Parkinsonism and Related Disorders 63 (2019) 42–45

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Point of view

A significant inflation in *TGM6* genetic risk casts doubt in its causation in spinocerebellar ataxia type 35

Jasmine L.F. Fung, Mandy H.Y. Tsang, Gordon K.C. Leung, Kit San Yeung, Christopher C.Y. Mak, Cheuk Wing Fung, Sophelia H.S. Chan, Mullin H.C. Yu**, Brian H.Y. Chung*

Department of Paediatric and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Spinocerebellar ataxia
TGM6
Clinical genetics
Whole exome sequencing
Inflation analysis

ABSTRACT

Spinocerebellar ataxia 35 (SCA35) has been associated with pathogenic mutations in the gene *TGM6*. In a Chinese exome sequencing cohort, we identified 8 families with reported *TGM6* variants sharing no features of SCA35. Considering this finding, we reviewed the public database gnomAD and found these variants to be significantly more common in the East Asians than in other ethnic groups ($P < 0.0001$). Gene constraint metrics showed that both missense and loss-of-function variants in *TGM6* are likely to be tolerated and there is no regional constraint. By performing inflation analysis, it demonstrated that the cumulative frequency of *TGM6* reported pathogenic variants is at least 111-fold inflated over disease prevalence of all autosomal dominant SCAs, indicating a high chance of misdiagnosis or low penetrance. Misclassification of benign or low penetrant variants as pathogenic is a significant problem that often results in genetic misdiagnosis. This highlights the necessity of evaluating variant pathogenicity with sequencing of genomes from diverse populations, both from asymptomatic controls and phenotypically different patients, in order to ensure accurate classification of variants.

An autosomal dominant type of spinocerebellar ataxia (SCA), SCA35, was first associated with mutations in *TGM6* in 2010, with 9 affected members in a four-generation Chinese family [1]. Subsequently, more unrelated affected individuals were identified. SCA35 is characterized by a slow, progressive course of trunk/limb ataxia and hand tremors. Age of onset varies from adolescence to the fifth decade. We have identified several individuals in our exome sequencing cohort with reported pathogenic *TGM6* variants but no clinical features of SCA35. Given these patients, we are questioning the pathogenicity of *TGM6* in SCA.

We reviewed our HKU undiagnosed disease exome sequencing cohort to identify (i) reported pathogenic *TGM6* variants and (ii) rare and predicted damaging *TGM6* variants which are defined as having an allele frequency of < 0.01 and CADD score ≥ 20. Variants with CADD score ≥ 20 is predicted to be the top 1% deleterious [2]. Our exome sequencing cohort consists of individuals with suspected pediatric-onset genetic diseases (including but not limited to neurological disorders, $n = 116$) and some of their asymptomatic parents ($n = 135$). We found 12 index patients (age: 0–36; 5 males and 7 females) and 13 asymptomatic parents (age: 35–71; 7 males and 6 females) with *TGM6* variants meeting these criteria (Table 1a). Among the 12 indexes, seven were found to have reported pathogenic *TGM6* mutations but they did not manifest clinical presentation of SCA35. Instead, these patients harbor a heterogeneous clinical presentations, including congenital myopathy (c.7+1G > T), suspected mitochondrial disorder (V391 M), maturity onset diabetes of the young (V391 M), infantile hypertrophic cardiomyopathy (V391 M), childhood onset dystonia and developmental delay (D510H), childhood onset spastic paraplegia (L517W) and recurrent febrile convulsions (L517W). Remarkably, mutations in these seven children were all inherited from their asymptomatic parents. An additional V391 M variant was identified in a healthy parent but was not found his child.

Majority of the previously reported subjects with *TGM6* mutations presented with SCA35 are of Chinese ancestry [1,3–10]. These may be a result of ascertainment bias. Using the publicly accessible population database gnomAD (v2.1) containing 141,456 exomes/genomes (East Asians = 9977), we found that the four reported pathogenic mutations (c.7+1G > T, V391 M, D510H and L517W) in *TGM6* are significantly more common in East Asians (allele frequency of 0.0019–0.0094) than in other ethnic groups ($P < 0.0001$). Further, bioinformatics analysis

* Corresponding author.
** Corresponding author.
E-mail addresses: mullinyu@connect.hku.hk (M.H.C. Yu), bhychung@hku.hk (B.H.Y. Chung).

https://doi.org/10.1016/j.parkreldis.2019.01.013
Received 15 August 2018; Received in revised form 28 November 2018; Accepted 13 January 2019
1353-8020/

Table 1a
Individuals with rare and predicted damaging *TGM6* variants in HKU undiagnosed disease exome sequencing cohort (N_{index} = 12; $N_{parents}$ = 13).

Family	Variant	Clinical diagnosis of the index patient	Variant identified in			CADD[b]	Allelic frequency[b] in	
			Index (sex/age)	Father (age)	Mother (age)		East Asian	other population
1	c.7+1G > T	Congenital myopathy	✓ (F/36)	✓ (71)	–	22.7	1.91×10^{-3}	0
2	c.631C > T, p.(R211C)	Pancytopenia	–	–	✓ (35)	33	2.01×10^{-4}	1.29×10^{-4}
3	c.940G > A, p.(V314 M)	Noonan syndrome	✓ (M/0d)	✓ (N/A)	–	24.2	5.36×10^{-3}	3.42×10^{-5}
4	c.1040C > T, p.(P347L)	Leigh's syndrome	✓ (M/3)	N/A	N/A	25.2	6.72×10^{-3}	6.85×10^{-5}
5	c.1040C > T, p.(P347L)	Premature aging	✓ (M/21)	–	✓ (57)	25.2	6.72×10^{-3}	6.85×10^{-5}
6	c.1040C > T, p.(P347L)	Hypotonia, severe global developmental delay	✓[a] (M/3 m)	N/A	N/A	25.2	6.72×10^{-3}	6.85×10^{-5}
7	c.1154C > T,p.(A385V)	Fetal akinesia	✓ (F/8d)	–	✓ (36)	26.4	0	8.58×10^{-6}
8	c.1171G > A, p.(V391 M)	Maturity onset diabetes of the young (MODY)	✓ (F/18)	✓ (59)	–	31	9.42×10^{-3}	1.26×10^{-4}
9	c.1171G > A, p.(V391 M)	Mitochondrial disorder	✓ (F/3y)	✓ (38)	–	31	9.42×10^{-3}	1.26×10^{-4}
10	c.1171G > A, p.(V391 M)	Noonan syndrome	–	✓ (43)	–	31	9.42×10^{-3}	1.26×10^{-4}
11	c.1171G > A, p.(V391 M)	Infantile hypertrophic cardiomyopathy	✓ (F/86d)	✓ (37)	–	31	9.42×10^{-3}	1.26×10^{-4}
12	c.1528G > C, p.(D510H)	Dystonia, global developmental delay	✓ (F/7)	–	✓ (41)	25.9	2.11×10^{-3}	0
13	c.1550T > G, p.(L517W)	Spastic paraplegia	✓ (M/28)	–	✓ (51)	26.9	2.01×10^{-3}	3.85×10^{-6}
14	c.1550T > G, p.(L517W)	Recurrent febrile convulsion	✓ (F/34 m)	–	✓ (35)	26.9	2.01×10^{-3}	3.85×10^{-6}
15	c.2009T > C, p.(F670S)	Mitochondrial disorder	–	✓ (60)	–	29.6	0	0

[a] Variant found in homozygous state.
[b] Rare variants are defined as variants with allelic frequency < 0.01 in publicly available database (gnomAD); and predicted damaging variants as loss-of-function mutations or rare missense mutation with a CADD score ≥20. Variants with CADD score ≥20 are predicted to be the top 1% deleterious in human genome [2].

Table 1b
Inflation analysis of the 16 pathogenic/likely pathogenic *TGM6* variants reported in literature and ClinVar (as per November 2018).

Variant	Reference	Variant classification	Found in our exome cohort?	CADD	Allelic count in East Asians [a]	Allelic count in all populations [a]
c.7+1G > T	Yang et al., 2018 [6]	Pathogenic	Yes	26.1	37	37
c.115A > T, p.(Ser39Cys)	Farwell et al., 2015 [7]	Pathogenic	No	14.13	0	120
c.331C > T, p.(Arg111Cys)	Guo et al., 2014 [3]; ClinVar [RCV000170474.3]	Pathogenic	No	24.7	4	7
c.543G > T, (p.Gln181His)	Tripathy et al., 2017 [10]	Pathogenic	No	33	0	0
c.841delC, p.(Leu281fs)	Lin et al., 2018 [9]; ClinVar [RCV000516766.1]	Likely Pathogenic	No	N/A	0	0
c.842_843insC, p.(Cys282fs)	ClinVar [RCV000658907.1]	Likely Pathogenic	No	N/A	0	0
c.844_850+3del	ClinVar [RCV000658908.1]	Likely Pathogenic	No	N/A	0	0
c.980A > G, p.(Asp327Gly)	Wang et al., 2010 [1]; Guan et al., 2013 [12]; ClinVar [RCV000024081.2]	Pathogenic	No	27.4	0	24
c.1171G > A. p.(Val391Met)	Tripathy et al., 2017 [10]	Pathogenic	Yes	27.1	188	221
c.1322A > G, p.(Tyr441Cys)	Tripathy et al., 2017 [10]	Pathogenic	No	27.9	0	3
c.1478C > T, p.(Pro493Leu)	Choi et al., 2017 [8]; Yang et al., 2018 [6]	Pathogenic	No	24.8	19	20
c.1501T > A, p.(Leu502Gln)	Tripathy et al., 2017 [10]	Pathogenic	No	24	0	2
c.1528G > C, p.(Asp510His)	Li et al., 2013 [4]; Guo et al., 2014 [3]; ClinVar [RCV000034865.4]	Pathogenic	Yes	24.1	42	42
c.1550T > G, p.(Leu517Trp)	Wang et al., 2010 [1]; Guan et al., 2013 [12]; Pan et al., 2015 [5]; ClinVar [RCV000024080.2; RCV000077795.1]	Pathogenic	Yes	26.9	40	41
c.1722_1724delAGA, p.(Glu574del)	Guo et al., 2014 [3]; ClinVar [RCV000170475.3]	Pathogenic	No	N/A	1	1
c.1951_1952insAAC, p.(Gln652dup)	Tripathy et al., 2017 [10]	Pathogenic	No	N/A	0	361
			Total Count:		331	879
			Genetic risk[b]:		0.033176306	0.006213946
			Inflation[c]:		592.4340268	110.9633283

Note: Reported pathogenic mutation with population frequency > 0.01 were excluded in the inflation analysis.
[a] Data is obtained from publicly available gnomAD database with $N_{EastAsians}$ = 9977 and $N_{AllPopulation}$ = 141,456.
[b] Calculated using total count divided by $N_{EastAsians}$ and $N_{AllPopulation}$ respectively.
[c] Inflation = Observed genetic risk/Disease prevalence (5.6×10^{-5}).

showed that variation in *TGM6* gene is likely to be tolerated. This is supported by the low constraint metrics of loss-of-function (LoF) o/e = 1.02 (90%CI 0.78–1.34) and missense o/e = 1.09 (90%CI 1.01–1.18) in gnomAD [11]. o/e score shows the observed over expected number of variants in the population, lower o/e value shows stronger intolerance to variation. An o/e value > 1 means the observed number of variants exceeds the expected value. Both missense and LoF variants in the gene are unlikely to be disease-causing. There is no evidence that certain mutation classes are more pathogenic than the other. In addition, by reviewing the CADD score of *TGM6* variants in GnomAD, one third (420 out of 1280) have CADD score ≥20. These variants spread across the gene, a high CADD score may have little predictive value of the pathogenicity of *TGM6* variant.

Earlier *in vitro* and *in vivo* studies illustrated that *TGM6* mutations

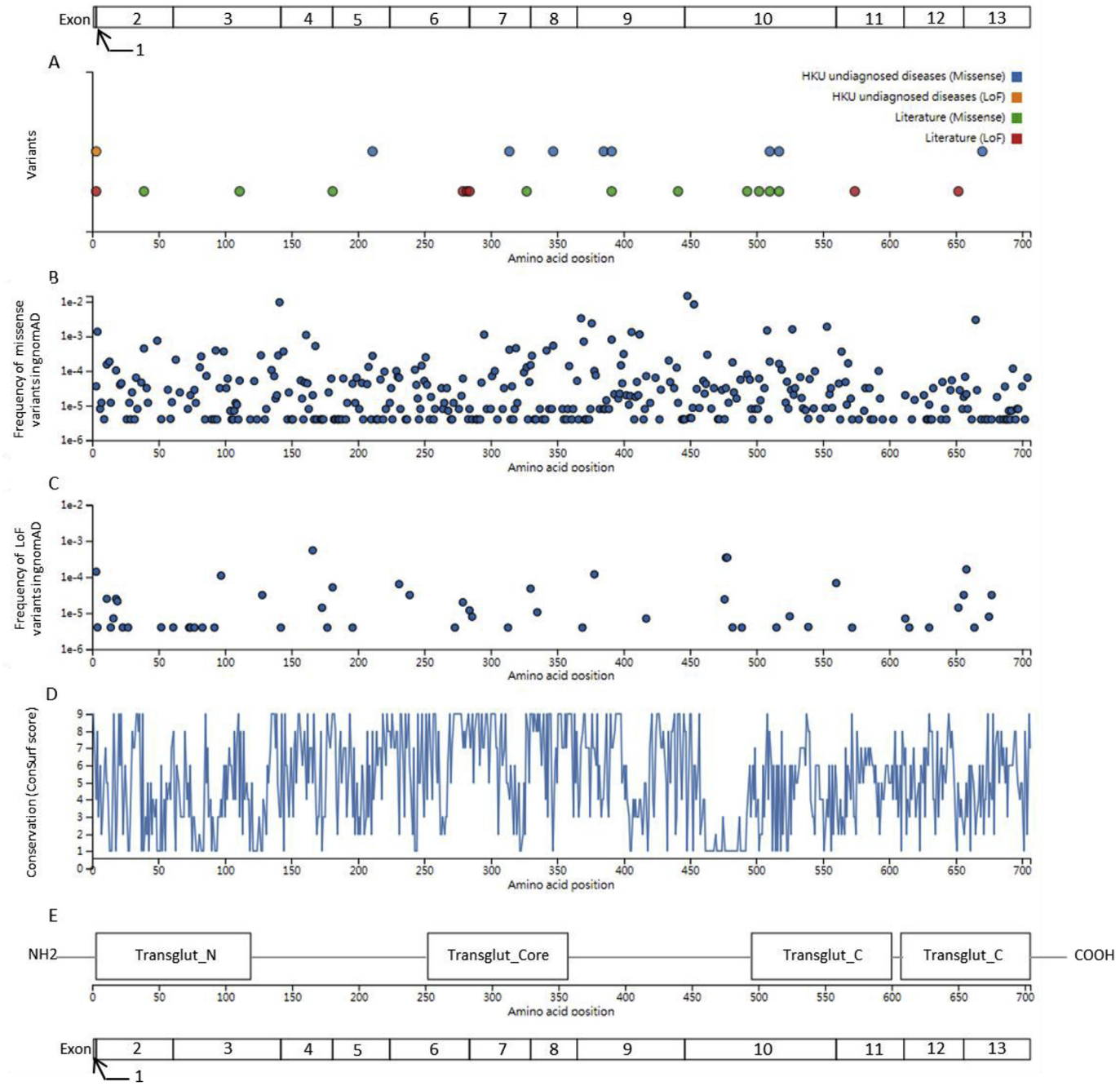

Fig. 1. Distribution of *TGM6* variants with regards to protein domain and conservation score. Panel A: The distribution of rare (allelic frequency < 0.01) and predicted damaging (CADD score ≥20) *TGM6* variants in the HKU undiagnosed disease exome sequencing cohort and variants reported pathogenic in medical literature and ClinVar in respective of the amino acid position. Panel B: The frequency of *TGM6* missense variants in gnomAD against the amino acid position (in log10 scale). Panel C: The frequency of *TGM6* loss-of-function variants in gnomAD against the amino acid position (in log10 scale). Panel D: The conservation of the amino acids across the protein TG6. Conservation is measured using Consurf score, with score 1 being the least conservative and score 9 being the maximum. Panel E: The functional domains of TG6. Transglut_N: transglutaminase N-terminal domain; Transglu_Core: transglutaminase catalytic core domain Transglut_C: transglutaminase C-terminal domain.

lead to decreased transglutaminase (TG6) activity and stability, increased cellular apoptosis and/or increased ER stress response [3,10,12]. However, unlike the expression in mouse, the GTEx expression level of *TGM6* in human is zero TPM across the central nervous system including the cerebellum.

TG6 has a similar structure to other members of the TG family, consisting of four protein domains: a transglutaminase N-terminal domain, a catalytic core, and two transglutaminase C-terminal domains [1,10]. Fig. 1 illustrates the distribution of variants in TG6 with regards to the functional domains and the amino acid conservation [13]. Both missense and LoF variants span across the gene in the disease and healthy cohorts. Of note, a few reported pathogenic LoF variants cluster at the end of exon 6 (c.841delC, c.842_843insC and c.844_850 + 3del), putting forward an argument whether this is a mutation hotspot for SCA35. However, the localization of these variants is not evident to represent a hotspot for disease-causing variants as similar LoF variants are observed in nearby positions in the healthy population.

Further to the variant analysis, we performed an inflation analysis

as suggested by Shah et al. [14] to empirically evaluate the pathogenicity. The working principle of an inflation analysis is that a completely penetrant genetic variant should not be more common in the population than the prevalence of the disease that it causes. An inflation greater than 1 indicates that the genetic disease risk, or frequency of pathogenic variants in the population, exceeds the expected prevalence of the disease. Therefore, a 10-fold inflation indicates a high chance of misdiagnosis/misclassification of the variant or very low penetrance. Autosomal dominant SCAs have a combined prevalence of < 5.6 per 100,000 [15]. We took the maximum prevalence into calculation with the assumption that all AD SCAs are caused by *TGM6* mutations with full penetrance. The number of individuals in gnomAD carrying reported pathogenic or likely pathogenic *TGM6* variants from ClinVar and the medical literature (n = 16) was used to estimate genetic risk. With the calculation shown in Table 1, we found that there is a 111-fold and 592-fold inflation in the overall population and the East Asian population respectively. This level of inflation is far beyond the threshold of 10-fold, suggesting a very high chance of misdiagnosis or a very low penetrance. Since SCA is highly genetically heterogeneous (43 OMIM genes with SCA type 1, 2, 3 constituting 40% of AD forms of SCA), the analysis likely underestimates the inflation of *TGM6* in SCA35.

The over-representation of variants in the population makes it questionable that all *TGM6* mutations result in highly penetrant Mendelian disorders, but we cannot rule out the possibilities that certain genuine pathogenic *TGM6* variants exist. Further study with more convincing mechanistic evidence, using the curation framework recommended by ClinGen [16], is needed to confirm the gene-disease association. It is unclear if variants in *TGM6* may serve as a genetic risk factor/modifier of ataxia.

Misclassification of benign or low penetrant variants as pathogenic is a significant problem that often results in genetic misdiagnosis. With the previous reports, *TGM6* is already frequently included in NGS ataxia panels, potentially leading to a premature ending of the diagnostic odyssey. This highlights the necessity of including sequencing of genomes from diverse populations, both from asymptomatic controls and phenotypically different patients, in order to ensure accurate classification of variants.

Acknowledgements

We would like to thank The Society for the Relief of Disabled Children, Hong Kong for their financial support.

References

[1] J. Wang, J.L. Wang, X. Yang, K. Xia, Z.M. Hu, L. Weng, X. Jin, H. Jiang, P. Zhang, L. Shen, J.F. Guo, N. Ii, Y.R. Li, L.F. Lei, J. Zhou, J. Du, Y.F. Zhou, Q. Pan, J. Wang, R.Q. Li, B.S. Tang, TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing, Brain 133 (2010) 3510–3518, https://doi.org/10.1093/brain/awq323.

[2] M. Kircher, D.M. Witten, P. Jain, B.J. O'roak, G.M. Cooper, J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet. 46 (2014) 310–315, https://doi.org/10.1038/ng.2892.

[3] Y.C. Guo, J.J. Lin, Y.C. Liao, P.C. Tsai, Y.C. Lee, B.W. Soong, Spinocerebellar ataxia 35: novel mutations in TGM6 with clinical and genetic characterization, Neurology 83 (2014) 1554–1561, https://doi.org/10.1212/WNL.0000000000000909.

[4] M. Li, S. Pang, Y. Song, M. Kung, S.-L. Ho, P.-C. Sham, Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family, Clin. Genet. 83 (2013) 269–273, https://doi.org/10.1111/j.1399-0004.2012.01895.x.

[5] L.L. Pan, Y.M. Huang, M. Wang, X.E. Zhuang, D.F. Luo, S.C. Guo, Z.S. Zhang, Q. Huang, S.L. Lin, S.Y. Wang, Positional cloning and next-generation sequencing identified a TGM6 mutation in a large Chinese pedigree with acute myeloid leukaemia, Eur. J. Hum. Genet. 23 (2015) 218–223, https://doi.org/10.1038/ejhg.2014.67.

[6] Z. hua Yang, M. meng Shi, Y. tao Liu, Y. lin Wang, H. yang Luo, Z. lei Wang, C. he Shi, Y. ming Xu, TGM6 gene mutations in undiagnosed cerebellar ataxia patients, Park. Relat. Disord. 46 (2018) 84–86, https://doi.org/10.1016/j.parkreldis.2017.07.001.

[7] K.D. Farwell, L. Shahmirzadi, D. El-Khechen, Z. Powis, E.C. Chao, B. Tippin Davis, R.M. Baxter, W. Zeng, C. Mroske, M.C. Parra, S.K. Gandomi, I. Lu, X. Li, H. Lu, H.M. Lu, D. Salvador, D. Ruble, M. Lao, S. Fischbach, J. Wen, S. Lee, A. Elliott, C.L.M. Dunlop, S. Tang, Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions, Genet. Med. 17 (2015) 578–586, https://doi.org/10.1038/gim.2014.154.

[8] K.D. Choi, J.S. Kim, H.J. Kim, I. Jung, S.H. Jeong, S.H. Lee, D.U. Kim, S.H. Kim, S.Y. Choi, J.H. Shin, D.S. Kim, K.P. Park, H.S. Kim, J.H. Choi, Genetic variants associated with episodic ataxia in Korea, Sci. Rep. 7 (2017) 1–11, https://doi.org/10.1038/s41598-017-14254-7.

[9] C. Lin, S. Gan, D. Gupta, A. Alaedini, P.H. Green, S. Kuo, Hispanic Spinocerebellar Ataxia Type 35 (SCA35) with a Novel Frameshift Mutation vol. 35, (2018), pp. 5–8.

[10] D. Tripathy, B. Vignoli, N. Ramesh, M.J. Polanco, M. Coutelier, C.D. Stephen, M. Canossa, M.L. Monin, P. Aeschlimann, S. Turberville, D. Aeschlimann, J.D. Schmahmann, M. Hadjivassiliou, A. Durr, U.B. Pandey, M. Pennuto, M. Basso, Mutations in TGM6 induce the unfolded protein response in SCA35, Hum. Mol. Genet. 26 (2017) 3749–3762, https://doi.org/10.1093/hmg/ddx259.

[11] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A.H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D.P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B.P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.H. Won, D. Yu, D.M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J.C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S.M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M.T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, Analysis of protein-coding genetic variation in 60,706 humans, Nature (2016), https://doi.org/10.1038/nature19057.

[12] W.J. Guan, J.L. Wang, Y.T. Liu, Y.T. Ma, Y. Zhou, H. Jiang, L. Shen, J.F. Guo, K. Xia, J. Da Li, B.S. Tang, Spinocerebellar ataxia type 35 (SCA35)-associated transglutaminase 6 mutants sensitize cells to apoptosis, Biochem. Biophys. Res. Commun. 430 (2013) 780–786, https://doi.org/10.1016/j.bbrc.2012.11.069.

[13] S. Ellard, E.L. Baple, M. Owens, D.M. Eccles, S. Abbs, C. Zandra, ACGS best practice guidelines for variant classification 2017, Assoc. Clin. Genet. Sci. (2017) 1–12.

[14] N. Shah, Y.C.C. Hou, H.C. Yu, R. Sainger, C.T. Caskey, J.C. Venter, A. Telenti, Identification of misclassified ClinVar variants via disease population prevalence, Am. J. Hum. Genet. 102 (2018) 609–619, https://doi.org/10.1016/j.ajhg.2018.02.019.

[15] L. Ruano, C. Melo, M.C. Silva, P. Coutinho, The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies, Neuroepidemiology (2014), https://doi.org/10.1159/000358801.

[16] N.T. Strande, E.R. Riggs, A.H. Buchanan, O. Ceyhan-Birsoy, M. DiStefano, S.S. Dwight, J. Goldstein, R. Ghosh, B.A. Seifert, T.P. Sneddon, M.W. Wright, L. V Milko, J.M. Cherry, M.A. Giovanni, M.F. Murray, J.M. O'Daniel, E.M. Ramos, A.B. Santani, A.F. Scott, S.E. Plon, H.L. Rehm, C.L. Martin, J.S. Berg, Evaluating the clinical Validity of gene-disease Associations: an evidence-based framework developed by the clinical genome resource, Am. J. Hum. Genet. 100 (2017) 895–906, https://doi.org/10.1016/j.ajhg.2017.04.015.