Results

Trinucleotide or pentanucleotide repeat distributions

Figure 1 shows the frequency distributions of (CAG/CTG/ATTCT) repeat lengths at the nine loci in neurodegenerative patients and controls. Expansions of SCA types 1, 3, and 6 were detected in clinically defined patients with dominant SCA. Excluding known SCAs, the frequency distributions of normal alleles in patient group were not significantly different from those in the control group at the nine loci studied ($P > 0.05$) (**Table 3**), and the CAG repeats at SCA types 1, 2, 3, 6, and DRPLA are generally polymorphic within the region smaller than 40 repeats in each group. In the case of *SCA8* locus, the frequency distribution subclasses into three groups. The small group closely distributed around an allele with 18 CTGs; the second class comprised CTG repeat sizes of 22 to 39 units, and the third one of extremely large normal alleles were found in one SCA3 patient (65 units) and four PD patients (75-92 units). At the *SCA10* locus, almost all alleles with the pentanucleotied ATTCT repeats range form 10 to 20 repeats and 14 repeat is the most frequent in each group. The frequencies of most *SCA12* alleles distribute smaller than 20 repeats, with one group distributing around 10 repeats and the other ranging from 13 to 18 repeats. At the *SCA17* locus, alleles vary from 30 to 43 in the control group and from 28 to 46 in the patient group, while alleles with 36 are predominant (>50%) in each group.

Frequency of SCAs and genetic and clinical features of patients

Analysis of loci involved in the SCAs showed that 26 unrelated patients (67%) had ataxia due to CAG expansion. SCA3 was the most common types of dominant SCA in Taiwan, accounting for 18 cases (46%), followed by SCA6 (7 cases, 18%) and SCA1 (1 case, 3%). The genes responsible for remaining 13 cases (33%) of dominant SCA remain to be determined. For SCA3, the mean age at onset was 36 years (SD, 12.74) and mean expanded allele sizes was 71 repeats (SD, 4.99) in 18 patients from 10 families. For SCA6, the mean age at onset and mean expanded allele sizes in 7 patients from 5 families were 48 years (SD, 10.02) and 23 repeats (SD, 0.79), respectively. For the SCA1 patient, the age at onset was 39 years and the expanded allele was 49 repeats.

Frequency of large normal alleles

When alleles corresponding to 5~10% of the upper tails (large normal alleles) at the nine loci were compared, no significant difference was found between the patient group and the control group. Nevertheless, we observed close associations between prevalences and frequencies of large normal alleles of SCA types 1, 2, 3, 6, and DRPLA in Taiwanese, Japanese and Caucasian families (**Table 4**).

SCA8 alleles with 75 to 92 repeats in PD patients

Although no expansion of SCA type 8 was detected in patients with dominant SCA, abnormal expansions were detected in four patients with PD (1.5%) (**Table 5**). The four patients met the criteria for PD, which included the presence of two of the four cardinal signs (resting tremor, cogwheel rigidity, bradykinesia or postural instability), improvement of symptoms with L-dopa therapy, and no evidence of secondary parkinsonism caused by

other neurologic disease or known drugs or toxins, or of atypical parkinsonism. DNA sequencing analysis revealed that these potentially pathogenic alleles had 75, 82, 88, and 92 combined repeats with a pure uninterrupted CTG repeat tract or interruptions (**Table 5**).

SCA17 allele with 46 repeats in PD patient

Abnormal CAG expansion in the SCA17 *TBP* gene was detected in one patient with PD (**Table 5**). The patient met the criteria for PD as described above. DNA sequencing analysis revealed that allele in the patient had CAG/CAA combined 46 repeats (**Table 5**).

Regulation of KLHL1 and CAG repeat-containing RNA expression by SCA8

The natural overlapping organization of the *SCA8* and *KLHL1* coding regions (Benzow and Koob 2002), and the similar distributions has been suggested that *SCA8* may function as a natural antisense regulator of *KLHL1* (Nemes et al. 2000). To test this hypothesis, we constructed the human *SCA8* cDNA with 0, 23, 88, or 157 combined repeats driven by the EF promoter (pEF-SCA8) as well as *KLHL1* cDNA tagged with EGFP (pEF-KLHL1-EGFP) (**Figure 3A**, **3B**). Since human embryonic kidney-derived HEK293 cells express many neuron-specific mRNAs (Shaw et al. 2002), and were frequently used to study other repeat expansion diseases (Handa et al. 2005), they were used in this study as a *in vitro* cell model. After co-transfecting of pEF-SCA8 and pEF-KLHL1 into HEK293 cells for two days, FACS was performed to evaluate the expression of the *KLHL1* fusion gene. The KLHL1-EGFP fusion protein production from cells co-transfected with *SCA8* carrying 0 ~ 157 combined repeats was significantly reduced (29%, 35%, 38% and 55% of the levels of KLHL1 fusion gene, $P < 0.05$) (**Figure**

5B). The difference between co-transfecting *SCA8* carrying 0 and 157 repeats constructs were significant (29% vs. 55%, P = 0.03). The results suggest that *SCA8* may function as a negative regulator of *KLHL1* in an inversely CUG repeats-dependent manner.

The expanded CUG repeats within *DMPK* transcripts were able to interact with CAG repeats located within the *TBP* or androgen receptor mRNA (Hamshere and Brook 1996; Sasagawa et al. 1999). To examine if *SCA8* RNA could pair with CAG repeats from the *TBP* gene transcript, we placed the 5' *TBP* $(CAG)_{36}$-containing cDNA fragment upstream to the IRES-mediated translation of EGFP gene [pCMV-$(CAG)_{36}$-IRES-EGFP] (**Figure 3C**). After co-transfecting equal amount of pCMV-$(CAG)_{36}$-IRES-EGFP and pEF-SCA8-0R, -23R, -88R, or 157R constructs into HEK293 cells for two days, EGFP protein production detected was 97%, 84%, 80% (P > 0.05), and 71% (P = 0.01) of the levels in cells transfecting pCMV-$(CAG)_{36}$-IRES-EGFP and pEF vector (**Figure 5C**). The difference between co-transfecting *SCA8* carrying 0 and 157 repeats constructs were also significant (97% vs. 71%, P = 0.04). These results suggested that *SCA8* RNA may down regulate the protein expression of CAG repeat-containing RNA gene, and the length of CUG repeats affects this down-regulation.

To examine the specificity of SCA8 *trans* RNA interference, pEGFP-N1 (**Figure 3D**) was used to co-transfect with pEF-SCA8 constructs. As shown in **Figure 5A**, 93%, 98%, 98% and 106% of the levels of the EGFP protein expression were detected as compared to that of co-transfecting pEGFP-N1 and pEF vector (P > 0.05). The difference between co-transfecting *SCA8*

carrying 0 and 157 repeats constructs were not significant (93% vs. 106%, *P* = 0.09). The results demonstrated the specific regulation of *SCA8* on *KLHL1*.

SCA8 encodes translable ORF1 and ORF3

Although reported non-coding (Koob et al. 1999), small ORFs in the *SCA8* transcripts were noted (**Figure 2B**). Among them, a 102 amino acids containing-ORF1 and a 41 amino acids plus a polyleucine tract containing-ORF3 may be translated if *SCA8* RNA possesses a cap independent IRES activity. To investigate if indeed the *SCA8* ORF1 and ORF3 are translated, *EGFP* gene was fused in-frame with the C terminal of the *SCA8* ORF1 (pCMV-ORF1-EGFP) as well as ORF3 carrying 23 combined repeats (pCMV-SCA8-23R-EGFP) (**Figure 4A**, **4B**). The predicted ORF1 and ORF3 contain 102 amino acids and 41 amino acids plus a polyleucine tract (23 leucines), respectively (**Figure 4D**). The constructs were expressed in HEK293 cells driven by the cytomegalovirus (CMV) promoter. After two days the levels of EGFP protein were evaluated by FACS analysis. As shown in **Figure 6A**, the 3%, 38% and 52% IRES-dependent EGFP production was seen in cells transfected with pCMV-SCA8-23R-EGFP, pCMV-ORF1-EGFP and pIRES2-EGFP constructs as compared to the control pEGFP-N1 construct (cap-dependent EGFP expression). Additionally, equal amounts of cell lysates were separated and immunoblotted with GFP antibody. As shown in **Figure 6B**, while no specific polypeptide was detected in mock-transfected HEK293 cells, a 50 kDa protein was detected in pCMV-ORF1-EGFP transfected cells, as compared to a 27 kDa protein in pEGFP-N1-transfected cells. Probably due to very low IRES-dependent expression of ORF3-23R-EGFP (**Figure 6A**), no ORF3-EGFP fusion protein was detected in Western blot.

ORF1 and expanded polyleucine-containing ORF3 proteins form aggregates

To further investigate the expression of *SCA8* ORF1 and ORF3 proteins, confocal microscopy examination of ORF-GFP fluorescence was carried out after transfection of pCMV-ORF1-EGFP and pCMV-SCA8-23R-EGFP constructs (**Figure 4A**, **4B**) into HEK293 cells. In addition, in order to investigate the effects of expanded CTG repeats on protein expression and distribution, ORF3 carrying 0, 88 and 157 combined repeats were prepared too. As shown in **Figure 7**, strong GFP fluorescence was distributed diffusely in cells expressing EGFP-N1. With *SCA8* ORF1 fused at the N terminus of GFP (ORF1-EGFP), small and dispersed aggregates appeared both in the nucleus and cytoplasm (accounting for 80 ± 8% of transfected cells), in addition to showing diffuse cytoplasm expression. Cells expressing ORF3-0R-EGFP gave much weaker but similar fluorescence pattern to that of GFP only. For cells expressing ORF3-23R-EGFP, weak but more or less unevenly distributed cytoplasmic fluorescence was observed. Moreover, cytoplasmic microaggregates, mostly perinuclear, were seen in cells transfected with ORF3-88R-EGFP (41 amino acids plus $Leu_8ProLeu_{79}$) and ORF3-157R-EGFP ($Leu_8ProLeu_{148}$). These results demonstrate that in addition to be translated, the expressed GFP-tagged ORF1 and polyleucine-expansion ORF3 proteins formed aggregates.

Aggregated ORF1 and ORF3 proteins are localized in both nuclei and cytoplasm

As IRES activity of *SCA8* is too low to detect ORF3-EGFP fusion protein in immunoblot, constructs were designed to force expression of ORF3-EGFP proteins by cloning the ORF3 immediately downstream of an eukaryotic

translation initiation Kozak consensus sequence (pCMV-K-ORF3-23R-EGFP and pCMV-K-ORF3-157R-EGFP) (**Figure 4C**). Fluorescence observation of K-ORF3 showed intensely expressing EGFP pattern, and most cells expressing either pCMV-K-ORF3-23R-EGFP or pCMV-K-ORF3-157R-EGFP formed aggregates, probably due to overexpression of ORF3-EGFP (data not shown). To examine the subcellular localization of *SCA8* ORF proteins, HEK293 cells were transfected with pCMV-ORF1-EGFP, pCMV-K-ORF3-23R-EGFP and pCMV-K-ORF3-157R-EGFP constructs and performed cell fractionation studies. Protein blot analysis (**Figure 8**) shows that the GFP-tagged ORF3-157R protein migrates as a smear at the top of the gel. The ORF3-157R-EGFP fusion protein is found mostly in the nuclei. Although ORF3-23R-EGFP protein also migrates as a larger protein product, a band of around 30 kDa could be detected. The expressed ORF3-23R-EGFP protein localized in both nuclear and cytoplasmic extracts. The expressed 50 kDa ORF1-EGFP protein localized mostly in the cytoplasm. In addition, confocal microscopy analysis of ORF1 and ORF3 distributions from continuous focal planes also demonstrated that ORF1 was abundantly cytoplasmic (**Figure 9**), while ORF3-157R aggregates were formed closely associated with the nucleus (**Figure 10**). Aggregates formation was also observed when neuroblastoma IMR-32 cells were transfected ORF1 and expanded *SCA8* (data not shown). And overexpression of ORF3-157R-EGFP fusion protein led to stalling in the stacking gel as well (**Figure 11**).

Generation and characterization of isogenic and inducible SCA8 cell lines

Transient transfection often results in overexpression of gene of interest since multiple copies are gained in transfected cells. Additionally, copy

numbers of transfected gene vary from cell to cell, and it could integrate at different loci on chromosomes in the presence of antibiotics selection. In order to establish isogenic and inducible cell line expressing *SCA8* gene carrying 0~157 CTA/CTG combined repeats, we adopted the Flp-In T-REx system (Invitrogen) to generate stable HEK293 cell lines. As shown in **Figure 12**, the FRT locus ensures a single defined site of chromosomal integration via homologous recombination mediated by the Flp recombinase, and in the T-REx form of these cells, expression from this locus is controlled in a Tet-on-inducible fashion by the addition of either tetracycline or the related antibiotic doxycycline. The integration of *SCA8* gene into Flp-In T-REx HEK293 cells was verified by PCR (**Figure 13A**), and the RNA expression level was determined by real-time PCR assays (**Figure 13B**). The addition of doxycycline (1 μg/ml, 2 days) resulted in 20-30 times induction of *SCA8* RNA, indicating that the expression of integrated SCA8 was tetracycline/doxycycline-dependent regulated. To find out whether *SCA8* isogenic cell lines with more CTA/CTG combined repeats are more susceptible to the stress, these cell lines were treated with various concentrations of staurosporine (apoptotic stimulus), MG-132 (proteasome inhibitor), or paraquat (source of free radicals) for one day and cell proliferation was assessed with WST-1 assay. As shown in **Figure 14**, the addition of low dosages of staurosporine (0~15 nM) or MG-132 (0~0.5 μM) did not cause significant differences of death among cell lines. However, when isogenic HEK293 cell lines with expanded *SCA8* combined repeat tracts were exposed to high dosages of staurosporine (30~50 nM) or MG-132 (0.75~1.0 μM), the viability of these cell lines decreased significantly. Unexpectedly, treatment of the high dosage of staurosporine or MG-132 also caused more death to the cell line with *SCA8* carrying 0 CTG repeats. In

addition, neither low nor high dosage had significant effects on all cell lines. Taken together, isogenic HEK293 cell lines with mutant *SCA8* were more sensitive to staurosporine and MG-132, but not paraquat.

Generation and characterization of SCA8 transgenic mice

To generate transgenic murine models expressing human *SCA8* within the cerebellum, we subcloned human *SCA8* cDNA with exons D, C2, C1, B, and A bearing either normal-range 23 or expanded 157 CTG/CTA combined repeats downstream to the cerebellar Purkinje-specific pcp2/L7 promoter (**Figure 15A**). The transgene genotyping was performed with PCR reaction specifically amplified the region from pcp2/L7 promoter to the exon D of human *SCA8* gene (**Figure 15B**). The transgene copy numbers of two SCA8-23R (17 and 24) and five SCA8-157R (7, 11, 17, 19 and 62) transgenic lines were estimated by comparison of transgene PCR products to the PCR products of copy number standards (**Figure 15C**). As summarized in the figure, the copy numbers are ranged from 1 to 30 among these transgenic lines.

Expression of SCA8 transcripts in SCA8 transgenic lines

From each group we chose 2 lines (23R-17, 23R-24, 157R-11, 157R-62) for further breeding and analyses. The human *SCA8* transgene can be successfully transmitted from generations 1 to 5 in these lines. Since *SCA8* gene is considered not having translatable peptide products, we assessed the expression of transgene in the level of transcripts by RT-PCR. As the result shown in **Figure 16**, all the transgenic lines we examined expressed the human *SCA8* in the cerebellum (lane C). In the brain region other to the cerebellum (lane B), however, transgene transcripts also could be detected,

implying that expression of human *SCA8* transgene might not be restricted to the cerebellum.

Behavior assessment of transgenic mice by performance on the accelerating rotarod

To evaluate the motor performance and the equilibrium ability, the trained mice were placed on the accelerating rotarod form 2 to 20 rpm over the first 5 min and maintaining at a maximum speed for another 5 min to score their latency on the rod. However, at the early life, neither SCA8-23R (**Figure 17A**) nor SCA-157R (**Figure 17B**) lines showed a different latency compared to their wild-type littermates. Although we also tested the older ones, the sample size was too small to tell the difference from sample variance. We also examined the body weights and the grabbing force of these transgenic mice, but no significant difference was detected (data not shown).

SCA8 transgene does not cause the morphological abnormalities of Purkinje cells

To examine if the *SCA8* transgene causes the morphological changes of the cerebellum, the histological analysis was performed with the calbindin D28k antibody to assess the numbers and the arrangement of the cerebellar Purkinje cells. However, at the age of 2-year-old, no markedly loss or shrinkage of Purkinje cells was observed in 23R-24 (**Figure 18A, 18C**) or 157R-62 (**Figure 18B, 18D**) lines. All these cells formed a continuous cellular sheet along the outer margin of granular layer of the cerebellar cortex, indicating that the ectopic expression of human *SCA8* bearing either

normal or expanded CTG repeats in the cerebellum of the mouse does not cause unusual orientation, or ectopic localization of Purkinje cells.

SCA8 transgenic mice exhibit the clasping phenotype

Although *SCA8* transgenic mice did not show abnormalities in both histolopathological and behavioral analyses, it is striking that our SCA8 transgenic mice exhibit a special paw clasping phenotype when suspended by the tail (**Figure 19A**), similar to that observed in HD transgenic murine model R6/2 (**Figure 19B**) (Mangiarini et al. 1995; Juvonen et al. 2000; Worth et al. 2000). The clasping phenotype of our *SCA8* transgenic mice progresses an alternating clasping and releasing of the feet, instead of maintaining holding all feet together. The founders of 157R-11 and -62 first exhibited this clasping phenotype at about 1 year of age, and some of their offspring showed the same phenotypes at younger age. However, less than 50% of the transgenic offspring displayed this phenotype, suggesting that a incomplete penetrance of SCA8 transgene might also exist in our transgenic model as in the human patients.

To investigate whether human *SCA8* gene with expanded CTG tract would affect global RNA expression, total RNAs from cerebella of SCA8-23 and 157R transgenic mice at the age of 20 month and age-matched wild-type littermate control were genome-wide analyzed using MOE430A chips from Affymetrix. After normalization, differentially expressed genes were selected using one-way ANOVA and fold-change analysis. All the probe sets that showed greater than 1.5- fold or smaller than 0.7-fold changes with were considered significantly differentially expressed (**Table 7**). The candidate genes found form the microarray assays were further validated by

real-time PCR. Among 11 candidate genes, the expression patterns of *Calml4*, *Igf2*, *Igfbp2*, *Ttr*, *Pthlh*, and *Rbp1* were in accordance with the microarray analyses, although the patterns of remaining 5 genes seemed no significant differences (**Figure 20**, **21**, and **22**).

Proteomic analysis of SCA8 transgenic mice

Although RNA plays an intermediate role between DNA and protein synthesis, differences of transcript level could not account for overall changes in different disease states or treatments. Since most cellular functions are exerted by the protein units, the protein patterns of SCA8 transgenic mice was performed to search deranged proteins. The resulting 2-D DIGE gel images were matched and analyzed, and expression levels of 13 protein spots were significantly changed (**Figure 23)**, and relative changes of protein abundance are summarized in **Table 9**. Using MALDI-TOF spectrometry and MASCOT search engine, 5 out 13 protein spots were identified as mitochondrial ATP synthase D chain (spot 1565), calbindin (spot 1525), calretinin (spot 1473), heterogeneous nuclear ribonucleoprotein K (spot 841), and serum albumin precursor (spot 761) (**Table 10**).