Expansion of Intronic GGCCTG Hexanucleotide Repeat in NOP56 Causes SCA36, a Type of Spinocerebellar Ataxia Accompanied by Motor Neuron Involvement.

AUTHOR(S):

Kobayashi, Hatasu; Abe, Koji; Matsuura, Tohru; Ikeda, Yoshio; Hitomi, Toshiaki; Akechi, Yuji; Habu, Toshiyuki; Liu, Wanyang; Okuda, Hiroko; Koizumi, Akio

CITATION:

Kobayashi, Hatasu ...[et al]. Expansion of Intronic GGCCTG Hexanucleotide Repeat in NOP56 Causes SCA36, a Type of Spinocerebellar Ataxia Accompanied by Motor Neuron Involvement.. American journal of human genetics 2011, 89(1): 121-130

ISSUE DATE:

2011-06-15

URL:

http://hdl.handle.net/2433/141931

Expansion of Intronic GGCCTG Hexanucleotide Repeat in *NOP56* Causes a Type of Spinocerebellar Ataxia (SCA36) Accompanied by Motor Neuron Involvement

Hatasu Kobayashi[1,4], Koji Abe[2,4], Tohru Matsuura[2, 4], Yoshio Ikeda[2], Toshiaki Hitomi[1], Yuji Akechi[2], Toshiyuki Habu[3], Liu Wanyang[1], Hiroko Okuda[1] and Akio Koizumi[1*]

[1]Department of Health and Environmental Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

[2]Department of Neurology, Graduate School of Medicine, Dentistry and Pharmaceutical Science, Okayama University, Okayama, Japan.

[3]Radiation Biology Center, Kyoto University, Kyoto, Japan.

[4]These authors contributed equally to this work

[*]Correspondence: Prof. Akio Koizumi M.D., Ph.D.

Department of Health and Environmental Sciences, Graduate School of Medicine, Kyoto University, Konoe-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan.

Tel: +81-75-753-4456; Fax: +81-75-753-4458

E-mail: koizumi.akio.5v@kyoto-u.ac.jp

Abstract

Autosomal dominant spinocerebellar ataxias (SCAs) are a heterogeneous group of neurodegenerative disorders. In this study, we performed genetic analysis of a unique form of SCA (SCA36) that is accompanied by motor neuron involvement. Genome-wide linkage analysis and subsequent fine mapping for three unrelated Japanese families in a cohort of SCA cases, in whom molecular diagnosis had never been done, mapped the disease locus to the region of a 1.8 Mb stretch (LOD score of 4.60) on 20p13 (D20S906– D20S193) harboring 37 genes with definitive open reading frames. We sequenced 33 of these and revealed a large expansion of an intronic GGCCTG hexanucleotide repeat in *NOP56* and an unregistered missense variant (Phe265Leu) in *C20orf194*, but no mutations in *PDYN* and *TGM6*. The expansion showed complete segregation with the SCA phenotype in family studies, whereas Phe265Leu in *C20orf194* did not. Screening the expansions in the SCA cohort cases revealed additional four occurrences, but none in the cohort of 27 Alzheimer's cases, 154 ALS cases, or 300 controls. Totally nine unrelated cases were found in 251 cohort SCA patients (3.6%). A founder haplotype was confirmed in these cases. RNA foci formation was detected in lymphoblastoid cells from affected subjects by fluorescence *in situ* hybridization. Double-staining and gel shift assay showed that (GGCCUG)n binds the RNA-binding protein SRSF2, but that $(CUG)_6$ did not. In addition, transcription of *MIR1292*, a neighboring microRNA, was significantly decreased in lymphoblastoid cells of SCA patients. Our finding suggests that SCA36 is caused by hexanucleotide repeat expansions through RNA gain-of-function.

Autosomal dominant spinocerebellar ataxias (SCAs) are a heterogeneous group of neurodegenerative disorders characterized by loss of balance, progressive gait, and limb ataxia[1-3]. We recently encountered two unrelated patients with intriguing clinical symptoms from a community in the Chugoku region in western mainland Japan[4]. These patients both showed complicated clinical features with ataxia as the first symptom, followed by characteristic late-onset involvement of the motor neuron system, with symptoms similar to amyotrophic lateral sclerosis (ALS[MIM 105400])[4]. Some SCAs (SCA1 [MIM 164400], SCA2 [MIM 183090], SCA3 [MIM 607047], and SCA6 [MIM 183086]) are known to slightly affect motor neurons; however, their involvement is minimal and the patients usually do not develop skeletal muscle and tongue atrophies[4]. Of particular interest is that RNA foci have been recently demonstrated in hereditary disorders caused by microsatellite repeat expansions/insertions in the non-coding region of their responsible gene [5-7]. The unique clinical features in these families have seldom been described in previous reports; therefore, we undertook a genetic analysis.

A similar form of SCA was observed in five Japanese cases from a cohort of 251 patients with SCA, in whom molecular diagnosis had not been performed, and who were followed by the Department of Neurology, Okayama University Hospital. These five cases originated from a city of 450,000 people in the Chugoku region. Thus, we suspected the presence of a founder mutation common to these five cases, prompting us to recruit these five families (Pedigrees 1–5) (**Figure 1, Table 1**). This study was approved by the Ethics Committee of Kyoto University and the Okayama University Institutional Review Board. Written informed consent was obtained from all subjects. An index of cases per family was investigated in some depth: IV-4 in Pedigree 1, II-1 in Pedigree 2, III-1 in Pedigree 3, II-1 in Pedigree 4, and II-1 in Pedigree 5. Mean age at

onset of cerebellar ataxia was 52.8 ± 4.3 years, and the disease was transmitted by an autosomal dominant mode of inheritance. All affected individuals started their ataxic symptoms, such as gait and truncal instability, ataxic dysarthria, and uncoordinated limbs, in their late forties to fifties. Magnetic resonance imaging revealed relatively confined and mild cerebellar atrophy (**Figure 2A**). Unlike previously known SCAs, all affected individuals with longer disease duration showed obvious signs of motor neuron involvement (**Table 1**). Characteristically, all affected individuals exhibited tongue atrophy with fasciculation to a greater or lesser extent (**Figure 2B**). Despite severe tongue atrophy in some cases, their swallowing function was relatively preserved, and they were allowed oral intake even at a later point after onset. In addition to tongue atrophy, skeletal muscle atrophy and fasciculation in the limbs and trunk appeared in advanced cases[4]. Tendon reflexes were generally mild-to-severely hyperreactive in most affected individuals, without severe lower limb spasticity and extensor plantar response. Electrophysiological studies were performed in an affected individual. Nerve conduction studies revealed normal findings in all the cases examined; however, an electromyogram showed neurogenic changes only in cases with skeletal muscle atrophy, indicating that lower motor neuropathy existed in this particular disease. Progression of motor neuron involvement in this SCA was typically and limited to the tongue and main proximal skeletal muscles in both upper and lower extremities, which is clearly different from typical ALS, which usually involves most skeletal muscles in a few years, leading to fatal results within several years.

We conducted genome-wide linkage analysis for nine affected subjects and eight unaffected subjects in three informative families (Pedigrees 1–3; **Figure 1**). For genotyping, we used an ABI Prism Linkage Mapping Set (Version 2; Applied

Biosystems, Foster City, CA, USA) with 382 markers, 10 cM apart, for 22 autosomes. Fine-mapping markers (approximately 1 cM apart) were designed according to information from the uniSTS reference physical map in the NCBI database. A parametric linkage analysis was carried out using GENEHUNTER[8], assuming an autosomal dominant model. The disease allele frequency was set at 0.000001 and a phenocopy frequency of 0.000001 was assumed. Population allele frequencies were assigned equal portions for individual alleles. We performed multipoint analyses for autosomes and obtained logarithm of the odds (LOD) scores. We considered LOD scores above 3.0 to be significant[8]. Genome-wide linkage analysis revealed a single locus on chromosome 20p13 with a LOD score of 3.20. Fine mapping increased the LOD score to 4.60 (**Figure 3**). Haplotype analysis revealed two recombination events in pedigree 3, delimiting a1.8-Mb region (D20S906– D20S193) (**Figure 1**). We further tested whether the five cases shared the haplotype. As shown in **Figure 1**, pedigrees 4 and 5 were confirmed to have the same haplotype as pedigrees 1, 2, and 3, indicating that the 1.8 Mb region is very likely to be derived from a common ancestor.

The 1.8-Mb region harbors 44 genes (NCBI, Build 37.1). We eliminated two pseudogenes and five genes (LOC441938, LOC100289473, LOC100288797, LOC100289507 and LOC100289538) from the candidates. Evidence view showed that the first, fourth, and fifth genes were not found in the contig in this region, while the second and third of these genes had mismatches over the mouse genes. Sequence similarities among paralogue genes defied direct sequencing of four genes: *SIRPD* [NM 178460.2], *SIRPB1* [NM 603889], *SIRPG* [NM 605466], and *SIRPA* [NM 602461]. Thus, we sequenced 33 of 37 genes (*PDYN* [MIM 131340], *STK35* [MIM 609370], *TGM3* [MIM 600238], *TGM6* [NM_198994.2], *SNRPB* [MIM 182282],

SNORD119 [NR_003684.1], *ZNF343* [NM_024325.4], *TMC2* [MIM 606707], *NOP56* [NM_006392.2], *MIR1292* [NR_031699.1], *SNORD110* [NR_003078.1], *SNORA51* [NR_002981.1], *SNORD86* [NR_004399.1], *SNORD56* [NR_002739.1], *SNORD57* [NR_002738.1], *IDH3B* [MIM 604526], *EBF4* [MIM 609935], *CPXM1* [NM_019609.4], *C20orf141* [NM_080739.2], *FAM113A* [NM_022760.3], *VPS16* [MIM 608550], *PTPRA* [MIM 176884], *GNRH2* [MIM 602352], *MRPS26* [MIM 611988], *OXT* [MIM 167050], *AVP* [MIM 192340], *UBOX5* [NM_014948.2], *FASTKD5* [NM_021826.4], *ProSAPiP1* [MIM 610484], *DDRGK1* [NM_023935.1], *ITPA* [MIM 147520], *SLC4A11* [MIM 610206], and *C20orf194* [NM_001009984.1]) (**Figure 2C**). All noncoding and coding exons, and the 100 bp up- and down-stream of the splice junctions of these genes were sequenced in two index cases (IV-4 in pedigree1 and III-1 in pedigree 3) and in three additional cases (II-1 in pedigree 2, II-1 in pedigree 4 and II-1 in pedigree 5) using specific primers (**Supplemental Table 1**). Eight unregistered variants were found among the two index cases. Among these, there was a coding variant (Phe265Leu), g.3324373 C > G of C20orf194, while the other seven included one synonymous variant (Leu565Leu in ZNF343; g.2463912 T>A) and six non-splice-site intronic variants (**supplemental Table 2**). We tested segregation by sequencing exon 11 of *C20orf194* in IV-2 and III-5 in the pedigree 1. Neither IV-2 nor III-5 had this variant. We thus eliminated *C20orf194* as a candidate. Missense mutations in *PDYN* and *TGM6*, which have been recently reported as causes of SCA, mapped to 20p12.3-p13[9; 10], but none were detected in the five index cases studied here (**Supplemental Table 2**).

Possible expansions of repetitive sequences in these 33 genes were investigated when intragenic repeats were indicated in the database (UCSC Genome

Bioinformatics). Expansions of the hexanucleotide repeat GGCCTG (rs68063608) were found in intron 1 of *NOP56* (**Figure 2D**) in all five index cases, using a repeat-primed PCR method[11-13]. An outline of the repeat-primed PCR experiment is described in **Figure 2D.** Briefly, the fluorescent dye-conjugated forward primer corresponded to the region upstream of the repeat of interest. The first reverse primer consisted of four units of the repeat (GGCCTG) and a 5′-tail used as an anchor. The second reverse primer was an "anchor" primer. These primers are described in **Supplemental Table 3.** Complete segregation of the expanded hexanucleotide was confirmed in all pedigrees, and the maximum repeat size in nine unaffected members was eight (data not shown).

In addition to the SCA cases in five pedigrees, four unrelated cases (SCA#1–SCA#4) were found to have a (GGCCTG)n allele by screening in the cohort SCA patients (**Table 1**). Neurological examination was reevaluated in these four cases, revealing both ataxia and motor neuron dysfunction with tongue atrophy and fasciculation (**Table 1**). Totally nine unrelated cases were found in the 251 cohort patients with SCA (3.6%). To confirm the repeat expansions, Southern blot analysis was conducted in six affected subjects (Ped2_II-1, Ped3_III-1, Ped3_III-2, Ped5_I-1, Ped5_II-1 and SCA#1). The data showed >10 kb of repeat expansions in the lymphoblastoid cell lines (LCLs) obtained from the SCA patients (**Figure 2E**). Furthermore, the numbers of GGCCTG repeat expansion were estimated by Southern blotting in other 11 cases. The expansion analysis revealed approximately 1500 to 2500 in 17 cases (**Table 1**). There was no negative association between age of onset and the number of GGCCTG repeats (n=17, r=0.42, p=0.09; **Supplemental Figure 1**), and no obvious anticipation in the current pedigrees.

To investigate the disease specificity and disease spectrum of the hexanucleotide repeat expansions, we tested the repeat expansions in an Alzheimer's disease [MIM 104300] cohort and an ALS cohort followed up by the Department of Neurology, Okayama University Hospital. We also recruited Japanese controls, who were confirmed to be free from brain lesions by magnetic resonance imaging and magnetic resonance angiography, as described previously[14]. Screening of the 27 Alzheimer's disease cases and 154 ALS cases failed to detect further cases with repeat expansions. The GGCCTG repeat sizes ranged from three to eight in 300 Japanese controls (5.9 ± 0.8 repeats), suggesting that the >10-kb repeat expansions were mutations.

Expression of *Nop56*, an essential component of the splicing machinery[15], was examined by RT-PCR using primers for wild-type mouse *Nop56* cDNA (**Supplemental Table 3**). Expression of *Nop56* mRNA was detected in various tissues including central nervous system, while a very weak signal was detected in spinal cord (**Figure 4A**). Immunohistochemistry using an anti-mouse Nop56 antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) detected the Nop56 protein in Purkinje cells of the cerebellum as well as motor neurons of the hypoglossal nucleus and the spinal cord anterior horn (**Figure 4B**), suggesting that these cells may be responsible for tongue and muscle atrophy in the trunk and limbs, respectively. Western blotting also confirmed the presence of the Nop56 protein in neural tissues (**Figure 4C**), where Nop56 is localized in both the nucleus and cytoplasm.

Alterations of *NOP56* RNA expression and protein levels in LCLs from patients were examined by real-time RT-PCR and western blotting. The primers for quantitative PCR of human *NOP56* cDNA are described in **Supplemental Table 3**. Immunoblotting was performed using an anti-human NOP56 antibody (Santa Cruz Biotechnology, Santa

Cruz, CA, USA). We found no decrease in *NOP56* RNA expression or protein levels in LCLs from these patients (**Figure 5A**). To investigate abnormal splicing variants of *NOP56*, we performed RT-PCR using the primers covering the region from the 5' UTR to exon 4 around the repeat expansion (**Supplemental Table 3**); however, no splicing variant was observed in LCLs from the cases (**Figure 5B**). Furthermore, immunocytochemistry for NOP56 and coilin, a marker of the Cajal body, where NOP56 functions[16], was carried out. NOP56 and coilin distributions were not altered in LCLs of the SCA patients (**Figure 5C**), suggesting that qualitative or quantitative changes in the Cajal body did not occur. These results indicated that haploinsufficiency could not explain the observed phenotype.

We performed fluorescent *in situ* hybridization to detect RNA foci containing the repeat transcripts in LCLs from patients, as previously described[17; 18]. Lymphoblastoid cells from two SCA patients (Ped2_II-2 and Ped5_I-1) and two control subjects were analyzed. An average of 2.1 ± 0.5 RNA foci/cell were detected in 57.0% of LCLs (n = 100) from the SCA subjects using a nuclear probe targeting the GGCCUG repeat, whereas no RNA foci were observed in control LCLs (n = 100) (**Figure 6A**). In contrast, a probe for the CGCCUG repeat, another repeat sequence in intron 1 of *NOP56*, detected no RNA foci in either SCA or control LCLs (n = 100 each) (**Figure 6A**) indicating that the GGCCUG repeat was specifically expanded in the SCA subjects. The specificity of the RNA foci was confirmed by sensitivity to RNase A treatment and resistance to DNase treatment (**Figure 6A**).

Several reports have suggested that RNA foci play a role in the etiology of SCA through sequestration of specific RNA-binding proteins[5-7]. *In silico* searches (ESEfinder 3.0) predicted an RNA-binding protein, SRSF2 [MIM 600813], as a strong candidate

for binding the GGCCUG repeat. Double-staining with the probe for the GGCCUG repeat and an anti-SRSF2 antibody (Sigma-Aldrich Inc., Tokyo, Japan) was performed. The results showed co-localization of RNA foci with SRSF2, while NOP56 and coilin were not co-localized with the RNA foci (**Figure 6B**), suggesting a specific interaction of endogenous SRSF2 with the RNA foci *in vivo*.

To further confirm the interaction, gel-shift assays were carried out to investigate the binding activity of SRSF2 with $(GGCCUG)_n$. Synthetic RNA oligonucleotides (200 pmol), $(GGCCUG)_4$ or $(CUG)_6$, which is the latter part of the hexanucleotide, as well as the repeat RNA involved in myotonic dystrophy type 1 (DM1) [MIM 160900][18] and SCA8 [MIM 608768][5],were denatured and immediately mixed with different amounts (0, 0.2, or 0.6 μg) of recombinant full-length human SRSF2 protein (Abcam, Cambridge, UK). The mixtures were incubated and the protein-bound probes were separated from the free forms by electrophoresis on 5–20% native polyacrylamide gels. The separated RNA probes were detected with SYBR Gold staining (Invitrogen, Carlsbad, CA, USA). We found a strong association of $(GGCCUG)_4$ with SRSF2 *in vitro* in comparison to $(CUG)_6$ (**Figure 6C**). Collectively, we concluded that (GGCCUG)n interacts with SRSF2.

It is notable that *MIR1292* is located just 19 bp 3' of the GGCCTG repeat (**Figure 2D**). MicroRNAs such as *MIR1292* are small non-coding RNAs that regulate gene expression by inhibiting translation of specific target mRNAs[19; 20]. MicroRNAs are believed to play important roles in key molecular pathways by fine-tuning gene expression[19; 20]. Recent studies have revealed that microRNAs influence neuronal survival and are also associated with neurodegenerative diseases[21; 22]. *In silico* searches (Target Scan Human 5.1) predicted glutamate receptors (*GRIN2B* [MIM 138252] and

GRIK3 [MIM 138243]) as potential target genes. Real time RT-PCR using TaqMan probes for miRNA (Invitrogen, Carlsbad, CA, USA) revealed that the levels of both mature and precursor *MIR1292* were significantly decreased in SCA LCLs (**Figure 6D**), indicating that the GGCCTG repeat expansion decreased the transcription of *MIR1292*. A decrease in *MIR1292* expression may upregulate glutamate receptors in particular cell types, *e.g. GRIK3* in stellate cells in the cerebellum[23], leading to ataxia because of perturbation of signal transduction to the Purkinje cells. In addition, it has been suggested, based on ALS mouse models[24; 25], that excitotoxicity mediated by a type of glutamate receptor, the NMDA receptor including *GRIN2B*, is involved in loss of spinal neurons. A very slowly progressing and mild form of the motor neuron disease, i.e., mostly limited to fasciculation of tongue, limbs and trunk, may also be compatible with such a functional dysregulation rather than degeneration

In the present study, we have conducted genetic analysis to find a genetic cause for the unique SCA with motor neuron disease. With extensive sequencing the 1.8 MB linked region, we found a large hexanucleotide repeat expansions in *NOP56,* which were completely segregated with SCA in five pedigrees and was found in four unrelated cases with the similar phenotype. The expansion was neither found in 300 controls or other neurodegenerative diseases. We further proved that repeat expansions of *NOP56* induce RNA foci and sequester SRSF2. Taken together, we thus concluded that hexanucleotide repeat expansions are considered to cause SCA by a toxic RNA gain-of-function mechanism and name this unique SCA as SCA36. Haplotype analysis indicates that hexanucleotide expansions are derived from a common ancestor. The prevalence of the SCA36 was estimate 3.6% in the SCA cohort in Chugoku district, suggesting that prevalence of SCA36 may be geographically limited to the western part

of Japan and is rare even in Japanese SCAs.

Expansion of tandem nucleotide repeats in different regions of respective genes (most often the triplets CAG and CTG) has been shown to cause a number of inherited diseases over the past decades. An expansion in the coding region of a gene causes a gain of toxic function and/or reduces the normal function of the corresponding protein at the protein level. RNA-mediated noncoding repeat expansions have been also been identified to cause eight other neuromuscular disorders, namely DM1, DM2 [MIM 602668], fragile X tremor/ataxia syndrome (FXTAS) [MIM 300623], Huntington's disease-like 2 (HDL2) [MIM 606438], SCA8, SCA10 [MIM 603516], SCA12 [MIM 604326], and SCA31 [MIM 117210] [26]. The repeat numbers in affected alleles of SCA36 are among the largest seen in this group of diseases (i.e. thousands of repeats). Moreover, SCA36 is not merely a non-triplet repeat expansion disorder after SCA10, DM2, and SCA31, but is now proven to be a human disease caused by a large hexanucleotide repeat expansion. In addition, no or only weak anticipation has been reported for non-coding repeat expansion in SCA, while clear anticipation has been reported for most polyglutamine expansions in SCA[2]. As such, absence of anticipation in SCA36 is in accord with previous studies on SCAs with noncoding repeat expansions. The common hallmark in these noncoding repeat expansion disorders is transcribed repeat nuclear accumulations with respective repeat RNA-binding proteins, which are considered to primarily trigger and develop the disease at the RNA level. However, multiple different mechanisms are likely to be involved in each disorder. There are at least two possibilities to explain motor neuron involvement of SCA 36: gene and tissue specific splicing specificity of *SRSF2* and involvement of microRNA. In SCA36, there is the possibility that the adverse effect of the expansion mutation is mediated by

downregulation of microRNA expression. The biochemical implication of microRNA involvement cannot be evaluated in this study, because availability of tissue samples from affected cases was limited to LCLs. Given definitive downregulation of microRNA 1291 in LCLs, we should await further study to substantiate its involvement in affected tissues. Elucidating which mechanism(s) play a critical role in the pathogenesis will be required to determine whether cerebellar degeneration and motor neuron disease occur with a similar scenario.

In conclusion, expansion of the intronic GGCCTG hexanucleotide repeat in *NOP56* causes a unique form SCA (SCA36), which shows not only ataxia, but also motor neuron dysfunction. This characteristic disease phenotype can be explained by the combination of RNA gain-of-function and *MIR1292* suppression. Further studies are required to investigate the roles of each mechanistic component in the pathogenesis of SCA36.

Acknowledgments

This work was supported mainly by grants to AK and partially by grants to MT, IY, HK and KA. We thank Mr. Norio Matsuura, Dr. Kokoro Iwasawa, and Dr. Kouji H. Harada (Kyoto University Graduate School of Medicine).

Web Resources

NCBI, http://www.ncbi.nlm.nih.gov/

UCSC Genome Bioinformatics, http://genome.ucsc.edu

ESEfinder 3.0, http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home

Target Scan Human 5.1, http://www.targetscan.org/

References

1. Harding, A.E. (1982). The clinical features and classification of the late onset autosomal dominant cerebellar ataxias. A study of 11 families, including descendants of the 'the Drew family of Walworth'. Brain 105, 1–28.

2. Matilla-Duenas, A., Sanchez, I., Corral-Juan, M., Davalos, A., Alvarez, R., and Latorre, P. (2010). Cellular and molecular pathways triggering neurodegeneration in the spinocerebellar ataxias. Cerebellum 9, 148–166.

3. Schols, L., Bauer, P., Schmidt, T., Schulte, T., and Riess, O. (2004). Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. Lancet Neurol 3, 291–304.

4. Ohta, Y., Hayashi, T., Nagai, M., Okamoto, M., Nagotani, S., Nagano, I., Ohmori, N., Takehisa, Y., Murakami, T., Shoji, M., et al. (2007). Two cases of spinocerebellar ataxia accompanied by involvement of the skeletal motor neuron system and bulbar palsy. Intern Med 46, 751–755.

5. Daughters, R.S., Tuttle, D.L., Gao, W., Ikeda, Y., Moseley, M.L., Ebner, T.J., Swanson, M.S., and Ranum, L.P. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. PLoS Genet 5, e1000600.

6. Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., Takahashi, M., Matsuura, T., Flanigan, K.M., Iwasaki, S., et al. (2009). Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. Am J Hum Genet 85, 544–557.

7. White, M.C., Gao, R., Xu, W., Mandal, S.M., Lim, J.G., Hazra, T.K., Wakamiya, M., Edwards, S.F., Raskin, S., Teive, H.A., et al. (2010). Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCdelta to mitochondria in spinocerebellar ataxia 10. PLoS Genet 6, e1000984.

8. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., and Lander, E.S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58,

1347–1363.

9. Bakalkin, G., Watanabe, H., Jezierska, J., Depoorter, C., Verschuuren-Bemelmans, C., Bazov, I., Artemenko, K.A., Yakovleva, T., Dooijes, D., Van de Warrenburg, B.P., et al. (2010). Prodynorphin mutations cause the neurodegenerative disorder spinocerebellar ataxia type 23. Am J Hum Genet 87, 593–603.
10. Wang, J.L., Yang, X., Xia, K., Hu, Z.M., Weng, L., Jin, X., Jiang, H., Zhang, P., Shen, L., Guo, J.F., et al. (2010). TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 133, 3510–3518.
11. Cagnoli, C., Michielotto, C., Matsuura, T., Ashizawa, T., Margolis, R.L., Holmes, S.E., Gellera, C., Migone, N., and Brusco, A. (2004). Detection of large pathogenic expansions in FRDA1, SCA10, and SCA12 genes using a simple fluorescent repeat-primed PCR assay. J Mol Diagn 6, 96–100.
12. Matsuura, T., and Ashizawa, T. (2002). Polymerase chain reaction amplification of expanded ATTCT repeat in spinocerebellar ataxia type 10. Ann Neurol 51, 271-272.
13. Warner, J.P., Barron, L.H., Goudie, D., Kelly, K., Dow, D., Fitzpatrick, D.R., and Brock, D.J. (1996). A general method for the detection of large CAG repeat expansions by fluorescent PCR. J Med Genet 33, 1022–1026.
14. Hashikata, H., Liu, W., Inoue, K., Mineharu, Y., Yamada, S., Nanayakkara, S., Matsuura, N., Hitomi, T., Takagi, Y., Hashimoto, N., et al. (2010). Confirmation of an association of single-nucleotide polymorphism rs1333040 on 9p21 with familial and sporadic intracranial aneurysms in Japanese patients. Stroke 41, 1138–1144.
15. Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718.
16. Lechertier, T., Grob, A., Hernandez-Verdun, D., and Roussel, P. (2009). Fibrillarin and Nop56 interact before being co-assembled in box C/D snoRNPs. Exp Cell Res 315, 928–942.
17. 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 293, 864–867.

18. 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. J Cell Biol 128, 995–1002.
19. Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 11, 228–234.
20. Zhao, Y., and Srivastava, D. (2007). A developmental view of microRNA function. Trends Biochem Sci 32, 189–197.
21. Eacker, S.M., Dawson, T.M., and Dawson, V.L. (2009). Understanding microRNAs in neurodegeneration. Nat Rev Neurosci 10, 837–841.
22. Hebert, S.S., and De Strooper, B. (2009). Alterations of the microRNA network cause neurodegenerative disease. Trends Neurosci 32, 199–206.
23. Tsuzuki, K., and Ozawa, S. (2005). Glutamate Receptors. Encyclopedia of life sciences. John Wiley and Sons, Ltd., http://onlinelibrary.com/doi/10.1038/npg.els.0005056
24. Nutini, M., Frazzini, V., Marini, C., Spalloni, A., Sensi, S.L., and Longone, P. (2010). Zinc pre-treatment enhances NMDAR-mediated excitotoxicity in cultured cortical neurons from SOD1(G93A) mouse, a model of amyotrophic lateral sclerosis. Neuropharmacology 60, 1200–1208.
25. Sanelli, T., Ge, W., Leystra-Lantz, C., and Strong, M.J. (2007). Calcium mediated excitotoxicity in neurofilament aggregate-bearing neurons in vitro is NMDA receptor dependant. J Neurol Sci 256, 39–51.
26. Todd, P.K., and Paulson, H.L. (2010). RNA-mediated neurodegeneration in repeat expansion disorders. Ann Neurol 67, 291–300.

Figure Legends

Figure 1. Pedigree Charts of the Five SCA Families (Pedigrees 1–5)

Haplotypes are shown for nine markers from D20S906 (1,505,576 bp) to D20S193 (3,313,494 bp), spanning 1.8 Mb on chromosome 20p13. *NOP56* is located at 2,633,254–2,639,039 bp (Build 37.1). Filled and unfilled symbols indicate affected and unaffected individuals, respectively. Squares and circles represent males and females, respectively. A slash indicates a deceased individual. The putative founder haplotypes among patients are shown in boxes constructed by GENHUNTER[8]. Arrows indicate the index case. The pedigrees were slightly modified for privacy protection.

Figure 2. Motor Neuron Involvement and (GGCCTG)n Expansion in the First Intron of *NOP56*

(A) Magnetic resonance imaging of an affected subject (SCA#3) showed mild cerebellar atrophy (arrow), but no other cerebral or brainstem pathology. (B) Tongue atrophy (arrow) was observed in SCA#1. (C) Physical map of the 1.8-Mb linkage region from D20S906 (1,505,576 bp)–D20S193 (3,313,494 bp), with 33 candidate genes shown, as well as the direction of transcription (arrows). (D) The upper portion of the panel shows the scheme of primer binding for repeat-primer PCR analysis. In the lower portion, sequence traces of the PCR reactions are shown. Red lines indicate the size markers. The vertical axis indicates arbitrary intensity levels. A typical saw tooth pattern is observed in an affected pedigree. (E) Southern blotting of lymphoblastoid cell lines (LCLs) from SCA cases and three controls. Genomic DNA (10 μg) was extracted from Epstein-Barr virus immortalized LCLs derived from six affected subjects (Ped2_II-1, Ped3_III-1, Ped3_III-2, Ped5_I-1, Ped5_II-1 and SCA#1) and digested with 2 units of

AvrII overnight (New England Biolabs Inc., Beverly, MA, USA). A probe covering exon 4 of *NOP56* (452 bp) was PCR amplified from human genomic DNA using primers (**Supplemental Table 3**), and labeled with ^{32}P-dCTP.

Figure 3. Multipoint Linkage Analysis with 10 Markers on Chromosome 20p13.

Figure 4. Nop56 in Mouse Nervous System

(A) RT-PCR analysis of Nop56 (422 bp) in various mouse tissues. cDNA (25 ng) collected from various organs of C57BL/6 mice were purchased from GenoStaf (Tokyo, Japan). (B) Immunohistochemical analysis of Nop56 in cerebellum, hypoglossal nucleus, and spinal cord anterior horn in a wild male Slc:ICR mice at 8 weeks of age (Japan SLC Inc., Shizuoka, Japan). The arrows indicate anti- Nop56 antibody staining. The negative control was the cerebellar sample without the Nop56 antibody treatment. Bar, 100 µm (C) Western blotting of Nop56 protein (66 kDa) in cerebellum and cerebrum. Protein sample (10 µg) was subjected to immunoblotting. LaminB1, a nuclear protein, and beta-tubulin, were used as loading controls.

Figure 5. Analysis of NOP56 in LCLs from SCA patients.

(A) mRNA expression (upper panel) and protein levels (lower panel) in LCLs from cases (n = 6) and controls (n = 3) were measured by RT-PCR and western blotting, respectively. cDNA (10 ng) was transcribed from total RNA isolated from LCLs and used for RT-PCR. Western blotting was performed using protein sample (40 µg) extracted from LCLs. The data indicate the mean ± SD relative to the levels of *PP1A* and GAPDH, respectively. There was no significant difference between LCLs from

controls and cases. (B) Analysis for splicing variants of NOP56 cDNA. RT-PCR with 10 ng cDNA and primers corresponding to the region from 5' UTR to exon 4 around the repeat expansion were performed. The PCR product has an expected size of 230 bp. (C) Immnunocytochemistry for NOP56 and coilin. Green signals represent NOP56 or coilin. Shown are representative samples from 100 observations of controls or cases.

Figure 6. RNA Foci Formation and Decreased Transcription of *MIR1292*

(A) Cells were fixed on coverslips and then hybridized with solutions containing either a Cy3-labeled C(CAGGCC)$_2$CAG or G(CAGGCG)$_2$CAG oligonucleotide probe (1 ng/μl). For controls, the cells were treated with 1000 U/ml DNase or 100 μg/ml RNase for 1 h at 37°C prior to hybridization, as indicated. After a wash step, coverslips were placed on the slides in the presence of ProLong Gold with DAPI mounting media (Molecular Probes, Tokyo, Japan), and photographed with a fluorescence microscope. The upper panels indicate LCLs from an SCA case and a control hybridized with C(CAGGCC)$_2$CAG (left) or G(CAGGCG)$_2$CAG (right). Red and blue signals represent RNA foci and the nucleus (DAPI staining), respectively. Similar RNA foci formation was confirmed in LCLs from another index case. The lower panels show RNA foci in SCA LCLs treated with DNAse or RNAse. (B) Double-staining was performed with the probe for (GGCCUG)$_n$ (red) and anti-SRSF2, NOP56, or coilin antibody (green). (C) Gel-shift assays revealed specific binding of SRSF2 to (GGCCUG)$_4$ but little t to (CUG)$_6$. (D) RNA samples (10 ng) were extracted from LCLs of controls (n = 3) and cases (n = 6). MicroRNAs were measured using a TaqMan probe for precursor (Pri-) and mature *MIR1292*. The data indicate the mean ± SD, relative to the levels of *PP1A* or *RNU6*. *: $P < 0.05$.

Table 1. Clinical characteristics of affected subjects

Pedigree No.	Patient ID	Gender	Onset age (y)	Current age (y)	Ataxia	Motor neuron involvement			Genotype of GGCCTG repeats
						Skeletal muscle atrophy	Skeletal muscle fasciculation	Tongue atrophy/fasciculation	
1	III-5	M	50	70	+++	N.D.	N.D.	N.D.	g.263397_263402[6]+(1800)
	III-6	F	52	68	++	+	+	+	g.263397_263402[6]+(2300)
	IV-2	F	57	63	+	-	-	+	g.263397_263402[6]+(2300)
	IV-4	M	50	59	+	-	-	+	g.263397_263402[6]+(2300)
2	II-1	M	55	77	+++	++	+	+	g.263397_263402[6]+(2200)
	II-2	F	53	70	++	N.D.	N.D.	N.D.	g.263397_263402[6]+(2200)
3	II-3	M	58	77	++	++	+	+	g.263397_263402[3]+(2300)
	III-1	M	56	62	+	-	-	±	g.263397_263402[8]+(2200)
	III-2	M	51	61	++	+	+	+	g.263397_263402[6]+(1800)
4	I-1	M	57	died in 2001 at 83	++	N.D.	N.D.	N.D.	g.263397_263402[5]+(1800)
	II-1	F	48	61	++	+	±	++	g.263397_263402[6]+(2000)
5	I-1	M	57	86	++	+++	+	+	g.263397_263402[5]+(2000)
	II-1	F	47	58	++	+	+	+	g.263397_263402[8]+(1700)
	SCA#1	M	52	69	+++	+++	+++	+++	g.263397_263402[5]+(2200)
	SCA#2	F	43	53	+++	-	-	+	g.263397_263402[6]+(1800)
	SCA#3	M	55	60	++	-	-	++	g.263397_263402[8]+(1700)
	SCA#4	M	57	81	+++	+	+	+++	g.263397_263402[5]+(2200)
Mean			52.8						
SD			4.3						

N.D.: not determined

Figure 1

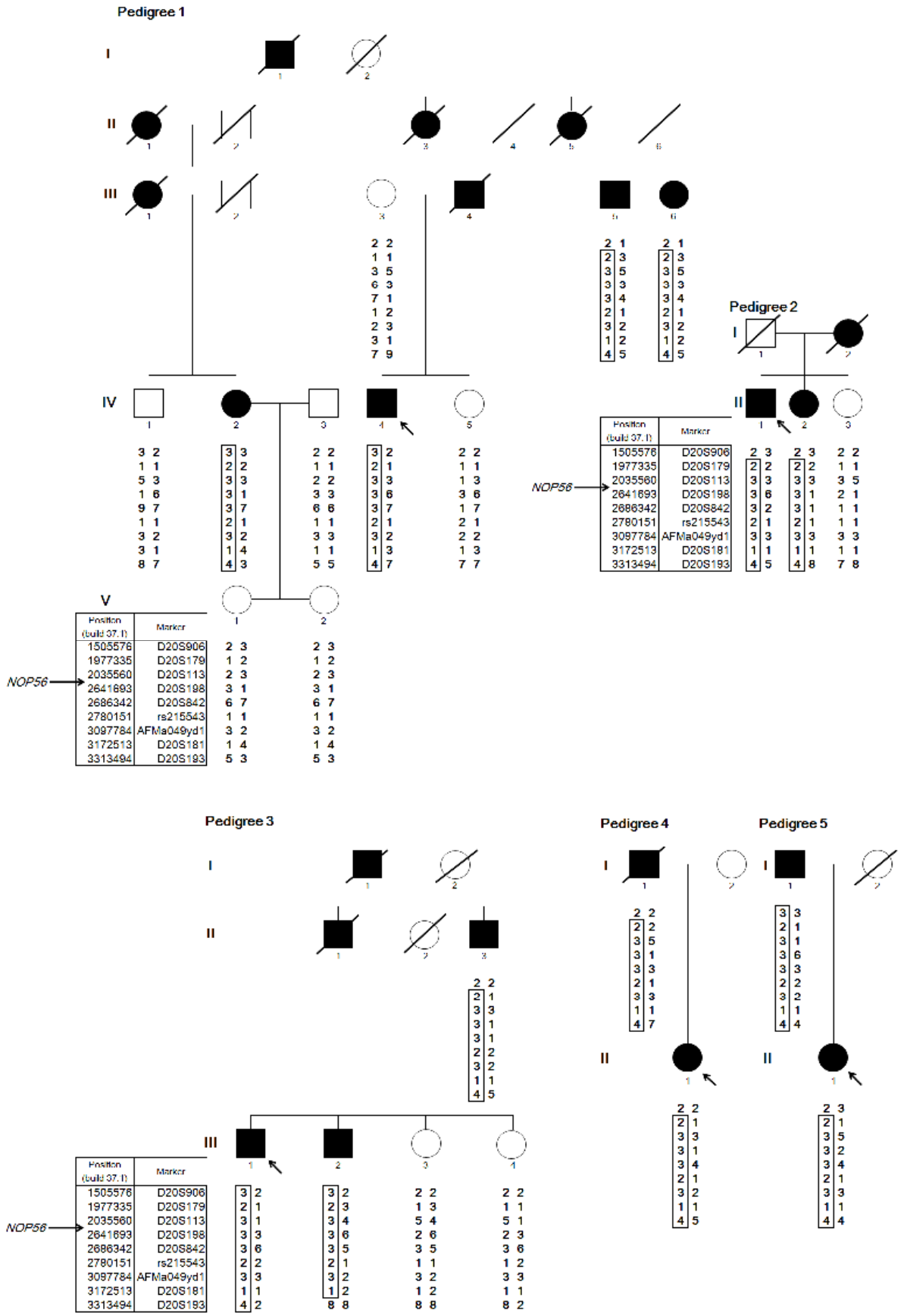


Figure 2

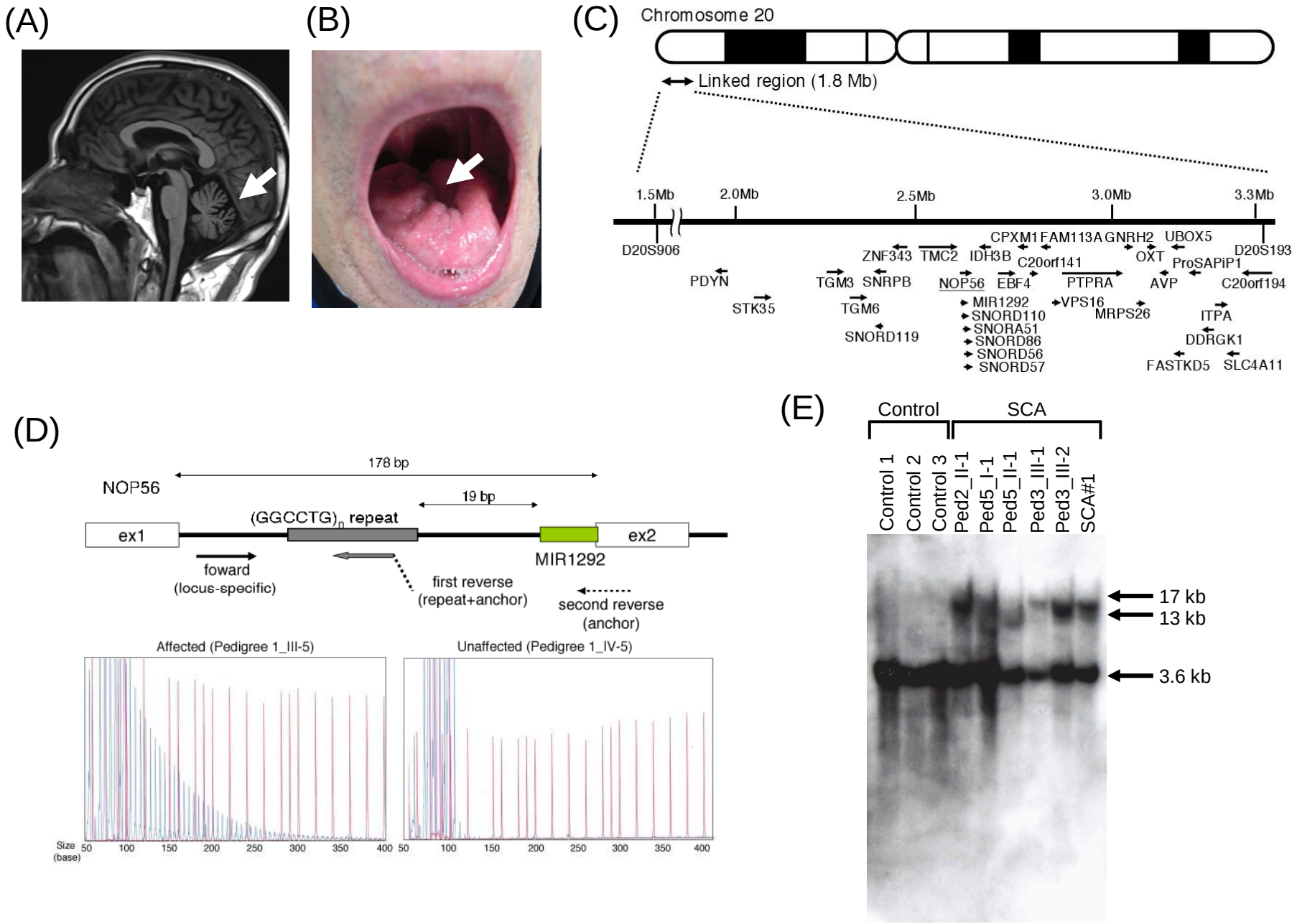

(A)
(B)
(C)
Chromosome 20
Linked region (1.8 Mb)
1.5Mb
2.0Mb
2.5Mb
3.0Mb
3.3Mb
D20S906
D20S193
CPXM1
FAM113A
GNRH2
UBOX5
ZNF343
TMC2
IDH3B
C20orf141
OXT
ProSAPiP1
PDYN
TGM3
SNRPB
NOP56
EBF4
PTPRA
AVP
C20orf194
STK35
TGM6
MIR1292
VPS16
MRPS26
ITPA
SNORD110
SNORA51
SNORD86
SNORD56
SNORD57
SNORD119
DDRGK1
FASTKD5
SLC4A11
(D)
178 bp
NOP56
19 bp
(GGCCTG)n repeat
ex1
ex2
MIR1292
foward
(locus-specific)
first reverse
(repeat+anchor)
second reverse
(anchor)
Affected (Pedigree 1_III-5)
Unaffected (Pedigree 1_IV-5)
Size
(base)
50
100
150
200
250
300
350
400
(E)
Control
SCA
Control 1
Control 2
Control 3
Ped2_II-1
Ped5_I-1
Ped5_II-1
Ped3_III-1
Ped3_III-2
SCA#1
17 kb
13 kb
3.6 kb

Figure 3

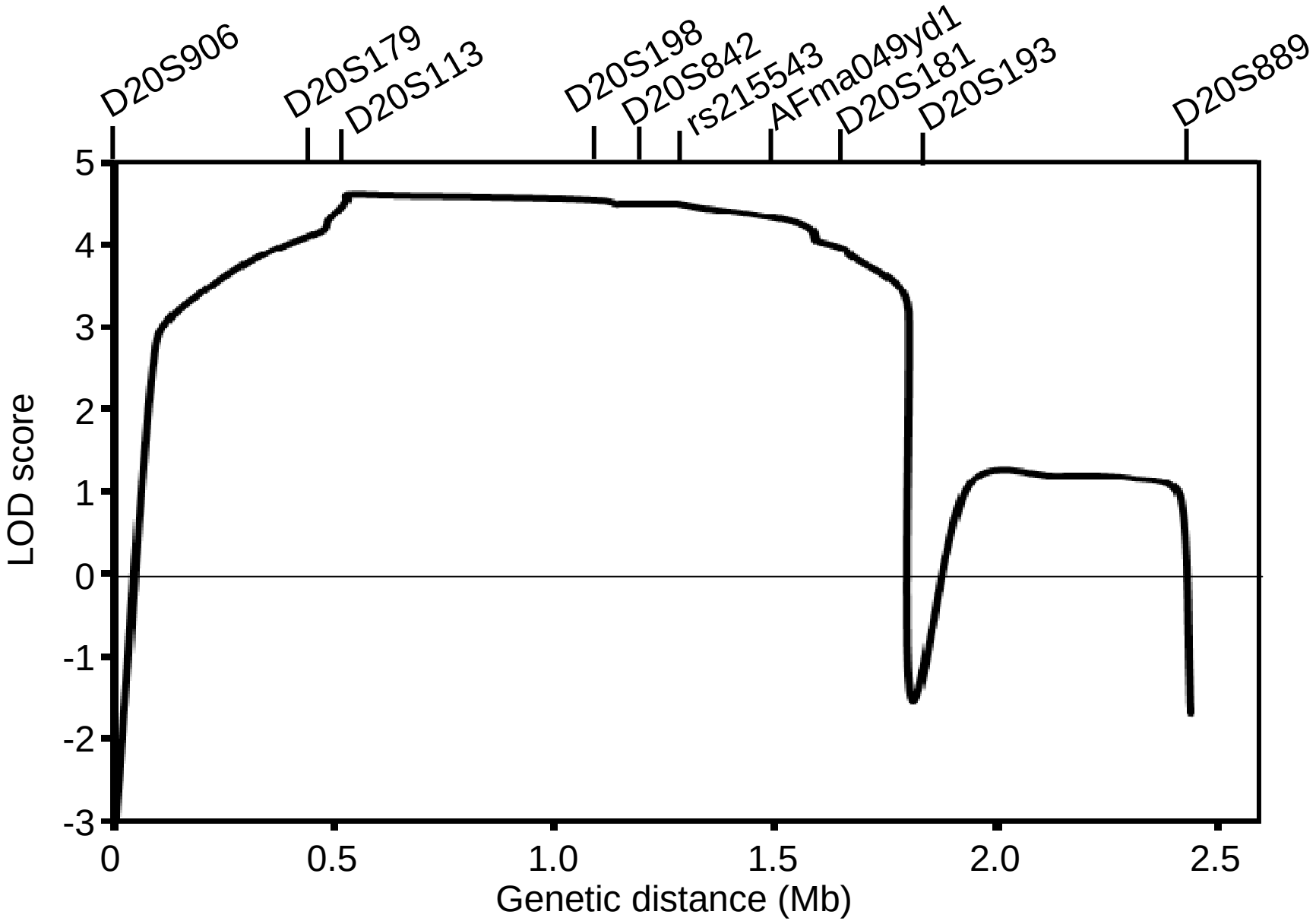

D20S906
D20S179
D20S113
D20S198
D20S842
rs215543
AFma049yd1
D20S181
D20S193
D20S889
5
4
3
2
1
0
-1
-2
-3
LOD score
0
0.5
1.0
1.5
2.0
2.5
Genetic distance (Mb)

Figure 4

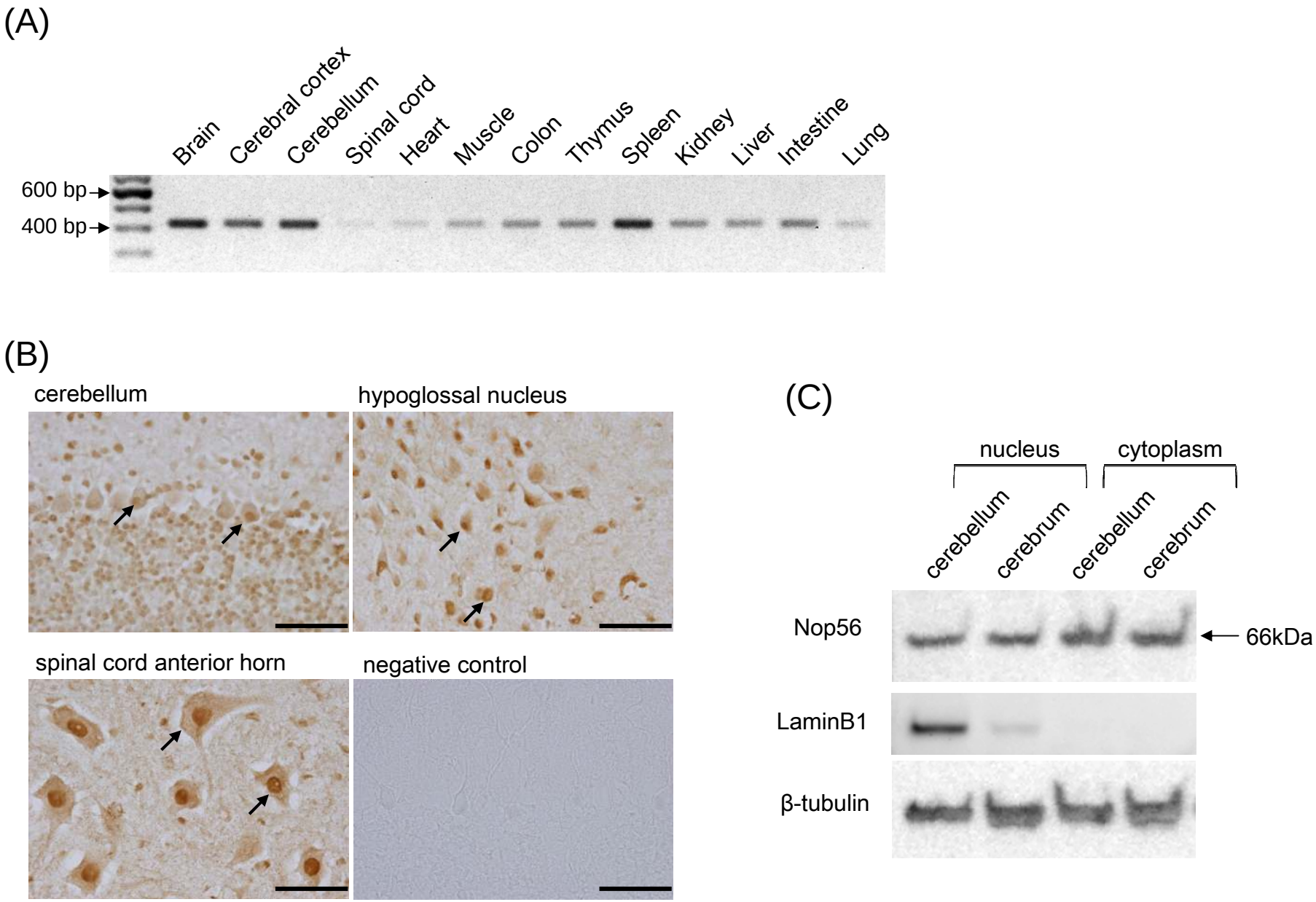

(A)
Brain
Cerebral cortex
Cerebellum
Spinal cord
Heart
Muscle
Colon
Thymus
Spleen
Kidney
Liver
Intestine
Lung
600 bp
400 bp
(B)
cerebellum
hypoglossal nucleus
spinal cord anterior horn
negative control
(C)
nucleus
cytoplasm
cerebellum
cerebrum
cerebellum
cerebrum
Nop56
66kDa
LaminB1
β-tubulin

Figure 5

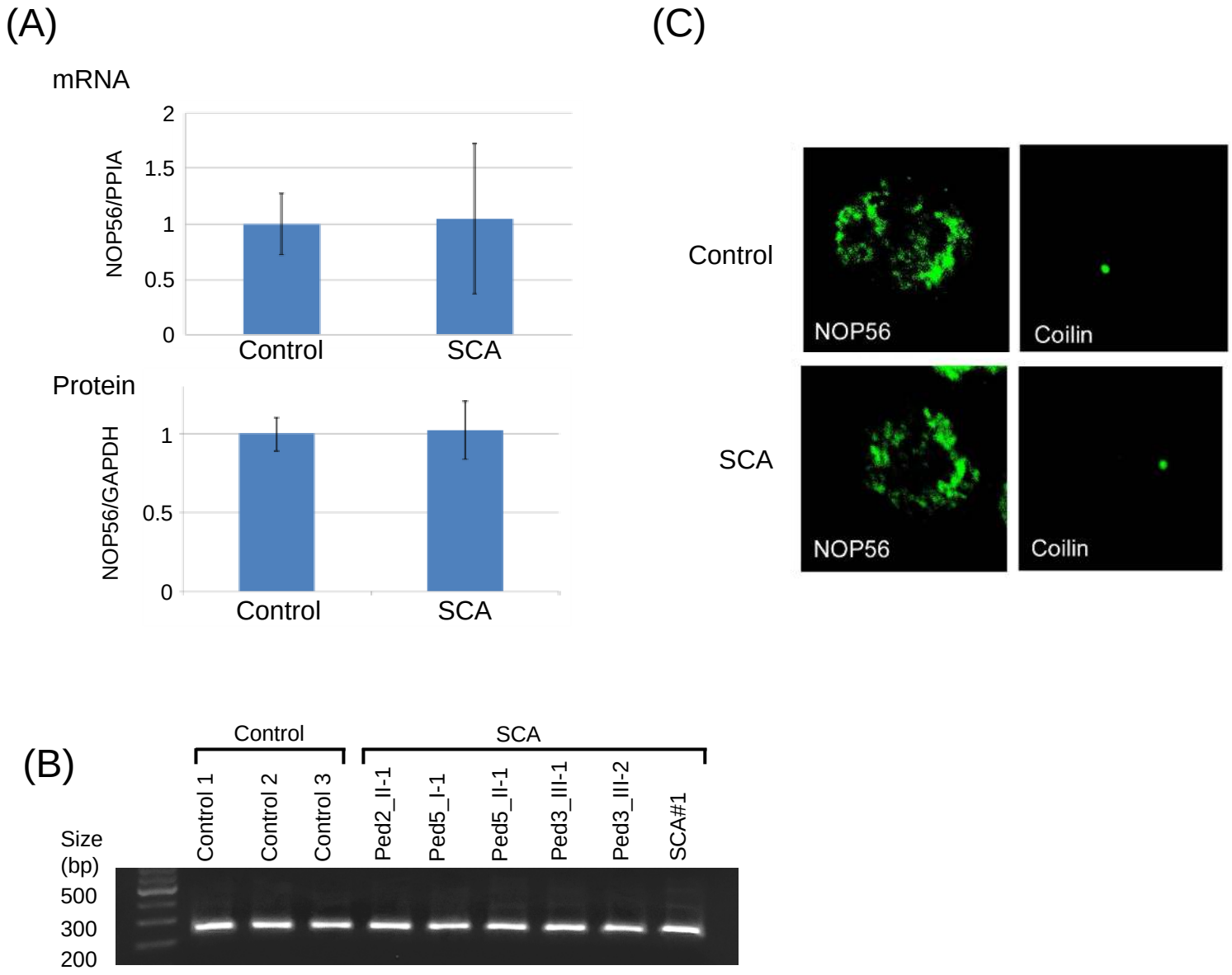
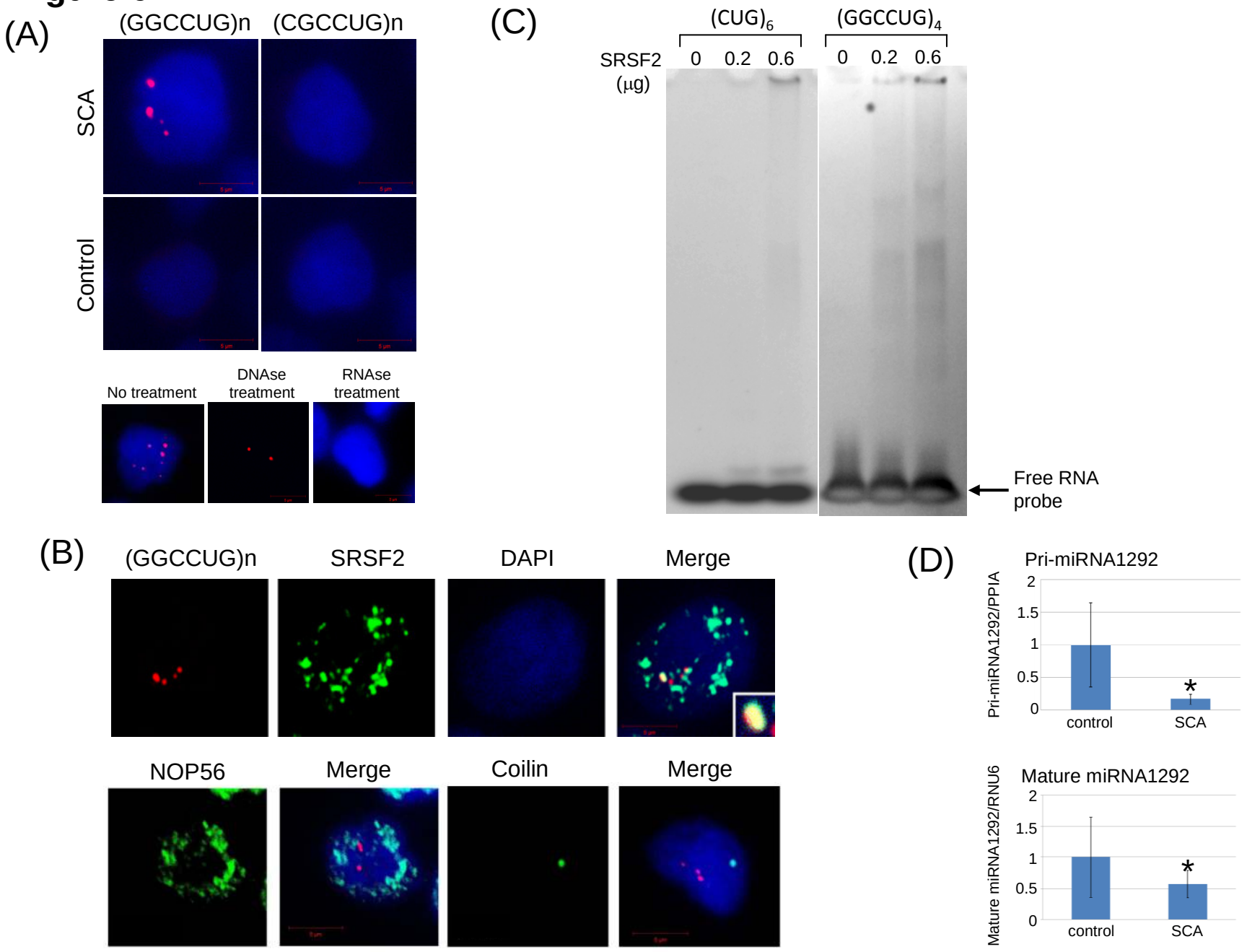

(A)
mRNA
NOP56/PPIA
2
1.5
1
0.5
0
Control
SCA
Protein
NOP56/GAPDH
1
0.5
0
Control
SCA
(C)
Control
SCA
NOP56
Coilin
NOP56
Coilin
(B)
Control
SCA
Control 1
Control 2
Control 3
Ped2_II-1
Ped5_I-1
Ped5_II-1
Ped3_III-1
Ped3_III-2
SCA#1
Size
(bp)
500
300
200

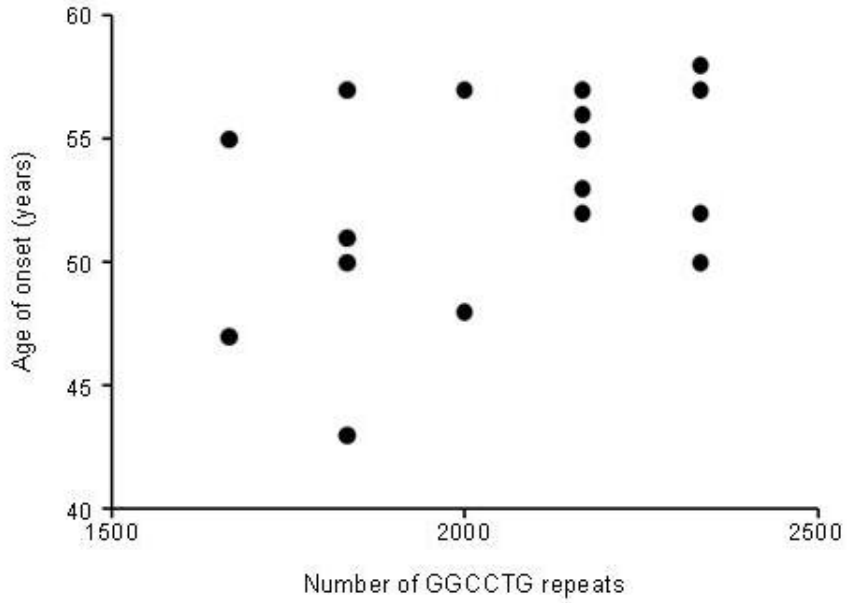
Figure 6


(A)
(GGCCUG)n
(CGCCUG)n
SCA
Control
No treatment
DNAse treatment
RNAse treatment
(B)
(GGCCUG)n
SRSF2
DAPI
Merge
NOP56
Merge
Coilin
Merge
(C)
(CUG)6
(GGCCUG)4
SRSF2 (μg)
0
0.2
0.6
0
0.2
0.6
Free RNA probe
(D)
Pri-miRNA1292
Pri-miRNA1292/PPIA
2
1.5
1
0.5
0
control
SCA
*
Mature miRNA1292
Mature miRNA1292/RNU6
2
1.5
1
0.5
0
control
SCA
*

Supplemental Data

Expansion of Intronic GGCCTG Hexanucleotide Repeat in *NOP56* Causes a Type of Spinocerebellar Ataxia (SCA36) Accompanied by Motor Neuron Involvement

Hatasu Kobayashi, Koji Abe, Tohru Matsuura, Yoshio Ikeda, Toshiaki Hitomi, Yuji Akechi, Toshiyuki Habu, Liu Wanyang, Hiroko Okuda and Akio Koizumi



Supplemental Figure 1. Correlation of the Number of Repeats with Age of Onset.

A scatter plot shows no negative correlation between GGCCTG repeat number and onset age (n=17, r=0.42, p=0.09).

Supplemental Table 1. Primers used for amplification of candidate genes (Human build 37.1)

PDYN	exon 3	CTTTGGGCCTCTGCTTTACCT	TCCAGGCCATCTATAGGGCA
	exon 4	TCCCCTACCTTTATGCACCA	AACATACTCCCACGCAGAAGA
STK35	exon 1	CGGATCACGGGAATTTCG	ATTGGCTGAAAAGTTCGGCT
	exon 2	TGGCTTCCGTCTTAAAGGG	CAGGAAAGGAGGGTGTGTCA
	exon 3	GTCCCCTTGGAGCAGTTGTTT	AATCACTTGAACTCGGGAGGT
	exon 5	TCTCTTTAGAGCTCTGCCCCA	TTGCCCACATTGAATTTCTT
TGM3	exon 1	TTATTATCTGCCCCCTTCTCC	CTCTGGCTAGCACCCTAAAAT
	exon 3	AACAGGATGCACAGAGGTTCA	TCCCTCTTGATTTGAGGATGG
	exon 4	TGGCCTGTATGTTTGTTCCA	TTGGGGCTTGGAGAGATAGAA
	exon 5	TCAGGGGAGGGCTAAAGGT	AGGTGGCCAATGAAAGTCTT
	exon 6	TTCCTGTGGTTCTTGCCAGT	TGTAAAGAGTGTGCGTGCCTA
	exon 7,8	TTCAATCATGGCCTTTGGCT	AATCAGCATTGCCAGCAGTT
	exon 9	TGTTGTCATGCTGCACTGTTG	TTGTTTTAATCCCATCATGCA
	exon 10	GGTTGCAGTGGTCCTGGAA	GCAATCCTATCATTCAGGCA
	exon 11	TGAAAGTCGAATGCCTGCTA	TCCAAAGCATTAATACATGGC
	exon 12	AGATCCTCCCACCAGCTCA	AAAACTCTCCTTTGCCCTCTG
	exon 13	TCTCCCCTTCTTCATCCTCA	CCAAACCAAATGCAAAAGCAG
	exon 14	CTCCATCAGAACAGGACAGGA	CACTCCCTTTGGACATTGAA
TGM6	exon 1	TGATTTTGTGTCTCGTGGGTG	AGTTCATGTGTTCATGGTGGA
	exons 2,3	ATGAACAAATGACTGGCCGA	TAAGTTCTTGCCCAGCTCTTG
	exon 4	AAGCCCCCTCTTGACCTCT	CCTGGCCCAGTGAATAGTAGA
	exon 5	TTGAGGAAGGGTTTCCAAGAC	CAGCGAATTAAAACAAGGGG
	exons 6,7	AAAAGCAAGAGTGACCCCGAT	AGATTCAGGAGAGCTGGGCT
	exons 8,9	GATTCACAACATGCAGCCACA	AATAAAGGCACTTGGCTCAGA
	exon 10	GAGAATCAAACACAAGGCATG	AAGGCACTGAGCACAATGCT
	exons 11,12	TGGGCCTTAGGTTTCTTCAT	ATAGTCTGTGGGCTGGTTCCT
	exon 13	ATGTCAAGCCACAAGGTGAA	AGATGAAGGTTGGAGAGGCTC
SNRPB	exon 1	AGCAGCTCTCAGTACGGATTT	CGCCAAGGTCCTGGTCTTT
	exon 2	GTGGCATGGGAGAATTCCTA	AGCGTTCAATGTCCCCATTT
	exon 3	GGGCTATCTTGGGAAAGTTTG	CCTCTTGCAGGGTAACCTCTT
	exon 4	AATGGTGTTGGCACACATGGA	CTTTTCACTTGCTTCTAGGGC
	exons 5,6 (*SNORD119*)	TGGTGCTGAGAGTTGTAGCAT	GCCCCAAGAATATAGCTCAAG
	exon 7	ACATAGGAACCCAGTCAGCCT	GGGAAGAAGGTAAGATCAGGG
ZNF343	exons 3,4	GATGGAGACTCCGTGTTTGTT	TCAAGAAAGCCTAATGCTTCA
	exon 5	TGTCAAGGGCCTAGAATGTGT	TAGAAAGTAAGCCCCATGGAG
	exon 6	TTTGCAAGGGGGAGAGATG	TTATGGCTGCAACTCAATGG
	exon 7-1	GGCATGGCAGTAAAACATTC	TCACAAAAGCTTCGTCCACA
	exon 7-2	GTGTGGGCAAAGCTTTAGAA	TACACAGGGTCTACTCCCCAT
TMC2	exons 1,2	AATCTCAAACCAAGAAGCCCC	TCCCATGTTAAACACCTTGCC
	exon 3	TTTGGGGTGTCCTGTTCTGA	TTCTGGGATGAAGAACCCAC
	exon 4	TAGAATTTTGGCCTCCGCT	AGTCCTCCGAGTCTTCCATG
	exon 6	TTCTGCCTACCATGCCTGTTA	AATTGGCAGGATACGATTGG
	exon 7	CGTTGGGGAAGTAAAGGTTTG	TGGCCTGACAAGTTCAAAAC
	exon 8	TTCACTTTCTGACTGTGGGCA	CAATTACTTTTGCACCAGCC
	exon 10	AGCTGAGACTGTGCTATTGCA	TGCAGAGACATGGTTTGCA
	exon 11	AGGAGGTGAAGGGAAGAACAA	ATGGCCTGGTTCATGTCTTT
	exon 12	TGTCCTGGCTATCCTCAAAGC	ACCAGCCACGAACATTCTTA
	exon 13	ACCACGTCTGACACATGGTGA	GTTGTAAAAGCCAGTTCCCAA
	exon 14	CATGTTGATGCCTTTGCCA	AACAATGGTCATCTTTGGGTC
	exon 15	TGAAAGCCTGAGAAGGGAAT	CAAAGAGATTTGGAGTTCCCA
	exon 16	TCGCCTCTCTTCACACACAA	AGGAGCTCAGCAGACTTGTTT
	exon 17	TGCAAGGCCTGAAAGATAGA	CAGAAACTGTCCTGCTTCTCA
	exon 18	ATGGTTCTGGCTTGATGATCG	ATTAGTTGGGCATTGTGGTG
	exon 20	TCTGGTTCTTCCAGGAAAGCA	CCATGCATTTTCCTTTCCCT
	exon 21	CAGGACCTTCTCCACATTGA	TGCATGCTGTTGAATCTCACA
	exon 22	AATCGCTTGAACCCAGGA	ATTCACCTGCCCAACTTCAT
	exon 23	AGTGCAATCCGACAGCTCTCA	TCAAGCGATCCACCCACCTT
NOP56	exons 1,2 (*MIR1292*)	TTCCCAAGTCGTTTCGCC	ATCTAGAGCTTTCCAGGCCC
	exons 3,4	TGAAGGAAGTGGAGGAGATCA	CCTTGAGCTCTGTGAAGACAA
	exons 6,7	TGATGGGGAGGGATCTAGGTA	AACACAGCCTGTGGTAAGCA
	exon 9-1	TGGATCTTTGTCCCATTTCC	TGGTCAGCCATCACCGTGA
	exon 9-2 (*SNORD86* , *SNORD56*)	ATGCTGGCAGCCTCACCAA	CAGACAGTTCATCACCTCCAA
	exon 9-3 (*SNORD57*)	GGAGAAGCTTCGAGAACAAGT	AAAAAACACCCACCATGCTG
	exon 9-4	TGGGCTGAGGTAATTTCTCAT	ACTGAGGCTGTCATTGCTGC
SNORD110	exon 1	TCTGCTTTCTGTTCGATTGG	TCAGGGGAAAGAACACAGTTC
SNORA51	exon 1	CCACCCATAATACTGGAGCCT	TGCAAAGAGCCACAGTCACT
IDH3B	exons 1,2	AAAAGGAGAAACAGGCGTGA	ACGGATCCTGGGAGTAGAGAA
	exons 3,4	AATCTGGCTGGGGTCTCTCT	TGGTTGCCCTGGAGTTAATA
	exons 5,6	TTACTGATGTGGGAATGGGA	TCACAAGCACATCAAACTGGT
	exons 7,8,9	CCCCAAAATCAAATTTGAGAC	AGATGAAGAACACGCCCTGAGA
	exon 12	ATCCTGGCTCCTCCTTCCATT	AAGAGGCGGTTGGCAAGA
EBF4	exon 1,2	GGAAATGCGGGAGTACAGTCA	TCAGAAATCTACCGGGGCA
	exon 3	TCCAACATTCAAGGCCTATCA	TTGAGTCTTCAGGGGAGTCAG
	exon 4	TTTTTGCCCAAACTCTTGGC	GGACAAGATGGGAGGATGCT
	exon 5,6	AAGTTGGGGTTAGGAGAAGGG	TAGAACAGAGGGCAAAAGCCA
	exon 7	TAGGCTTGGGAGATGCCA	TAAACAGGCCAGGCTAATGG
	exon 8	ACATCAGCACCCTCAGCTCA	TCCAGAAAAGTTGGCCACCT
	exon 9	CCATGATGGGAATATGGGAT	ACAAGTTGAGAAGGAGCTGCC
	exon 10,11,12	TTTTTGTAGCGCCTGGGGA	AATGTTCAGGAGGTGAGATGG
	exon 13,14	GAGTTTTCCGGGGGACTTG	TGCTGAAGGCGTTGATGC
	exon 15,16	AGTAACCAGGTATGGCGCCTC	CGGCAAAGAAGGCTAAAAGT
	exon 17	AACAAAGTACCCCAGGGTGCA	AAGAGCAGGTTAGCCAGCAT
	exon 18	AAAGTGCTATATCCCCTGCC	AGGCTTGAGCACAGGATGAA
CPXM1	exon 1	TCCTGTTGGTCGACTTGATG	TGTGTGAATGTGTGTGAGTGC
	exon 2	TGCTGTGGGCTCACATGTC	TAAGGTTGCTCCTGCGCTA
	exon 3	AACCTTAGGCTCAGCTTCCCA	GACACAGGACATGGTGGTCA
	exon 4,5	ATGGTCTCAGGGTAGGGAAGG	AAGGCAAGAAGTGATGTGGA
	exon 6	TCTAGCTGAGCCCACTAGGGT	AAAGGGTGTGTCCACAGTGAA
	exon 7	AGTCAGGCCAGGGTTGGTT	TCGATGCTCTGGTTGTTCCA
	exon 8,9	ACTCTGTCCTTCTGCCCTGGT	GAACAGACGCCAGCACTGTA
	exon 10	TGAAGTGTCCCTCAGAGAAGG	CCTGTGTGCTTCCAAGACAAT
	exon 11,12	TTCATGTTCCATGGAGCTCA	TGGTACCTGTAGCAGAAGCTG
C20orf141	exon 1	CACCAGCTGCTTATGAGGTCA	TGCTGCCCACTTACCTATGGA
	exon 2	TGGCAGGTGGGCATTGTA	TTGGGCTCCCTGGCTAGTAT
FAM113A	exon 2	TCACCTCCTCCCTTTAGCATT	ACTGGACGGAACAGACCAGAA
	exon 3	TCTGTTCCGTCCAGTAGGTTT	TTGAGCAGCGACCATATCTGT
	exon 4-1	ATCAACTCCTGCCTCTGGGAT	TTTCCACTCCTCACACCCAT
	exon 4-2	CGGAGCAATATTCTTGTCCA	TCAAAGGCCTAAGCCATCAA
	exon 4-3	TGTAACACTTGGTAGCCAGGA	TTGTTTTAGGTAGGCTTGGGA
	exon 4-4	AAGGAACACACTTTGGGCTTG	TCCAGGTGAGCTTCTCTGTTT
	exon 4-5	TCCTTGTGCCCCTAGCACT	GGAGGGCACATTCAATGATT
VPS16	exon 1	AAGTGAGGCTGCCCACAGT	TGTGCGCTAAGTGGCAGA
	exon 2,3,4	AGCCTTGTGGAAGACAAATGGA	CGGAACCAAACTCAGTGTGAAA
	exon 5,6,7,8	GACACTTCAGCATGGGCAATGTA	CGAATCAAGGAACTGATTGTCC
	exon 9,10,11	GCTGTCCCGAGACAAAGGATTA	TGGAGGACATAGTGTCTCTTCA
	exon 12	TGGGTTACTATTGGGAGGAGTTCT	TGGAGAATAAGCCCCGCTT
	exon 13,14	TATAGCCAGTATCCCTGTGCACG	TGTTGGGATTACAGGCATGA
	exon 15,16,17,18	TAAGGGCTTGCAGGAGTGGA	AACACGAAGGCCTAGATTCCT
	exon 19,20,21	ATCCCTCTAGGACATCAGAGTGG	AGCTGAACAGGAGCATGAA
	exon 22,23	GGGGTTGGGGGATTATATGTACT	GGAACACATGGAGTTTTGCTGT

PTPRA	exons 1,2,3,4	ATCCAGATGTTTGTGACACCC	ACAGTGAGGACCAGATGGAGT
	exons 5,6,7	AGCCATCCCTCTAGGACATCA	TGCCTGCCCACAAATGTGTAT
	exons 8,9	TGGTTTAGGTGATTTCTGCCC	GCTTTCCTTGGTAACTGTGGA
	exon 12	TGCCTGGCTACTTTTTGTGGA	ATGCCACCACATCTGGCTAAT
	exon 15	TGAGGAAGCATGCATATCAGG	CAATGCTGAGCATTCAATTCC
	exon 16	TGTTGAGGGGGATTGGTCT	CTTCACTCATGCTAACCCAAA
	exon 17	CCAGACCACTGTCCAAAGTTT	AGGGGAAAAACAACACAAAGA
	exon 20-1	CACCTCAATAGCCCTGGCAT	TGGGCTTGGACAGATGGAA
	exon 20-2	TTCCAGTGTGCCAAGGGTAA	TGGAGGCTAAACGGGGTTCTA
	exon 21	TCTTACAGGCTTGGTCCATGA	GGTGAGGCAAATCTCACTTCA
	exon 24	TAAGGAGCTTGTGGCTGTTTC	ACCTTGGCCTTCCAAAGTTCT
	exon 25	TGGCATCTTTATACAAGCGTG	ACATGGGAAACCATAGGGAA
	exon 27	CCTGGCCTGGATTTCTTATT	CAAGGACAGAGGGCCTTATTA
	exon 28	CCTGGCACATACATGGTAGAA	TCTAGGCACACACCTGAGGTT
	exon 29	AAGAGGTCAAAGGGCCTTCT	TTTCCACAGTGCTTGGTCAT
	exon 30	TTTTAGTTCTGACCCTGCCA	TAATCTTGGAGGACTGCCCTA
	exon 31	TTCTAGCTGGAGGTCAGGATT	TTGGGGCTAGGGTTACAGATT
	exon 32	TGCATTTCAAGTCCCACTTCA	AGTGATTCCAACGTGCAGGAT
	exons 33,34	CCAGCTGAACATATGGGAACA	AGTGGGTGGGTGAGTATCAAT
	exon 35	CAGAAAGCAACCAGCTTGTCA	TTGGGAGAAGCTAATGACTGC
	exons 36,37	ACGAAGCGTATCAGCGTAAGA	TAGCATCCAATCCTGTCTTGG
	exon 38	TGAGTCCCTCCACAGCACAT	TATGCTCCCATTGCTGCCAT
	exon 39	CAGAGCTCAGGTGAAAGTTCA	TTTCTCGGCTGAGGATTTCA
GNRH2	exon 2, 3	GCAGAGAGGGAAGGGCATAA	TGAGAAATGGCTGGGGGT
	exon 4	TAGCTGGATCCTCAGGCTTCT	GGGGCCATCCCTTAGTTACT
MRPS26	exon 1	TTCGGTTCCAGAGGCCACA	TTCCTCTGCACCTCGGACA
	exon 2, 3	TTACCAGCACTACCGCCAGA	TTTGCGCCTGACTGGCACT
	exon 4	AGAGCAGGAGCTGCTTTCTCA	TGCGTTTGGAAGTTTCTGAC
OXT	exon 1	AATGAAGAGGAAAGCCCGTA	TCAAAATCCGCTCAGCTCCT
	exon 2, 3	GGAGCTGAGCGGATTTTGA	AGAACAGCACCCGCTCTGT
AVP	exon 1	TGTCCCCAGATGCCTGAAT	ATGCCATGCCTCCCTCTT
	exon 2, 3	AAACCAAGGTGCCGAGCAGAT	TCCCACCTCTCTCCCTTTC
UBOX5	exon 4	AAGGAAAGTCAGTGTGGACCG	TGATTCTAGAGGTGAGCCTGC
	exon 5-1	CCTGATCTGGGACAATTCAGT	TGCGGTTACACTTCTCCAGT
	exon 5-2	AGAAGCTGGCCGAGATCATT	TCTGAAGACAAAGCTGAGGGG
	exon 6	TACCTCCCGGTGTTCTGTCAT	AGGGTGGGTGTTGGAACTGA
	exon 7	CCACTCCCCTACCTGATCAGA	AGCGCAGAAGCAATGTGCTAT
FASTKD5	exon 2-1	TCATTTGTGATCCCTGGCTC	ATGCTGCTCTGCAGGCAAA
	exon 2-2	CCGTGTTCACAGCTATAATGC	ATGTGATCCACGTGAGTGAAA
	exon 2-3	AGGTTGGTACCATCTGTTTGG	ATACCACAGCAATTCAGACCC
	exon 2-4	AGGGTTTGTCAGGTTAGCTCA	AACTCTACCACACGGTAGCCA
	exon 2-5	TAGCAGATAAATCAGGGGCCA	TAGCTCTAACCTGCCTTGGAT
ProSAPiP1	exon 1	ATTCCTTCACCTTGGATGCCT	CACAACCCAACCTCCAAGAAA
	exon 2-1	GGGAACCTCAGGGTGGAAAT	AGCTCCTGGATGAGTGCACTG
	exon 2-2	TCCAGCAAGAGTGGGTCGT	ATTGATTTTTGTCCCCCCTG
	exon 3	TTGAGTCCAGGCAGGGAAT	AGGGAGGAACCTGGTCACA
DDRKG1	exon 1	GGACATACCGTCTGCTATAATTTCC	TTGGAGTCGAGAGAAAGGGGTA
	exon 2	CCTTGCCAGTCAGACTGAGA	AACAAATGCCAGGTCCCAA
	exons 3,4	AGTGACATTTGCAGGTGGGT	AGGGACCAAATAAACCAGGA
	exons 5,6	TTGGGGGAATGGAGAAATG	GGGTTGGAGGGAGAGAAACT
	exons 7,8	TACAGTGTTTTTCCAGCCACC	TCCTCCTTGTAACTGCATCCA
	exon 9	TGATTCGACTCTCCTAGCAGG	TTATCTAGGTCTTGGGGGCA
ITPA	exon 1	AGAGAAGAGCGAAAGCAGGG	TTCTTGCGCCCCAGCTTTT
	exon 2, 3	GTAAGCTTTAGGAGATGGGCA	CGGTCCTAGAAAGCTCAACAA
	exon 4	CCAAAGTTAAGAGATTGGCCG	AAAGAAAGGCATGCTTCTCC
	exon 5	TGCTGGGATTATAGGCGGGA	TACAGGGTACGAGCTGCAGGT
	exon 6	CCGCTACCCCAATTGAGA	TGAAAAGCTGGAAAGGCTGA
	exon 7	AGCAAACATTTGCAGGTGCT	AGATTCCTAGTGTCCACCCCA
	exon 8	ACTCCCCTTTCCTTGGGGT	TCCACTTGCCAGAGTTTCTCA
SLC4A11	exon 1	AGTCGAACGTTTTCCCAGAAG	CAGAGCCCTAATGAAACCA
	exon 2,3	TTTTGGACCAACGGCTCTG	AGATAGGCGAGCAAAGCCA
	exon 4,5	TTCCTCGCCTATGGGATG	TCCTGGAGGCATGGGAAGA
	exon 6,7	TGATGGCTTCCCTGAGAAT	TCTTCTCCCAAGTTGGTTGG
	exon 8	TTTTCCCTCCCTAGCAGAGGT	CAACATGTTTCTGACACACCCA
	exon 9,10,11	AAAACCTGCTGCCAGTTCATG	AATGGCTGCCCAGAGAAGA
	exon 12	ATCGCTTTCGGGTCTCTCAA	TTGGGGCAGCAATATGGT
	exon 13	ACCATATTGCTGCCCCAA	TTGATCACGGGCACACACT
	exon 14,15,16	TTGATCACGGGCACACACT	TTCACCAGCCTGCAGCAGA
	exon 17,18	TTGGTGAATGCACCGGAGAA	ACCCTCCGGATGTAGTGTGT
	exon 19,20	CTCTATGGCCTCTTCCTCTACAT	AGTCACCCACACACCTACACCT
C20orf194	exon 2	TTAAGAAGTGGGGTCCCTGT	TGAGCCGTTCAGCAAAGAA
	exon 3	ATCCCCAGCAAAGTCATTCCT	ACAAATTTCGGGGGAACAAG
	exon 4	GTTTCTGGCACAAAACTGGT	ACTGGGGCTTTGGACTATTT
	exon 5	TTTGGTTTACAAGGCACAGTG	TCCTTTTGTCTCCACAGGCATA
	exon 6	CTTTGGAAACACCTCCTTGGGT	TTGCACAGAAAAGTCCCCAT
	exon 7	TTTTGCTAGGTGAGGCCCT	TAACTTGCTCCATGCCCTTGT
	exon 9	TCTGTATGTGCATGTGTGTATGT	TGACCGAAGCAAACTAAAATATCC
	exon 10	GAGATTCATCCATAGGAGTAGCA	GGGGGGAACTACTTTATGGTAT
	exon 11	CACTCTCCTGAAGCATGTGTGTA	GCAGAGAAACAGACACATTTACAG
	exon 12	CTAGCTTGAGTTTAGTTATGTCCC	GGGTTAAATCAACACACTAACTGG
	exon 13	CCATCCTACATGACAGAGTAAGAC	ATCCAGAAGTAATCAGAGAGGGAG
	exon 14	TGGGCGACAAAGCTAAACTG	TTATGTTGCCCAGGCTGGT
	exon 15	GCCCAGGCATGTGACTTTT	TGGTGTCAAAGAGGCCCAA
	exon 17	TCATGCCTACCAAGTAGTCACAT	TTTGGCTTCTAAGTGAAGGCTG
	exon 18,19	AAGGATGACACACACCTCACTGT	TGGGATAGGACTGAGAGAAGATCA
	exon 20	AAAGTCACCTCCCAGTTCAAAGA	AGTGCAGTGCTGTGATCATCA
	exon 21	AGCTCCTGAGAAGGGCATTT	AACAGCTAGTTCAGGACCTGACAT
	exon 22	ACATGGACATGGTGGAAGGA	AGGGGAGGAATGCAAATAGGAA
	exon 23	ATACATTGGGCATATCTGAGCCT	TCATCTACAAAGTGGGTGGGT
	exon 24	CACAGGGCTCAGCATACAAATC	TGTGCTGGTTCCTGACATACTG
	exon 25	AGCACATCTATACTGAACCACAG	ATTAGAAGCAGTCACCCCACA
	exon 26	TCAGGCCTCTATTTTTGAAGCA	GCAAGTTGGCAGCATTGAAA
	exon 27,28	GTGTTTTGGAAGAGTTGACTCTG	AAGGAAGTGGAGAGTCCTGTAA
	exon 29	TGGCCTAAGGTCACAGAGTTAGT	TACAGGAAGTGCTCAGAAGAGCAT
	exon 30	CCAAGATGGCTTTCTCCTGAATG	CTGCTCATGCATGTAGAGTCTGTT
	exon 31	AGATTGCGGGCTCTTCCTTTT	TCCAAAGGGGTCTATTGAGGA
	exon 32	CATGGCTATCCTTAGTGCTCAGT	GCTCTTGGAAGAAATGTGCCTA
	exon 33	TATTGGGGTGTAAGGGTTGTCT	TGGCTCAAAAACTGACTTCTCC
	exon 34	ATTACCTGGGTATGATGGCACA	TTGCTGCCTACAGGATGATT
	exon 35,36	GACGGGGTTTCACCTTGTTAG	TTTCTGAGAGTCTGAGCAGCAT
	exon 37	ACCTACTCCATCCTTTCTAAGCTG	TCTCTCAAAGTGTCCCTGCAA
	exon 38	TTATATCCAGGCCATAGCGCA	AATTCTGTACGCAGCTTCCCT
	exon 39	AGGCAGGGCTTGAGGTTT	TTTGCCCCTGTGCATTCCT
	exon 40	ACAAAGAGTCCATCAGGTTCCCT	ACAACATGCACTCAAGCCA

Supplemental Table 2. Variants identified by candidate gene sequencing

Gene	Position (NCBI 37.1)		region	position		rs	SNPs	Wild-type	Ped1_IV-4	Ped3_III-1	Ped2_II-1	Ped4_II-1	Ped5_II-1
	Start	End		NT_011387.8	Chr (V37.1)								
PDYN NM_024411.2	1959402	1974702							No variation	No variation	No variation	No variation	No variation
STK35 NM_080836.3	2082528	2129201	exon 1	2022732	2082732	rs6112857	Arg69Gly	CC	GG	GC			
				2022767	2082767	rs6106228	Gln80Gln	GG	AG	AA			
			exon 3	2037688	2097688	rs1891227	Ala423Ala	TT	CT	CC			
TGM3 NM_003245.3	2276613	2321725	intron 12	2255929	2315929	rs2076406	IVSI2+10	GG	AA	AA			
			exon 7	2237790	2297790	rs214814	Ser249Asn	GG	GG	GA			
TGM6 NM_198994.2	2361554	2413399	5' near gene	2301505	2361505	rs9680025	5' near gene	AA	AG	GG	AG	AG	GG
			intron 1	2301684	2361684	rs2422753	IVSI+63	CC	CT	CC	CT	CC	CC
			exon 2	2315262	2375262	rs2076405	M58V	TT	TC	CC	CC	TC	CC
			intron 2	2315440	2375440	rs7266902	IVS2+169	GG	GT	TT	GT	GT	TT
			exon 6	2320323	2380323	rs6114033	Lys263Lys	GG	GG	GA	GA	GG	GG
			intron 6	2320396	2380396	rs2076404	IVS6+12	CC	TT	CT	CT	CT	CT
				2320628	2380628	-	IVS6+242	CC	CC	CC	CT	CC	CC
				2320629	2380629	-	IVS6+243	GG	GA	GG	GG	GG	GG
			intron 8	2324151	2384151	rs6137891	IVS8+5	GG	GA	GA	GA	GA	GA
				2338017	2398017	rs2295077	Lys492Lys	CC	CT	CT	CT	TT	CT
				2351368	2411368	rs2076648	IVSI2+121	CC	CC	CC	CC	CC	CT
			intron 2	2351737	2411737	rs11470465	IVSI2+64	GA/GA	-/GA	-/GA	-/GA	-/GA	GA/GA
				2353125	2413125	rs2076653	IVSI2−11	GG	GA	GA	GA	GA	GG
				2353126	2413126	rs6036467	IVSI2−10	CC	TT	TT	TT	TT	TT
			exon 13	2353320	2413320	rs2076652	3'-UTR	AA	AG	AG	AG	AG	GG
				2353472	2413472	rs45610835	3'-UTR	TT	TT	TT	TT	TT	TG
SNRPB NM_003091.3	2442281	2451499	5' near gene	2391504	2451504	rs6049290	5' near gene	CC	CC	TC			
			5' near gene	2391503	2451503	rs4815262	5' near gene	AA	AA	GA			
			exon 1	2391451	2451451	rs6049288	5'-UTR	GG	GG	TG			
			intron 3	2384665	2444665	rs73606142	IVS3−120	TT	CT	TT			
SNORD119 NR_003684.1	2443598	2443693							No variation	No variation			
ZNF343 NM_024325.4	2462463	2489778	intron 3	2414176	2474176	-	IVS3−11	TT	TT	TC			
			intron 4	2413587	2473587	rs41308639	IVSI−22	AA	AT	AT			
			exon 7	2403921	2463912	-	Leu565Leu	TT	TT	TA			
TMC2 NM_080751.2	2517253	2622430	exon 3	2539387	2479387	rs6050063	Arg123Lys	AA	AG	AG			
			intron 3	2539568	2479568	rs4815320	IVS3+148	TT	TA	TA			
			intron 4	2542669	2482669	rs7270277	IVSI+13	CC	CT	CC			
				2542747	2482747	rs4815323	IVSI+91	GG	GT	GG			
			intron 6	2552926	2492926	rs1883980	IVS6+11	AA	AG	GG			
				2559778	2499778	rs6083735	IVS6−14	GG	GC	CC			
			intron 9	2572816	2512816	rs6050433	IVS9−139	TT	TG	GG			
			intron 13	2591005	2531009	rs1883978	IVSI3−59	TT	GG	GG			
			exon 14	2591232	2531232	rs6050576	Asp527Asp	CC	TT	TT			
			intron 15	2593006	2533006	rs1015159	IVSI5+20	AA	AG	GG			
			exon 16	2593863	2533869	rs6515646	Ser589Ser	TT	TC	TC			
				2594254	2534254	rs6115181	non-coding	AA	AG	AG			
			intron 16	2596762	2536762	rs6050622	IVSI6−21	GG	GA	GA			
			intron 17	2596969	2536969	rs4621228	IVSI7+118	TT	TT	TA			
			exon 18	2597978	2537978	rs4815428	non-coding	GG	GG	GA			
				2598019	2538019	rs6083566	non-coding	CC	CC	CT			
				2598405	2538405	rs13040075	non-coding	GG	GA	GA			
			intron 20	2616556	2556556	rs910271	IVS20−26	TT	TC	CC			
			intron 21	2616679	2556679	rs2422808	IVS21+29	CC	CG	GG			
				2616776	2556776	rs13038659	IVS21+126	TT	TG	TG			
				2618094	2558094	rs6037181	IVS21−26	GG	GG	AA			
			exon 22	2618140	2558140	rs6083915	Ser802Ser	TT	TT	CC			
			intron 22	2618308	2558308	rs6050771	IVS22+71	CC	CC	TT			
			exon 23	2561998	2621998	rs6050798	3'-UTR	TT	TT	CT			
NOP56 NM_006392.2	2633254	2639039	exon 1	2573296	2633296	rs6138678	5'-UTR	GG	GG	CG	GG	GG	GG
			intron 1	2573397	2633397	rs68063608	IVSI−25	g.263397_263403[5]	g.263397_263403[6]+(2300)	g.263397_263403[8]+(2200)	g.263397_263403[6]+(2200)	g.263397_263403[6]+(2000)	g.263397_263403[8]+(1700)
			exon 9	2577071	2637071	rs8958	Thr345Thr	TT	CT	TT	CT	CT	TT
MIR1292 NR_031699.1	2633423	2633488							No variation	No variation	No variation	No variation	No variation
SNORD110 NR_003078.1	2634858	2634932							No variation	No variation			
SNORA51 NR_002981.1	2635713	2635844							No variation	No variation			
SNORD86 NR_004399.1	2636743	2636828							No variation	No variation			
SNORD56 NR_002739.1	2637270	2637340							No variation	No variation			
SNORD57 NR_002738.1	2637585	2637656							No variation	No variation			
IDH3B NM_174856.1	2639041	2644843	intron 2	2584407	2644407	rs2073193	IVS2−3	GG	CC	CC			
EBF4 *NM001110514.1*	2673524	2740754	intron 1	2614174	2674174	rs55820831	IVSI-133	AA	AG	AG			
			intron 2	2617329	2677329	rs874688	IVS2-11	TT	TC	CC			
			exon 3	2617566	2677566	rs2325900	non-coding	TT	TT	TC			
			intron 3	2618202:2618203	2678202:2678203	rs11474226	IVS3+5:8	-/-	-/AAAG	AAAG/AAAG			
			intron 12	2670870	2730870	rs6138883	IVSI2+236	CC	GC	CC			
			intron15	2672934	2732934	rs60014511	IVSI5+49	GG	AG	GG			
			exon17	2676104	2736104	rs13042767	non-coding	GG	CG	GG			
CPXM1 *NM_019609.4*	2774715	2781292	intron 6	2717828	2777828	rs742707	IVS6+10	GG	AG	GG			
C20orf141 *NM_080739.2*	2795657	2735657	exon 1	2736007	2796007	rs12625619	Leu59Leu	GG	AG	AA			
FAM113A NM_022760.3	2815971	2821332	exon 4	2758801	2818801	rs751899	non-coding	TT	TC	CC			
				2758480	2818480	rs2325970	non-coding	AA	AG	GG			
				2757100	2817100	rs78139021	non-coding	TT	TG	TT			
				2756821	2816821	rs2274669	Pro372Pro	CC	CT	CT			

VPS16 NM_022575.2	2821373	2847378	exon 3	2780773	2840773	rs3818605	Ser72Ser	CC	CT	CT	
			intron 10	2782685	2842685	rs6051449	IVS10-15	AA	AC	CC	
			intron 11	2783053	2843053	rs632080	IVS11-49	AA	AG	GG	
			intron 16	2784983	2844983	-	IVS16-28	AA	AC	AC	
			intron 17	2785130	2845130	rs730819	IVS17+22	GG	GA	AA	
PTPRA NM_002836.3	2844841	3019315	intron 3	2785130	2845130	rs730819	IVS3+22	GG	AG	AA	
			intron 26	2936423	2996423	rs1178015	IVS26−73	TT	AA	AA	
			exon 27	2936497	2996497	rs1178016	Gly303Gly	CC	TT	TT	
			intron 27	2936589	2996589	rs1178017	IVS27+58	GG	TT	TT	
			intron 31	2942889	3002889	rs544090	IVS31+23	TT	GG	GG	
				2943027	3003027	rs2277756	IVS31+161	TT	CT	CT	
			intron 37	2957674	3017674	rs561843	IVS31−126	TT	GG	GG	
GNRH2 NM_001501.1	3024268	3026391	exon 2	2965107	3025107	rs6051545	Ala16Val	CC	CT	CT	
			intron 3	2966238	3026238	rs6138994	IVS3-94	GG	GT	GT	
			exon 4	2966359: 2966363	3026359: 3026363	rs71195814	frameshift	-/-	-/CCCCG	-/CCCCG	
				2966415	3026415	rs8184100	3'-UTR	CC	CT	CT	
MRPS26 NM_030811.3	3026675	3028896	inton 2	2967220	3027220	rs2277757	IVS3-40	GG	GA	GA	
OXT NM_000915.2	3052266	3053162							No variation	No variation	
AVP NM_000490.4	3063202	3065370							No variation	No variation	
UBOX5 NM_014948.2	3088219	3140540	exon 7	3030848	3090848	rs708973	Arg510Arg	TT	TT	TG	
FASTKD5 NM_021826.4	3127165	3140532	exon 2	3068403	3128403	rs3746698	Gly438Gly	AA	AG	AA	
ProSAPiP1 NM_014731.2	3143273	3149207	exon 1	3087468	3147468	rs17853865	Asn114Asn	GG	GG	GA	
			intron 1	3087024	3147024	rs7260750	IVS1-13	AA	AA	AG	
DDRKG1 NM_023935.1	3171012	3185295	intron 1	3124134	3184134	rs2295553	IVS1−73	CC	TC	CC	
			intron 4	3116072	3176072	rs7263489	IVS4−73	CC	AC	CC	
			intron 5	3115556	3175556	rs2295549	IVS5−42	CC	GC	CC	
			intron 7	3112246	3172246	rs2295547	IVS7−16	TT	GT	TT	
ITPA NM_033453.2	3190056	3204506	5' near gene	3130039	3190039	rs45620433	5' near gene	CC	CG	CC	
			exon 3	3133978	3193978	rs8362	Gln46Gln	GG	GA	GG	
			intron 5	3144176	3204176	rs75609817	IVS5-17	GG	GA	AA	
			exon 8	3144084	3204084	rs9101	Glu187Glu	GG	GA	GG	
SLC4A11 NM_001174089.1	3208063	3219887	5' near gene	3159894	3219894	rs6107260	5' near gene	TT	TC	CC	
			exon 1	3159698	3219698	-	5'-UTR	GG	GG	GC	
			exon 1	3159692	3219692	rs6084314	5'-UTR	GG	GA	AA	
			exon 2	3158634	3218634	rs3810562	Pro26Arg	CC	CG	GG	
			exon 3	3158355	3218355	rs3827076	non-coding	GG	GC	CC	
			intron 3	3155548	3215548	rs3803957	IVS3-8	GG	GT	GT	
			intron 5	3155088	3215088	rs6133022	IVS5+126	GG	GA	AA	
			exon 6	3154819	3214819	rs3827075	Arg145Arg	AA	AC	CC	
			intron 8	3154126	3214126	rs3803955	IVS8+34	GG	GA	GA	
			intron 10	3151719	3211719	rs3803953	IVS10-15	AA	AC	AC	
			intron 17	3149371	3209371	rs2281575	IVS17-18	CC	CT	CT	
			intron 17	3149357	3209357	rs10048856	IVS17-4	GG	GG	GA	
C20orf194 NM_001009984.1	3229948	3388255	exon 4	3302033	3362033	rs6051818	Val92Val	AA	AG	AG	
			intron 4	3302009	3362009	rs2008730	IVS4+17	AA	AA	AG	
			intron 6	3295567	3355567	rs2208030	IVS6+110	AA	AG	AG	
			exon 11	3264373	3324373	-	Phe265Leu	CC	CG	CC	
			exon 12	3261244	3321244	rs80335486	Thr296Thr	TT	CT	CT	
			intron 19	3239054	3299054	rs6139071	IVS19-7	AA	AG	AG	
			intron 22	3236279	3296279	rs2236115	IVS22+103	GG	AG	AG	
			intron 23	3236428	3296428	-	IVS23-99	-/-	T/-	T/-	
			exon 24	3225140	3285140	rs2422864	Arg577Gly	TT	CC	CC	
			intron 25	3217602	3277602	rs2236104	IVS25-6	CC	CA	CA	
			exon 27	3215239	3275239	rs2281504	Ser685Ser	GG	AG	GG	
			intron 34	3180723	3240723	rs6051685	IVS34-49	CC	CG	CG	
			intron 37	3176554	3236554	rs117680746	IVS37+98	GG	AG	AG	
			exon 38	3175916	3235916	rs2281496	Lys1092Lys	TT	CT	TT	
			intron 40	3173142	3233142	rs6037516	IVS40+76	GG	GC	GC	

Supplemental Table 3. Primers used for repeat-primed PCR, Southern blotting, and RT-PCR.

For repeat-primed PCR

Primer name	Primer sequence
Forward primer	TTTCGGCCTGCGTTCGGG
First reverse primer	TACGCATCCCAGTTTGAGACGCAGGCCCAGGCCCAGGCCCAGGCC
Second reverse primer	TACGCATCCCAGTTTGAGACG

For probes for Southern blot analysis.

Primer name	Primer sequence
Forward primer	TTTAAGAGCTTCCAAGGCTGA
Reverse primer	AGTGCCCACAAGGAAACCGTTA

For quantitation of mouse NOP56 cDNA

Primer name	Primer sequence
mouse NOP56 F	GTTGGCGCTGAAGGAAGTGG
mouse NOP56 R	CTTTGGCACGAGAGTAGCTG

For quantitation of human NOP56 cDNA

Primer name	Primer sequence
human NOP56 cex4F	TTGCCTTGGAAAATGCCAAC
human NOP56 cex6R	TGTATTGCGGCACCAATCTT

For investigation of human NOP56 cDNA splicing variants

Primer name	Primer sequence
human NOP56 cex1F	TAGCCGCATTGCGAGCCGAA
human NOP56 cex4R	GTTGCCTTGGAAAATGCCAA

AJHG Conflict of Interest Form

Please complete this form electronically and upload the file with your final submission.

AJHG requires all authors to disclose any financial interest that might be construed to influence the results or interpretation of their manuscript.

As a guideline, any affiliation associated with a payment or financial benefit exceeding $10,000 p.a. or 5% ownership of a company or research funding by a company with related interests would constitute a financial interest that must be declared. This policy applies to all submitted research manuscripts and review material.

Examples of statement language include: AUTHOR is an employee and shareholder of COMPANY; AUTHOR is a founder of COMPANY and a member of its scientific advisory board. This work was supported in part by a grant from COMPANY.

Please disclose any such interest below on behalf of all authors of this manuscript.

Please check one of the following:

☒ None of the authors of this manuscript have a financial interest related to this work.

☐ Please print the following Disclosure Statement in the Acknowledgments section:

Please provide the following information:

☒ Please check this box to indicate that you have asked every author of this work to declare any conflicts of interest. Your answers on this form are on behalf of every author of this work.

Manuscript #: AJHG-D-11-00147

Article Title: Expansion of Intronic GGCCTG Hexanucleotide Repeat in NOP56 Causes a Type of Spinocerebellar Ataxia (SCA36) Accompanied by Motor Neuron Involvement

Author List: Hatasu Kobayashi, Koji Abe, Tohru Matsuura, Yoshio Ikeda, Toshiaki Hitomi, Yuji Akechi, Toshiyuki Habu, Liu Wanyang, Hiroko Okuda and Akio Koizumi

Your Name: Akio Koizumi

Date: 2011/5/15