行政院國家科學委員會專題研究計畫　成果報告

(CAG)n 三聯核酸重複序列之基因轉殖動物模式建立及研究

(3/3)

計畫類別：個別型計畫
計畫編號：NSC93-2320-B-040-054-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：中山醫學大學生物醫學科學學系

計畫主持人：潘惠錦

報告類型：完整報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 10 月 20 日

行政院國家科學委員會專題研究計劃成果報告

(CAG)n 三聯核酸重複序列之基因轉殖動物模式建立及研究(3/3)

Establishment and studies of transgenic animal models expressing (CAG)n trinucleotide repeat expansion (3/3)

計劃編號 : NSC 91-2320-B-040-054

執行期限 : 93 年 8 月 1 日至 94 年 7 月 31 日

主持人 : 潘惠錦　中山醫學大學 生命科學系

共同主持人: 蕭光明 中山醫學大學 生命科學系

計畫參與人員: 張佑誠、許仁駿、李秋燕、、林明忠

中文摘要

在人類疾病中發現的三聯核酸重複序列絕大多數為 CAG 及其互補序列 CTG。CAG 重複序列主要位於轉譯區內而 CTG 則多位於非轉譯區內。為研究 CAG 三聯核酸重複序列是否會造成活體內的病理作用，我們製造基因轉殖小鼠使其表現肌肉特異性的 EGFP 基因，並在其 3'端不轉譯區含有不同長度的 CAG 重複序列。EGFP RNA 在所有的基因轉殖小鼠系中的表現量接近，但是在表現 200 次 CAG 重複的小鼠中 EGFP 的蛋白質表現量顯著下降。肌肉的組織化學分析顯示表現 200 次 CAG 重複的小鼠其肌肉細胞型態異常，並且其琥珀醯酸去氫酶和 NADH 還原酶的活性也不正常。藉由抓力及籠內活動力測試，我們觀察到 CAG_{200} 的成鼠其肌肉衰弱。CAG_{200} 的公鼠其精子數目正常但是游動力顯著下降，且其精子尾部的微管及粒腺體結構有缺陷，粒腺體功能也不正常。更進一步 CAG_{200} 的成鼠其肌肉在葡萄糖存在下會發生神經性的欒縮現象，在正常及對照組老鼠不會產生此現象。這些結果是首次證明，位於一個不相關基因的 3' 不轉譯區的 CAG 重複序列擴增會造成小鼠的肌肉病理現象，此與 RNA gain-of-function 機制符合。這也暗示非轉譯的 CAG 三聯核酸重複序列可能在人類疾病扮演一個角色。

關鍵詞：三聯核酸重複序列擴增、CAG、RNA 致病、基因轉殖小鼠

ABSTRACT

The most frequent trinucleotide repeats found in human disorders are CAG and its complementary sequence, CTG. CAG repeats are mostly found in coding regions whereas CTG expansions are located in untranslated regions. To investigate whether an untranslated CAG repeat expansion has pathogenic effects *in vivo*, we generated transgenic mice expressing muscle-specific *EGFP* transcripts with different sizes of CAG repeats in its 3'-untranslated region. While the expression levels of the *EGFP* transcripts were comparable in all transgenic lines, the EGFP protein levels were significantly reduced in mice expressing 200 CAG repeats. Histological analysis of the muscle revealed atypical cell morphology, as well as altered activities of succinate dehydrogenase and NADH reductase in CAG_{200} mice. These mice showed signs of low grip strength, less cage activity, and reduced fertility. The sperm counts of male CAG_{200} mice were normal but the sperm motility was significantly decreased. The reduction of sperm motility was associated with structural and mitochondrial defects in the sperm tails. Furthermore, abnormal muscle contracture was elicited in the skeletal muscle of CAG_{200} mice, suggesting an impaired energy metabolism. These results demonstrate, for the first time, that untranslated CAG repeat expansion in an unrelated mRNA can have pathogenic effects *in vivo*, consistent with the RNA gain-of-function model. It also suggests a possible role for untranslated CAG expansion in human disorders.

KEYWORDS: trinucleotide repeat expansion, CAG, RNA pathogenesis, transgenic mice

INTRODUCTION

A growing number of human neuromuscular disorders have been found to be caused by expansion of unstable trinucleotide repeats (1, 2). According to the location of the expansion, these trinucleotide repeat disorders are grouped into two categories. In the first category, the repeat expansion is located in the coding region of the affected gene. Usually these are caused by CAG repeats which result in mutant proteins with long track of polyglutamines. Examples include Huntington's disease (HD), spinal and bulbar muscular atrophy (SBMA), dentatorubral pallidoluysian atrophy (DRPLA), and spinocerebella ataxias (SCAs) -1, 2, 3, 6, 7 (3-7). In the second category, different repeat expansions are within the non-coding or untranslated region (UTR). Diseases associated with this category include fragile X syndrome (FRAX), Friedreich's ataxia (FRDA), myotonic dystrophy type 1 (DM1), SCA8 and SCA12 (8).

Although the expansions occur in different genetic loci, the trinucleotide repeat disorders share some common features. First, most of these disorders are dominantly inherited and the clinical manifestation develops progressively. Second, the size of the expanded repeat usually correlates positively with the clinical severity and reversely with the age of onset. Moreover, the repeats are unstable in both mitotic and meiotic divisions, resulting in somatic heterogeneity and germline instability. Since the repeat length frequently increases as it is transmitted, subsequent generations of the affected families usually experience an earlier age of onset and increase in clinical severity.

The mechanism by which expanded repeats lead to the pathogenic phenotype is complex and depends on the location of the repeat within a gene. Most CAG expansions are less than 150 repeats, located in coding regions and are translated into polyglutamine tracts within the corresponding protein. Recent lines of evidence suggest that aggregates of polyglutamine, protein misfolding and transcriptional dysregulation result in neuronal cell toxicity (3-6). In addition, loss of neurotrophic support due to reduced protein activity by polyglutamine expansions may also contribute to the pathogenesis of neurodegeneration (6, 9).

On the other hand, CTG expansions are usually very large and located exclusively outside coding sequences. The most intensively studied disorder in this category is DM1. DM1 is a multisystemic disorder characterized by skeletal muscle wasting and myotonia, cardiac conduction defects, insulin resistance and cataracts (10). It is caused by an expansion of CTG repeat in the 3'-UTR of the *DMPK* gene (11-13). This region overlaps the 5' end of a neighboring homeodomain-encoding gene, *SIX5* (14). Haploinsufficiency of the DMPK protein and altered expression of *SIX5* have been demonstrated to contribute to partial DM1 phenotypes, such as progressive myopathy, atrioventricular conduction abnormalities (15-17) and cataracts (18, 19). However, absence of the major feature of DM1, myotonia, suggests

that additional mechanisms must be involved. In support of this notion, mice expressing mRNA with long CUG repeats in the 3'-UTR of either *DMPK* or an unrelated transgene produce myotonia and myopathy (20, 21), indicating that transcripts with expanded CUG repeats are pathogenic. Moreover, expansion of CCTG repeats in a second locus (DM2) also leads to a clinical presentation that is strikingly similar to DM1 (22). Together these findings point to an important role for RNA gain-of-function in DM pathogenesis.

CUG repeats in RNA can form highly stable hairpin structures (23-25). Several RNA-binding proteins, including PKR, CUGBP1 and muscleblind, have been suggested to interact with CUG repeats both *in vitro* and *in vivo* (24, 26-28). Inappropriate activation of CUGBP1 disrupts alternative splicing regulation and leads to the pathological features of DM1 (29-31). By fluorescent in situ hybridization, mutant DMPK transcripts or chimeric RNA with expanded CUG repeats were shown to form nuclear foci in DM1 cells and model systems (20, 28, 32-35). Similar foci were also identified in DM2 cells (36, 37). Remarkably, muscleblind and its related proteins not only colocalize to the nuclear foci in DM1 and DM2 cells (28, 35-37), but *muscleblind* knockout mice also present abnormalities that are characteristic of DM disease (38). These data argue that expanded CUG repeats sequester and/or alter specific RNA-binding proteins that are required for proper cellular function, leading to cytotoxic effects.

Although current data suggest that mutant proteins with expanded polyglutamine tracts are key players in CAG repeat disorders, they do not rule out a possible role for expanded CAG RNA in pathogenesis. Indeed, CAG repeat RNAs are predicted to form stable secondary structures similar to those detected in DM1 (2, 24), and RNA-binding proteins that specifically interact with CAG-repeat sequences have been reported (39). Screening for genes that modify SCA1-induced neurodegeneration in a Drosophila model led to the identification of genes that encode RNA-binding proteins (40), suggesting that alteration of RNA processing is relevant to SCA1 pathogenesis.

To address whether untranslated CAG expansions are pathogenic *in vivo* in a trans-dominant manner, we generated transgenic mice expressing muscle-specific transcripts of the EGFP gene, of which no CAG repeat (CAG_0), 23 repeats (CAG_{23}) or 200 repeats (CAG_{200}) were inserted in the 3'-untranslated region. While the CAG_0 and CAG_{23} mice behaved normally, the CAG_{200} mice showed reduced cage activity and muscle weakness. We present histological, molecular and electrophysiological evidence showing that muscle differentiation and metabolism in CAG_{200} mice are impaired. These results are the first to show that expanded CAG RNA can result in pathogenic effects *in vivo* through a trans-dominant mechanism.

MATERALS AND METHODS

Transgene construction

A 1.5 kb 5'-flanking sequence of mouse gamma-sarcoglycan (gsg) gene was excised by *Kpn*I and *Xma*I from plasmid pGSG-EGFP-1 (kindly provided by Dr. Noguchi) (41), and inserted at the same sites upstream of the EGFP sequence in vector pEGFP-1 (Clonetech). This vector was used as a backbone for insertion of CAG repeat sequence. The CAG repeat sequence was generated by polymerase chain reaction (PCR) using two complementary primers, $(CTG)_{10}$ and $(CAG)_{17}$, which were denatured at 94℃ for 5 min, and then amplified by 30 cycles of 94℃, 30 sec; 37℃, 1 min; 72℃, 6 min, and a final extension at 72℃ for 10 min. The PCR products were separated on a gel and the appropriate sized fragments were recovered from the gel and ligated to pGEM-T Easy vector (Promega). The number and integrity of CAG repeats in each clone was then determined by sequencing. Two fragments containing 23 and 200 CAG repeats were cut out by *Eco*RI at the multiple cloning site of pGEM-T Easy, and inserted at a modified *Not*I site which is downstream of the EGFP's stop codon and upstream of the SV40 polyadenylation sequence in the pGSG-EGFP vector. These constructs were mapped and sequenced to confirm the correct position, orientation, integrity and number of CAG repeats.

Production of transgenic mice

The transgene fragments containing a 1.3-kb 5'-flanking sequence of the GSG gene (Fig 1A) were cut out from vectors with *Afl*II. Transgenic mice were generated by microinjecting the purified transgene fragments into the male pronuclei of one-cell FVB/N embryo. Mice were genotyped initially by PCR using tail DNA with primers f2 (5'-CCACATGAAGCAGCACGAC-3') and r1 (5'-GCTTTACTTGTACAGCTCGTC-3'). PCR results were further confirmed by Southern blotting using 10 μg of tail DNA (65), digested with *Bam*HI and hybridized with either an *EGFP*- or *GSG* promoter-specific probe. To confirm the length of CAG repeat sequence in mice, PCR-based Southern blotting was carried out by primers f1 (5'-CTGAAGTTCATCTGCACCAC-3') and r2 (5'-CTACAAATGTGGTATGGCTG-3'), using a labeled $(CAG)_{10}$ probe as described previously (66).

RT-PCR and Northern blotting

Total RNA was isolated from various tissues using the TRIREAGENT (Molecular Research Center Inc) according to manufacturer's protocol. Routinely, 5μg of RNA was reverse transcribed with SuperScript II (Invitrogen) and one twentieth of the cDNA was used for PCR amplification using primers f2 and r1 as described above. Amplification of *GAPDH*

(nt 458-774, accession no. NM_008084) was used as internal controls. For Northern blotting, 20μg of total RNA from muscle was fractionated on a 1% agarose gel containing 3.7% formaldehyde, transferred to a nylon membrane (NEN), and hybridized with a ^{32}P- labeled *EGFP* probe for 16 hr in 6XSSC, 2XDenhard's, 0.1% SDS, and 100 μg/ml denatured salmon sperm DNA, at 60℃. The membrane was then washed and exposed to IMAGING PLATE and read by a Phospho Image Analyzer (Fujifilm).

Western Blot

Fresh tissues were homogenized in protein lysis buffer (15 mM Tris, 250 mM sucrose, 1 mM EDTA and 2 mM PMSF) and sonicated for 20 min. After centrifugation, about 100 μg of supernatants, as determined by the Bradford protein assay (Bio-Rad), were fractionated by 12.5% SDS-PAGE. The proteins were transferred to PolyScreen PVDF membranes (NEN), which were then blocked in 3% skim milk in Tris-buffered saline (TBS). The membranes were incubated in blocking solution with primary anti-EGFP (Living Colors A.v. peptide Antibody, 1: 500 dilution, BD Biosciences Clontech) or anti-tubulin b (1:1000 dilution, MD Bio) overnight, followed by washing in TBS and incubation with HRP-conjugated secondary antibodies (1:3000 dilution, Santa Cruz) in 3% skim milk/TBS for 1 hr. After a final wash, bound antibodies were visualized by SuperSignal West Pico chemiluminescent substrate kit (Pierce).

Histology and histochemistry

Muscle tissues of soleus and diaphragm were fixed in 4% paraformaldehyde/PBS overnight at 4℃, washed in PBS, dehydrated in increasing concentrations of ethanol and embedded in paraffin. Sections of 5 μm were cut and stained with hematoxylin and eosin. For histochemical reactions, fresh soleus muscle tissues were coated with OCT and frozen at -80 ℃. Cryostat sections of 8 μm were cut and immediately soaked in incubation medium. The succinate dehydrogenase, NADH-tetrazolium reductase and ATPase enzyme histochemistry were carried out as described by Cash and Blumbergs (67).

Phenotype analysis

A total of 60 mice at 2, 4 and 6 months of age were analyzed. Each 15 mice of heterozygous CAG_{200}, CAG_{23}, CAG_0 and nontransgenics were divided into three age-matched groups. For grip strength test, the mice were placed with their forelimbs on a narrow bar and the amount of time taken for the mouse to fall was assessed. For cage activity test, mice were placed on to a grid of squares (3x 3 cm) in a cage, and were allowed 5 min to settle before testing began. The mouse was observed for 3 min and the number of squares crossed by the

mouse per minute was recorded. Both tests were repeated three times with at least 10 min between each test.

Sperm counts, Rh123 staining, and electron microscopy

Male mice at 3 to 4 months of age were sacrificed and the cauda epididymides were excised and placed in 4ml of phosphate-buffered saline (PBS). Following incubation at 37℃ for 1 hour, the spermatozoa were observed and counted by a hemacytometer under a light microscope. Sperm from 6 males of each group was counted and analyzed, and the numbers were averaged. For Rh123 staining, the spermatozoa were incubated with Rh123 (5 g/ml in PBS) at 21℃ for 30 min, submitted to a centrifugation step (500 x *g* 5 min) and then incubated with fresh PBS at 21℃ for 45 min to eliminate nonspecific binding of the dye. After recentrifugation at 500 x *g* for 5 min, the spermatozoa were counterstained with PI (0.0ṣ g/ml in PBS) for 5 min and then subjected to flow cytometric analysis. For electron microscopy, the spermatozoa released from epididymides were washed with PBS for three times, pelleted by centrifugation at 1000 x *g* for 20 min, and fixed in 4% paraformaldehyde plus 2.5% glutaraldehyde in 0.1M cacodylate. Following post fix in 2% osmium tetroxide, the pellets were embedded in spurr. The thin sections were examined in a Joel 1200EX electron microscope after double staining with uranyl acetate and lead citrate.

Muscle action potential and contracture recordings

Conventional microelectrode recording techniques were used. The glass microelectrodes were filled with 3 M KCl and had resistance in the range 5-15 MΩ. The mouse diaphragm was placed in modified Krebs solution (130.6 mM NaCl, 4.8 mM KCl, 1.2 mM $MgSO_4$, 12.5 mM $NaHCO_3$, 2.5 mM $CaCl_2$, 10 mM glucose, pH 7.2–7.4) and gased with 95% + 5 % CO_2. Muscle fiber action potentials were elicited by intracellularly injected depolarizing current (12 nA, 100 ms), the pulses being delivered through the bridge mode of the Axoclamp-2B (Axon Instruments). The waveforms were recorded and analysis on a computer with P-Clamp 9.0 software (Axon Instruments). For contracture recording, the phrenic nerve-hemidiaphragm was isolated as previously described (68). The diaphragm was suspended in 10 ml Krebs solution at 37.0 ± 0.5 ℃ and constantly gassed with 95% O_2 + 5% CO_2. Subsequently the solution was substituted by either fresh Krebs solution or glucose-free Krebs solution (glucose substituted by equal moles of NaCl). The twitches of diaphragm were elicited by indirect stimulation of the phrenic nerve with supramaximal rectangular pulses of 0.05 ms duration at 5 Hz, and recorded with an isometric transducer (Grass FT.03) on a Gould Model TA240 polygraph. The muscle resting tension was adjusted to 2 g, and the preparation was then

allowed to balance for 15-20 min with 0.1 Hz before starting the 5 Hz stimulation experiment.

RESULTS

Generation of transgenic mice

To investigate the physiological effect of the untranslated CAG trinucleotide repeat expansion, we generated constructs containing the enhanced green fluorescent protein (*EGFP*) gene with different lengths of CAG repeats (no repeats, 23 and 200 CAG repeats, designated as CAG_0, CAG_{23}, and CAG_{200}, respectively) inserted in its 3'-untranslated region (Fig. 1A). The *EGFP* gene was placed downstream of the gamma-sarcoglycan (*gsg*) promoter, which has been shown to direct strong gene expression in the skeletal muscle during mouse development (41). Fragments of DNA containing these transgenes were microinjected into fertilized FVB/N mouse eggs. Transgenic mice were screened initially by PCR and the results were subsequently confirmed by Southern blot analyses. The presence of correct length of CAG repeat sequence in transgenic mice were also checked by PCR-based Southern blot, using $(CAG)_{10}$ oligomer as a probe (Fig. 1B). Three transgenic founder animals with CAG_0, seven with CAG_{23} and six with CAG_{200} transgenes were produced and were bred to establish independent lines.

Expression of the transgenes

To verify expression of the transgenes, RNA isolated from various tissues was subjected to RT-PCR analysis. The *EGFP* transcripts were present in heart, skeletal muscle, diaphragm, testis and ovary (Fig. 2A). This expression pattern was consistent with that described previously in a transgenic mouse model using the 1.5 kb -SG promoter (41). All transgenic lines generated exhibited the same patterns of tissue distribution. The sizes and the levels of EGFP transcripts were examined by Northern blotting using RNA isolated from the skeletal muscle (Fig. 2B). The length of the EGFP transcript from CAG_{200} lines was about 1.7 kb, which was 600 bp longer than that from CAG_0 lines as expected. There was no noticeable difference in the levels of these two transcripts (Fig. 2B). However, Western blotting revealed that EGFP protein levels were significantly reduced in tissue extracts of all CAG_{200} lines (Fig. 2C). This decreased protein expression was also reflected in the low fluorescence intensity of EGFP in CAG_{200} lines, when muscle sections were observed under a fluorescent microscope (Fig. 2D). Thus, expression of the *EGFP* transcript with long CAG repeats in the 3'UTR resulted in decreased protein expression.

Histological analysis of muscle

Skeletal muscle of the soleus taken from adult non-transgenic (NT), CAG_0 and CAG_{200}

mice was prepared for histological examination to evaluate the muscle morphology. There were no signs of fibrosis, inflammation or regeneration in muscle fibers from any of the mice observed. In the control mice, the shape and size of muscle fibers were homogeneous and most nuclei were peripherally situated (Fig. 3A, C, E, G). In CAG_{200} mice we observed muscle fibers of various size and shape, increased intermyofibril connective tissues, split fibers and centronucleated fibers (Fig. 3B, D, F, H). The nuclei appeared larger and the relative nucleus/fiber ratio was higher (Fig. 3I; the nucleus/fiber ratios for NT and CAG_0 mice in a cross section were 1.049±0.025 and 1.023±0.025, respectively, versus 1.413±0.044 and 1.424±0.058 for two CAG_{200} lines, $p<0.001$).

Further histochemical staining of the oxidative reaction by succinate dehydrogenase (SDH) and nicotinamide adenine dinucleotide reductase (NADHR) also revealed abnormal staining patterns in the muscle of CAG_{200} mice. By a modified SDH reaction using phenazine methylsulfate, fibers with heavy staining known as "ragged blue" fibers were observed in sections from CAG_{200} lines (Fig. 4D). The heavy staining reflected a marked excess of mitochondria and was not seen in the muscle of control or CAG_{23} mice (Fig. 4A-C). The fibrillar network of NADHR reactivity is shown in Fig. 4 E-J. The normal intermyofibrillar lattice pattern and subsarcolemmal pockets of NADHR reaction corresponded to the normal distribution of mitochondira (Fig. 4E, I). Both CAG_0 and CAG_{23} mice showed similar patterns to those of NT (Fig. 4F, G). In CAG_{200} mice, however, the network was irregular. Many fibers (both type I and type II) displayed an unusual "hollow" lattice pattern where NADHR activity was lacking in one or more areas within the fibers (Fig. 4H, J). These "moth-eaten" fibers were not artificial, because they were reproducible in different mice of three CAG_{200} lines. Finally, staining for ATPase activity revealed that type I fibers were predominant and grouped in the soleus muscle of CAG_{200} mice (Fig. 4L). In contrast, in the same area of muscle from NT, CAG_0 (data not shown) and CAG_{23} mice (Fig. 4K), the two fiber types appeared to be present in similar numbers and randomly distributed.

Phenotype analysis

Mice carrying the CAG_{200} construct did not show an overt phenotypic difference from the CAG_0 and NT mice. However, some of the CAG_{200} mice occasionally displayed bizarre postures and intermittent convulsions. When assayed by a locomotor activity test and a narrow bar hang assay (42), CAG_{200} mice showed reduced cage activity and muscle weakness (Fig. 5), as they crossed fewer squares per minute and became fatigue and fell in less time than control animals. The mean litter size in all CAG_{200} lines was close to control (7.6 pups per litter for 20 litters versus 8.4 pups per litter for 31 litters in CAG_0); however, breeding efficiency was greatly reduced. The time needed for a litter to be born from a heterozygous

CAG_{200}-NT cross ranged from 6 to 8 weeks, compared with 3 to 4 weeks from a normal or CAG_0-NT cross. In heterozygous CAG_{200} self cross, it took 9 to 12 weeks for a litter to be born if mating did take place. About 30% of such mating resulted in no offspring at all.

Sperm function and mitochondrial activity

To understand the cause of reduced fertility in CAG_{200} males, we first examined the testes and sperm. No significant differences were found in the sizes and histological features of testes and epididymis between the control and CAG_{200} males (data not shown). The numbers of sperm per epididymis were comparable in all four groups, whereas the motility of sperm of CAG_{200} males was greatly reduced to only 0.3% over a 1-hour time frame (Fig. 6A). Surprisingly, sperm motility of CAG_{23} males was also reduced to 6.5%, compared to 57-59% motility in CAG_0 and NT males. When examined by electron microscopy, some of the CAG_{200} sperm showed structural defects in axonemes, with loss of one outer doublet from the normal 9+2 microtubule arrangement (Fig. 6B, top panels). In addition, empty and abnormally shaped mitochondria along the midpiece were observed (Fig. 6B, lower panels). To evaluate mitochondrial function, spermatozoa were stained with rhodamine123 (Rh123) and propidium iodide (PI) followed by flow cytometric sorting. Rh123 specifically accumulates in the mitochondria and is recognized as an indicator of mitochondrial membrane potential (MMP), which has been shown to correlate positively with sperm motility (43, 44). As shown in Fig 6C, most of the sperm cells from control males had relatively high Rh123 fluorescence, whereas most sperm cells from CAG_{200} males were gated with low Rh123 fluorescence. Sperm cells from CAG_{23} males had intermediate Rh123 fluorescence, consistent with the sperm motility count. All the spermatozoa had low PI fluorescence indicating that they were viable. The result of this analysis confirmed that there was a defect in mitochondrial activity which led to low sperm motility and reduced fertility in CAG_{200} mice.

Muscle electrophysiology

Previously, repetitive electromyographic discharges were observed in a mouse model in which expression of an untranslated CUG expansion was directed to the skeletal muscle (20, 45). To determine if there were defects in muscle membrane conductance in CAG_{200} mice, sharp microelectrode recordings were performed on excised diaphragm muscle fibers. A single action potential was triggered in both NT and CAG_{200} fibers when 12 nA current was applied (Fig. 7A). There was no difference in the current threshold or latency, indicating that muscle membrane conductance was not impaired by CAG expansion. Because the patterns of mitochondrial enzyme activity were altered in muscle of CAG_{200} mice (Fig. 4), we then measured muscle contracture by nerve-evoked stimulation. As shown in Fig. 7B, contracture

of the isolated phrenic-nerve diaphragm was induced significantly by 5 Hz nerve-evoked stimulation for 20 min in CAG_{200} but not NT, CAG_0, or CAG_{23} mice. Table I summarizes the contracture force detected in different lines. In contrast, under metabolic stress condition (glucose-free Krebs solution), contracture was induced in all mice including controls by the same stimulation.

DISCUSSION

Our data demonstrated that a long tract of untranslated CAG trinucleotide repeat can exert a length-dependent pathological effect in transgenic mice. The mice expressing the *EGFP* transcript with 200 CAG repeats showed altered muscle histology and function, impaired sperm motility and defective mitochondrial activity. In contrast, transgenic mice expressing EGFP without CAG repeats or with a short tract of 23 repeats did not display abnormal phenotypes in most aspects. This mouse model shows that noncoding CAG repeat expansion, which is transcribed into RNA but not translated into protein, may have a pathogenic role.

The pathogenic mechanism of a non-coding triplet repeat is best studied in DM, in which CTG expansion primarily affects muscle function. To compare the effect of a non-coding CAG expansion, we used a *gsg* promoter to direct the transgene expression in skeletal muscle (41). In addition to muscle, we also observed *EGFP* RNA and protein expression in diaphragm, heart, testis and ovary (Fig 2A and data not shown). It has been reported that insertion of a CAG repeat sequence in the 3'-UTR of a reporter gene in yeast resulted in transcription of RNA which was longer then the expected size (46). The long RNA was proposed to result from transcription slippage. In our case, however, the size of the 200 CAG repeat-containing transcript was as expected (Fig 2B), arguing against the mechanism of transcription slippage playing a role in this transgenic system.

Interestingly, while the expression level of *EGFP* RNA was not affected by the length of the CAG repeat (Fig 2B), the expression level of EGFP protein was significantly reduced in all tissues expressing the long CAG-containing transcript, including testis and ovary (Fig 2C, D and data not shown). Thus, the data suggest that the long CAG repeat RNA may somehow interfere with protein translation or with nucleocytoplasmic transport of the transcripts. Similar observations have been made in other studies. In two cell culture models, CTG expansions in the 3'-UTR and CTG/CAG expansions in the 5'-UTR of reporter constructs were both shown to reduce reporter protein translation (34). Additionally, CGG repeat expansion in the premutation range was shown to impede FMR1 mRNA translation efficiency without reducing the mRNA levels (47, 48). More interestingly, animals carrying a human ataxin-1 transgene with 82 CAG repeats failed to produce detectable ataxin-1 protein, despite

the abundant transgene RNA levels (49). Since CUG, CAG and CGG repeats in RNA are all predicted to form stable hairpin structures (2, 50), it is possible that this hairpin structure may lead to a common mechanism of translation impediment. On the other hand, we have observed nuclear RNA foci formation in cultured cells from CAG_{200} neonatal mice (data not shown), raising another possibility that reduced protein production results from nuclear retention of long CAG-containing RNAs. In this respect, a recent study also showed that CAG repeats were able to form nuclear RNA foci (51).

Although there was very low mutant protein production, the transgenic mice expressing the long CAG repeat RNA were found to display several pathological features. In muscle, myopathic changes such as internalized nuclei, fiber splitting, increased connective tissues and fiber-type grouping were observed. Histochemical analysis revealed an increase in staining for succinate dehydrogenase. The dark-staining fibers are the result of large peripheral and intermyofibrillar aggregates of abnormal mitochondria. These so-called ragged blue fibers are the most typical morphologic change in mitochondrial myopathies. On the other hand, the moth-eaten fibers revealed by NADHR staining are somewhat reminiscent of central core disease or oligomitochondrial disease, which are both characterized by focal lack of mitochondria in myofibers (52). Besides muscle, these mice also showed reduced breeding efficiency. Since the testis histology and sperm counts were normal, the reduced breeding efficiency was attributable to poor sperm motility which in turn could result from defects in sperm mitochondria and axoneme structures (Fig. 6). Most CAG_{200} mice analyzed in this study were heterozygous for the transgene because breeding the transgene to homozygosity was difficult due to their reduced fertility. Moreover, the frequency of homozygous offspring from heterozygous parents in the CAG_{200} lines was less then 25%. This raises the possibility that homozygosity of the long CAG-containing transgene causes a deleterious effect during embryonic development. Taken together, the pathological features of the CAG_{200} mice are suggestive of mitochondrial dysfunction.

Recent evidence has shown that mitochondrial dysfunction may play a role in the pathogenesis of several neurodegenerative disorders (53). For example, the involvement of energetic defects and oxidative damage in muscle has been shown in patients with Huntington's disease. It is of particular interest because the defect is correlated with the expansion of CAG repeats (54). The fact that our CAG_{200} mice displayed lower grip strength and reduced locomotor activity also indicates that these mice may have muscle weakness and altered energy metabolism. Consistent with this notion, significant contracture was measured when the diaphragm muscle of CAG_{200} mice was induced in a condition that normally does not trigger contracture. Normal muscle fiber can sustain prolonged stimulation without eliciting contracture if the energy supply is not limited, i.e., in the presence of glucose. Thus,

the result suggests that CAG expansion may interfere with unknown factors associated with energy metabolism and mitochondrial function. We did not observe overt heart phenotype although the transgene was expressed in heart as well. The heteroplasmy of mitochondria and somatic heterogeneity of trinucleotide repeats may account for the complexity of phenotype expression.

In most human CAG repeat disorders, the pathological threshold for the length of translated CAG repeat is about 40. However, as few as 20 CAG repeats are responsible for SCA6, and as many as 55 repeats are required to confer the full symptoms of SCA3 or DRPLA (6, 7). In this study, the muscle phenotypes of mice that expressed a transgene with 23 CAG repeats were normal. Unexpectedly, however, these mice exhibited reduced fertility and decreased sperm motility. While this number of repeats is within the size range that rarely causes disease phenotype, it is possible that mouse sperm is more sensitive to the detrimental effects of CAG repeat expansions. Given that sperm motility depends heavily on energy supply from mitochondria, mutations causing a subtle change in mitochondrial activity could impede its motility dramatically.

Intergenerational instability of expanded trinucleotide repeats has been described in humans and a number of animal models (2). In this study, however, the *EGFP* transgene containing 200 CAG repeats did not show variation in repeat length upon parent to offspring transmission. The number of CAG repeats is stable up to the seventh (F7) generation (data not shown). In other animal models, transgenes with an expanded number of trinucleotide repeats that failed to manifest intergenerational instability have also been reported (49). It seems that instability is only apparent when the repeat sequence is introduced within its appropriate genomic context. Within the cDNA or a reporter construct the repeat sequence usually remains stable (54-57).

Although the widely accepted pathogenic mechanism of polyglutamine diseases involves mutant proteins, our data suggest that expanded CAG RNA may also play a pathogenic role. Concordantly, there are several lines of evidence suggesting an involvement of RNA-binding proteins in the pathogenesis of polyglutamine diseases. First, McLaughlin et. al. have reported the observation of a length-dependent association of cytoplasmic proteins with CAG repeat RNA (39). These proteins were present in the cortex and striatum of human brain, sites where neuronal degeneration occurs. Second, in a Drosophila model, screening for genes that modify the SCA1-induced neurodegenerative phenotype resulted in the identification of genes encoding RNA-binding proteins (40). These genes encode proteins that are both suppressors and enhancers, suggesting that alteration of RNA processing is relevant to SCA1 pathogenesis. These data together with the fact that CAG repeat RNA forms length-dependent stable hairpin structures both *in vitro* and *in vivo*, strongly suggest a role for CAG repeat RNA pathogenesis.

More recently, a double-stranded RNA binding protein, PKR, was shown to preferentially bind to a mutant Huntingtin transcript and to be activated in affected tissues of Huntington's disease (58). Since PKR also binds to expanded CUG repeat RNA and has been implicated in DM pathogenesis (24), it is possible that an RNA-mediated mechanism is involved in Huntington's disease.

On the other hand, studies showing untranslated CAG expansions which were non-toxic have also been documented. Expression of 79 repeats of CAG in the 3'-UTR of a truncated cDNA did not have any effect in COS cells, whereas expression of a cDNA containing an expanded polyglutamine tract resulted in significant cell death (59). Furthermore, transgenic mice expressing ataxin-1 with 82 CAG repeats and a mutated nuclear localization signal showed no altered phenotype (60). Since these mice express levels of RNA comparable to that found in SCA-1 transgenic mice (49), the data argue against a role for expanded CAG RNA in the pathogenesis of SCA1. However, it should be noted that these studies used expanded alleles with 79 and 82 CAG repeats. Although these sizes of CAG repeats in the coding region are clearly able to induce ataxia, they are still shorter than the 200 repeats used in this study. We previously used *C. elegans* to test the effect of untranslated CAG repeats and it appears that the pathogenic threshold of the repeat is longer than those frequently observed in polyglutamine diseases (unpublished result). Thus, the length of CAG repeats may be critical to its pathogenicity.

So far, the only identified human disorder caused by untranslated CAG repeat expansion is SCA12, in which expansion of the CAG repeat occurs in the 5'-UTR of the gene *PPP2R2B* (61). While preliminary data suggest that the repeat-containing region can function as a promoter and the mutation may affect gene expression, evidence showing obvious association of the gene function with the disease is still lacking. Thus, a role for RNA-mediated pathogenesis can not be excluded. Other CAG repeat loci have been found in the non-coding regions of a number of genes (ex STC, MAB21L1, MAP-2) (62-64). Whether expansion of these loci is related to human disorders is not yet clear. The pathologic findings in multiple transgenic lines expressing 200 CAG repeats in the untranslated region of a reporter gene clearly demonstrated the gain-of-function effect of the expanded CAG RNA. Our data strongly suggest that there may be unknown disorders that are linked to CAG expansion in RNA. Moreover, in addition to toxic protein gain-of-function, a role for an RNA-mediated mechanism may not be excluded in the phathogenesis of polyglutamine diseases.

TABLE

Table 1. Nerve-evoked muscle contracture force of isolated mouse phrenic nerve diaphragm

Line	Age (months)	Sex	Muscle contracture (g)
NT	6	F/M	0 (n=3)
CAG_0-10	6	F/M	0 (n=3)
CAG_{23}-16	2	F/M	0 (n=3)
CAG_{200}-62	5	M	0.62
	7	F	1.75
	7	M	1.25
	8	F	0.62
	9	F	0.38
	9	F	1.75
CAG_{200}-32	6	M	1.02
	10	F	0.38
	10	F	0.93

NT, non-transgenic; F, female; M, male; n, number of mice analyzed.

All transgenic mice used in this assay are heterozygous.

FIGURES

Fig.1

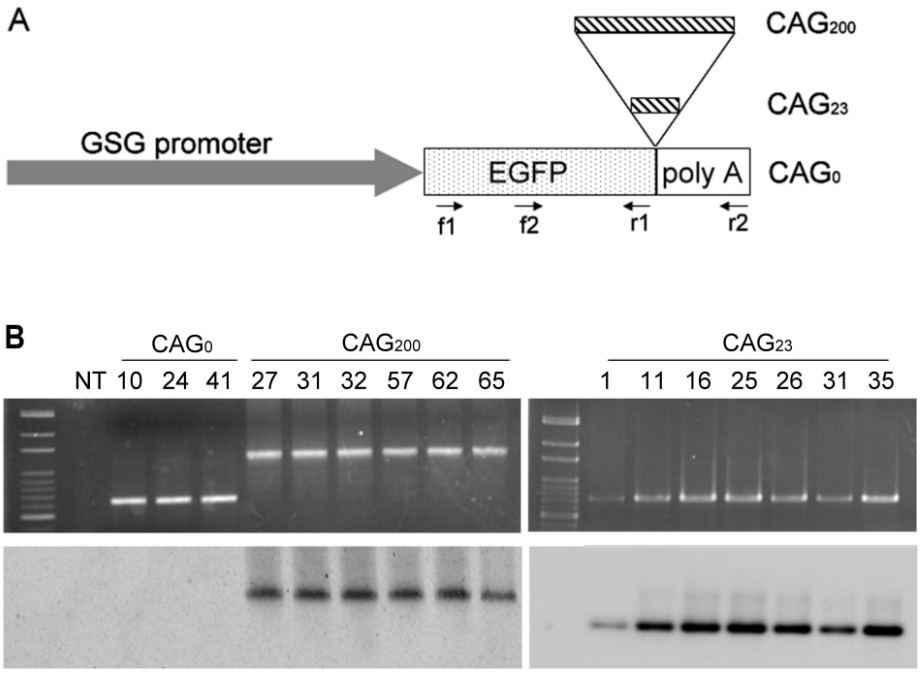

Fig.1. Generation of transgenic mice. (A) Diagram of the transgene constructs. The transgenes contain gamma-sarcoglycan (GSG) promoter fused with the EGFP coding sequence, followed by SV40 polyadenylation signals (poly A). Insertion of 23 or 200 CAG repeats was made downstream of the EGFP stop codon and before the poly A sequence. Locations of primers (f1、f2、r1、r2) used for PCR are marked. (B) PCR-based Southern blot analysis. Tail DNA from different founder animals (as indicated by numbers) of three constructs were PCR amplified by primers f1 and r2, which generate fragments of 632 bp, 701 bp and 1232 bp from CAG_0, CAG_{23} and CAG_{200} transgenes, respectively. Upper panels, ethidium bromide stained agarose gels; lower panels, blots of PCR products hybridized with a CAG_{10} probe.

Fig. 2

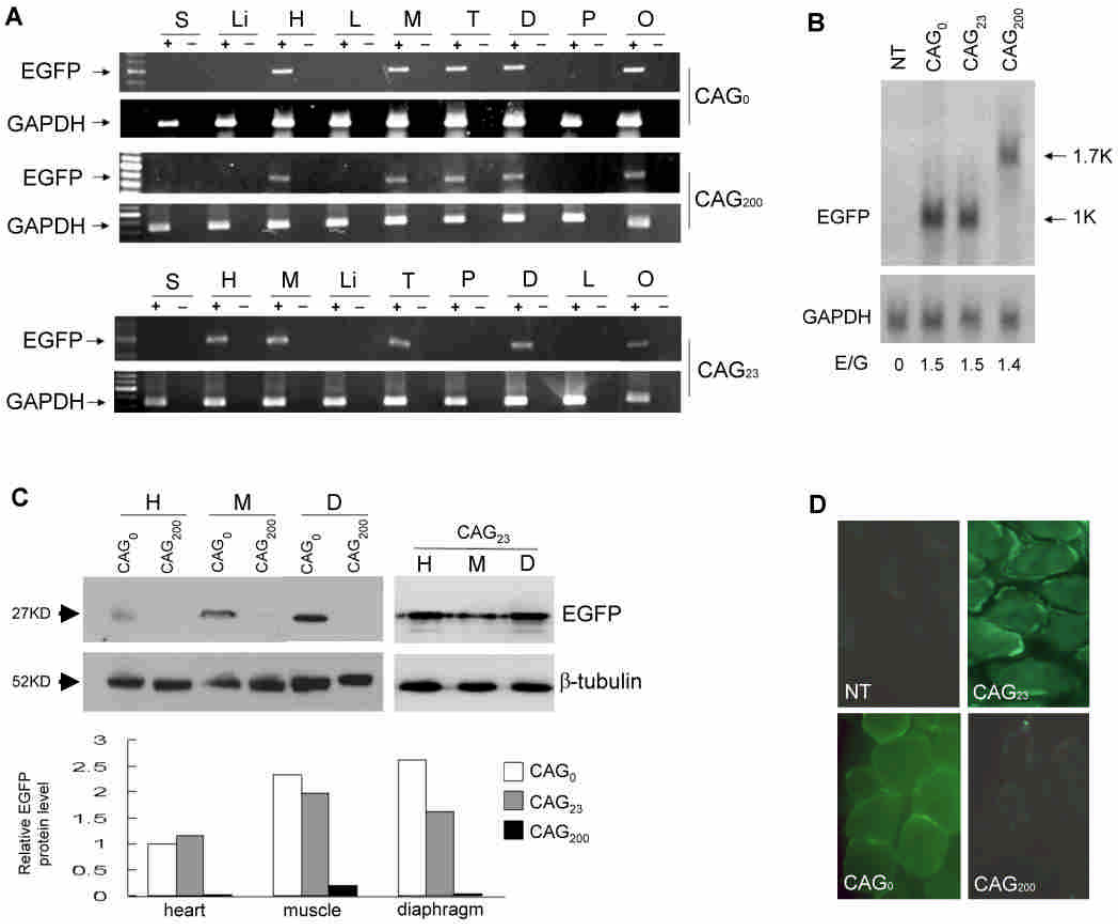

Fig. 2. Expression of the transgene. (A) RT-PCR on RNA isolated from different adult tissues. Representative gels showing expression of EGFP in CAG_0, CAG_{23} and CAG_{200} lines. Amplification of GAPDH was used as an internal control. +, with RT; -, without RT. S, skin; Li, liver; H, heart; L, lung; M, muscle (soleus); T, testis; D, diaphragm; P, pancreas; O, ovary.

(B) Northern blot analysis. Representative blot showing skeletal muscle RNA from non-transgenic (NT), CAG_0, CAG_{23} and CAG_{200} animals hybridized with an EGFP or a control GAPDH probe. Relative RNA expression levels (E/G, EGFP divided by GAPDH) are shown below. (C) EGFP protein expression in different transgenic lines by Western blotting using an anti-EGFP antibody. H, heart; M, muscle; D, diaphragm. By normalizing to the expression of -tubulin, the relative protein expression levels are calculated and shown by histogram below. (D) EGFP fluorescence expression was observed using cryostat sections of the soleus muscle under a fluorescence microscope. Reduced fluorescence in CAG_{200} muscle is consistent with reduced protein expression as shown in (C).

Fig.3.

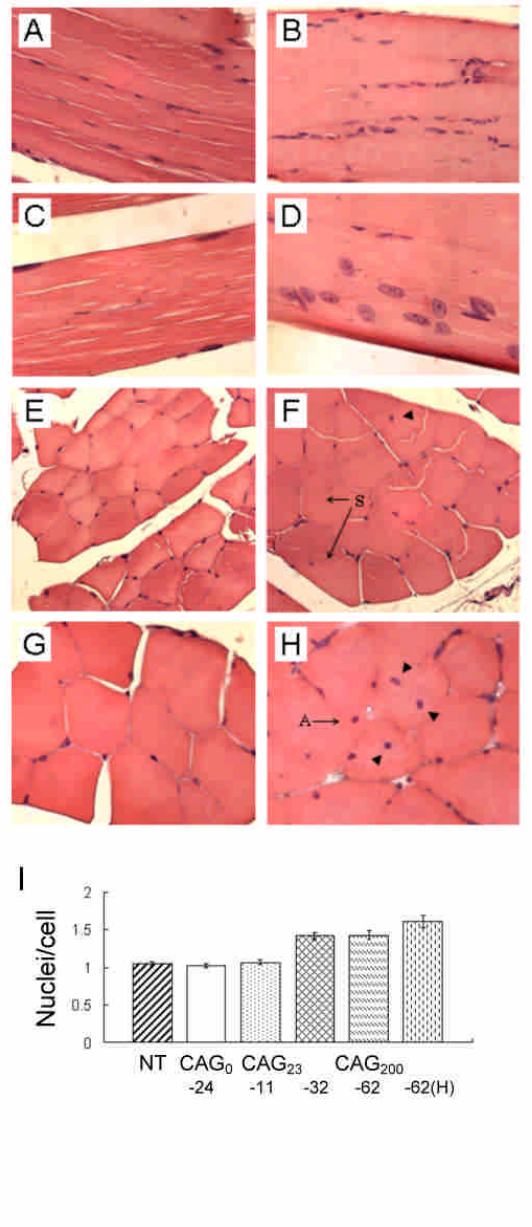

Fig3. Muscle morphology. Hematoxylin and eosin-stained paraffin sections of the soleus muscle from 2-month-old non-transgenic (A, E), CAG_0 (C), CAG_{23} (G), and CAG_{200} (B, D, F, H) animals in longitudinal (A-D) and transverse sections (E-H). Note that the fiber diameters in CAG_{200} animal (B) are not as uniform as in control (A) and that multiple rounded nuclei are observed (D). Some nuclei are located internally (arrowheads, F, H) instead of peripherally (C, E, G), and there are signs of split fibers (arrow S in F) and angular fibers (arrow A in H). (I) Quantification of nuclei (expressed by nuclei to cell ratio). Four sections from each lines (NT, CAG_0 line 24, CAG_{23} line 11, CAG_{200} line 32, 62 and one homogygous mouse) were counted in 10 fields under 400x magnification and numbers averaged. *P<0.001. (A-B), x200; (C-D), x400; (E-F), x250; (G-H), x400.

Fig. 4.

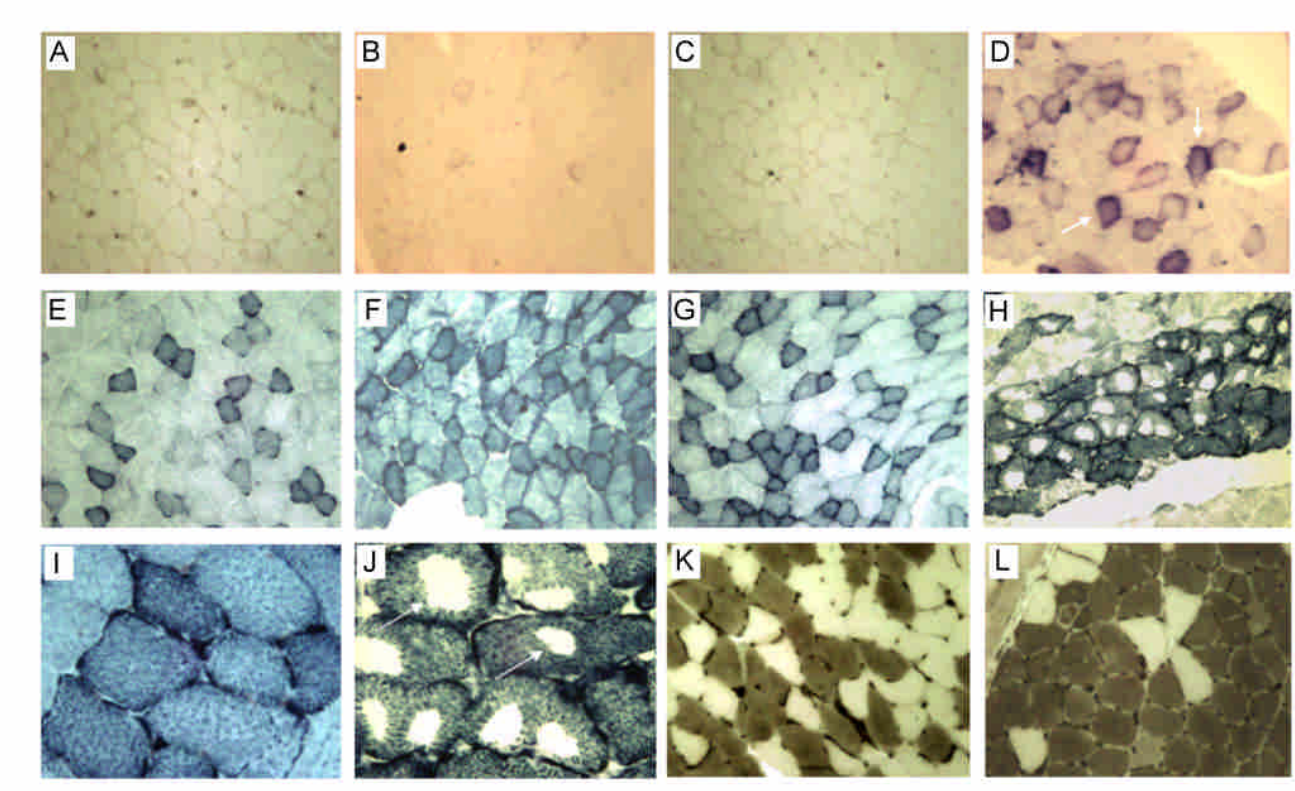

Fig. 4. Histochemical analysis of muscles. Representative results showing cryostat sections of the soleus muscles from 8-month-old nontransgenic (A, E, I), CAG_0-24 (B, F), CAG_{23}-11 (C, G, K) and CAG_{200}-32 (D, H, J, L) animals. (A-D) Staining for succinate dehydrogenase (SDH). Note that normal fibers show very little reaction by the modified SDH (with

phenazine methylsulfate) while “ragged blue” fibers (arrows) are detected in lines CAG_{200}-32 (D) and CAG_{200}-62(data not shown). (E-J) Staining for NADH-tetrazolium reductases. Note that fiber-type grouping and moth-eaten patterns or focal lack of the intermyofibrillar network of enzyme activity (H, J, arrows) in CAG_{200} mice in comparison to clear fiber-type distinctions (E-G) and uniform lattices (I) in control animals. (K,L) Staining for ATPase activity at pH 4.3. Type I fibres are black and type II fibers are cream. Type I predominance is seen in CAG_{200} mice (L). A-H, x160; I-J, x400; K-L, x250.

Fig. 5.

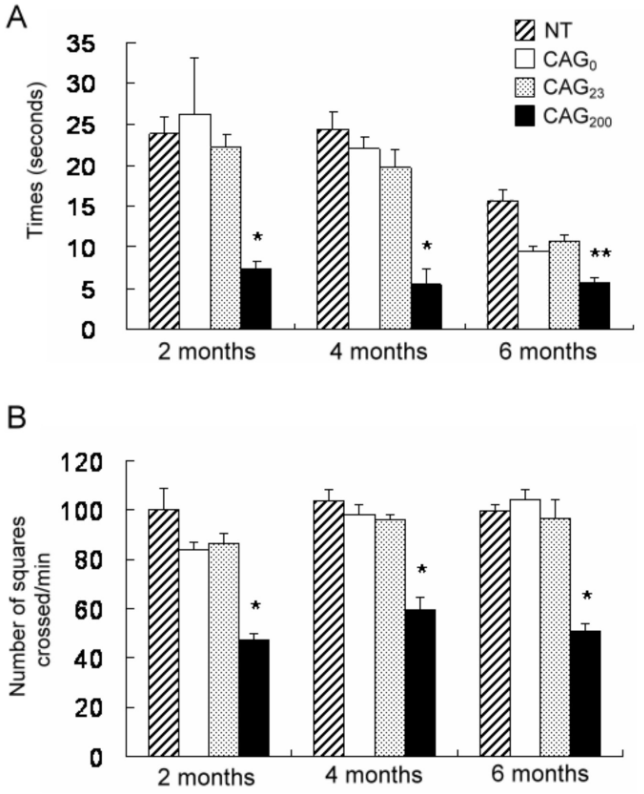

Fig. 5 Phenotypic analysis of the transgenic mice. (A) Grip strength test. Mice were suspended by the forelimbs on a narrow bar and the amount of time taken to fall was recorded. (B) Cage activity test. CAG_{200} lines displayed reduced locomotion activity as measured by a grid assay. All data represent averages from 5 mice, each measured for 3 times (A and B). *, P <0.001; **, $P < 0.01$ (compared to NT, CAG_0, or CAG_{23}; F test).

Fig. 6

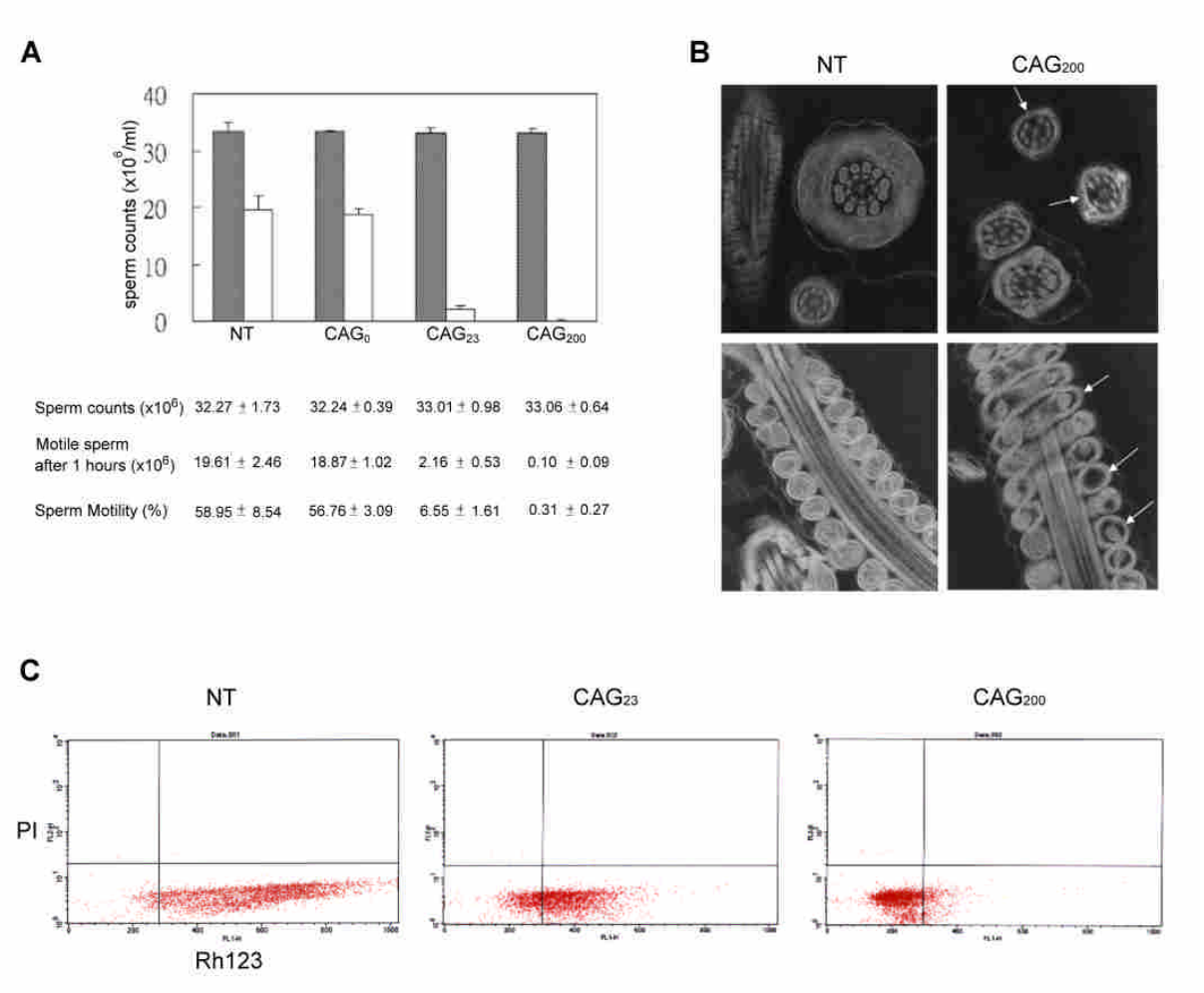

Fig. 6. Sperm motility and mitochondrial function. (A) Counts of total sperm (gray) and motile sperm (white) after 1 hr incubation at 37℃.(*n* = 6 per group). The average counts are shown below the histogram. Percent sperm motility was determined as counts of motile sperm divided by counts of total sperm. (B) Ultrastructures of the sperm tails. Electron microscopy revealed structural defects in the microtubule arrangement of axoneme (loss of one outer doublet; upper right panel, arrows) and in the mitochondria along the midpiece (lower right panel, arrows) in some sperm tails of CAG_{200} mice. (C) Flow cytometric sorting of Rh123- and PI-stained sperm cells. Horizontal scale, intensity of Rh 123; vertical scale, intensity of PI. Note that most sperm cells from CAG_{200} males are gated with low PI and low Rh123 fluorescence, indicating that they are viable yet with low mitochondrial activity.

Fig. 7.

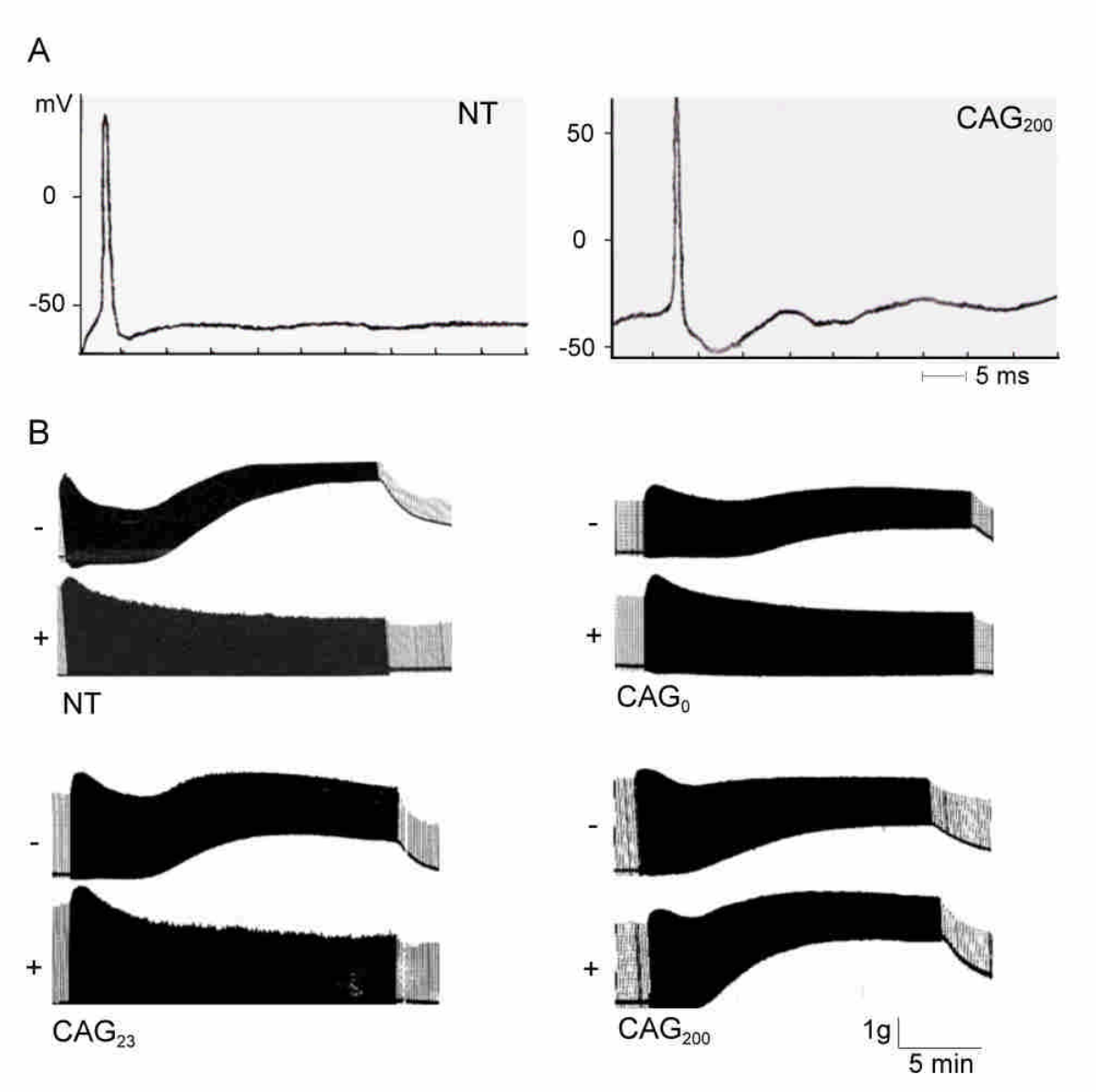

Fig. 7 Electrophysiology of the muscle. (A) Action potential. A single action potential was triggered by depolarizing current of 12 nA at 50 ms duration both in NT and CAG_{200} mice. (B) Phrenic nerve-evoked contracture of isolated diaphragm. In the presence of glucose, significant contracture was induced by 5 Hz stimulation for 20 min, in CAG_{200} but not NT, CAG_0, or CAG_{23} mice. As a control, all mice produced muscle contracture by the same stimulation with glucose-free Krebs solution. +, With glucose; -, without glucose.

成果自評

A manuscript of this study was submitted. This study also resulted in three master theses and five symposium abstracts. Two of these poster abstracts were awarded at the 11th and 12th Symposium on Recent Advances in Cellular and Molecular Biology. We have set up the screening of possible downstream molecules involved in the pathogenesis process, using microarray chips and protein 2D gel combined with proteomics We will keep progress in analyzing the possible candidate genes responsible for the phenotypes observed.