[Title]

Prevalence of repeat expansions causing autosomal dominant spinocerebellar ataxias in Hokkaido, the northernmost island of Japan

[Short running title]

AD-SCD in Hokkaido

[Authors]

Keiichi Mizushima, MD [1 #], Yuka Shibata, MD. PhD[2 #], Shinichi Shirai, MD. PhD [1], Masaaki Matsushima, MD. PhD [1,2], Satoko Miyatake, MD. PhD [3,4], Ikuko Iwata, MD. PhD [1]. Hiroaki Yaguchi, MD. PhD [1], Naomichi Matsumoto, MD. PhD [3], Ichiro Yabe, MD. PhD [1,2*]

1. Department of Neurology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan

2. Division of Clinical Genetics, Hokkaido University Hospital, Sapporo, Japan

3. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan

4. Department of Clinical Genetics, Yokohama City University Hospital, Yokohama 236-0004, Japan

Equality to work

*Corresponding author: Ichiro Yabe

Department of Neurology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, N-15 W-7, Kita-ku, Sapporo, Japan

Tel: +81-11-706-6028

Fax: +81-11-700-5356

E-mail; yabe@med.hokudai.ac.jp

Manuscript word count: 1905 words

Tables: 0,

Figure: 4,

Supplementary Figure: 2

[Conflict of Interest]

None declared.

Funding: This work was supported by Ministry of Health, Labor and Welfare - 23FC1010.

Author contributions

KM, YS, SS, MM, II, HY and IY contributed to the conception and design of the study. KM, YS, SS, MM, SM, II, HY, NM and IY contributed to the acquisition and analysis of data. KM, YS, SS, MM, SM, II, HY, NM and IY contributed to drafting the text or preparing the figures.

Abstract (207/250 words)

In Japan, approximately 30% of spinocerebellar degeneration (SCD) is hereditary, and more than 90% of hereditary SCD is autosomal dominant SCD (AD-SCD). We have previously reported the types of AD-SCD in Hokkaido, twice. In this study, we investigated the status of AD-SCD mainly due to repeat expansions, covering the period since the last report. We performed genetic analysis for 312 patients with a clinical diagnosis of SCD except for multiple system atrophy at medical institutions in Hokkaido between January 2007 and December 2020. The median age at the time of analysis was 58 (1-86) years. Pathogenic variants causing AD-SCD due to repeat expansion were found in 61.5% (192 cases). Spinocerebellar ataxia (SCA) 6 was the most common type in 25.3% (79 cases), followed by Machado-Joseph disease (MJD) / SCA3 in 13.8% (43), SCA1 in 6.4% (20), SCA2 in 5.1% (16), SCA31 in 4.8% (15), dentatorubral-pallidoluysian atrophy in 4.8% (15), SCA7 in 0.6% (2), and SCA8 in 0.6% (2). SCA17, 27B, 36, and 37 were not found. Compared to previous reports, this study found higher prevalence of SCA6 and lower prevalence of MJD/SCA3. An increasing number of cases identified by genetic testing, including cases with no apparent family history, accurately revealed the distribution of disease types in Hokkaido.

Introduction (/5000 words)

Spinocerebellar degeneration (SCD) is a general term for neurodegenerative disorders whose main symptom is cerebellar or spinal ataxia. About 30% of SCD is hereditary, and about 27% of hereditary SCD is of the autosomal dominant type. Autosomal dominant cerebellar ataxias (ADCAs) are classified into three types [1]. ADCA type 1 is a non-pure cerebellar type, which is associated with variable additional signs, e.g., ophthalmoplegia, dementia, extrapyramidal signs, optic atrophy, and amyotrophy. Typical types include spinocerebellar ataxia (SCA) 1, 2, 3 (Machado-Joseph disease: MJD), 17, 28, and dentatorubral-pallidoluysian atrophy (DRPLA). ADCA type 2 is associated with retinal symptoms with cerebellar symptoms and classified as SCA7. Lastly, ADCA type 3 is a pure cerebellar type, characterized mainly by cerebellar ataxia and slow progression. Typical types include SCA5, 6, 31, 36, 37, and 42 [2]. With regard to the forms of SCA described above and except for a few non-repeat diseases such as SCA14 and SCA42, the majority of causative genetic mutations are repeat expansions. The distribution of disease types varies regionally [3]. For example, it has been reported that SCA1 is more common in Italy, South Africa, and Australia; SCA2 in Italy, India, Mexico, and Cuba; and MJD in France, the United States, Portugal, Brazil, and China [4]. In Japan, the most common type is MJD, followed by SCA6, DRPLA, SCA31, and SCA1; but there are regional differences even in Japan [5,6]. We have previously reported the types of autosomal dominant spinocerebellar degeneration (AD-SCD) twice [7,8], and suggested that SCA6 is the most common type in Hokkaido, the northernmost island of Japan, whereas DRPLA is less common than in other regions. Since these reports, the number

of genetic analyses has increased, and new disease types have been reported one after another. For example, SCA27B was reported to be caused by a GAA repeat expansion deep in the first intron of *FGF14* gene in 2023 [9,10]. In this study, we report the current status of AD-SCD due only to repeat expansions, which are the main cause of AD-SCD in Hokkaido. In Japan, because genetic testing for eight types of AD-SCD, SCA1, 2, MJD, 6, 17, DRPLA, 8, and 31, have been covered by public health insurance since 2020, it would be difficult to conduct a study like this study again.

Materials and Methods

Study participants

The participants were patients with suspected SCD, but not multiple system atrophy (MSA), at medical institutions in Hokkaido from January 2007 to December 2020. This clinical study was approved by the Ethics Committee of School of Medicine, Hokkaido University (protocol number: 16-032).

Genetic analysis

Genomic DNA was extracted from blood lymphocytes and amplified using known primers. Conditions for polymerase chain reaction (PCR) amplification were the same as in previous studies [7,8]. All the PCR products were separated by electrophoresis in 1～3% agarose. Genotypes were determined using computer software (GeneScan® Analysis Ver.2.0, Perkin Elmer), as described

previously [11].

SCA1, 2, 3, 6, 7, 17, 31, and DRPLA, which are major polyglutamine diseases, were detected by agarose gel electrophoresis which was used to evaluate repeat number expansion, and if positive, fragment analysis and Sanger sequencing were performed to confirm repeat numbers. SCA8 and 27B were detected by fragment analysis and repeat primed PCR (RP-PCR) after electrophoresis. For SCA27B, on suspected cases detectable by RP-PCR, the diagnosis was confirmed by performing targeted long-read sequencing at the Department of Human Genetics, Yokohama City University Graduate School of Medicine. Targeted long-read sequencing uses the adaptive sampling option on the nanopore GridION sequencer (Oxford Nanopore Technologies, Oxford, UK) with the time-lag method as described previously [12], and with a browser extensible data file targeting the *FGF14* repeat locus and surrounding region as previously reported [13]. Only those with *FGF14* GAA expansion of 250 repeats or more were considered significant as previously reported [9]. In cases of SCA36, 37, if only one band appeared on electrophoresis, RP-PCR was performed.

Statistical analysis

We used JMP® Pro 16.2.0 as statistical software. For the investigation of family history and age at analysis, the Wilcoxon rank sum test was performed. For the investigation of clinical presentation and family history, the Fisher exact test was performed.

Ethical considerations

This study was conducted with approval of the Hokkaido University Hospital Institutional Review Board. Written consent to participate in this study was obtained from all participants.

Results

Distribution of disease types

A total of 312 patients were participants in our study. In all, 58.3% (182 cases) were cases in our hospital and 41.7% (130 cases) were cases in other hospitals in Hokkaido. The median age at time of analysis was 58 (1-86) years and the proportion of females was 54.8%. Pathogenic variants causing repeat diseases were found in 61.5% (192 cases). SCA6 was the most common type in 25.3% (79 cases), followed by MJD / SCA3 in 13.8% (43 cases), SCA1 in 6.4% (20 cases), SCA2 in 5.1% (16 cases), SCA31 in 4.8% (15 cases), DRPLA in 4.8% (15 cases), SCA7 in 0.6% (2 cases), and SCA8 in 0.6% (2 cases) (Figure 1). SCA17, 27B, 36, and 37 were not found. The negative correlation between the number of repeats and age at time of onset was suggested for the major disease types similar to previous reports [14] (Figure 2).

Next, we focused on family history and analyzed cases with family history. In our hospital, 65.4% (119/182 cases) had a family history. Pathogenic variants occurred in 79.8% (95/119 cases) with family history and 25.4% (15/59 cases) without family history but suspected AD-SCD based on phenotypes (Figure 3). Comparing the age at the time of analysis of cases with pathogenic variants with or without

family history, the median age was 57 (17-84) years in cases with family history and 67 (25-80) years in cases without, indicating that the age was significantly higher in cases without family history ($p = 0.01$) (Figure 4). Compared with cases classified as non-pure types, cases of pure types tended to have no family history ($p = 0.17$).

Discussion

There were three important points to note about this study. First, this study is the third report on AD-SCD, following those in 2000 and 2007, and represents a continuation of earlier research on AD-SCA in Hokkaido. Second, we have been investigating AD-SCD for 40 years and have performed genetic analysis in 847 patients with a clinical diagnosis of SCD. Third, there was a higher percentage of pure cerebellar types, such as SCA6, compared to our previous study.

As the majority of causative genetic variants in AD-SCD are repeat expansions, in this study we only investigated the distribution of repeat disease, excluding AD-SCD caused by other variants. Compared to our previous studies [7, 8], this study found the percentage of SCA6 increased from 20.3% to 25.3%, while that of MJD decreased from 29.8% to 24.1% and then to 13.8% and SCA7 was newly detected. SCA6 was the most frequent AD-SCD, while MJD had been reported to be the most frequent type in Japan as mentioned above [5]. One reason for this difference may be that many previous studies included only cases with family history, and it is possible that incidence of pure cerebellar types such as SCA6, for which there is no clear family history, has been underestimated. In fact, pathogenic variants were

detected in 25.4% of cases without family history in this study. In addition, cases without family history tended to be significantly older at the time of analysis and were more frequently classified as pure cerebellar type. A possible reason for the lack of apparent family history is the presence of relatives who die before diagnosis because of the later onset type. It has been reported that MJD is more frequent in cases with family history, while SCA6 is more frequent in cases without family history [6,15]. On the other hand, there are also reports that SCA6 is more frequent when cases with family history are targeted [16], no consensus has been reached. Moreover, a potential bias may be introduced when physicians choose not to perform genetic testing on affected individuals already identified within the patient's family, and patients with non-pure cerebellar type SCD who avoid marriage, pregnancy, and childbirth due to having knowledge of disease progression in affected individuals of the family. To elucidate these possibilities, further studies, including interviews, are needed. Since 2008, no national survey to detect AD-SCD cases has been conducted in Japan. Genetic testing should be performed when AD-SCD is suspected based on phenotype even in the absence of family history, and a nationwide survey should be planned again.

SCA27B, a type of AD-SCD reported in 2023, is caused by a deep intronic GAA repeat expansion in the *FGF14* gene [9]. In the original study, cases with *FGF14* GAA expansions of 250 or more repeats were considered SCA27B cases. In this study, the same conditions were considered and no cases were applicable. However, in two cases, between 200 and 250 GAA repeats were observed in this study (Supplement, Supplementary figure). These cases presenting with pure cerebellar ataxia and gaze-

evoked horizontal nystagmus are similar to the cases in the original study. A resent study reported that a case with *FGF14* GAA expansions of 234 repeats presented with slowly progressive cerebellar ataxia, and her child was diagnosed with SCA27B with 256 GAA repeats [17]. Therefore, it may be pathogenic in cases of GAA expansions below 250 repeats, which is a subject for further study.

In 36.2% of the patients, no pathogenic variant in the target genes was observed even in familial cases. There are several reasons for this: it is a new type of AD-SCD, it is an SCD with a different type of inheritance pattern, there are issues with the analysis technique. Recently, whole exome/whole genome sequencing technologies, including long-read sequencing, have been performed for SCA, and it is expected that more variants will be discovered in the future.

In conclusion, the percentage of pure cerebellar types, such as SCA6, was higher compared to our previous studies [7,8]. For more than 40 years, we have continuously performed genetic analysis of spinocerebellar degeneration, which revealed that genetic testing in cases of AD-SCD, including cases without apparent family history, may more accurately reveal the distribution of disease types in Hokkaido, Japan. Despite some newly reported forms of the disease, there is still a significant number of cases of AD-SCD for which pathogenic variants have not yet been found. Continuous improvement of genetic analysis research is essential.

Study limitations

First, candidates for participation in this study were selected by attending physicians. Therefore, it is possible that cases in which pathogenic variants were already recognized in their relatives were not included in this study because genetic analysis is less important for diagnosis. Second, new types of AD-SCD were reported one after another, and new types were analyzed more carefully. Therefore, detection bias may have influenced the results. Third, no analysis of AD-SCDs not caused by repeat expansions was carried out in this study. However, as SCA14 and SCA42 had been reported in Japan in the past, investigations of them were also conducted. Finally, in Japan, genetic testing for eight types of AD-SCD, SCA1, 2, MJD, 6, 17, DRPLA, 8, and 31 have been covered by public health insurance from 2020, making it easier to perform genetic analysis. It may increase the number of patients with genetic diagnoses, but their data cannot be accessed because of personal information protection. So, this is the final report concerning the distribution of SCA types in Hokkaido, based on genetic examinations conducted in our laboratory.

Acknowledgements

We thank all patients for their cooperation. We thank Rie Nomachi for data processing. We would like to thank the neurologists of the Department of Neurology, Hokkaido University for their attentive care of the patients during the period of this study.

This work was supported by the Japan Agency for Medical Research and Development (AMED) under grant numbers JP23ek0109527h0003 (I. Yabe), JP23ek0109674, JP23ek0109549, JP23ek0109617,

JP23ek0109648 (N. Matsumoto); JSPS KAKENHI under grant numbers JP23H02829 (S.Miyatake); the Takeda Science Foundation (N. Matsumoto), and by a Grant from the Research Committee on the Medical Basis of Motor Ataxias, Health and Labor Sciences Research Grants, The Ministry of Health, Labor and Welfare, Japan under grant numbers 23FC1010 (to I.Yabe.)

Conflicts of interest

None

References

1. Harding AE. Clinical features and classification of inherited ataxias. Adv Neurol. 1993;61:1-14.

2. Dueñas AM, Goold R, Giunti P. Molecular pathogenesis of spinocerebellar ataxias. Brain. 2006;129:1357-70.

3. Susan Perlman MD. Hereditary Ataxia Overview. GeneReviews ® [Internet]. Initial posting: 1998, Last Update: 2022.

4. Sun YM, Lu C, Wu ZY. Spinocerebellar ataxia: relationship between phenotype and genotype - a review. Clin Genet. 2016;90:305-14.

5. Tsuji S, Onodera O, Goto J, Nishizawa M; Study Group on Ataxic Diseases. Sporadic ataxias in Japan--a population-based epidemiological study. Cerebellum. 2008;7:189-97.

6. Maruyama H, Izumi Y, Morino H, et al. Difference in disease-free survival curve and regional

distribution according to subtype of spinocerebellar ataxia: a study of 1,286 Japanese patients. Am J Med Genet. 2002;114:578-83.

7. Sasaki H, Yabe I, Yamashita I, Tashiro K. Prevalence of triplet repeat expansion in ataxia patients from Hokkaido, the northernmost island of Japan. J Neurol Sci. 2000;175:45-51.

8. Basri R, Yabe I, Soma H, Sasaki H. Spectrum and prevalence of autosomal dominant spinocerebellar ataxia in Hokkaido, the northern island of Japan: a study of 113 Japanese families. J Hum Genet. 2007;52:848-55.

9. Pellerin D, Danzi MC, Wilke C, Renaud M, Fazal S, Dicaire MJ, et al. Deep Intronic *FGF14* GAA Repeat Expansion in Late-Onset Cerebellar Ataxia. N Engl J Med. 2023;388:128-41.

10. Rafehi H, Read J, Szmulewicz DJ, Davies KC, et al. An intronic GAA repeat expansion in *FGF14* causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14. Am J Hum Genet. 2023;110:105-19.

11. Yabe I, Sasaki H, Matsuura T, Takada A, Wakisaka A, Suzuki Y, et al. SCA6 mutation analysis in a large cohort of the Japanese patients with late-onset pure cerebellar ataxia. J Neurol Sci. 1998;156:89-95.

12. Miyatake S, Koshimizu E, Fujita A, Doi H, et al. Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. NPJ Genom Med. 2022;7:62.

13. Rafehi H, Read J, Szmulewicz DJ, et al. An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA27B/ATX-FGF14. Am J Hum Genet. 2023;110:1018.

14. Stevanin G, Dürr A, Brice A. Clinical and molecular advances in autosomal dominant cerebellar ataxias: from genotype to phenotype and physiopathology. Eur J Hum Genet. 2000;8:4-18.

15. Onodera Y, Aoki M, Mizuno H, Warita H, Shiga Y, Itoyama Y. Clinical features of chromosome 16q22.1 linked autosomal dominant cerebellar ataxia in Japanese. Neurology. 2006;67:1300-2.

16. Nozaki H, Ikeuchi T, Kawakami A, Kimura A, Koide R, Tsuchiya M, et al. Clinical and genetic characterizations of 16q-linked autosomal dominant spinocerebellar ataxia (AD-SCA) and frequency analysis of AD-SCA in the Japanese population. Mov Disord. 2007;22:857-62.

17. Hengel H, Pellerin D, Wilke C, et al. As Frequent as Polyglutamine Spinocerebellar Ataxias: SCA27B in a Large German Autosomal Dominant Ataxia Cohort. Mov Disord. 2023;38:1557-8.

Titles and legends to figures

Figure 1. Distribution of disease types

The distribution of disease types is in 312 patients with suspected SCD but not MSA at medical institutions in Hokkaido. The data collected from April 1994 to December 2006,originally from familial cases [8], was converted to a case-based format and reanalyzed. The proportion of MJD/SCA3 is gradually decreasing, while that of SCA6 is increasing, and there are new diagnoses of SCA31, resulting in an increased proportion of pure cerebellar type cases.

Figure 2. Correlation between age of onset and number of repeats

The number of repeats and age of onset in each case are described. The number of repeats refers to the number in alleles with repeat expansion. The disease types are represented by the following symbols. Circle: SCA6, Square: SCA2, Cross: SCA1, Triangle: DRPLA, Diamond: MJD.

Figure 3. Percentage of pathogenic variants by family history

The figure shows the percentage of cases in which a pathogenic variant was detected. The cases have been divided into two groups: one with family history and one without.

Figure 4. Age at the time of analysis by family history

The age distribution at time of analysis of cases is shown for the group with a family history and group without.

Tables

None

Supplementary information

Supplement. Information about cases with *FGF14* GAA expansion between 200 and 250 repeats.

Supplementary figure. Image findings in cases with *FGF14* GAA expansion between 200 and 250 repeats

Figure 1. Distribution of disease types

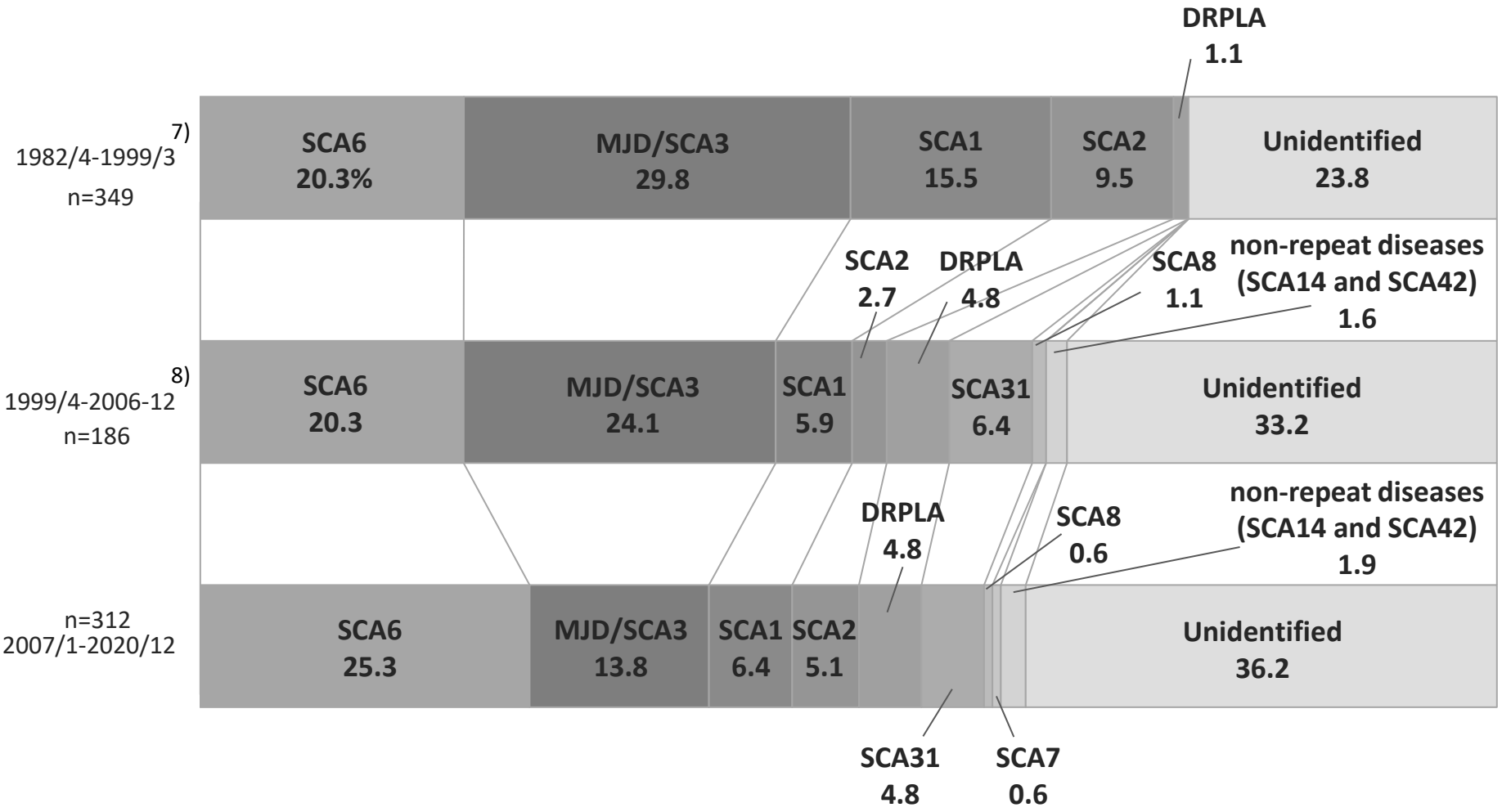

Figure 2. Correlations between age of onset and number of repeats

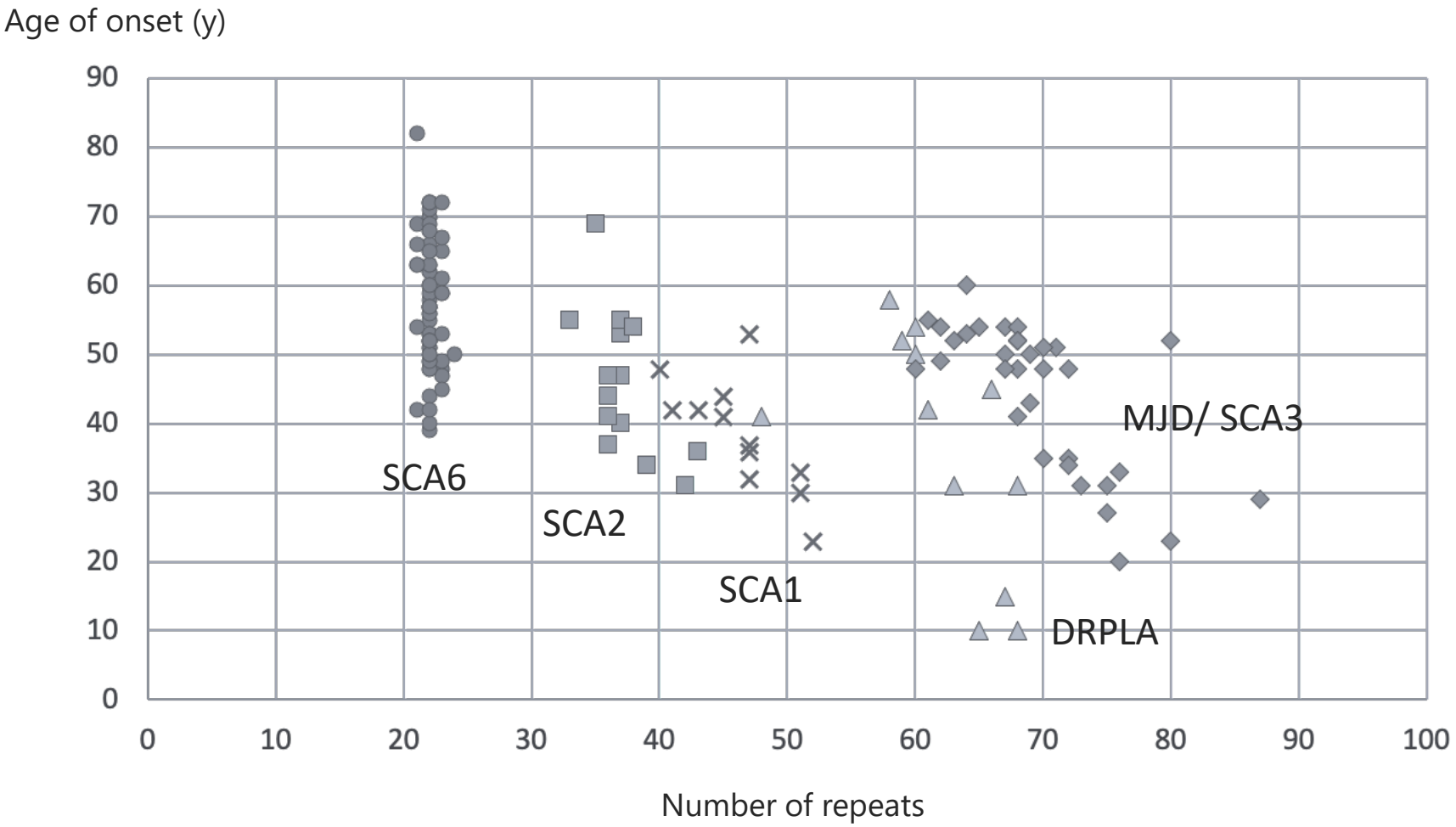

Figure 3 (clean). Percentage of pathogenic variants by family history

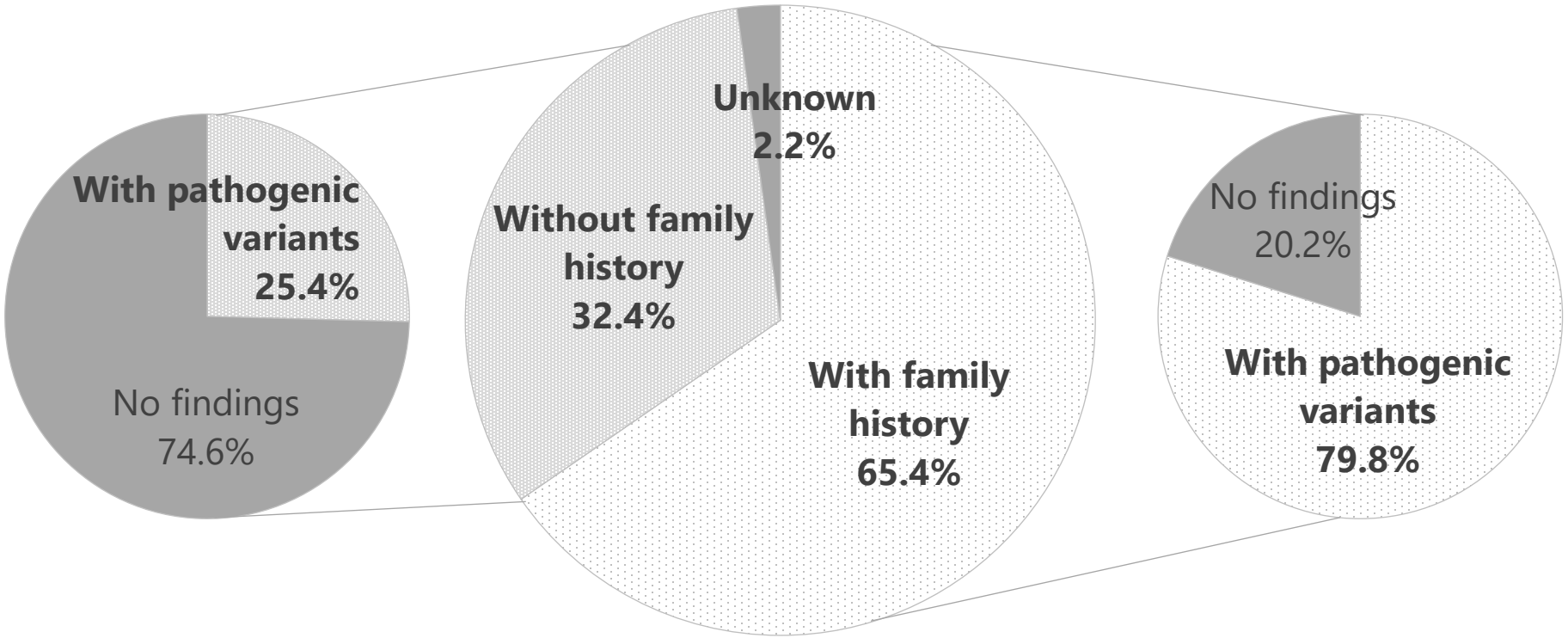

Figure 4. Age at the time of analysis by family history