). Polymorphisms in the glutathione S-transferase mu cluster are associated with tumour progression and patient outcome in colorectal cancer. International journal of oncology 28, 231-236.

Holmes, S.E., O'Hearn, E.E., McInnis, M.G., Gorelick-Feldman, D.A., Kleiderlein, J.J., Callahan, C., Kwak, N.G., Ingersoll-Ashworth, R.G., Sherr, M., Sumner, A.J., et al. (1999). Expansion of a novel CAG

trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nature genetics 23, 391-392.

Houseley, J.M., Wang, Z., Brock, G.J., Soloway, J., Artero, R., Perez-Alonso, M., O'Dell, K.M., and Monckton, D.G. (2005). Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila. Human molecular genetics 14, 873-883.

Hui, J., Reither, G., and Bindereif, A. (2003a). Novel functional role of CA repeats and hnRNP L in RNA stability. RNA (New York, NY 9, 931-936.

Hui, J., Stangl, K., Lane, W.S., and Bindereif, A. (2003b). HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. Nature structural biology 10, 33-37.

Ikeda, Y., Dalton, J.C., Moseley, M.L., Gardner, K.L., Bird, T.D., Ashizawa, T., Seltzer, W.K., Pandolfo, M., Milunsky, A., Potter, N.T., et al. (2004). Spinocerebellar ataxia type 8: molecular genetic comparisons and haplotype analysis of 37 families with ataxia. Am J Hum Genet 75, 3-16.

Ikeda, Y., Dick, K.A., Weatherspoon, M.R., Gincel, D., Armbrust, K.R., Dalton, J.C., Stevanin, G., Durr, A., Zuhlke, C., Burk, K., et al. (2006). Spectrin mutations cause spinocerebellar ataxia type 5. Nature genetics 38, 184-190.

Ikeda, Y., Shizuka-Ikeda, M., Watanabe, M., Schmitt, M., Okamoto, K., and Shoji, M. (2000a). Asymptomatic CTG expansion at the SCA8 locus is associated with cerebellar atrophy on MRI. J Neurol Sci 182, 76-79.

Ikeda, Y., Shizuka, M., Watanabe, M., Okamoto, K., and Shoji, M. (2000b). Molecular and clinical analyses of spinocerebellar ataxia type 8 in Japan. Neurology 54, 950-955.

Ito, H., Kawakami, H., Wate, R., Matsumoto, S., Imai, T., Hirano, A., and Kusaka, H. (2006). Clinicopathologic investigation of a family with expanded SCA8 CTA/CTG repeats. Neurology 67, 1479-1481.

Izumi, Y., Maruyama, H., Oda, M., Morino, H., Okada, T., Ito, H., Sasaki, I., Tanaka, H., Komure, O., Udaka, F., et al. (2003). SCA8 repeat expansion: large CTA/CTG repeat alleles are more common in ataxic patients, including those with SCA6. Am J Hum Genet 72, 704-709.

Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. Journal of virology 62, 2636-2643.

Jardim, L.B., Silveira, I., Pereira, M.L., Ferro, A., Alonso, I., do Ceu Moreira, M., Mendonca, P., Ferreirinha, F., Sequeiros, J., and Giugliani, R. (2001). A survey of spinocerebellar ataxia in South Brazil - 66 new cases with Machado-Joseph disease, SCA7, SCA8, or unidentified disease-causing mutations. J Neurol 248, 870-876.

Juvonen, V., Hietala, M., Paivarinta, M., Rantamaki, M., Hakamies, L., Kaakkola, S., Vierimaa, O., Penttinen, M., and Savontaus, M.L. (2000). Clinical and genetic findings in Finnish ataxia patients with the spinocerebellar ataxia 8 repeat expansion. Ann Neurol 48, 354-361.

Juvonen, V., Kairisto, V., Hietala, M., and Savontaus, M.L. (2002).

Calculating predictive values for the large repeat alleles at the SCA8 locus in patients with ataxia. Journal of medical genetics 39, 935-936.

Kao S.H. (2008) Spinocerebellar ataxia type 8: genetic and promoter functional studies. Master thesis. National Taiwan Normal University.

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

Kim, J.K., Samaranayake, M., and Pradhan, S. (2009). Epigenetic mechanisms in mammals. Cell Mol Life Sci 66, 596-612.

Klar, A., Jessell, T.M., and Ruiz i Altaba, A. (1992). Control of floor plate identity and function in the embryonic nervous system. Cold Spring Harbor symposia on quantitative biology 57, 473-482.

Koch, K.S., and Leffert, H.L. (1998). Giant hairpins formed by CUG repeats in myotonic dystrophy messenger RNAs might sterically block RNA export through nuclear pores. Journal of theoretical biology 192, 505-514.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nature genetics 6, 9-13.

Koide, R., Kobayashi, S., Shimohata, T., Ikeuchi, T., Maruyama, M., Saito, M., Yamada, M., Takahashi, H., and Tsuji, S. (1999). A neurological

disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? Human molecular genetics 8, 2047-2053.

Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W., and Ranum, L.P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nature genetics 21, 379-384.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. The Journal of biological chemistry 266, 19867-19870.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. The EMBO journal 16, 2482-2492.

Larkin, K., and Fardaei, M. (2001). Myotonic dystrophy--a multigene disorder. Brain research bulletin 56, 389-395.

Lee G.C. (2008) Spinocerebellar ataxia: Epigenetic and cell model studies of SCA type 8. Master thesis. National Taiwan Normal University.

Lin H.Y. (2007) Genetic studies of spinocerebellar ataxias and molecular impacts of CTG trinucleotide expansion. Doctoral thesis. National Taiwan Normal University.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science (New York,

NY 293, 864-867.

Lock, P., Ralph, S., Stanley, E., Boulet, I., Ramsay, R., and Dunn, A.R. (1991). Two isoforms of murine hck, generated by utilization of alternative translational initiation codons, exhibit different patterns of subcellular localization. Molecular and cellular biology 11, 4363-4370.

Marouga, R., David, S., and Hawkins, E. (2005). The development of the DIGE system: 2D fluorescence difference gel analysis technology. Analytical and bioanalytical chemistry 382, 669-678.

Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L., et al. (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nature genetics 26, 191-194.

Mitchell, S.A., Spriggs, K.A., Coldwell, M.J., Jackson, R.J., and Willis, A.E. (2003). The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr. Molecular cell 11, 757-771.

Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nature genetics 38, 758-769.

Mosemiller, A.K., Dalton, J.C., Day, J.W., and Ranum, L.P. (2003). Molecular genetics of spinocerebellar ataxia type 8 (SCA8). Cytogenet Genome Res 100, 175-183.

Munoz, U., Bartolome, F., Bermejo, F., and Martin-Requero, A. (2008). Enhanced proteasome-dependent degradation of the CDK inhibitor p27(kip1) in immortalized lymphocytes from Alzheimer's dementia patients. Neurobiology of aging 29, 1474-1484.

Nemes, J.P., Benzow, K.A., Moseley, M.L., Ranum, L.P., and Koob, M.D. (2000). The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). Human molecular genetics 9, 1543-1551.

Ngoka, L.C. (2008). Sample prep for proteomics of breast cancer: proteomics and gene ontology reveal dramatic differences in protein solubilization preferences of radioimmunoprecipitation assay and urea lysis buffers. Proteome science 6, 30.

Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J., Jr., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P., and Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature genetics 4, 221-226.

Otten, A.D., and Tapscott, S.J. (1995). Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. Proceedings of the National Academy of Sciences of the United States of America 92, 5465-5469.

Pato, C.N., Macedo, A., Ambrosio, A., Vincent, J.B., Bauer, A., Schindler, K., Xu, J., Coelho, I., Dourado, A., Valente, J., et al. (2000). Detection of expansion regions in Portuguese bipolar families. American journal of medical genetics 96, 854-857.

Philips, A.V., Timchenko, L.T., and Cooper, T.A. (1998). Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science (New York, NY 280, 737-741.

Polyak, K., Lee, M.H., Erdjument-Bromage, H., Koff, A., Roberts, J.M., Tempst, P., and Massague, J. (1994). Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. Cell 78, 59-66.

Pulst, S.M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., et al. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nature genetics 14, 269-276.

Pyronnet, S., Pradayrol, L., and Sonenberg, N. (2000). A cell cycle-dependent internal ribosome entry site. Molecular cell 5, 607-616.

Ranum, L.P., and Day, J.W. (2002). Myotonic dystrophy: clinical and molecular parallels between myotonic dystrophy type 1 and type 2. Current neurology and neuroscience reports 2, 465-470.

Ranum, L.P., and Day, J.W. (2004). Pathogenic RNA repeats: an expanding role in genetic disease. Trends Genet 20, 506-512.

Robinson, D.N., and Cooley, L. (1997). Drosophila kelch is an oligomeric ring canal actin organizer. The Journal of cell biology 138, 799-810.

Sala, S.G., Munoz, U., Bartolome, F., Bermejo, F., and Martin-Requero, A. (2008). HMG-CoA reductase inhibitor simvastatin inhibits cell cycle

progression at the G1/S checkpoint in immortalized lymphocytes from Alzheimer's disease patients independently of cholesterol-lowering effects. The Journal of pharmacology and experimental therapeutics 324, 352-359.

Sasagawa, N., Takahashi, N., Suzuki, K., and Ishiura, S. (1999). An expanded CTG trinucleotide repeat causes trans RNA interference: a new hypothesis for the pathogenesis of myotonic dystrophy. Biochemical and biophysical research communications 264, 76-80.

Saveliev, A., Everett, C., Sharpe, T., Webster, Z., and Festenstein, R. (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. Nature 422, 909-913.

Schols, L., Bauer, I., Zuhlke, C., Schulte, T., Kolmel, C., Burk, K., Topka, H., Bauer, P., Przuntek, H., and Riess, O. (2003). Do CTG expansions at the SCA8 locus cause ataxia? Ann Neurol 54, 110-115.

Shaw, G., Morse, S., Ararat, M., and Graham, F.L. (2002). Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. Faseb J 16, 869-871.

Sheline, C.T., and Choi, D.W. (1998). Neuronal death in cultured murine cortical cells is induced by inhibition of GAPDH and triosephosphate isomerase. Neurobiology of disease 5, 47-54.

Shih, S.C., and Claffey, K.P. (1999). Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. The Journal of biological chemistry 274, 1359-1365.

Silveira, I., Alonso, I., Guimaraes, L., Mendonca, P., Santos, C., Maciel, P., Fidalgo De Matos, J.M., Costa, M., Barbot, C., Tuna, A., et al. (2000). High germinal instability of the (CTG)n at the SCA8 locus of both expanded and normal alleles. Am J Hum Genet 66, 830-840.

Sobrido, M.J., Cholfin, J.A., Perlman, S., Pulst, S.M., and Geschwind, D.H. (2001). SCA8 repeat expansions in ataxia: a controversial association. Neurology 57, 1310-1312.

Sonenberg, N. (1994). Regulation of translation and cell growth by eIF-4E. Biochimie 76, 839-846.

Stevanin, G., Herman, A., Durr, A., Jodice, C., Frontali, M., Agid, Y., and Brice, A. (2000). Are (CTG)n expansions at the SCA8 locus rare polymorphisms? Nature genetics 24, 213; author reply 215.

Stoneley, M., and Willis, A.E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. Oncogene 23, 3200-3207.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature 403, 41-45.

Sulek, A., Hoffman-Zacharska, D., Bednarska-Makaruk, M., Szirkowiec, W., and Zaremba, J. (2004). Polymorphism of trinucleotide repeats in non-translated regions of SCA8 and SCA12 genes: allele distribution in a Polish control group. J Appl Genet 45, 101-105.

Sultana, R., Perluigi, M., Newman, S.F., Pierce, W.M., Cini, C., Coccia, R., and Butterfield, D.A. Redox proteomic analysis of carbonylated brain

proteins in mild cognitive impairment and early Alzheimer's disease. Antioxidants & redox signaling 12, 327-336.

Taneja, K.L., McCurrach, M., Schalling, M., Housman, D., and Singer, R.H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. The Journal of cell biology 128, 995-1002.

Tang, H.L., Yeh, L.S., Chen, N.K., Ripmaster, T., Schimmel, P., and Wang, C.C. (2004). Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons. The Journal of biological chemistry 279, 49656-49663.

Tazon, B., Badenas, C., Jimenez, L., Munoz, E., and Mila, M. (2002). SCA8 in the Spanish population including one homozygous patient. Clin Genet 62, 404-409.

Tian, B., White, R.J., Xia, T., Welle, S., Turner, D.H., Mathews, M.B., and Thornton, C.A. (2000). Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. RNA (New York, NY 6, 79-87.

Timchenko, N.A., Cai, Z.J., Welm, A.L., Reddy, S., Ashizawa, T., and Timchenko, L.T. (2001). RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. The Journal of biological chemistry 276, 7820-7826.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001). Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics 1, 377-396.

Topisirovic, I., Dragasevic, N., Savic, D., Ristic, A., Keckarevic, M., Keckarevic, D., Culjkovic, B., Petrovic, I., Romac, S., and Kostic, V.S. (2002). Genetic and clinical analysis of spinocerebellar ataxia type 8 repeat expansion in Yugoslavia. Clin Genet 62, 321-324.

Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.C., and Vagner, S. (2003). Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biology of the cell / under the auspices of the European Cell Biology Organization 95, 169-178.

Tsai, H.F., Lin, S.J., Li, C., and Hsieh, M. (2005). Decreased expression of Hsp27 and Hsp70 in transformed lymphoblastoid cells from patients with spinocerebellar ataxia type 7. Biochemical and biophysical research communications 334, 1279-1286.

Tsuchida, T., Ensini, M., Morton, S.B., Baldassare, M., Edlund, T., Jessell, T.M., and Pfaff, S.L. (1994). Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. Cell 79, 957-970.

van Swieten, J.C., Brusse, E., de Graaf, B.M., Krieger, E., van de Graaf, R., de Koning, I., Maat-Kievit, A., Leegwater, P., Dooijes, D., Oostra, B.A., et al. (2003). A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]. Am J Hum Genet 72, 191-199.

Vincent, J.B., Yuan, Q.P., Schalling, M., Adolfsson, R., Azevedo, M.H., Macedo, A., Bauer, A., DallaTorre, C., Medeiros, H.M., Pato, M.T., et

al. (2000). Long repeat tracts at SCA8 in major psychosis. American journal of medical genetics 96, 873-876.

Waters, M.F., Minassian, N.A., Stevanin, G., Figueroa, K.P., Bannister, J.P., Nolte, D., Mock, A.F., Evidente, V.G., Fee, D.B., Muller, U., et al. (2006). Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes. Nature genetics 38, 447-451.

Wegrzyn, J.L., Drudge, T.M., Valafar, F., and Hook, V. (2008). Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites. BMC bioinformatics 9, 232.

Wen, F.C., Li, Y.H., Tsai, H.F., Lin, C.H., Li, C., Liu, C.S., Lii, C.K., Nukina, N., and Hsieh, M. (2003). Down-regulation of heat shock protein 27 in neuronal cells and non-neuronal cells expressing mutant ataxin-3. FEBS Lett 546, 307-314.

Worth, P.F., Houlden, H., Giunti, P., Davis, M.B., and Wood, N.W. (2000). Large, expanded repeats in SCA8 are not confined to patients with cerebellar ataxia. Nature genetics 24, 214-215.

Wu, Y.R., Chen, I.C., Soong, B.W., Kao, S.H., Lee, G.C., Huang, S.Y., Fung, H.C., Lee-Chen, G.J., and Chen, C.M. (2009). SCA8 repeat expansion: large CTA/CTG repeat alleles in neurological disorders and functional implications. Human genetics 125, 437-444.

Xue, F., and Cooley, L. (1993). kelch encodes a component of intercellular bridges in Drosophila egg chambers. Cell 72, 681-693.

Yabe, I., Sasaki, H., Chen, D.H., Raskind, W.H., Bird, T.D., Yamashita, I., Tsuji, S., Kikuchi, S., and Tashiro, K. (2003). Spinocerebellar ataxia type 14 caused by a mutation in protein kinase C gamma. Archives of neurology 60, 1749-1751.

Yamin, G., Glaser, C.B., Uversky, V.N., and Fink, A.L. (2003). Certain metals trigger fibrillation of methionine-oxidized alpha-synuclein. The Journal of biological chemistry 278, 27630-27635.

Yan, J.X., Devenish, A.T., Wait, R., Stone, T., Lewis, S., and Fowler, S. (2002). Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. Proteomics 2, 1682-1698.

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y., and Lee, C.C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nature genetics 15, 62-69.

Table 1. Lymphoblastoid cell lines used in real-time PCR

Lymphoblastoid lines		Age/sex	Diagnosis	Repeat no.
P1	CC4999	86/M	SCA	29/**88**
P2	ND844 (H600)	60/F	PD	28/**92**
P3	CC4998	72/M	SCA	18/**95~185**
P4	ND783 (H1415)	50/F	PD	21/**87**
P5	ND918 (H327)	71/F	PD	23/**82**
P6	ND7218 (H1410)	67/M	PD	18/**84**
C1	ND2955	80/M	NC	18/28
C2	ND2105	63/F	NC	18/23
C3	ND2106	68/M	NC	26/28
C4	ND1729	70/F	NC	25/30
C5	ND7219 (H1411)	63/M	NC	24/**85**

SCA, spinocerebellar ataxia; PD, Parkinson's disease; NC, normal control.

Table 2. Summary of the genes identified by microarray analysis

23 CR / WT ↓	
BFSP1	beaded filament structural protein 1, filensin
BTG3	BTG family, member 3
C20orf160	chromosome 20 open reading frame 160
C20orf72	chromosome 20 open reading frame 72
CABP7	calcium binding protein 7
CDC91L1	CDC91 cell division cycle 91-like 1 (S. cerevisiae)
CHGA	chromogranin A (parathyroid secretory protein 1)
CHGB	chromogranin B (secretogranin 1)
CKM	creatine kinase, muscle
CSRP2BP	CSRP2 binding protein
DECR1	2,4-dienoyl CoA reductase 1, mitochondrial
DKFZP434B044	hypothetical protein DKFZp434B044
ELF1	E74-like factor 1
EYA2	eyes absent homolog 2 (Drosophila)
GK	glycerol kinase
KCNN4	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4
KIAA0998	KIAA0998
KLF1	Kruppel-like factor 1
LOC114990	Vasorin
LOC283352	hypothetical protein LOC283352
LOC440030	LOC440030
MGC24133	hypothetical protein MGC24133
MGC4655	hypothetical protein MGC4655
RAB32	RAB32, member RAS oncogene family
RIBC2	RIB43A domain with coiled-coils 2
SNRPB2	small nuclear ribonucleoprotein polypeptide B
SNX5	sorting nexin 5
THBS4	thrombospondin 4
TNNC2	troponin C2, fast
WBSCR5	Williams-Beuren syndrome chromosome region 5
ZNF217	zinc finger protein 217
ZNF239	zinc finger protein 239
ZP3	zona pellucida glycoprotein 3

23 CR / WT ↑	
CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)
ISL1	ISL1 transcription factor, LIM/homeodomain, (islet-1)
157 CR / 23 CR ↓	
AMFR	autocrine motility factor receptor
CRTAP	cartilage associated protein
FKBP10	FK506 binding protein 10, 65 kDa
HBQ1	hemoglobin, theta 1
IGFBP7	insulin-like growth factor binding protein 7
KLHL1AS	kelch-like 1 antisense (Drosophila)
LITAF	lipopolysaccharide-induced TNF factor
LOC143903	layilin
PYCR1	pyrroline-5-carboxylate reductase 1
RHOB	ras homolog gene family, member B
TBL1X	transducin (beta)-like 1X-linked
157 CR / 23 CR ↑	
FKSG14	leucine zipper protein FKSG14
LOC440030	LOC440030
157 CR / WT ↓	
CEBPE	CCAAT/enhancer binding protein (C/EBP), epsilon
HBQ1	hemoglobin, theta 1
IGFBP7	insulin-like growth factor binding protein 7
IGFBPL1	insulin-like growth factor binding protein-like 1
LOC389772	similar to Osteotesticular phosphatase; protein tyrosine phosphatase receptor
LOC400523	hypothetical LOC400523
MYD88	myeloid differentiation primary response gene (88)
PYCR1	pyrroline-5-carboxylate1reductase 1
RAB32	RAB32, member RAS oncogene family
SYTL1	synaptotagmin-like 1
TBL1X	transducin (beta)-like 1X-linked
157 CR / WT ↑	
CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)
FKSG14	leucine zipper protein FKSG14
GSTP1	glutathione S-transferase pi
ISL1	ISL1 transcription factor LIM/homeodomain, (islet-1)
SPON1	spondin 1, extracellular matrix protein

Table 3. Summary of the proteins identified by proteomic analysis

Identified protein	Gene name	Score	MW (kDa)	pI	157 CR/ 23 CR ratio
Peptidylprolyl isomerase A isoform 2	*PPIA*	61	11411	7.74	-- [a] ↓
Chain B, Human Triosephosphate isomerase	*TPI*	109	26666	6.50	0.47 ↓
ACTB protein	*ACTB*	75	40194	5.55	-- [a] ↓
Proteasome 26S ATPase subunit 4 isoform 1	*PSMC4*	69	47337	5.09	3 ↑
Enolase 1	*ENO1*	93	47139	7.01	0.43 ↓
Heterogeneous nuclear ribonucleoprotein	*HNRNPL*	67	60195	6.65	0.3 ↓
Phosphoglycerate mutase 1 (brain)	*PGAM1*	70	28786	6.67	0.35 ↓
Proline 4-hydroxylase β-subunit	*P4HB*	130	57081	4.76	-- [b] ↑

[a] No signal was detected on 157 CR gel

[b] No signal was detected on 23 CR gel

Table 4. Summary of the RNA binding proteins of *ATXN8OS*

Identified protein	Gene name	Score	MW (kDa)	pI
Zinc finger, FYVE domain containing 26	*ZFYVE26*	74	54794	8.90
Hypothetical protein	--	59	55787	9.01
Chromosome X open reading frame 34	*CXorf34*	52	51453	8.82
Inositol polyphosphate multikinase	*IPMK*	44	47192	7.27
Enolase 1	*ENO1*	51	47139	7.01
2-phosphopyruvate-hydratase alpha-enolase; carbonate dehydratase	*CA*	50	47079	7.01
p53 cellular tumor antigen	*TP53*	48	43606	6.25
Unnamed protein product	--	34	45811	6.64
Poly(rC) binding protein 1	*PCBP1*	31	37502	6.66
Peptidylprolyl isomerase D	*PPID*	39	40738	6.77
Proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	*PSMD6*	46	40315	5.62
RAS guanyl releasing protein 4	*RASGRP4*	42	39135	7.28
Unknown (protein for IMAGE: 4338484)	--	39	39098	7.85
Zinc finger protein zfp47	*ZKSCAN3*	43	37660	8.22
PREDICTED: similar to Serine/threonine-protein kinase MARK2 (MAP/microtubule affinity-regulating kinase 2) (ELKL motif kinase) (EMK1)	*MARK2*	35	35054	8.37
Leucine rich repeat containing 10	*LRRC10*	31	28389	8.75
DEAH (Asp-Glu-Ala-His) box polypeptide 16	*DHX16*	28	26249	8.83

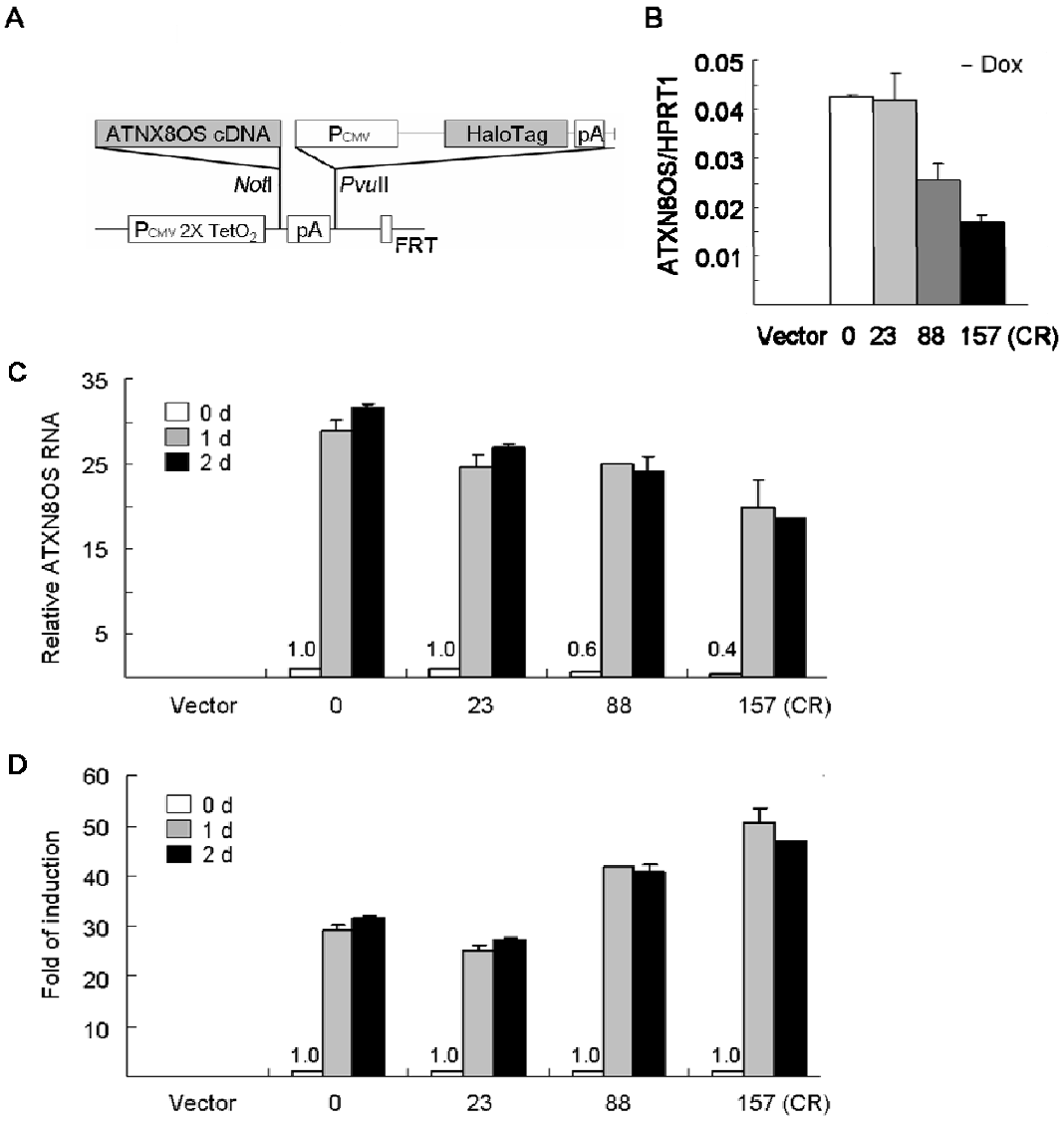

Figure 1. *ATXN8OS* 0~157 CR cells. (A) The features of pcDNA5/FRT/TO-ATXN8OS plasmids. *ATXN8OS* cDNA expression is driven by hybrid CMV/$TetO_2$ promoter (P_{CMV} 2X $TetO_2$). Also shown is the fragment containing CMV promoter, HaloTag open reading frame and SV40 late poly(A) signal placed at the *Pvu*II site between bovine growth hormone poly(A) signal and Flp recombination target (FRT) site. (B) Real-time PCR quantification of *ATXN8OS* RNA level relative to endogenous *HPRT1* RNA in *ATXN8OS* CR cells grown without doxycycline. (C) Real-time PCR quantification of relative

ATXN8OS RNA level after addition of doxycycline for 0~2 days. To normalize, the relative *ATXN8OS* RNA in 0 CR cells at the time of doxycycline addition (0 d) is set as 1.0. (D) The fold of induction after addition of doxycycline for 0~2 days. To normalize, the relative *ATXN8OS* RNA levels in each CR line at the time of doxycycline addition (0 d) are set as 1.0.

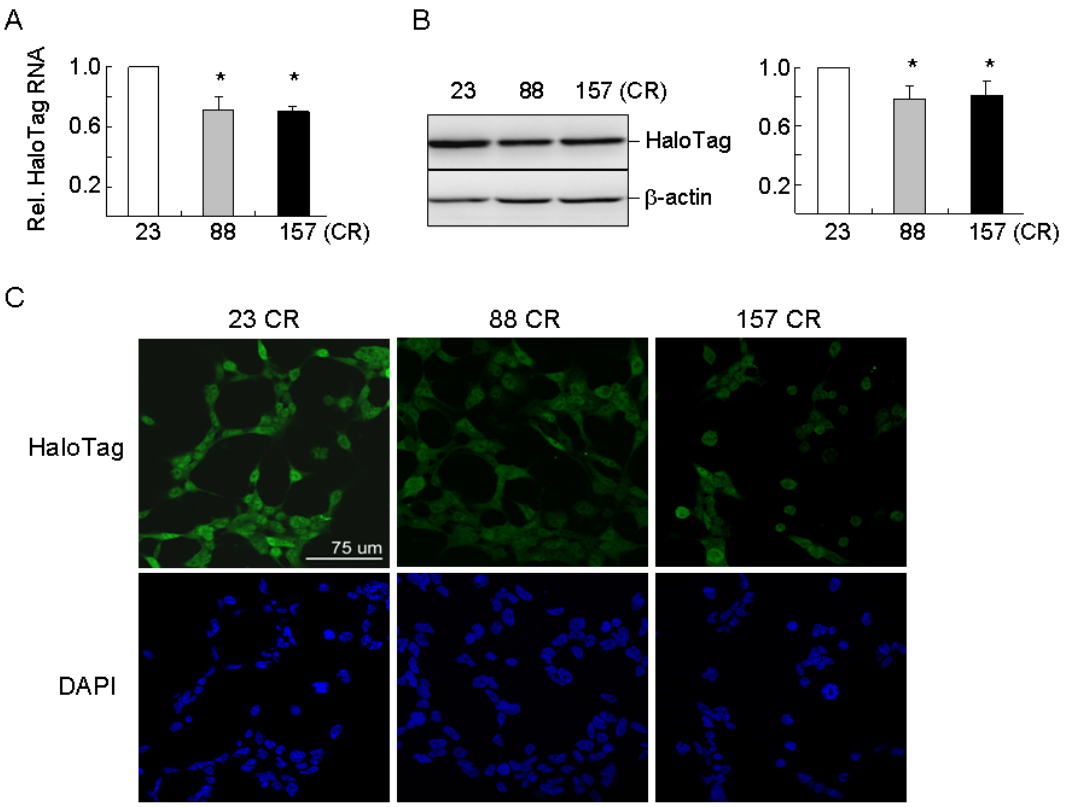

Figure 2. Expression of HaloTag gene located next to *ATXN8OS* cDNA gene. (A) Real-time PCR quantification of HaloTag RNA level relative to endogenous *HPRT1* RNA. To normalize, the relative HaloTag RNA in 23 CR cells is set as 1.0. (B) Representative western blot image of CR lines using HaloTag and β-actin antibodies. Levels of HaloTag were normalized with an internal control (β-actin). The relative immunoreactivity of HaloTag is shown in the right panel. For both (A) and (B), data are expressed as the mean ± SEM values from three independent experiments. * indicates $P < 0.05$. (C) Fluorescence microscopy examination after immunostaining using HaloTag antibody (green). Nuclei were counterstained with DAPI (blue).

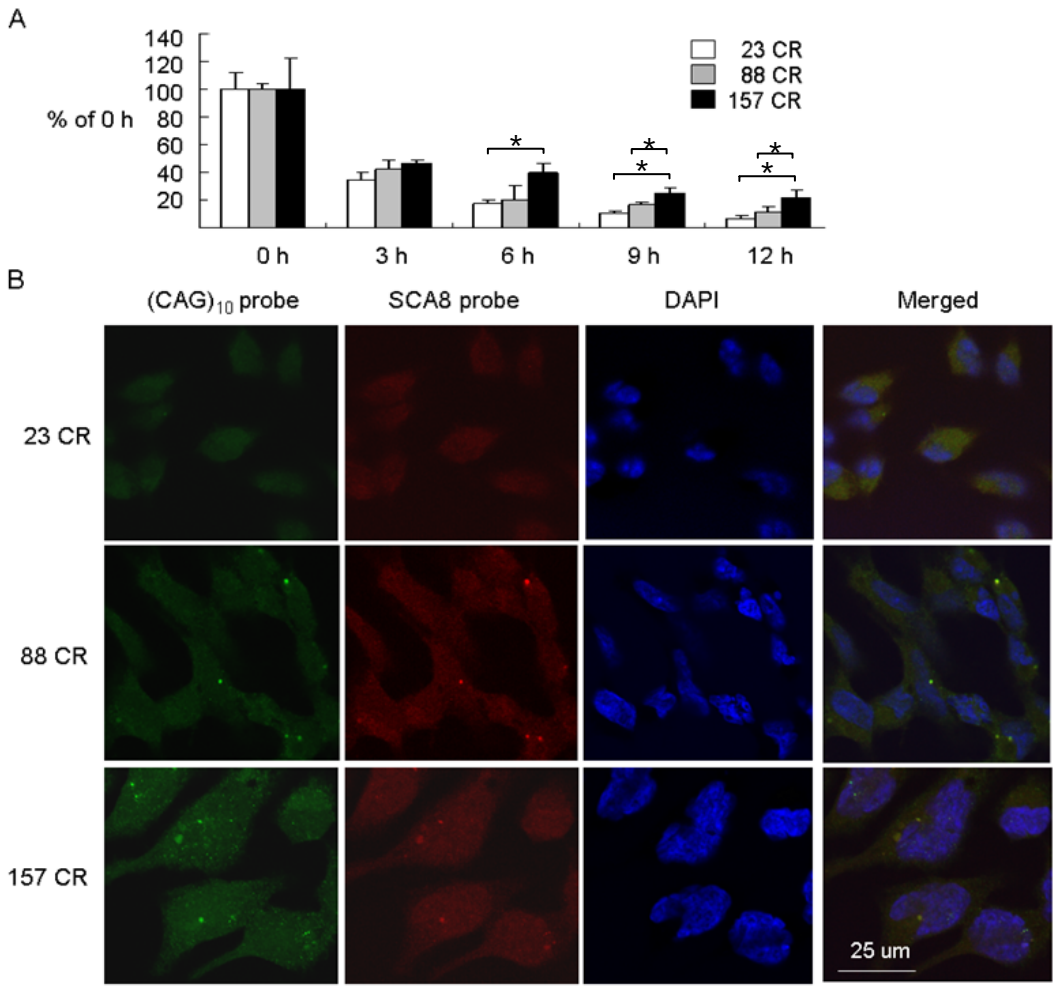

Figure 3. *ATXN8OS* transcript stability and ribonuclear foci formation with CUG repeat expansion. (A) Real-time PCR quantification of *ATXN8OS* RNA level relative to endogenous *HPRT1* RNA following addition of doxycycline for 2 days and actinomycin D treatment for 0, 3, 6, 9 or 12 hours. The relative *ATXN8OS* RNA in 23, 88 and 157 CR cells at the time of actinomycine D addition (0 h) is set as 100%. ***** indicates the difference between the indicated samples are statistically significant ($P < 0.05$) (B) Ribonuclear foci formation with *ATXN8OS* CUG expansion. *ATXN8OS* 23, 88 and 157 CR cells were grown with doxycycline for 2 days and analyzed by RNA-FISH using a Cy3 labeled $(CAG)_{10}$ (green) or Cy5 labeled *ATXN8OS*

unique sequence (red) oligonucleotide probe. Nuclei were counterstained with DAPI (blue).

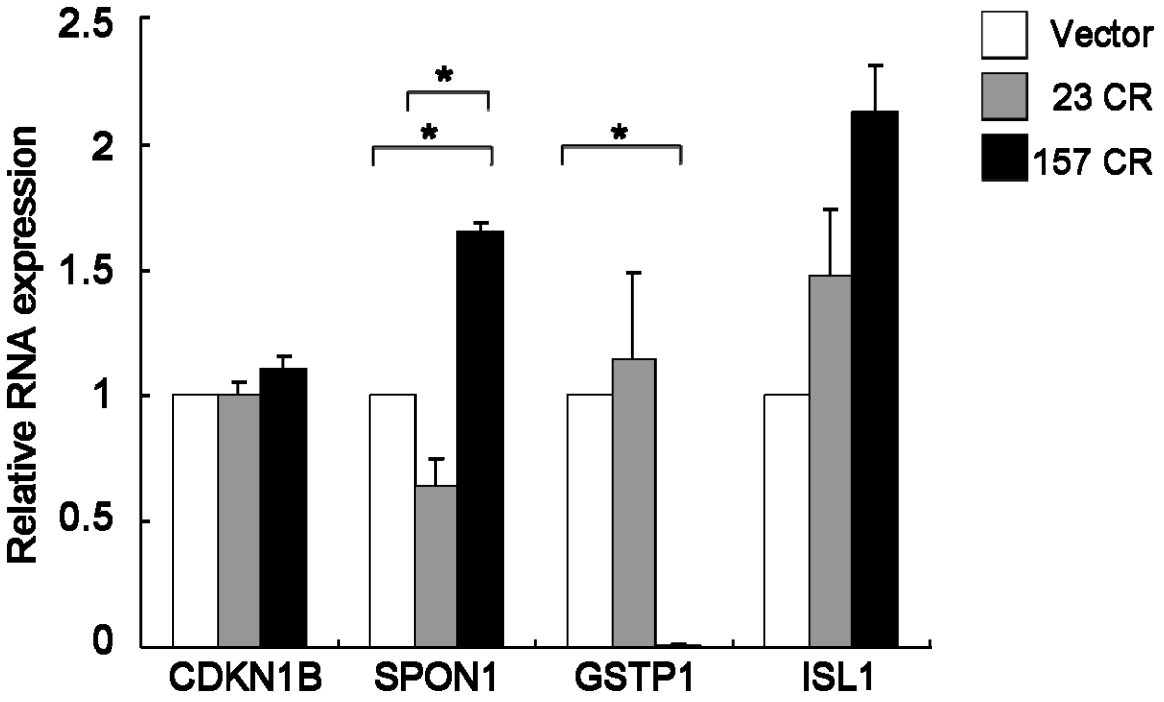

Figure 4. RNA expression of candidate genes validated by real-time PCR. No significant difference was observed in *CDKN1B* and *ISL1* expression level. Expression of *SPON1* is significantly higher in *ATXN8OS* 157 CR cells compared to vector cell line and *ATXN8OS* 23 CR cells. Contrarily, *GSTP1* expression was highly reduced in *ATXN8OS* 157 CR cells. ***** indicates the difference between the indicated samples are statistically significant ($P < 0.05$).

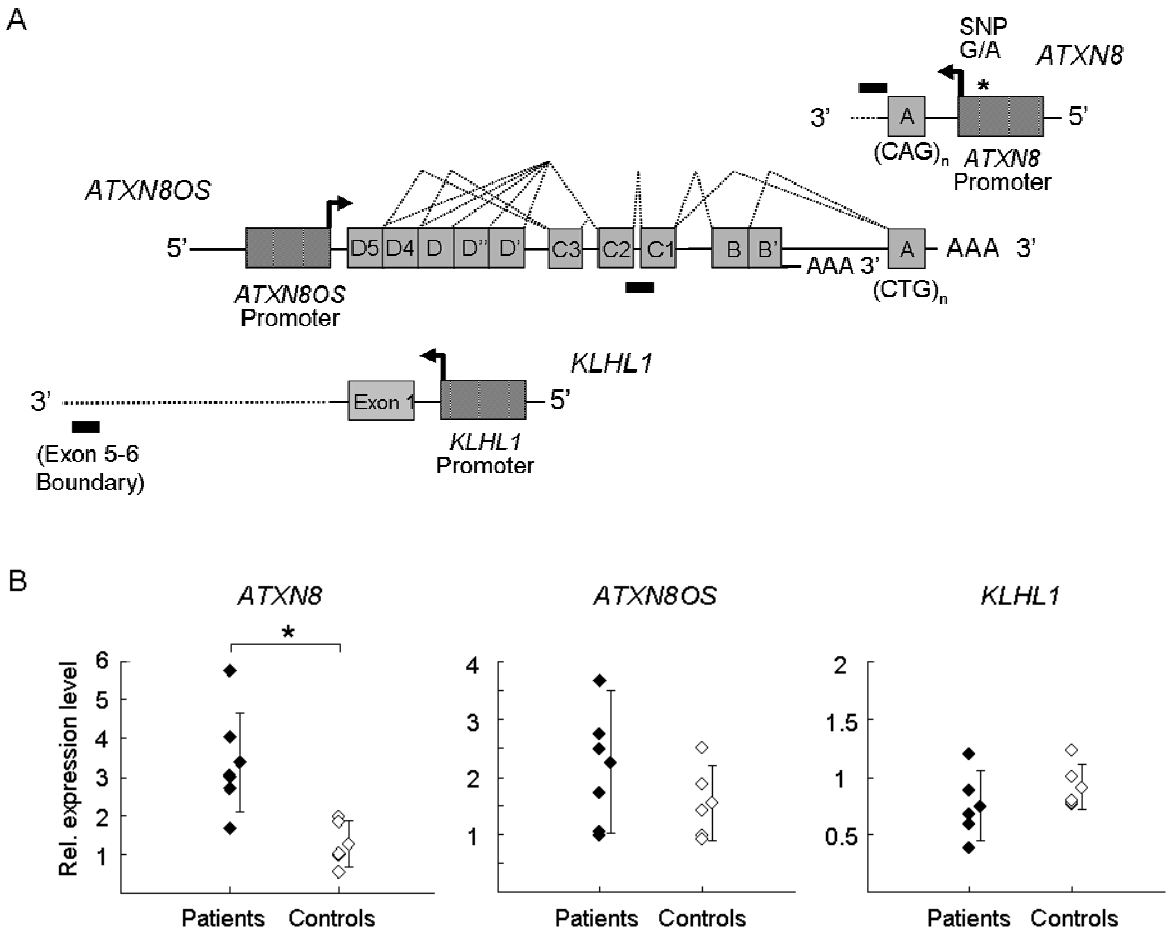

Figure 5. Analysis of *ATXN8OS*, *ATXN8*, and *KLHL1* expression. (A) Genomic organization of *ATXN8OS*, *ATXN8*, and *KLHL1* genes in humans. Straight line represents genomic DNA, gray boxes represent promoters and exons, black boxes represent gene-specific probes for real-time PCR quantification. (B) Real-time PCR quantification of *ATXN8*, *ATXN8OS* and *KLHL1* RNA level relative to endogenous *HPRT1* RNA in lymphoblastoid cells from controls and patients with SCA8 expansions. To normalize, expression level in control C1 is set as 1.0. The ***** indicates the difference between the indicated samples are statistically significant ($P < 0.05$).

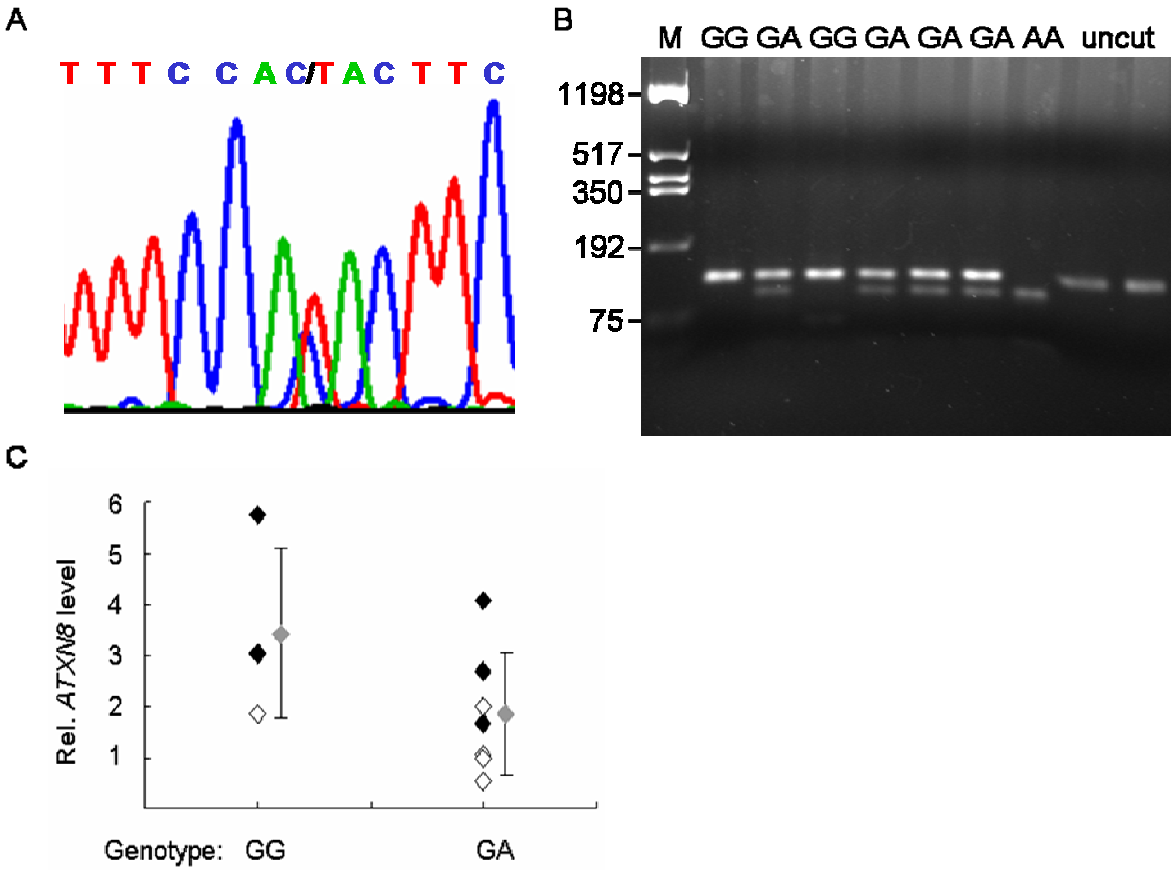

Figure 6. *ATXN8* -62 G/A SNP. (A) DNA sequence analysis of the novel SNP in CTG direction. (B) Mismatched PCR and *Mph*1103I restriction analysis of -62 G/A SNP. Lane M (*Hin*fI digest of pGEM4 DNA) is the size marker. (C) *ATXN8* RNA expression levels in lymphoblastoid cells (patients, black diamond; controls, white diamond; median, gray diamond) carrying -62 G/A GG (n = 4) and GA (n = 7) genotypes.

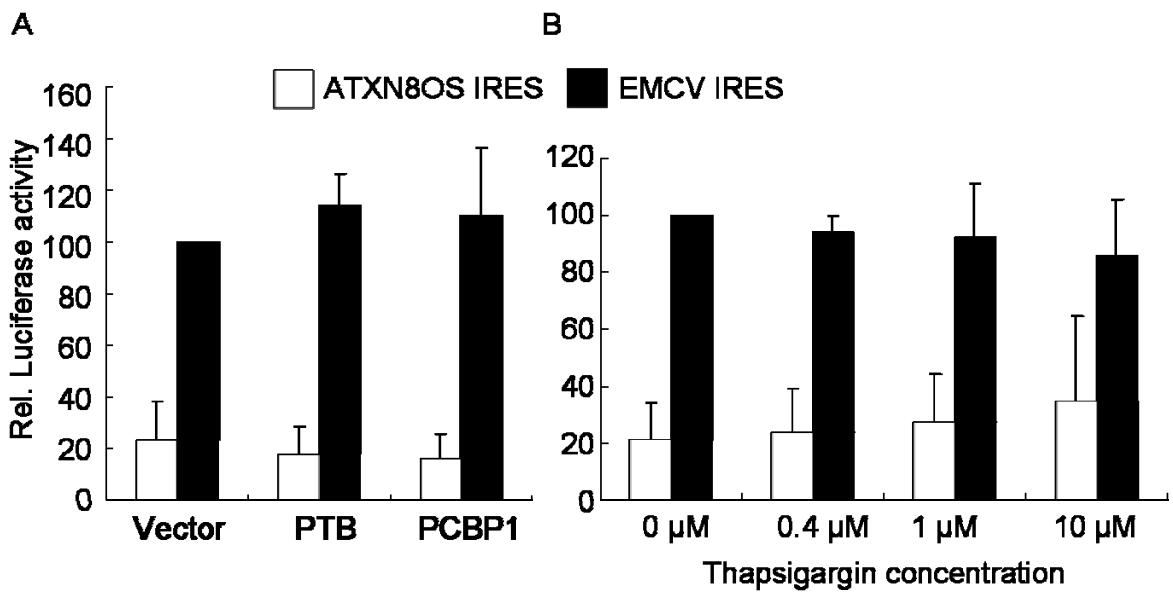

Figure 7. Dual luciferase reporter assay. (**A**) Co-tranfection of PTB or PCBP1 cDNA. (**B**) Addition of thapsigargin for 12 hours.

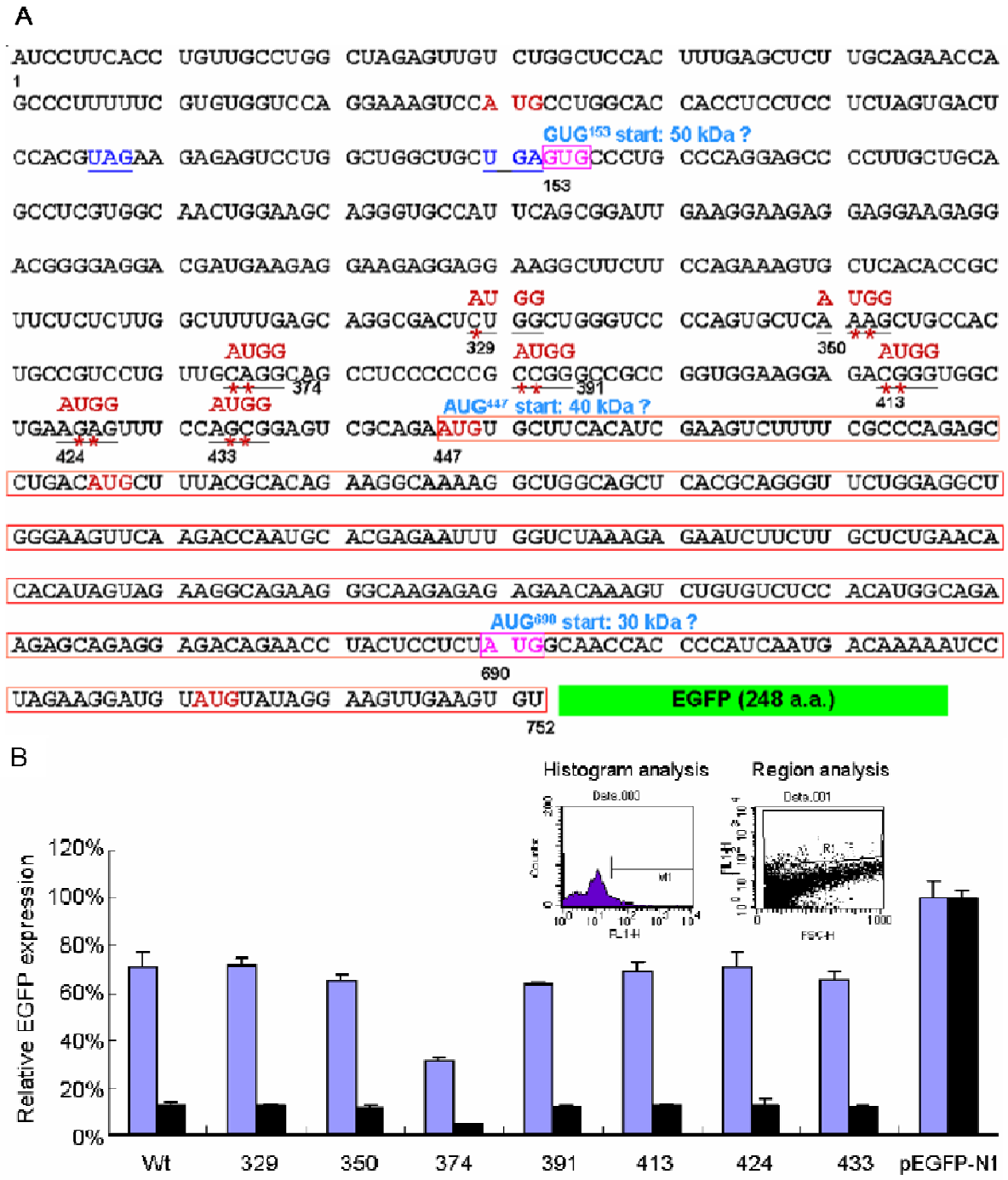

Figure 8. Locating the ribosome entry window in the *ATXN8OS* IRES. (**A**) The sequence of *ATXN8OS* 5' UTR indicating the positions at which initiation codons were introduced. (**B**) FACS analysis.

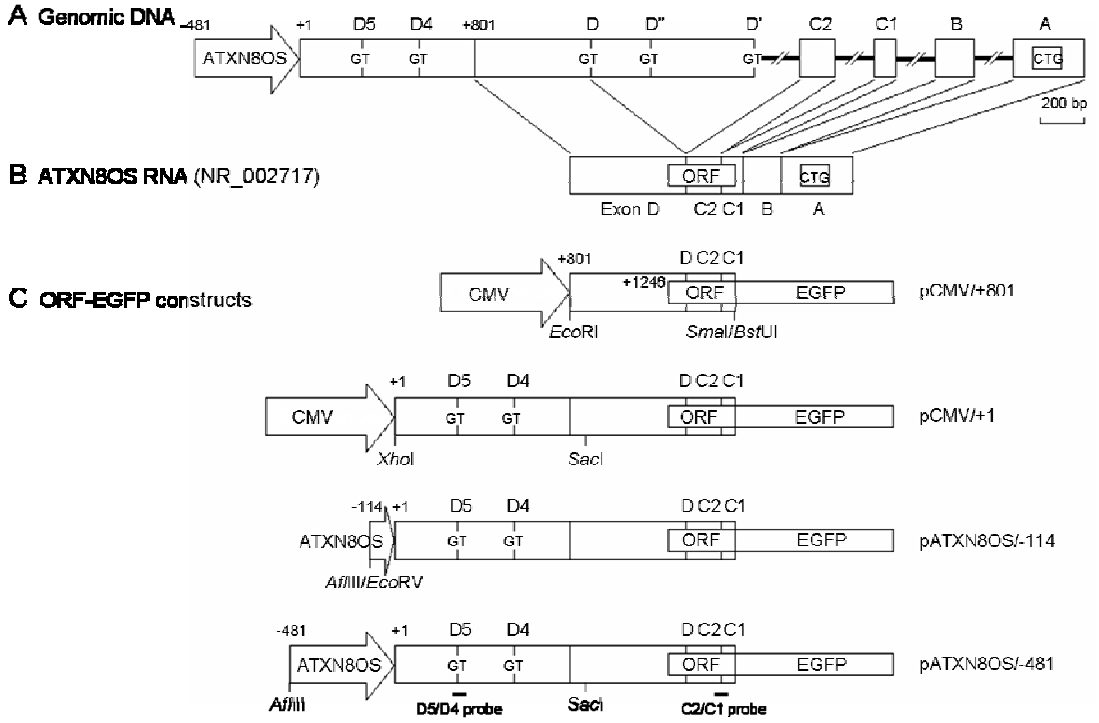

Figure 9. *ATXN8OS* genomic DNA, RNA and ORF-EGFP constructs. (A) *ATXN8OS* organization with promoter (open arrow), exons (open boxes) and functional splice donor sequences (GT) of D exons (D5, D4, D, D" and D') indicated. The CTG repeat tract is located in exon A. Transcription start site of exon D5 and exon D are represented by +1 and +801, respectively. (B) *ATXN8OS* RNA (NR_002717) generated from the splicing events represented by the wavy lines. The putative ORF is indicated by the open boxes inside the RNA. (C) ORF-EGFP constructs. A 752-bp cDNA fragment containing exon D, C2 and portion of C1 was inserted into pEGFP-N1 MCS so that *ATXN8OS* ORF was fused in-frame with the EGFP gene to generate pCMV/+801. A +1 ~ +800 *ATXN8OS* fragment was inserted between CMV promoter and exon D of pCMV/+801 to generate pCMV/+1. In pATXN8OS/-114 and -481, 114 and 481-bp *ATXN8OS* promoter fragments was used to replace the CMV promoter in pCMV/+1.

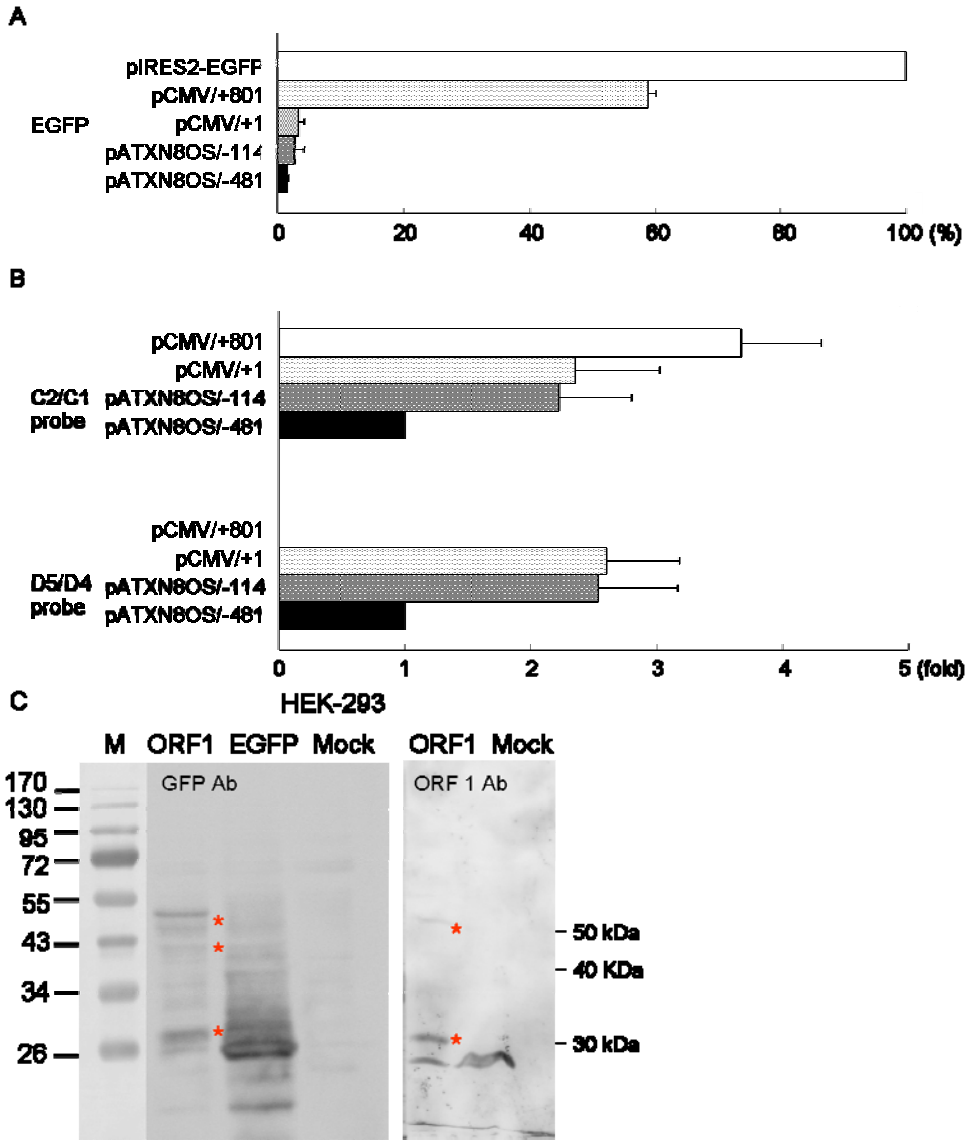

Figure 10. Transient expression of *ATXN8OS* ORF-EGFP constructs in HEK-293 cells. (A) FACS analysis of EGFP fluorescence. Levels of EGFP were expressed as percentages of the control vector pIRES-EGFP, which was set at 100%. Each value is the mean ± SD of three independent experiments each performed in duplicate. (B) Real-time PCR quantification of ORF-EGFP RNA level relative to endogenous *HPRT1* RNA. To normalize, expression level in pATXN8OS/-481 transfected cells is set as 1.0. (C) Western blot analysis of cells transfected with indicated plasmids using GFP and *ATXN8OS* ORF antibodies.

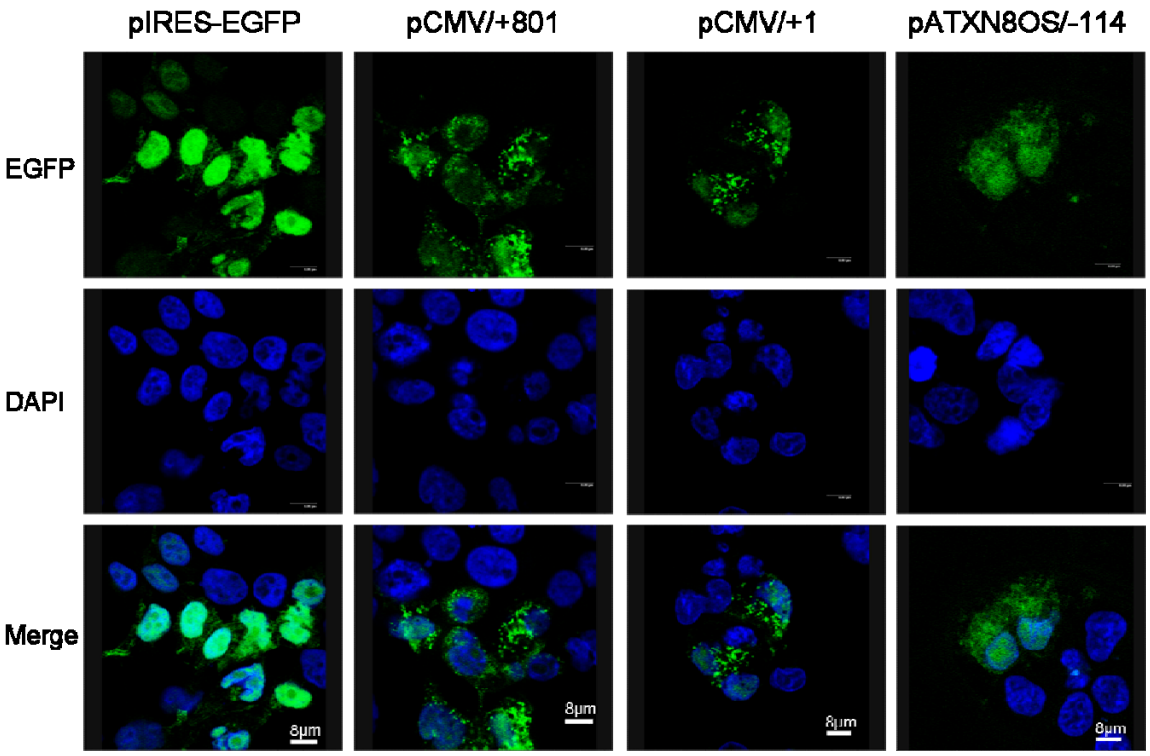

Figure 11. ORF-EGFP aggregation in transient expressed HEK-293 cells. Confocal images of cells expressing EGFP and ORF-EGFP (green). Nuclei were counterstained with DAPI (blue). (The scale bar = 8 μm)

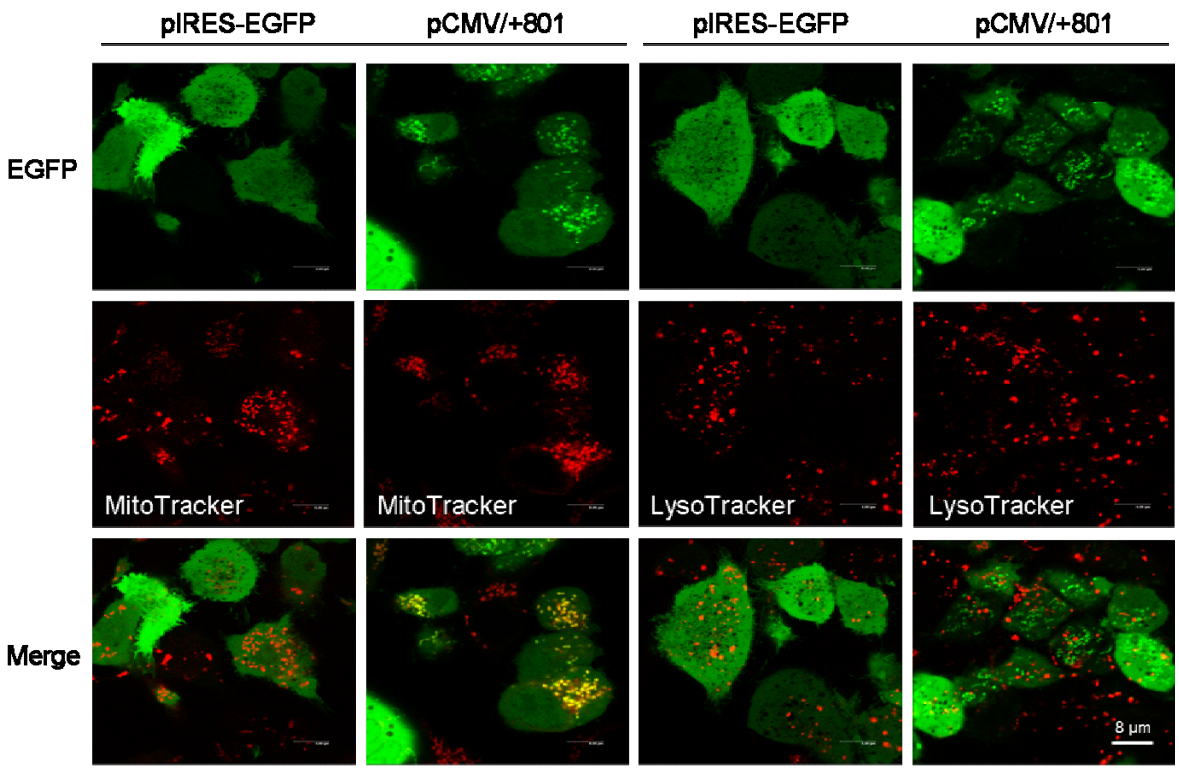

Figure 12. ORF-EGFP co-localized with mitochondria in HEK-293 cells. Confocal images of cells expressing EGFP and ORF-EGFP (green). Mitochondria and lysosome were stained with MitoTracker and LysoTracker (red). (The scale bar = 8 μm)

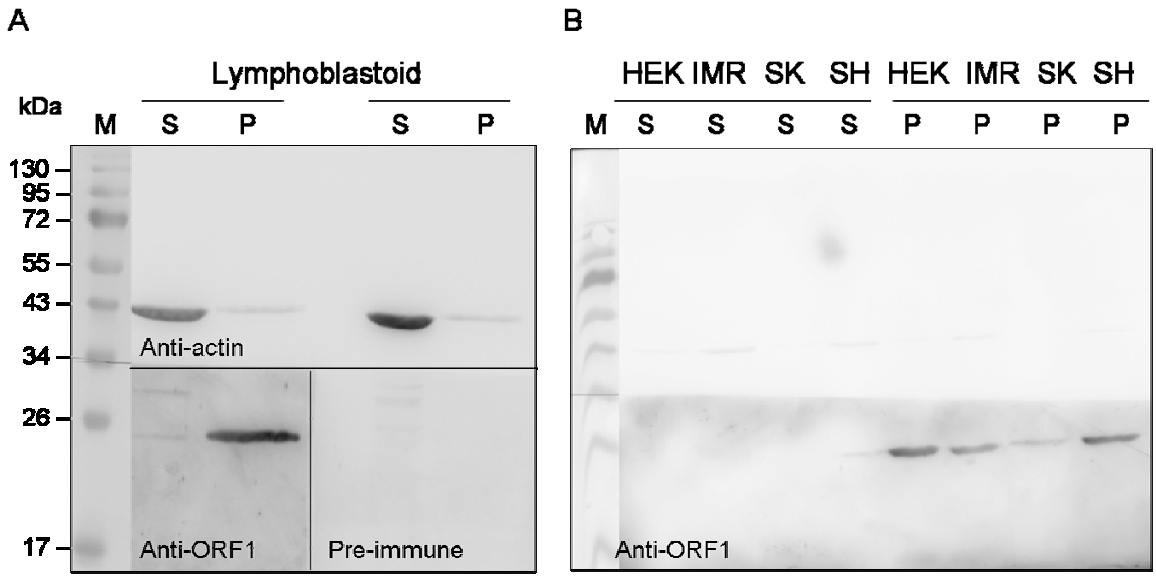

Figure 13. Western blot analysis of ORF protein expression in human cells using ORF antiserum and β-actin antibody. (A) In lymphoblastoid cells ORF protein was present in 8 M urea lysis buffer-insoluble fraction (P) whereas β-actin was enriched in supernatant fraction (S). (B) In addition to lymphoblastoid cells, ORF protein was present in HEK-293, IMR-32, SK-N-SH and SH-SY5Y (neuroblastoma) cells.

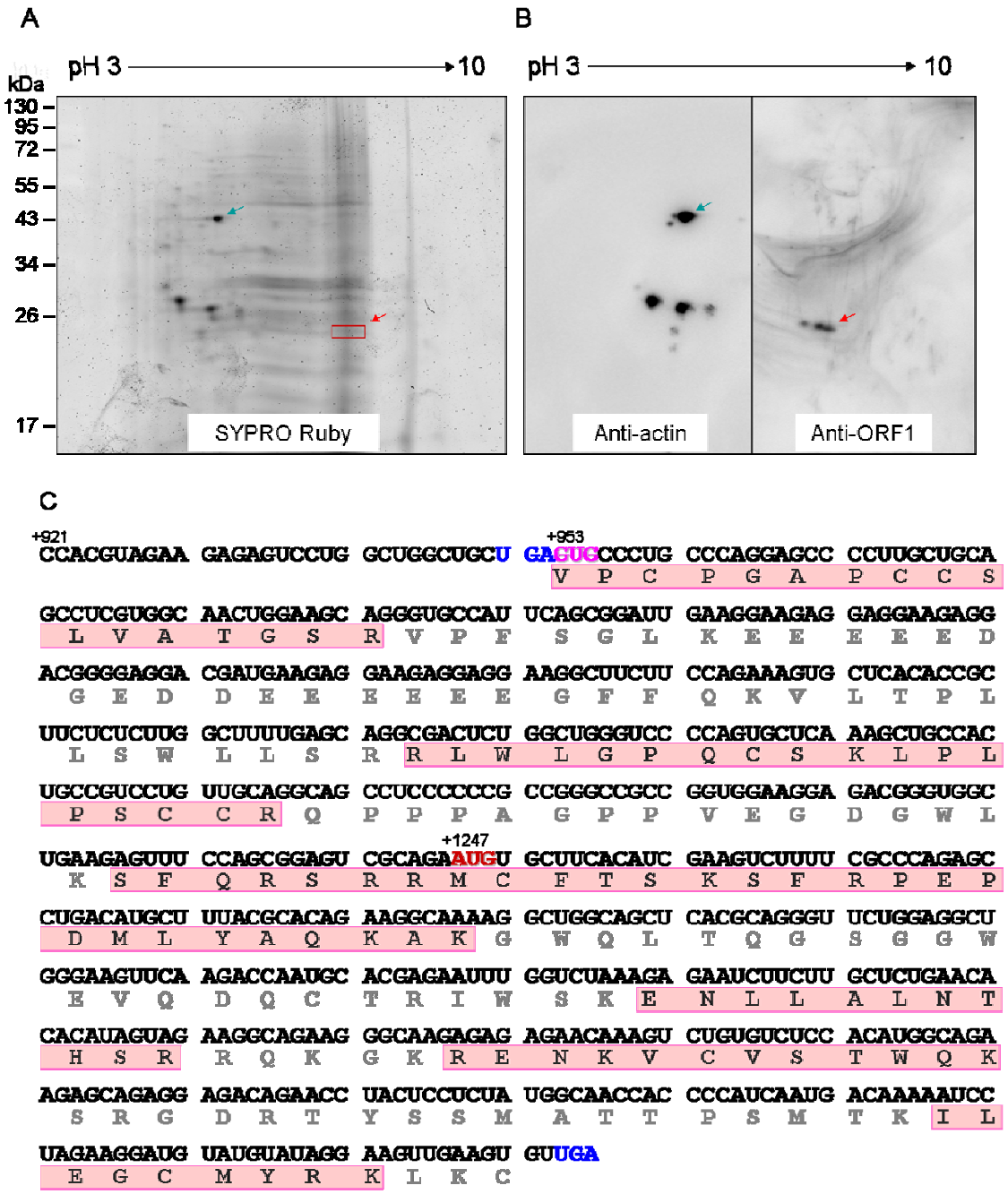

Figure 14. ORF protein identification. 2D image (A) and 2D immunoblot (B) of 8 M urea lysis buffer-insoluble fraction. (C) Six matched peptide with sequence coverage of 47% was obtained (shown in red box) by LC-MS/MS, including the N-terminal peptide VPCPGAPCCS LVATGSR which can only be generated from translation start.

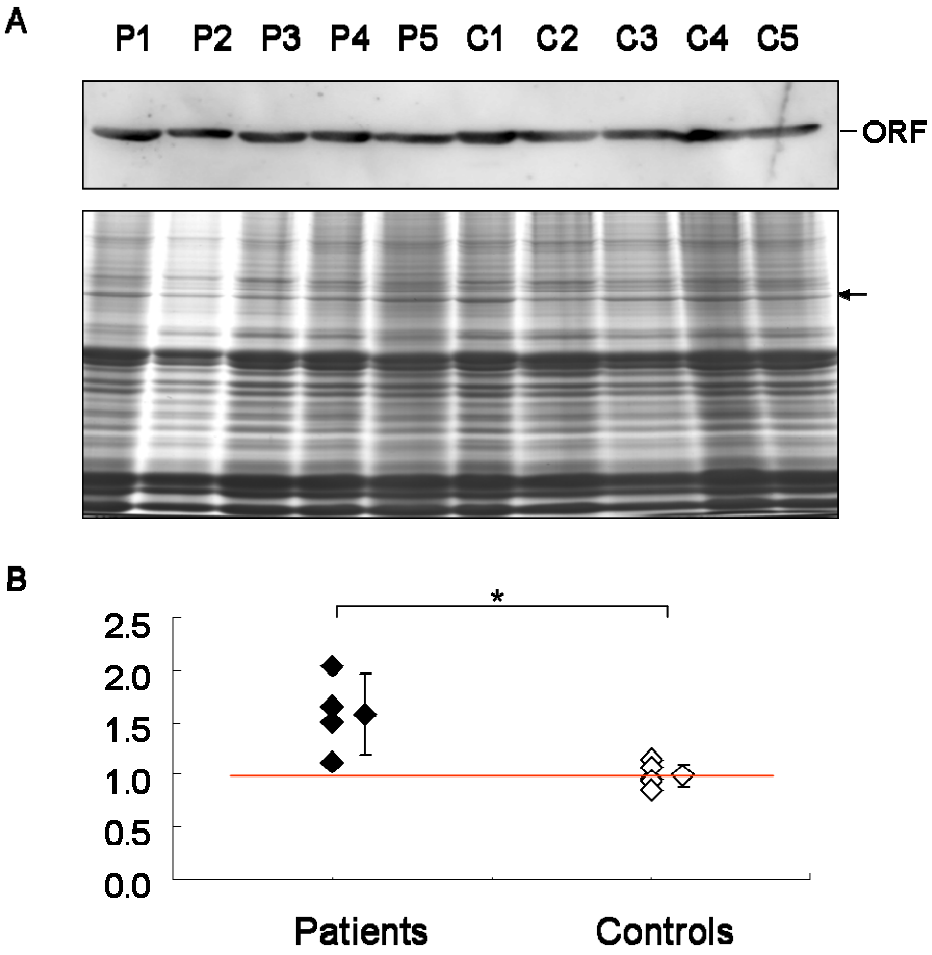

Figure 15. ORF protein expression in lymphoblastoid cells. (A) Representative image (top panel) and loading control (bottom panel, arrowhead) from controls (C1~C5) and patients with SCA8 expansions (P1, P3~P5). (B) The relative immunoreactivity of ORF protein in patients and controls. The * indicates a significant difference between the patient and the control samples ($P = 0.049$).