Molecular Genetics and Metabolism Reports 38 (2024) 101034

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism Reports

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

The mutation spectrum and ethnic distribution of Wilson disease, a review

Zahra Beyzaei [a], Arman Mehrzadeh [b], Niko Hashemi [b], Bita Geramizadeh [a,c,*]

[a] Transplant Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[b] School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
[c] Department of Pathology, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Keywords:
Wilson
Genotype
Variants
ATP7B
Mutation spectrum
Ethnics

ABSTRACT

Wilson's disease is a complicated medical condition caused by the accumulation of copper, mostly in the liver and brain. The genetic basis of Wilson's disease is attributed to the presence of pathogenic variants in the *ATP7B* copper-transporting gene, which prevents the excretion of copper through the biliary tract. To date, *ATP7B* remains the only identified gene that has been linked to the development of this disease. Our understanding of the disease has been associated with the identification of particular disease-causing variants that present specific impairments in copper transporters. It is crucial to identify the most frequent variant in terms of ethnicity to facilitate testing of its functionality. This study represents the initial comprehensive analysis of *ATP7B* variants, providing insights into the extensive range of disease-causing mutations. Here, we describe the 1275 distinct *ATP7B* variants documented so far, with particular emphasis on their regional and ethnic prevalence. The H1069Q missense variant is the most frequently reported in Europe, Northern America, and North Africa, whereas the R778L, C271*, and M645R variants are the most prevalent in the East Asian, Middle Eastern-South Asian, and South American populations, respectively. Acquiring such knowledge would facilitate the implementation of a selective mutation screening approach, targeting the most predominant variant identified within a specific ethnic group or geographic region for better diagnosis of the disease.

1. Introduction

Wilson disease, also known as hepatolenticular degeneration (HLD), is an autosomal recessive inherited disorder of copper metabolism [1]. It is caused by pathogenic variants in the *ATP7B* gene, which is located on chromosome 13q14.3 and consists of 21 exons and 20 introns, with a total genomic length of 80 kb. It encodes a copper-transporting ATPase type P, containing 1465 amino acids. The protein is synthesized in the endoplasmic reticulum (ER) and then relocated to the Golgi apparatus in hepatocytes [2]. ATP7B gene is mostly expressed in the liver but is also found in the kidney, mammary glands, placenta, lung, and brain. Mutations in the *ATP7B* gene disrupt the synthesis and function of the ATP7B protein, leading to further impairment of the copper excretion pathway. The deficient incorporation of copper into ceruloplasmin results in a decrease in biliary copper excretion and an increase in copper accumulation, mainly in the liver, brain, and eyes. The deposition of copper in the body is toxic and results in clinically heterogeneous manifestations, including liver function impairment, neurological disturbance, and developmental delay [3].

Regarding the management of Wilson patients and the detection of disease-causing variants, it would be beneficial to report all variants for diagnosis and future mutation-targeted therapeutic options [4]. Despite numerous papers on the subject, no clear analysis of the relationship between genotype, phenotype, and mutation distribution has been reported. Therefore, the review was focused on identifying and summarizing the reported variants in the *ATP7B* gene and their corresponding manifestations in related patients to date. To prepare this manuscript, a literature search was conducted by the authors using the electronic PubMed (Medline) database and the following search terms: "Wilson disease", "Hepatolenticular Degeneration", "mutation," and "genetic variation." The authors selected the most relevant English language articles.

2. Variants in the ATP7B gene

After cross-checking the literature and genetic databases on *ATP7B*

Abbreviations: ATP7B, adenosine triphosphatase 7B; ER, endoplasmic reticulum; MLPA, multiplex ligation-dependent probe amplification; TNG, Trans-Golgi network.

* Corresponding author at: Department of Pathology, Shiraz University of Medical Sciences, Shiraz 7193711351, Iran.
E-mail address: geramib@gmail.com (B. Geramizadeh).

https://doi.org/10.1016/j.ymgmr.2023.101034
Received 17 October 2023; Received in revised form 28 November 2023; Accepted 1 December 2023

from the oldest to the present day (from 1995 to 2022), we updated the number of variants with the recently reported articles. To date, 1275 unique variants in the *ATP7B* gene from 260 articles (16,183 patients) have been identified.

Of 1275 unique variants, 618 missense-reported variants illustrate the predominant type (48%) in the *ATP7B* gene of Wilson disease. Furthermore, 304 frameshift variants (24%), 111 nonsense variants (9%), 17 promoter variants (1%), 86 deletions (7%), and 15 large deletions (1%) are reported (Fig. 1). 124 intronic variants (10%) have been reported. These results suggest that intronic variants of *ATP7B* could potentially be an overlooked contributing factor in the development of Wilson's disease. However, it is worth noting that promoter mutations and large whole exon deletions/duplications of the ATP7B gene are rare occurrences. Thus, it appears that utilizing multiplex ligation-dependent probe amplification (MLPA) or other methods for *ATP7B* analysis may be time-consuming, costly, and of limited value in the molecular diagnosis of Wilson's disease.

The clustering of variants in specific subregions of the *ATP7B* gene across diverse populations has been well recognized [5]. In our study, in all exonic and intronic regions of the ATP7B gene, the variants are reported (Fig. 2). However, the frequency of regional clustering of variants was considerably higher in eight exons consisting of 2, 8, 11, 13, 14, 16, 18, and 20. It has been demonstrated that a restricted screen of eight exons can detect 56.3% (717/1275) of all variants, which offers a more convenient option for patients with distinct presentations of Wilson's disease. Additionally, identifying the causative region of the gene enables molecular genetic testing to be conducted without the need for a time-consuming and costly process. It helps clinicians decide on the appropriate time to begin proper treatment. Nevertheless, it should be noted that 43.7% (558/1275) of all variants were outside of the hotspot clustering regions, and due to the nature of Wilson's disease, compound heterozygote variants occur. Therefore, the decision on whether to sequence the entire or partial *ATP7B* coding region depends on the clinical and biochemical data of the patients and the population under study.

3. Geographical distribution of variants in the *ATP7B* gene

The frequency of reported alleles based on the origin of patients can be found in assisting clinicians in focusing on specific variants in certain regions, which facilitates the targeted detection of diseased alleles. Therefore, we decided to reclassify them into a unique list including the count of known alleles from patients and the predominant geographical distribution specific to each ethnic group (Fig. 3). More than 80% of the described variants are reported from East Asia and Europe, whereas the remaining (less than 20%) are from North/South- America, North Africa, and Middle East-South Asia.

In East Asia, the most prevalent variant was the missense R778L variant, which is represented in only 4% of the reported population i.e., China, South Korea, Taiwan, Vietnam, and Hong Kong. It is demonstrated that the R778L variant is located in exon 8 and the transmembrane domain 4 (TM4) region of the protein, which affects both its localization and trafficking of ATP7B [6]. Other common variants are P992L, A874V, I1148T, N1270S, and R919G, while each one is only 2% of reported variants.

In Europe, the H1069Q missense variant is the most common variant with a frequency of 6.9% among all reported variants. The observed variant induced a histidine to glutamic acid substitution within the SEHPL motif located in the N-domain of the ATP7B gene. This leads to protein misfolding, abnormal phosphorylation in the P-domain, diminished ATP binding affinity, compromised heat stability, and aberrant protein localization to the Trans-Golgi network (TGN) [7]. The H1069Q variant is found frequently in patients from a wide range of ethnic groups i.e., Middle East-Asia (Iran), North Africa (Egypt), North and South America (USA, Brazil, El Salvador, Cuba) over a large geographical distribution. Considering the notable prevalence and extensive distribution of this variant, it represents a highly ancient genetic variant, the origins of which remain unidentified so far, which needs further investigation. The other common variants were R969Q, A1135Qfs*, W779*, T977M, and R778G which occur in less than 2% of the population. Our findings have demonstrated that several disease-causing variants exhibit relatively low frequencies within specific populations, while the prevalence of these variants exhibits significant geographical variation across different regions [8].

In the Middle East-South Asia, many of the reported variants are typical to a specific region. The most common is a nonsense variant C271* (4.4%), which is reported in Indian and Pakistani patients. Other variants that are often detected in patients from the Middle East are N1270S, G1061E, A1003V, I1102T, and G1341S. In Iran, the H1069Q variant, which has not been reported from elsewhere in Middle Eastern Asia, accounts for 50% of the reported alleles. However, further studies are needed in this region to identify the most prevalent variants.

Given the extremely limited number of reported variants in North America, South America, and North Africa, the conclusion regarding prevalence should be considered with caution (Table 1).

To the best of our knowledge, no mutations in Wilson patients have

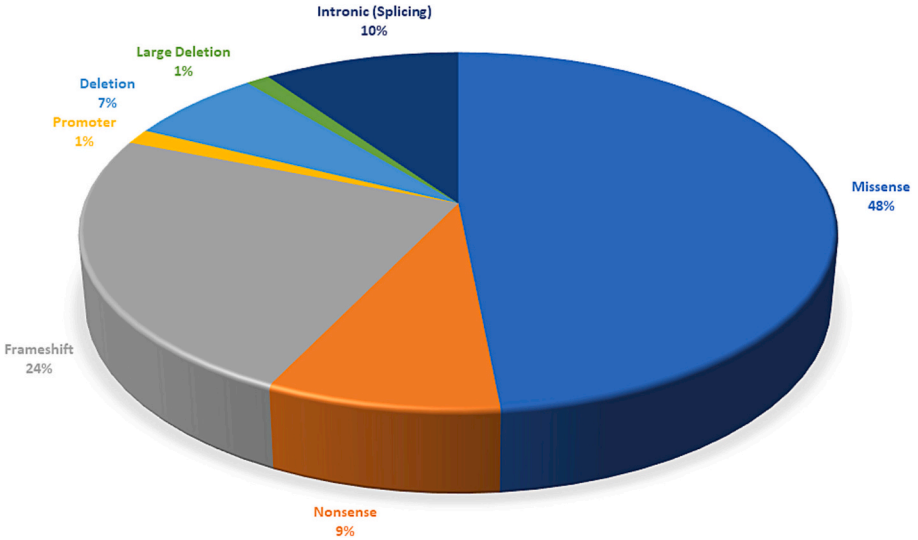

Fig. 1. Pie chart representing the distribution of different types of identified variants in Wilson patients.

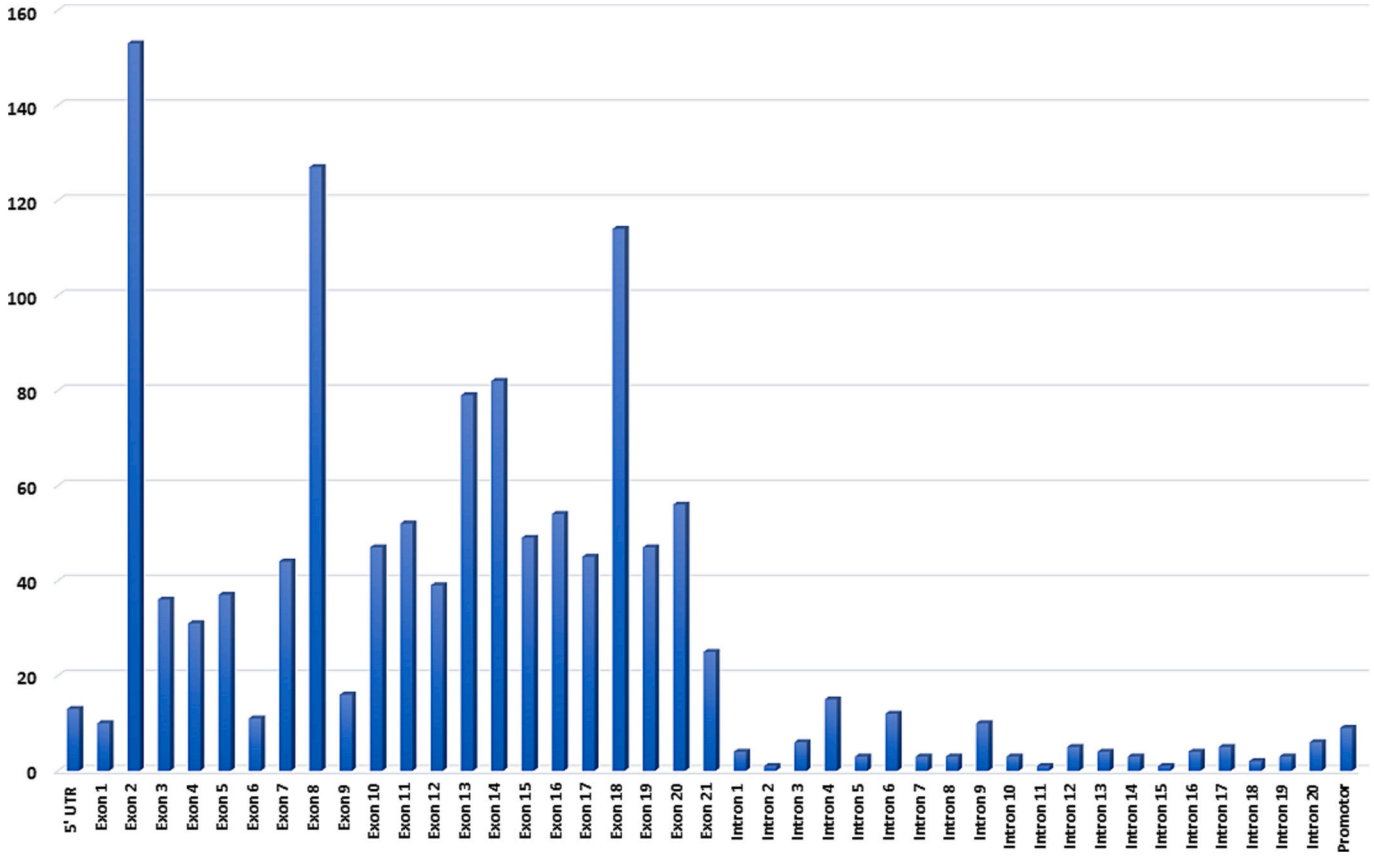

Fig. 2. The number and location of the 1275 unique variants, identified on the ATP7B gene.

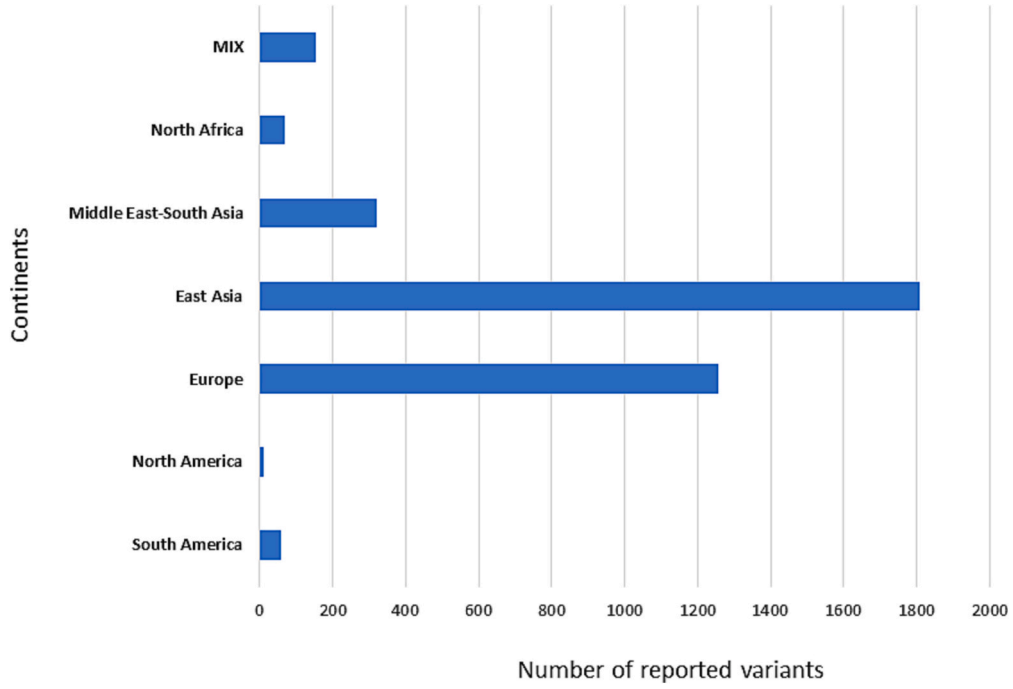

Fig. 3. Geographical distribution of the most common ATP7B alleles causing variants worldwide. It represents the distribution of ATP7B alleles in different continents. Where the paper's provenance was not clear or had different ethnic patients, the mutation is included in the mix on the graphic.

Table 1
The predominant geographical distribution of variants specific to each ethnic group and country.

East Asia	Europe	Middle East-South Asia	South America	North America*	North Africa**
Total number of Reported variants					
1808	1258	320	57	8	68
Most prevalent variant in the continent (%)					
R778L (4%) P992L (2%) A874V (2%) I1148T (2%) N1270S (2%) **Others (89%)**	H1069Q (6.9%) R969Q (1.8%) A1135Qfs*13 (1.8%) W779* (1.7%) T977M (1.5%) **Others (88.4%)**	C271* (4.4%) N1270S (3.8%) G1061E (2.8%) A1003V (2.2%) I1102T (2.1%) **Others (85%)**	M645R (7.1%) A1135Qfs*13 (5.3%) H1069Q (5.3%) IVS1 + 4 A > T (5.3%) L1088*(5.3%) **Others (72%)**	H1069Q (50%) K1248Tfs*83 (12.5%) M645R (12.5%) **Others (25%)**	H1069Q (4.4%) C703Y (2.9%) **Others (93%)**

Most prevalent variant in the country (%)							
China	R778L (2.8%) P992L (2.1%) A874V (1.9%)	France	H1069Q (2.8%) G1099S (1.4%) I1148T (1.4%)	India	C271* (4.9%) G1061E (3.4%) N1270S (3.1%)	Brazil	M645R (7.3%) S921N (4.8%) IVS1 + 4 A > T (4.8%)
Japan	R778L (8.1%) N958Tfs* (7.2%) A874V (4.1%)	Italy	H1069Q (6%) M645R (3%) R919W (3%)	Iran	H1069Q (50%) N1270S (12%) A874V (8%)	Venezuela	M645R (12.3%) A1135Qfs*13 (1.3%)
South Korea	R778L (7%) A874V (3%) N1270S (3%)	Germany	H1069Q (5.4%), R969Q (4.3%)	Saudi Arabia	G1341S (33%)		
Taiwan	R778L (5.8%) P992L (5.8%) G943D (4.8%)	Czech	H1069Q (6.3%) A1135Qfs*(3.2%) D1047fs (3.2%)				
Vietnam	S105* (4.8%) R778L (3%) P992L (3%)	Spain	L708P (3.7%) M645R (7.5%) A1135Qfs*13 (3.8%)				
		Poland	H1069Q (8.7%) S653Y (3.3%)				
		Greece	H1069Q (5.8%) I1148T (5.8%) IVS18-2 A > G (5.8%)				

* All study in North America was reported from the USA.
** All study in North Africa was reported from Egypt

been reported in Central America, South Africa, or the continent of Oceania.

4. Genotype-phenotype correlation

The correlation between genotype and phenotype is described as the association between a specific variant and/or class of variants with a particular clinical abnormality. Therefore, finding genotype-phenotype correlations in rare inherited metabolic disorders can be complicated due to a variety of factors. In Wilson disease, these factors include the large phenotypic heterogeneity associated with the same variant and a high proportion of subclinical variants. The majority of variants in the *ATP7B* gene are subclinical variants, so, genotype-phenotype correlations or correlations between type and location of the variant and clinical manifestation have not been established so far.

To achieve this goal, we have compiled a comprehensive list of patients identified from the literature, along with their clinical and biochemical data. This has enabled us not only to precisely estimate the number of published patients with Wilson disease but also to stratify

Table 2
Participant characteristics included Wilson patients.

Variable	Data
Presentation	
Hepatic	4302 (45.4%)
Neurologic	3631 (38.3%)
Mix (hepatic and neurologic)	1002 (10.7%)
Asymptomatic	538 (5.6%)
Liver biopsy	
Fibrosis	74 (9.7%)
Steatosis	178 (23.4%)
Cirrhosis	509 (66.9%)
Hepatic Cu (μg/g dry)	424.8 (113–1128)
Urine Cu 24 h (μg per 24 h)	435 (109–3084)
Serum Ceruloplasmin (mg/dL)	11.47 (1.34–19)

Data are given as n (%) or mean (range).
The mean age at presentation was 15.7 years (range, 0.8–40 years).
Normal ceruloplasmin serum: 20–35 (mg/dL); Normal urine cu 24 h: 10–30 μg per 24 h; Normal hepatic copper concentration: 10–35 μg/g dry weight.

patients for genotype-phenotype correlations. Data on clinical, biochemical, and pathological results of Wilson's disease were available for some of the patients (Table 2). Overall, of 16,183 patients, 5589 male, 4455 female, and 6139 unspecified genders were reported. The mean age of patients at the time of presentation was 15.7 years old. Out of the 9473 patients, the disease presentation was comprised of 45.6% hepatic, 38.3% neurologic, 10.7% mixed presentation, and 5.6% asymptomatic. Out of the 260 studies, 61 included a liver biopsy to measure hepatic copper content and/or pathological results. Among the patients who received a biopsy, 74 (9.7%) had fibrosis, 178 (23.4%) had steatosis, and 509 (66.9%) had cirrhosis. The mean Copper level in the liver was 424.8 (μg/g dry) among the 2319 patients. Additionally, the mean 24-h urine copper level was 435 (μg) and the mean serum ceruloplasmin level was 11.47 (mg/dL) among the 14,240 patients.

Regarding genetic variation, affected patients carrying an identical genotype in different families or even within the same family present extensive variation in hepatic and neurological manifestations [4,8–14]. It is unknown why a particular genotype is not associated with a specific behavior of the disease. In the most prevalent variant of East Asia, some articles report that Wilson's patients with the R778L variant present with an earlier onset of disease and predominantly hepatic symptoms [13,15–19]. Patients with the H1069Q variant, which is the most prevalent variant in Europe, North America, and North Africa, show a complete range of clinical manifestations; this mutation is not predominantly associated with a late or neurologic presentation [20–23]. The variants, C271*, and N1270S are the most common in Middle East-South Asia especially the Indian and Pakistani populations with allelic frequencies of 4.4% and 3.8%, respectively. The correlation between genotype and phenotype for these variants is controversial. While some papers document severe neurological impairment and K–F rings in patients during the first or early second decade of life, others report hepatic presentations [24–28]. It is worth mentioning that most studies have reported a lack of correlation between ATP7B gene variants and the phenotype.

5. Conclusion

Currently, available data illustrate that a single gene causes Wilson's disease. Many patients may not possess the characteristic findings and may present when their clinical disease is relatively mild. ATP7B molecular analysis can be particularly useful in patients with an atypical presentation and in identifying affected siblings including those without definitive symptoms. The most common variant in Europe, Northern America, and North Africa is the missense variant H1069Q, and the most prevalent ones in East Asian, Middle East-South Asian, and South American populations are the R778L, C271*, and M645R, respectively. Detecting a wide range of variants, noting the variability in disease manifestation despite carrying the same variants, and the predominance of compound heterozygosity are some of the difficulties encountered in establishing a genotype-phenotype correlation, premarital screening, prenatal testing, and pre-implantation genetic diagnosis of Wilson's disease in different population.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not required.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Zahra Beyzaei: Conceptualization, Data curation, Formal analysis, Writing – original draft. **Arman Mehrzadeh:** Data curation, Methodology, Writing – review & editing. **Niko Hashemi:** Data curation, Methodology, Writing – review & editing. **Bita Geramizadeh:** Conceptualization, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

All authors declare no competing interests.

Data availability

Data will be made available on request.

Acknowledgments

None.

References

[1] J. Bennett, S.H. Hahn, Clinical molecular diagnosis of Wilson disease, Semin. Liver Dis. 31 (2011) 233–238.
[2] I.J. Chang, S.H. Hahn, The genetics of Wilson disease, Handb. Clin. Neurol. 142 (2017) 19–34.
[3] V. Medici, J.M. LaSalle, Genetics and epigenetic factors of Wilson disease, Ann. Transl. Med. 7 (2019) S58.
[4] R. Cocoş, A. Şendroiu, S. Schipor, L.C. Bohîlţea, I. Şendroiu, F. Raicu, Genotype-phenotype correlations in a mountain population community with a high prevalence of Wilson's disease: genetic and clinical homogeneity, PLoS One 9 (2014), e98520.
[5] A.B. Shah, I. Chernov, H.T. Zhang, B.M. Ross, K. Das, S. Lutsenko, Identification and analysis of mutations in the Wilson disease gene (ATP7B): population frequencies, genotype-phenotype correlation, and functional analyses, Am. J. Hum. Genet. 61 (1997) 317–328.
[6] J.J. Xie, Z.Y. Wu, Wilson's disease in China, Neurosci. Bull. 33 (2017) 323–330.
[7] S. Caprai, G. Loudianos, F. Massei, L. Gori, M. Lovicu, G. Maggiore, Direct diagnosis of Wilson disease by molecular genetics, J. Pediatr. 148 (2006) 138–140.
[8] E. Couchonnal, L. Lion-François, O. Guillaud, D. Habes, D. Debray, T. Lamireau, Pediatric Wilson's disease: phenotypic, genetic characterization and outcome of 182 children in France, J. Pediatr. Gastroenterol. Nutr. 73 (2021) e80-e6.
[9] A. Antos, T. Litwin, M. Skowrońska, I. Kurkowska-Jastrzębska, A. Członkowska, Pitfalls in diagnosing Wilson's disease by genetic testing alone: the case of a 47-year-old woman with two pathogenic variants of the ATP7B gene, Neurol. Neurochir. Pol. 54 (2020) 478–480.
[10] W.O. Arruda, R.P. Munhoz, R.S. de Bem, M.M. Deguti, E.R. Barbosa, J.A. Zavala, Pathogenic compound heterozygous ATP7B mutations with hypoceruloplasminaemia without clinical features of Wilson's disease, J. Clin. Neurosci. 21 (2014) 335–336.
[11] M.S. Balashova, I.G. Tuluzanovskaya, O.S. Glotov, A.S. Glotov, Y.A. Barbitoff, M. A. Fedyakov, The spectrum of pathogenic variants of the ATP7B gene in Wilson disease in the Russian Federation, J. Trace Elem. Med. Biol. 59 (2020), 126420.
[12] J. Chaudhuri, S. Biswas, G. Gangopadhyay, T. Biswas, J. Datta, A. Biswas, Correlation of ATP7B gene mutations with clinical phenotype and radiological features in Indian Wilson disease patients, Acta Neurol. Belg. 122 (2022) 181–190.
[13] N. Cheng, H. Wang, W. Wu, R. Yang, L. Liu, Y. Han, Spectrum of ATP7B mutations and genotype-phenotype correlation in large-scale Chinese patients with Wilson Disease, Clin. Genet. 92 (2017) 69–79.
[14] E. Couchonnal, S. Bouchard, T.D. Sandahl, C. Pagan, L. Lion-François, O. Guillaud, ATP7B variant spectrum in a French pediatric Wilson disease cohort, Eur. J. Med. Genet. 64 (2021), 104305.
[15] Y.C. Chen, H. Yu, R.M. Wang, J.J. Xie, W. Ni, Y. Zhang, Contribution of intragenic deletions to mutation spectrum in Chinese patients with Wilson's disease and possible mechanism underlying ATP7B gross deletions, Parkinsonism Relat. Disord. 62 (2019) 128–133.
[16] Y. Dong, W. Ni, W.J. Chen, B. Wan, G.X. Zhao, Z.Q. Shi, Spectrum and classification of ATP7B variants in a large cohort of Chinese patients with Wilson's disease guides genetic diagnosis, Theranostics. 6 (2016) 638–649.
[17] Y. Fan, R. Yang, L. Yu, M. Wu, S. Shi, M. Ren, Identification of a novel missense mutation in Wilson's disease gene, Chin. Med. J. 110 (1997) 887–890.
[18] S. Jia, X. Li, W. Zhang, B. Zhang, Z. Wu, W. Duan, Laboratory and clinical evaluation of a microarray for the detection of ATP7B mutations in Wilson disease in China, J. Clin. Lab. Anal. 36 (2022), e24735.
[19] C.C. Lee, J.Y. Wu, F.J. Tsai, H. Kodama, T. Abe, C.F. Yang, Molecular analysis of Wilson disease in Taiwan: identification of one novel mutation and evidence of haplotype-mutation association, J. Hum. Genet. 45 (2000) 275–279.

[20] A. Krumina, J. Keiss, V. Sondore, A. Chernushenko, G. Cernevska, A. Zarina, From clinical and biochemical to molecular genetic diagnosis of Wilson disease in Latvia, Genetika. 44 (2008) 1379–1384.
[21] A.A. Sivtsev, I.Z. Zhalsanova, A.E. Postrigan, E.A. Fonova, O.Y. Vasilyeva, A. A. Zarubin, Analysis of mutations spectrum in the ATP7B gene in patients with Wilson disease using massively parallel sequencing, Klin. Lab. Diagn. 67 (2022) 250–256.
[22] A. Kluska, M. Kulecka, T. Litwin, K. Dziezyc, A. Balabas, M. Piatkowska, Whole-exome sequencing identifies novel pathogenic variants across the ATP7B gene and some modifiers of Wilson's disease phenotype, Liver Int. 39 (2019) 177–186.
[23] A. Sapuppo, P. Pavone, A.D. Praticò, M. Ruggieri, G. Bertino, A. Fiumara, Genotype-phenotype variable correlation in Wilson disease: clinical history of two sisters with the similar genotype, BMC Med. Genet. 21 (2020) 128.
[24] R. Aaron, A. Chapla, S. Danda, U. Zachariah, C.E. Eapen, A. Goel, Multiplex PCR-based sequencing of ATP7B gene in Wilson's disease - a preliminary study, J. Clin. Exp. Hepatol. 12 (2022) 711–713.
[25] A. Gupta, I. Chattopadhyay, S. Dey, P. Nasipuri, S.K. Das, P.K. Gangopadhyay, Molecular pathogenesis of Wilson disease among Indians: a perspective on mutation spectrum in ATP7B gene, prevalent defects, clinical heterogeneity and implication towards diagnosis, Cell. Mol. Neurobiol. 27 (2007) 1023–1033.
[26] H.N. Khan, M. Wasim, H. Ayesha, F.R. Awan, Molecular genetic diagnosis of Wilson disease by ARMS-PCR in a Pakistani family, Mol. Biol. Rep. 45 (2018) 2585–2591.
[27] N. Zali, S.R. Mohebbi, S. Esteghamat, M. Chiani, M.M. Haghighi, S.M. Hosseini-Asl, Prevalence of ATP7B gene mutations in Iranian patients with Wilson disease, Hepat. Mon. 11 (2011) 890–894.
[28] A. Aggarwal, G. Chandhok, T. Todorov, S. Parekh, S. Tilve, A. Zibert, Wilson disease mutation pattern with genotype-phenotype correlations from Western India: confirmation of p.C271* as a common Indian mutation and identification of 14 novel mutations, Ann. Hum. Genet. 77 (2013) 299–307.