Check for updates

DR. HIROSHI YAMAGUCHI (Orcid ID : 0000-0002-0787-964X)

DR. KANDAI NOZU (Orcid ID : 0000-0002-0290-3137)

Article type : Original Articles

Original article

Prevalence of Wilson Disease Based on Genome Databases in Japan

Short running title: Prevalence of Wilson Disease in Japan

Hiroshi Yamaguchi, M.D., D.V.M., PhD.[1], Hiroaki Nagase, M.D., PhD.[1], Shoichi Tokumoto, M.D.[1], Kazumi Tomioka, M.D., PhD.[1], Masahiro Nishiyama, M.D., PhD.[1], Hiroki Takeda, M.D., PhD.[1], Takeshi Ninchoji, M.D., PhD.[1], China Nagano, M.D., PhD.[1], Kazumoto Iijima, M.D., PhD.[1], Kandai Nozu, M.D., PhD.[1]

[1]Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan

***Corresponding author:**

Hiroshi Yamaguchi, MD, DVM, PhD

Department of Pediatrics, Kobe University Graduate School of Medicine, Hyogo, Japan

7-5-2 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan

Tel: +81-78-382-5111, Fax: +81-78-382-5050

E-mail: hiyamagu@med.kobe-u.ac.jp

Number of text pages: 12

Number of words: 2,277

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/PED.14565

This article is protected by copyright. All rights reserved

Accepted Article

Reference pages: 3

Tables: 2

Figures and legends to figures: 0

Accepted Article

This article is protected by copyright. All rights reserved

Abstract

Background: Wilson disease (WD) is an autosomal recessive disorder caused by mutations in the *ATP7B* gene. In 1984, Scheinberg and Sternlieb estimated the prevalence of WD to be 1:30,000. However, recent epidemiological studies have reported increasing prevalence rates in different populations. The carrier frequency of *ATP7B* variants and the prevalence of WD in the Japanese population have not been reported using multiple databases.

Methods: Multiple public databases were used. First, we included mutations in the *ATP7B* gene that were registered in the Human Gene Mutation Database (HGMD) Professional, where 885 *ATP7B* variants were identified as pathogenic. Next, we investigated the allele frequencies of these 885 variants in Japanese individuals using the Human Genetic Variation Database (HGVD) and the Japanese Multi Omics Reference Panel (jMorp).

Results: Of the 885 variants of *ATP7B*, 7 and 12 missense and nonsense variants, 0 and 3 splicing variants, and 0 and 2 small deletions were found in the HGVD and in jMorp, respectively. The total allele frequencies of the *ATP7B* mutations were 0.011 in the HGVD and 0.014 in the jMorp. According to these data, the carrier frequencies were 0.022 (2.2%) and 0.028 (2.8%), respectively, and patient frequencies were 0.000121 (1.21/10,000 individuals) and 0.000196 (1.96/10,000 individuals), respectively.

Conclusion: This is the first study to report the carrier frequency of *ATP7B* variants and the prevalence of WD in Japan using multiple databases. The calculated prevalence of WD was comparatively higher than that of previous reports, indicating previous underdiagnosis or the existence of less severe phenotypes.

Keywords: ATP7B; carrier frequency; Human Gene Mutation Database; Human Genetic Variation Database; Japanese Multi Omics Reference Panel

This article is protected by copyright. All rights reserved

Accepted Article

Accepted Article

Introduction

Wilson disease (WD) is an autosomal recessive disorder caused by mutations in the *ATP7B* gene that result in an impaired ability to excrete copper into the bile [1]. Consequently, in patients with WD, copper accumulates in the liver, ultimately leading to hepatitis and cirrhosis. Furthermore, copper accumulation in the brain causes neuropsychiatric symptoms [2]. A diagnosis can be made based on clinical symptoms, which include low serum ceruloplasmin levels and urinary and hepatic copper levels, and based on an ophthalmologic examination that can reveal the development of Kayser–Fleischer rings in the cornea [3]. Currently, common treatments for WD include chelators (d-penicillamine or trientine) that increase urinary copper excretion, zinc salts that reduce intestinal copper absorption [4], or even liver transplantation for cases of acute liver failure [5].

In 1984, Scheinberg and Sternlieb estimated the prevalence of WD to be 1:30,000, and this prevalence is still widely cited [6]. However, this estimation was based on limited available data. During the past decade, epidemiological studies of WD have reported increasing prevalence rates in different populations, as a large number of overlooked WD patients with fatal consequences who lack the typical symptoms or clinical findings have now been recognized [7,8]. Given that more effective treatments for this disease are now available, assessments using new genetic and epidemiological analyses have become increasingly important for identifying potentially fatal cases [9,10]. To the best of our knowledge, the carrier frequency of *ATP7B* variants and the prevalence of WD in the Japanese population based on the use of multiple databases have not been reported.

In the present study, we searched for single variants of the *ATP7B* gene that were classified as pathogenic in the Human Gene Mutation Database (HGMD) Professional [11]. The allele frequencies of these variants were analyzed using two public databases for single nucleotide polymorphisms (SNPs) in the Japanese population: the Human Genetic Variation Database (HGVD) [12] and the Japanese Multi Omics Reference Panel (jMorp) [13]. This study used public databases and therefore did not require ethical review.

Materials/Subjects, Methods

Databases

The present epidemiological study used multiple public databases. First, we included all single

This article is protected by copyright. All rights reserved

nucleotide mutations in the *ATP7B* gene (reference NM_000339.2) that were registered in the HGMD Professional 2020.3 (https://portal.biobase-international.com/hgmd/pro/start.php) and that represented published pathogenic gene variants for inherited diseases in humans [11]. Next, we investigated the allele frequencies of these variants in the Japanese population using HGVD version 2.3 (http://www.hgvd.genome.med.kyoto-u.ac.jp/) [12] and jMorp 202001 (https://jmorp.megabank.tohoku.ac.jp/202001/) [13]. HGVD is operated by the Kyoto University and is a database that analyzes the genome sequence of healthy Japanese individuals [14]. This database was constructed for the purpose of evaluating and analyzing the involvement of genetic variations in the onset of the disease. Information regarding the exome analysis of 1,208 samples and information on genetic variation obtained based on the results of cohort studies using 3,248 samples are posted. jMorp is a new database that contains metabolome and proteome data for plasma obtained from >5,000 healthy Japanese volunteers from the Tohoku Medical Megabank Cohort Study (https://jmorp.megabank.tohoku.ac.jp) [15]. This database includes large-scale data for healthy volunteers with various health records and genome data. From this database, we can also freely access the metabolome data that were obtained through the use of a single protocol at a single institute (measurement biases were significantly minimized).

Calculation methods

The frequencies of the single nucleotide mutations were summed to represent the total allele frequency of *ATP7B* mutations. The carrier frequency was assumed to be twice that of the total allele frequency, as each individual possesses two *ATP7B* alleles. The frequency of WD was calculated by squaring the carrier frequency and multiplying by 0.25 based on the knowledge that WD is an autosomal recessive inherited disease that develops in homozygous or compound heterozygous individuals for any pathogenic mutation. Multiplying by 0.25 shows that the probability of a child with heterozygous carrier parents possessing a homozygous or compound heterozygous mutation is 1 in 4. Finally, the prevalence and carrier frequency of WD in the Japanese population, obtained by the above calculations, were compared to values published in

This article is protected by copyright. All rights reserved

previous reports.

Results

We first identified all missense and nonsense mutations within the exons by searching the HGMD. Possible WD cases were excluded, and 570 WD-associated *ATP7B* variants reported as the cause of WD were included. Of the 570 variants of *ATP7B,* single nucleotide mutations were found in the alleles of 7 and 12 variants in the HGVD and in jMorp, respectively (Table 1). Second, we selected splicing mutations by searching the HGMD. Possible WD cases were excluded, and 73 WD-associated *ATP7B* variants reported as the cause of WD were included. Of the 73 *ATP7B* variants, splicing mutations were found in the alleles of 0 and 3 variants in the HGVD and in jMorp, respectively (Table 1). Third, we analyzed small deletions by searching the HGMD. Possible WD cases were excluded, and 156 WD-associated *ATP7B* variants reported as the cause of WD were included. Of the 156 *ATP7B* variants, small deletions were found in the alleles of 0 and 2 variants in the HGVD and in jMorp, respectively (Table 1). Fourth, we analyzed small insertions by searching the HGMD. Possible WD cases were excluded, and 75 WD-associated *ATP7B* variants reported as the cause of WD were included. Of the 75 *ATP7B* variants, small insertions were found in the alleles of 0 and 2 variants in the HGVD and in jMorp, respectively (Table 1). Finally, we analyzed small indels by searching the HGMD, and 11 WD-associated *ATP7B* variants reported as the cause of WD were included. However, no small indels were found in the HGVD or in jMorp. The total allele frequencies of the *ATP7B* mutations were 0.011 in the HGVD and 0.014 in jMorp (Table 2). Thus, according to these databases, the carrier frequencies were 0.022 (2.2%) and 0.028 (2.8%), respectively, and the patient frequencies were 0.000121 (1.2/10,000 individuals) and 0.000196 (1.96/10,000 individuals), respectively (Table 2).

Discussion

To the best of our knowledge, this is the first study to report the carrier frequency of *ATP7B* variants and the prevalence of WD in a Japanese population by analyzing multiple databases. We utilized two public Japanese databases, and the patient frequencies were strikingly similar between the two (1.21/10,000 [HGVD] and 1.96/10,000 [jMorp] individuals). The similarity of these results supports the accuracy of our analysis. However, the calculated prevalence of WD in the present study was comparatively higher than in previous reports.

As a result of improvements in genetic analysis technology, reports of WD based on genetic

This article is protected by copyright. All rights reserved

analysis have increased in prevalence, and the genetic prevalence of WD is reported to be higher than the clinical prevalence. Wallace *et al*. recently reported a similar approach for evaluating the pathogenicity of *ATP7B* variants using the Genome Aggregation Database (gnomAD) that included variant frequencies analyzed by whole-exome or whole-genome sequencing of over 120,000 people among eight ethnic groups [16]. They found 231 WD-related *ATP7B* variants in gnomAD and a prevalence of approximately 1 in 20,000. Gao *et al*. also recently reported the global prevalence of WD according to next-generation sequencing data [17]. They performed a meta-analysis that included all previously reported pathogenic *ATP7B* variants and annotated those variants with gnomAD allele frequencies, estimating the WD prevalence to be 1.27 in 10,000. Collet *et al*. studied the *ATP7B* gene using next-generation sequencing in a large French cohort [18]. Although they did not discuss the prevalence of WD, we were able to calculate the prevalence of WD based on their findings, and we determined this rate to be 2.5/10,000 individuals. When comparing our results to those of previous studies, the genetic prevalence of WD was found to be higher in our study than in previous estimates.

The discrepancy between the genetically estimated frequency and the clinical prevalence of WD may be explained by the recent development of advanced technology and the incomplete penetrance of WD. Coffey *et al*. conducted a genetic study of WD in the UK and reported that the prevalence of individuals predicted to carry two mutant pathogenic *ATP7B* alleles was 1:7,026 [8]. This was a much higher rate than reported in previous studies [16]. The authors stated that the high prevalence rate was a result of technological advances. Incomplete penetrance may also explain the discrepancy between the epidemiological and genetic prevalence of WD. The genetic prevalence is sometimes 3- to 4-fold higher than the clinical prevalence of WD [9]. Therefore, it is unclear if the penetrance rate of the disease is actually 100% [9, 19]. There is a possibility that decreased penetrance of *ATP7B* variants is responsible for the discrepancy between the genetic prevalence and the number of clinically diagnosed WD patients. Coffey *et al*. reported the existence of 116 different *ATP7B* mutations in 181 clinically and biochemically diagnosed WD patients [8]. In the study, they estimated that the higher frequency of single nucleotide variants in the broad area of the entire *ATP7B* coding region outside of the mutation hotspot may reflect reduced penetrance. Additionally, asymptomatic siblings of WD patients diagnosed by Sanger sequencing have been reported [19]. Therefore, caution should be exercised when interpreting these data. Wallace *et al*. suggested new criteria that included considering probable or possible low penetrance, and they reanalyzed the data presented by Gao *et al*. and Collet *et al*. [18]. They

This article is protected by copyright. All rights reserved

concluded that the prevalence of WD was 1 in 20,000 from the study by Gao *et al*. [17] and 1 in 47,000 in the UK and 1 in 54,000 in France from the study by Collet *et al*. [16]. Previously reported data suggest that human mutation databases should be analyzed with caution. However, there is no guarantee that such patients will not develop WD in the future, and such data should be considered as indicative of a high risk of developing WD. In the present study, we only included *ATP7B* variants that were reported as causes of WD, and we excluded possible WD cases. Our results suggest that the prevalence of WD in the Japanese population may be higher than reported in previous studies. In an Asian country (Korea), Jang *et al*. examined the carrier frequency of WD using a DNA-based approach, and they reported a rate of 1.3/10,000 people [7]. This result is very similar to our result and is comparable to the results of previous studies that reported much higher prevalence rates of WD in Asian countries [16].

Several previous nation-wide, large population-based, epidemiological studies examining the prevalence of WD have been conducted, and the prevalence rates reported in these studies likely differ due to methodological differences and the diversity of the assessed ethnic groups (0.15 cases per 10,000 in France [20], 1:49,500 in Netherlands [21], and 1:39,000 in Austria [22]). The previously reported estimated prevalence of WD in large cohorts is highly similar to the prevalence noted in the original report by Scheinberg and Sternlieb (1:30,000) [6]. Unfortunately, we could not find recent clinical data from Japan to estimate the prevalence according to nation-wide large population-based epidemiological studies. However, the Japanese Society of Pediatrics estimated that the incidence of WD was 1 in 35,000–45,000, and the number of patients in Japan was approximately 2,000 in 2016 according to the results of the WD national survey and the MC-Bank patient registration project (http://www.jpeds.or.jp/uploads/files/2016_ikotyosa_hokoku-32-34.pdf). Interestingly, a classical study in Japan estimated that the frequency of WD was 1 in 20,000 from 1945 to 1965. The frequency was expected to decrease to 1 in 30,000 as a result of a decrease in consanguineous marriages [23]. Therefore, the observed decrease in consanguineous marriages may contribute to the decreased number of WD patients. Taken together, the genetic prevalence has been reported to be higher than the clinical prevalence. The present study and previous studies suggest that there are still many undiagnosed cases. Based on the knowledge that irreversible damage to the brain or liver can be prevented by avoiding high levels of copper in food and through the use of copper-chelating medications, the early detection of WD is beneficial to individuals with this disease.

This article is protected by copyright. All rights reserved

The present study had some limitations. First, we did not consider the onset of *de novo* mutations, although their impacts are considered negligible. Additionally, we focused only on *ATP7B*. However, the overall mutation detection frequency of *ATP7B* was reported to be 98%, and the likelihood of mutations other than those in *ATP7B* is considered very low [8]. Furthermore, the possibility of sampling bias due to the locality of the sample provider in HGMD and jMorp cannot be ruled out for each database. Therefore, there is a possibility that the selection criteria of these databases may have biased our findings. Finally, it is possible that our analysis using these public databases may overestimate the prevalence of WD. Recent studies have raised the possibility of overestimation by computational analysis [24]. This may be due to the inclusion of benign variants and the prediction of too many *ATP7B* mutations as pathogenic [24]. This could result from the lack of clear genotype-phenotype relations in WD patients. Therefore, more research is needed to distinguish disease-causing mutations from benign variants. Taking these limitations into account, the prevalence of WD in the Japanese population is relatively higher than that reported in other ethnic groups. Further research is necessary to determine a more accurate prevalence of WD.

Conclusion

This is the first study to assess the prevalence and carrier frequency of WD in Japan using multiple databases. The carrier frequencies in the HGVD and jMorp databases were 0.022 (2.2%) and 0.028 (2.8%), respectively, and the patient frequencies were 0.000121 (1.21/10,000 individuals) and 0.000196 (1.96/10,000 individuals), respectively. The calculated prevalence of WD in the present study was comparatively higher than that of previous reports, thus indicating previous under-diagnosis or the existence of less severe phenotypes.

Disclosure

The authors declare no conflicts of interest.

Author contributions

H.Y. collected and analyzed the data and wrote the manuscript; H. N., S.T., K.T., M.N., H.T., T.N., and C.N. provided technical support and conceptual advice and critically reviewed the manuscript; K.I. and K.N. provided technical support and conceptual advice, contributed to the conception and design of this study, and critically reviewed the manuscript. All authors have read

This article is protected by copyright. All rights reserved

Accepted Article

and approved the final manuscript.

Accepted Article

This article is protected by copyright. All rights reserved

Accepted Article

References

1. Bandmann O, Weiss KH, Kaler SG. Wilson's disease and other neurological copper disorders. *Lancet Neurol.* 2015;14:103–13.
2. Ala A, Walker AP, Ashkan K, Dooley JS, Schilsky ML. Wilson's disease. *Lancet.* 2007;369:397–408.
3. European Association for Study of Liver. EASL clinical practice guidelines: Wilson's disease. *J Hepatol.* 2012;56:671–85.
4. Wiggelinkhuizen M, Tilanus ME, Bollen CW, Houwen RH. Systematic review: clinical efficacy of chelator agents and zinc in the initial treatment of Wilson disease. *Aliment Pharmacol Ther.* 2009;29:947.
5. Emre S, Atillasoy EO, Ozdemir S, Schilsky M, Rathna Varma CV, Thung SN, et al. Orthotopic liver transplantation for Wilson's disease: a single-center experience. *Transplantation.* 2001;72:1232.
6. Scheinberg IH, Sternlieb I. *Wilson's Disease*. Philadelphia: W.B. Saunders; 1984.
7. Jang JH, Lee T, Bang S, Kim YE, Cho EH. Carrier frequency of Wilson's disease in the Korean population: a DNA-based approach. *J Hum Genet*. 2017;62:815–8.
8. Coffey AJ, Durkie M, Hague S, McLay K, Emmerson J, Lo C, et al. A genetic study of Wilson's disease in the United Kingdom. *Brain.* 2013;136(Pt 5):1476–87.
9. Sandahl TD, Laursen TL, Munk DE, Vilstrup H, Weiss KH, Ott P. The prevalence of Wilson's disease: An update. *Hepatology.* 2020;71:722–32.
10. Walshe JM. Cause of death in Wilson disease. *Mov Disord.* 2007;22:2216–20.
11. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet.* 2017;136:665–77.
12. Higasa K, Miyake N, Yoshimura J, Okamura K, Niihori T, Saitsu H, et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. *J Hum Genet.* 2016;61:547–53.
13. Tadaka S, Katsuoka F, Ueki M, Kojima K, Makino S, Saito S, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum Genome Var.* 2019;6:28.
14. Narahara M, Higasa K, Nakamura S, Tabara Y, Kawaguchi T, Ishii M, et al. Large-

This article is protected by copyright. All rights reserved

scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. *PLoS One*. 2014;9:e100924.

15. Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, et al. jMorp: Japanese Multi Omics Reference Panel. *Nucleic Acids Res*. 2018;46:D551–D557.
16. Wallace DF, Dooley JS. ATP7B variant penetrance explains differences between genetic and clinical prevalence estimates for Wilson disease. *Hum Genet.* 2020. Apr 4.
17. Gao J, Brackley S, Mann JP. The global prevalence of Wilson disease from next-generation sequencing data. *Genet Med.* 2019;21:1155–63.
18. Collet C, Laplanche JL, Page J, Morel H, Woimant F, Poujois A. High genetic carrier frequency of Wilson's disease in France: discrepancies with clinical prevalence. *BMC Med Genet.* 2018;19:143.
19. Stattermayer AF, Entenmann A, Gschwantler M, Zoller H, Hofer H, Ferenci P. The dilemma to diagnose Wilson disease by genetic testing alone. *Eur J Clin Invest.* 2019;49:e13147.
20. Poujois A, Woimant F, Samson S, Chaine P, Girardot-Tinant N, Tuppin P. Characteristics and prevalence of Wilson's disease: A 2013 observational population-based study in France. *Clin Res Hepatol Gastroenterol.* 2018;42:57–63.
21. Møller LB, Horn N, Jeppesen TD, Vissing J, Wibrand F, Jennum P, et al. Clinical presentation and mutations in Danish patients with Wilson disease. *Eur J Hum Genet.* 2011;19:935–41.
22. Beinhardt S, Leiss W, Stättermayer AF, Graziadei I, Zoller H, Stauber R, et al. Long-term outcomes of patients with Wilson disease in a large Austrian cohort. *Clin Gastroenterol Hepatol.* 2014;12:683–9.
23. Saito T. An expected decrease in incidence of Wilson's disease due to decrease in Consanguinity. *Jinrui Idengaku Zasshi*. 1985;30:249–53.
24. Tang N, Sandahl TD, Ott P, Kepp KP. Computing the pathogenicity of Wilson's disease ATP7B mutations: Implications for disease prevalence. *J Chem Inf Model*. 2019;59:5230–43.

This article is protected by copyright. All rights reserved

Table1. Wilson disease-*ATP7B* variants and the allele frequencies of ATP7B gene mutations by Japanese database

Mutation Type	Number	Mutation	Amino acid change	rs number	Allele frequency by Japanese Data bases	
					HGVD	jMORP
Missense/nonsense	1	c.2267C>T	Ala756Val	rs769927137		0.0001
	2	c.2333G>A	Arg778Gln	rs28942074	0.0012	0.0008
	3	c.2333G>T	Arg778Leu	rs28942074	0.0012	0.0008
	4	c.2621C>T	Ala874Val	rs121907994	0.0004	0.0007
	5	c.2755C>G	Arg919Gly	rs121907993	0.0004	0.0005
	6	c.2755C>T	Arg919Trp	rs121907993	0.0004	0.0005
	7	c.2975C>T	Pro992Leu	rs201038679	0.0008	0.0001
	8	c.3104G>T	Gly1035Val	rs753594031		0.0003
	9	c.3182G>A	Gly1061Glu	rs764131178		0.0001
	10	c.3809A>T	Asn1270Ile	rs121907990		0.0003
	11	c.3859G>A	Gly1287Ser	rs762866453		0.0005
	12	c.3886G>A	Asp1296Asn	rs199821556	0.0066	0.0085
Splicing	1	c.1708-5T>G		rs770829226		0.0003
	2	c.3060+5G>T		rs1353373400		0.0002
	3	c.3556G>A		rs786204547		0.0001
Small deletion	1	c.2513delA	p.(Lys838Serfs*35)	rs777362050		0.0001
	2	c.3787delG	p.(Ala1263Profs*67)	rs758147392		0.0001

Annotation: Blank spaces indicate no reports. Only mutations reported in HGVD and jMorp are listed.

Table2. Allele, carrier frequency and prevalence of Wilson disease patients.

Deta bank	HGVD	jMorp
Allele frequency	0.011	0.014
Carrier frequency	0.022	0.028
Patient prevalence	0.000121	0.000196
	1.21/10,000 people	1.96/10,000 people