CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Whole-exome sequencing identifies novel pathogenic variants across the ATP7B gene and some modifiers of Wilson's disease phenotype
Authors	Kluska, Anna;Kulecka, Maria;Litwin, Tomasz;Dziezyc, Karolina;Balabas, Aneta;Piatkowska, Magdalena;Paziewska, Agnieszka;Dabrowska, Michalina;Mikula, Michal;Kaminska, Diana;Wiernicka, Anna;Socha, Piotr;Czlonkowska, Anna;Ostrowski, Jerzy
Publication date	2018-09-19
Original Citation	Kluska, A, Kulecka, M, Litwin, T, Dziezyc, K., Balabas, A., Piatkowska, M., Paziewska, A., Dabrowska, M., Mikula, M., Kaminska, D., Wiernicka, A., Socha, P., Czlonkowska, A. and Ostrowski, J. (2018) 'Whole-exome sequencing identifies novel pathogenic variants across the ATP7B gene and some modifiers of Wilson's disease phenotype', Liver International, 39(1), pp. 177-186. doi: 10.1111/liv.13967
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1111/liv.13967
Rights	© 2018, John Wiley & Sons A/S. . This is the accepted version of the following item: Kluska, A, Kulecka, M, Litwin, T, Dziezyc, K., Balabas, A., Piatkowska, M., Paziewska, A., Dabrowska, M., Mikula, M., Kaminska, D., Wiernicka, A., Socha, P., Czlonkowska, A. and Ostrowski, J. (2018) 'Whole-exome sequencing identifies novel pathogenic variants across the ATP7B gene and some modifiers of Wilson's disease phenotype', Liver International, 39(1), pp. 177-186, doi: 10.1111/liv.13967, which has been published in final form at: https://doi.org/10.1111/liv.13967. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.
Download date	2026-11-16 16:13:54

Item downloaded from	https://hdl.handle.net/10468/14935



Whole-exome sequencing identifies novel pathogenic variants across the *ATP7B* gene and some modifiers of Wilson's disease phenotype

Anna Kluska[1], Maria Kulecka[2], Tomasz Litwin[3], Karolina Dziezyc[3], Aneta Balabas[1], Magdalena Piatkowska[1], Agnieszka Paziewska[2], Michalina Dabrowska[1], Michal Mikula[1], Diana Kaminska[4], Anna Wiernicka[4], Piotr Socha[4], Anna Czlonkowska[3*], Jerzy Ostrowski[1,2*]

[1]Department of Genetics, Cancer Center-Institute, Warsaw 02-781, Poland

[2]Department of Gastroenterology and Hepatology, Centre of Postgraduate Medical Education, Warsaw 01-813, Poland

[3]Department of Neurology, Institute of Psychiatry and Neurology, Warsaw 02-957, Poland

[4]Department of Gastroenterology, Hepatology, Nutritional Disorders, and Pediatrics, The Children's Memorial Health Institute, Warsaw 04-730, Poland

***Corresponding authors**

JO: Department of Genetics, Cancer Center-Institute, Roentgena 5, Warsaw 02-781, Poland; jostrow@warman.com.pl, +48225462575

AC: Department of Neurology, Institute of Psychiatry and Neurology, Sobieskiego 9, Warsaw 02-957, Poland; czlonkow@ipin.edu.pl, +48224582537

Electronic word count: 4450

Figures and tables: 1 figure, 4 tables, 7 supplementary tables

Abbreviations:

WD – Wilson's disease

WES – whole-exome sequencing

VEP – Variant Effect Predictor

SNP – single nucleotide polymorphism

Conflict of interest: The authors declare that they have no conflict of interest.

Funding: This work was supported by the National Science Centre [2013/11/B/NZ2/00130]. JO was supported by the National Science Centre [2011/02/A/NZ5/00339].

Data Sharing Statement: Data on *ATP7B* gene mutation and rare variants in other genes associated with Wilson's disease, used in modeling are available as supplementary material in this manuscript (Online Resource: Dataset S1 and S2). Due to the ethical restrictions (possibility of identification), data from whole-exome sequencing are available from the Department of Genetics, Cancer Center-Institute, Warsaw

Abstract

Background & Aims. Wilson's disease (WD) is an autosomal recessive disorder associated with disease-causing alterations across the *ATP7B* gene, with highly variable symptoms and age of onset. We aimed to assess whether the clinical variability of WD relates to modifier genes.

Methods. 248 WD patients were included, of whom 148 were diagnosed after age of 17. Human exome libraries were constructed using AmpliSeq technology and sequenced using the IonProton platform.

Results. *ATP7B* p.His1069Gln mutation was present in 215 patients, with 112 homozygotes and 103 heterozygotes. Three other mutations: p.Gln1351Ter, p.Trp779Ter, and c.3402delC were identified in >10 patients. Among patients, 117 had a homozygous mutation, 101 were compound heterozygotes, 27 had one heterozygous mutation, and 3 other patients had no identifiable pathogenic variant of *ATP7B*. Sixteen mutations were novel, found as part of a compound mutation or as a sole, homozygous mutation. For disease phenotype prediction, age at diagnosis was a deciding factor, while frameshift allelic variants of *ATP7B* and being male increased the odds of developing a neurologic phenotype. Rare allelic variants in *ESD* and *INO80* increased and decreased chances for the neurologic phenotype, respectively, while rare variants in *APOE* and *MBD6* decreased the chances of WD early manifestation. Compound mutations contributed to earlier age of onset.

Conclusions. In a Polish population, genetic screening for WD may help genotype for four variants (p.His1069Gln, p.Gln1351Ter, p.Trp779Ter, c.3402delC), with direct sequencing of all *ATP7B* amplicons as a second diagnostic step. We also identified some allelic variants that may modify a WD phenotype.

Abstract electronic word count: 250

Keywords: Wilson's disease; exome sequencing; *ATP7B;* modifier genes

Lay summary

Wilson's disease is a genetic disorder, characterized by copper accumulation in different organs, caused by mutations in the *ATP7B* gene, which encodes one of the copper transporters. This disease has two distinct manifestations – hepatological and neurological. The aim of our study was to determine whether the appearance of those phenotypes is predetermined by genetic factors. We have discovered that frameshift mutations in the *ATP7B* gene may lead to developing a neurological phenotype of the disease, and compound mutations contributed to an earlier age of onset. Rare allelic variants in *ESD*, *INO80, APOE* and *MBD6* also modified the disease phenotype.

Introduction

Wilson's disease (WD) is an autosomal recessive metabolic disorder resulting from disease-causing *ATP7B* mutations (formally named allelic variants, according to the Human Genome Variation Society[1]) that lead to defective biliary excretion of copper and its accumulation, particularly in the liver and brain.[2–10] Single-gene disorders usually manifest in childhood, but the highly variable symptoms of WD develop mostly between ages 5 and 35 years; however, some individuals with these mutations have presented with symptoms as early as 2 years old or have lived into the eighth decade showing no symptoms at all.[9,11] Early onset of WD occurs with dominant hepatic manifestations (a hepatic phenotype) of any degree of severity, very rarely accompanied by neurologic or psychiatric symptoms (a neurologic phenotype). In turn, neurologic/psychiatric symptoms are common at onset in adults and can be present with liver injury or in isolation. Hepatic manifestations in WD range from abnormalities on liver function tests in apparently well individuals to symptoms suggesting acute liver failure, acute hepatitis, or liver cirrhosis. Symptoms may progress very rapidly or quite slowly.

As in other Mendelian diseases, clinical variability observed in WD may relate, at least in part, to the expression of modifier genes that vary in showing strong to low effects. Strong effects naturally relate to mutations with a high effect size. Mild effects, however, relate to common polymorphic susceptibility loci in different modifier genes which, by themselves, explain only a small proportion of the variability, and their clinical effects reflect interaction with other allelic variants and environmental factors.

Identification of highly penetrant allelic variants remains a possibility. Recently, whole-exome sequencing (WES) has allowed for the analysis of all exons of all protein-coding genes and is considered a cost-effective technology for detecting disease allelic variants underlying Mendelian disorders and cataloguing other disease-related genomic alterations.[12] However, pinpointing allelic variants that individually modify a disease phenotype remains challenging. Although polymorphisms in several genes have been postulated to be involved in the clinical outcome of WD (including apolipoprotein E, prion proteins, methylenetetrahydrofolate reductase, the copper metabolism gene Murr1, antioxidant 1, inhibitor of apoptosis linked to chromosome X, genes associated with iron metabolism, inflammatory processes and oxidative stress)[10,13–21] our WES-based study failed to confirm most of them. In turn, we identified a few novel allelic variants that may modify a WD phenotype.

Materials and methods

Patients

This study included 248 WD patients, of whom 148 were diagnosed after 17 years old. Diagnosis and clinical follow-up were conducted at two Polish centers, the Children's Memorial Health Institute and the Institute of Psychiatry and Neurology. The patients were classified into three disease phenotypes based on the following: 1) clinically asymptomatic or with mild liver injury (alanine aminotransferase serum activity levels below <200 U/I) (asymptomatic

phenotype); 2) hepatic features with clinical symptoms of liver injury (International Normalized Ratio >1.5 and/or cholestasis (hepatic phenotype), and 3) dominant of neurological features at onset with only mild or no liver symptoms (neurologic phenotype).[14,18,22] Patients with neurological symptoms on onset were later divided into groups with predominance of clinical sings of dystonia, tremor and parkinsonism. Patient selection was based on databases available at both institutions with information on children and adults with WD. All included patients met the criteria for a WD diagnosis with Ferenci score ≥4 points.[23] Additionally, 95 patients whose exomes were sequenced in parallel for other projects, including a group of patients with macroAST, reported by our previous study,[24] were used as controls. In all of them WD and other hepatological conditions were ruled out.

Ethics statement: The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki. All participants or their parents provided written informed consent for the genetic testing. The study was approved by the Institute of Psychiatry and Neurology Bioethics Committee.

See the Online Supplementary Materials and Methods for a detailed description of exome sequencing, variants filtering and their functional analyses.

Results

Sequencing results

In WD samples, 20 million reads per sample were obtained on average. Approximately 74% bases had coverage of 20× or more (median, 80%). After quality filtering, there were on average 17235 variants per sample, with 130 InDels. The average transition to transversion ratio was at 3.2. Each sample had on average 186 unique variants.

ATP7B gene mutations

Among 248 patients, the most frequent mutation, present in 215 patients, was *ATP7B* p.His1069Gln; 112 and 103 patients were carriers of the homozygous and heterozygous mutation, respectively (Fig. 1). In 18 patients, the heterozygous *ATP7B* p.His1069Gln mutation was the only mutation found, while in another 85 patients, this heterozygous mutation was associated with another heterozygous allelic variant. Patients with other mutations were mostly compound heterozygous. Of these, three other mutations, two nonsense changes (p.Gln1351Ter and p.Trp779Ter) and one frameshift deletion (c.3402delC), were each found in more than 10 patients (Table 1). In sum, 117 patients had a single homozygous mutation, 101 patients had two or more heterozygous mutations, and 27 had one heterozygous mutation, and in 3 other patients, no potentially pathogenic variant of the *ATP7B* gene was found (Table 2). Of all mutations identified, 41 had already been reported, and another 16 were novel (Table 3). Of 10 missense allelic variants, six had been reported as deleterious according to the Condel score. Most of these mutations also had a relatively high combined annotation-dependent deletion algorithm (CADD) score (≥20 on the PHRED scale), with the exception of c.1112C>T and c.712T>C. However, c.712T>C was present as part of a two-heterozygous mutations compound pairing, so its pathogenicity cannot be ruled out. In contrast, c.1112C>T accompanied a homozygous mutation, making its limited impact on protein function more plausible. Additionally, three large rearrangements were uncovered with CONTRA software in patients with heterozygous mutations (Online Resource Supplementary Table S1).

Modeling results for *ATP7B*

For disease subtype prediction, age at diagnosis emerged as a deciding factor, with the p-value smaller than 0.001 for this coefficient. The odds of having the neurological phenotype were increased in patients diagnosed after 18 years of age. Having a frameshift mutation of *ATP7B* also increased the odds of developing a neurological phenotype. On the other hand,

being female significantly decreased the odds of developing a neurological phenotype. Compound mutations contributed to earlier age at diagnosis (Online Resource Supplementary Tables S2, S3).

Variants in other disease-associated genes

There were 59 heterozygous rare allelic variants present in *HFE, SERPINA1, CFTR, NPC1* and *NPC2* among WD patients (Table 4). Two patients possessed two allelic variants in one gene - one patient, in *CFTR*: p.Glu695Gly and p.Arg1162Leu and another one in *SERPINA1*: p.Glu366Lys and p.Ala308Ser. The patient possessing *SERPINA1* variants was also heterozygous for the *ATP7B* mutation.

Results for genes that are known to associate with WD disease phenotype

No allelic variant in genes previously reported as associating with WD phenotype (*COMMD1, ATOX1, XIAP, APOE, DMT1 (SLC11A2), ATP7A, MTHFR* and *PRNP)* [25,26] presented a statistically significant difference (at a nominal p-value of 0.05) in allele frequency in the comparison between hepatological and neurological manifestations.

In the logistic regression model, possessing rare allelic variants in two genes – *ESD* and *INO80* was significantly associated with WD phenotype, by increasing and decreasing chances for the neurologic phenotype, respectively, while *APOE* narrowly missed statistical significance with a p-value of 0.054. Regarding age of onset, the presence of rare variants in *APOE* and *MBD6* decreased chances of early manifestation (Online Resource Supplementary Tables S4, S5).

Other allelic variants

We conducted several comparisons in our search for any other genetic differences between patients with early vs late onset of WD or patients with different clinical phenotypes of WD or different disease severity (dystonic form, regarded as more severe neurological manifestation) (Online Resource Supplementary Table S6). The comparisons involved the full

list of variants as well as rare variants, with and without accounting for the coding consequences of a given nucleotide change. However, no analyses yielded statistically significant results for disease phenotypes; the Fisher's exact test p-value did not reach the genome-wide cutoff of 5×10^{-8}, and the adjusted p-value in the CMC method was never <0.05. In age-of-onset comparison, one variant was overrepresented on allele level and passed additional filters: MTOR c.2863delC. This change, however, seems to be quite common in Polish population (or specific to our sequencing technique) with the frequency of 50% in control group. Additionally, when compared to the control group, patients with WD had no significant overrepresentation of rare, potentially deleterious variants (t-test statistic with mean of -1.8 in 1000 comparisons).

Functional analysis

In functional analysis, one pathway was significantly enriched – "TP53 Regulates Metabolic Genes" (p-value=0.035) in the comparison between different disease subtypes (Online Resource Supplementary Table S7).

Discussion

In portal blood, copper is carried as Cu^{2+}, loosely bound to albumin and aminoacids. Its intestinal absorption and biliary excretion are controlled respectively by two energy-requiring transport proteins, the enterocyte ATP7A, and the hepatocyte ATP7B. Located on chromosome 13, the *ATP7B* gene contains 80,000 base pairs divided into 21 protein coding exons and 20 non-coding introns. 767 mutations of *ATP7B,* associated with Wilson's disease, are as of August 2018 deposited in The Human Gene Mutation Database,[27] out of which H1069Q (His1069Glu) is the most common in Europe and North America; the majority of WD patients are compound heterozygotes with one mutation on each *ATP7B* allele.[28] In central Europe, H1069Q occurs at allele frequencies of 30–60% in series of WD patients from Poland,

Romania, Austria, and Saxony.[29–31] By contrast, H1069Q is rare in Asian, Japanese, and Chinese patients, and in the heterogeneous population of the UK, WD is associated with many mutations, mostly compound heterozygous. The H1069Q mutant binds ATP in an incorrect orientation with a reduced affinity, causing instability from temperature-dependent unfolding. For several other mutants, decreased ATP7B protein levels and mislocalization result in defective biliary excretion of copper and its hepatic and brain accumulation.[9]

In this study, we combined the clinical expertise in WD of two leading Polish institutes that are networked to a European Reference Network for WD, EuroWilson (http://www.eurowilson.org), together with an academic laboratory to perform and analyze exome sequencing data. Homozygous or compound heterozygous *ATP7B* mutations were identified in 218 patients, and the WES results were negative and inconclusive in three and 27 of WD patients, respectively. While heterozygote carriers are often clinically asymptomatic, detailed examination may reveal subtle abnormalities.[28] However, direct sequencing conducted in 44 Chinese WD patients presented with hepatic manifestations identified a heterozygous mutation in 5 cases, of whom 3 patients appeared to be also carriers of large deletions found by the multiplex ligation-dependent probe amplification assay.[32] Thus, while the clinical usefulness of WES is unquestionable, our WES results were negative and inconclusive in three and 27 of WD patients, respectively. These negative and inconclusive results likely reflect challenges in distinguishing structural variation of the DNA sequence, including multiplication and heterozygous deletions.[33] In addition, information from flanking noncoding regions, such as promoters, introns, and splice sites may be ignored by current analysis pipelines.[34] The AmpliSeq Exome kit employed in this study usually underestimates some insertions and deletions in exome enrichment products.[35]

In addition, 16 previously unreported mutations were discovered. Most were deemed pathogenic, at least according to CADD, which already has proven to perform well in the

clinical setting.[36] Most of the allelic variants were present as part of compound mutations or as sole, homozygous mutations; therefore, some effect on disease development seems likely. Their functional impact can be established only through *in vivo* high-throughput assays that are beyond the scope of this work.

Differing WD phenotypes are commonly found among siblings, and even for monozygotic twins, isolated populations show clinical variability among patients homozygous for the same mutation.[21,37,38] Studies performed on neonatal blood spots in Sardinia[39] and in the UK[40] have indicated that WD may be under-diagnosed and that the disease does not clinically manifests in many individuals with disease-causing WD mutations. Although individual susceptibility to WD depends on *ATP7B* variants, accurate mutation calls for WD patients do not explain all genuine genotype-phenotype associations, and other factors can influence the clinical outcome of the disease. The modifier gene concept was introduced in 1941 by Haldane.[41] Of 4900 Mendelian disorders, causative genes have been identified for more than 3300. Deleterious genetic variants do not act in isolation, so their clinical manifestations often relate to complex interactions among multiple genes and are additionally affected by environmentally related epigenetic modifications.[11,29] For this reason, identical manifestations within patients harboring even the same causal variant are unlikely.[30]

Searching for modifier genes differs from searching for disease-causing genes. In searching for genes involved in the disease, individuals are usually defined as affected or unaffected, but in searching for modifier genes, the affected individuals are related to the clinical symptom variability. The modifier gene effect can vary from strong effects, similar to those observed in a monogenic disease, to much milder effects typical for multifactorial, complex diseases.[42] The strong effect relates mostly to highly penetrant mutations, whereas mild-to-moderate effects relate to single nucleotide polymorphisms (SNPs). As a consequence, although WES-based studies may be suitable for searching an individual or shared variants with

a strong effect on a disease phenotype, associated SNPs can be uncovered using genome-wide association studies. However, magnitudes of associated SNPs are generally less than 1.5, and populations consisting of several hundreds to thousands of patients and healthy controls are needed for genome-wide association studies.

Although the variation in the WD phenotype may relate to the *ATP7B* genotype, high genetic heterogeneity and the disease rarity cause difficulties in assessing a relationship between WD genotypes and phenotypes.[25] In addition, the clinical effects of disease-causing *ATP7B* mutations can be modified by gene-gene and gene-environmental interactions. While several modifier effects on WD expression have been reported until now in clinical and experimental studies (as reviewed in [25] and [26]), we did not prove a statistically significant difference in the allele frequency across *COMMD1*, *ATOX1*, *XIAP*, *APOE*, *DMT1, ATP7A, MTHFR* and *PRNP* in comparison between hepatological and neurological manifestations. Rare allelic variants in *ESD* and *INO80* significantly increased and decreased chances for the neurologic phenotype, respectively, while rare variants in *APOE* and *MBD6* decreased chances of the WD early manifestation. Among genes, associating with age of onset or phenotype in our study, *APOE* and *ESD* were previously associated with either WD or variability in its manifestations. Common *APOE* genotypes were previously associated with delayed[43] or early onset,[14] albeit others failed to confirm these associations.[44,45] On the other hand, the *ESD* was previously linked with WD, probably due to its relative proximity to *ATP7B*.[46] Two other genes we found to modify WD phenotype, namely *INO80* and *MBD6,* encode chromatin-associated proteins. The INO80 is a catalytic ATPase subunit of the INO80 chromatin remodeling complex that functionally is involved in transcription, DNA repair, and DNA replication. [47] The MDB6 (Methyl-CpG-binding protein 6) contributes to the formation or function of heterochromatin, however does not bind directly to methylated DNA. [48] Functionally, it was found to control stemness of the cell and promoting human adipocytes differentiation to neuronal cells in the

SCID mouse brain. [49] We also found that *ATP7B* frameshift variants associated more often with a neurological subtype of WD and that their compound mutations contributed to an earlier age at diagnosis. To date, two valid findings have been reported: patients homozygous for H1069Q tend to present with WD later and with neurological disease,[44] and fulminant hepatic failure is more likely in patients with truncating mutations.[45] We also confirmed that WD patients with hepatic phenotype are diagnosed significantly earlier than those with neurological phenotype, and odds of developing a neurological symptoms at clinical onset in females are significantly decreased. The later observation of neurological phenotype is a well-known phenomenon,[48,49] attributed mostly to slower progression of the disease and often initial subtlety of symptoms.[7] Similarly, the prevalence of hepatic manifestations in women was described previously[18,50] and could be plausibly explained by the protective role of estrogens in brain, which is also visible in other neurodegenerative diseases such as Parkinson's disease.[52,53]

However, although the clinical variability of WD patients did not associate with a clear *ATP7B* mutation/phenotype relationship, uncovering the complete picture of the genetic determinants of the disease will require a catalogue of all relevant variants arranged in various combinations. This kind of investigation is technologically possible, but other factors need to be identified, including environmental effects, to ensure complete genetic and molecular surveillance for WD. This kind of identification is certainly difficult in terms of the accessibility of research methodologies.

To summarize, WES identified 41 already reported and 16 novel variants across the *ATP7B* gene. Most of the novel mutations were deemed pathogenic and found as part of a compound mutation or as a sole, homozygous mutation. For WD phenotype prediction, age at diagnosis was a deciding factor, while frameshift allelic variants of *ATP7B* and being male increased the odds of developing a neurological phenotype. Compound mutations contributed

to earlier age of onset. Rare allelic variants in *ESD* and *INO80* associated with a neurologic phenotype, while rare variants in *APOE* and *MBD6* with age of onset.

Acknowledgements

We acknowledge the help of Dr. Aldona Wierzbicka and Dr. Wojciech Janczyk provided during the research.

References

1. den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*. 2016;37:564–569.

2. European Association for Study of Liver. EASL Clinical Practice Guidelines: Wilson's disease. *J Hepatol*. 2012;56:671–685.

3. Scheinberg IH, Sternlieb I. Wilson's disease. Philadelphia: Saunders; 1984.

4. Gitlin JD. Wilson disease. *Gastroenterology*. 2003;125:1868–1877.

5. Tao TY, Gitlin JD. Hepatic copper metabolism: insights from genetic disease. *Hepatology*. 2003;37:1241–1247.

6. Lutsenko S, Petris MJ. Function and regulation of the mammalian copper-transporting ATPases: insights from biochemical and cell biological approaches. *J Membr Biol*. 2003;191:1–12.

7. Ferenci P, Członkowska A, Merle U, et al. Late-onset Wilson's disease. *Gastroenterology*. 2007;132:1294–1298.

8. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet*. 2013;14:295–300.

9. Socha P, Tanner S. Wilson's Disease. In: Textbook of Pediatric Gastroenterology, Hepatology and Nutrition. Springer, Cham; 2016. p. 743–752.doi:10.1007/978-3-319-17169-2_65

10. Gromadzka G, Schmidt HHJ, Genschel J, et al. p.H1069Q mutation in ATP7B and biochemical parameters of copper metabolism and clinical manifestation of Wilson's disease. *Mov Disord*. 2006;21:245–248.

11. Wiernicka A, Dądalski M, Jańczyk W, et al. Early Onset of Wilson Disease: Diagnostic Challenges. *J Pediatr Gastroenterol Nutr*. 2017;65:555–560.

12. Kulecka M, Habior A, Paziewska A, et al. Clinical Applicability of Whole-Exome Sequencing Exemplified by a Study in Young Adults with the Advanced Cryptogenic Cholestatic Liver Diseases. *Gastroenterol Res Pract*. 2017;2017:4761962.

13. Gromadzka G, Schmidt HH-J, Genschel J, et al. Frameshift and nonsense mutations in the gene for ATPase7B are associated with severe impairment of copper metabolism and with an early clinical manifestation of Wilson's disease. *Clin Genet*. 2005;68:524–532.

14. Litwin T, Gromadzka G, Członkowska A. Apolipoprotein E gene (APOE) genotype in Wilson's disease: impact on clinical presentation. *Parkinsonism Relat Disord*. 2012;18:367–369.

15. Merle U, Stremmel W, Gessner R. Influence of homozygosity for methionine at codon 129 of the human prion gene on the onset of neurological and hepatic symptoms in Wilson disease. *Arch Neurol*. 2006;63:982–985.

16. Gromadzka G, Rudnicka M, Chabik G, Przybyłkowski A, Członkowska A. Genetic variability in the methylenetetrahydrofolate reductase gene (MTHFR) affects clinical expression of Wilson's disease. *J Hepatol*. 2011;55:913–919.

17. Litwin T, Członkowska A. [Wilson disease - factors affecting clinical presentation]. *Neurol Neurochir Pol*. 2013;47:161–169.

18. Litwin T, Gromadzka G, Członkowska A. Gender differences in Wilson's disease. *J Neurol Sci*. 2012;312:31–35.

19. Gromadzka G, Członkowska A. Influence of IL-1RN intron 2 variable number of tandem repeats (VNTR) polymorphism on the age at onset of neuropsychiatric symptoms in Wilson's disease. *Int J Neurosci*. 2011;121:8–15.

20. Przybyłkowski A, Gromadzka G, Wawer A, Grygorowicz T, Cybulska A, Członkowska A. Intestinal expression of metal transporters in Wilson's disease. *Biometals*. 2013;26:925–934.

21. Członkowska A, Gromadzka G, Chabik G. Monozygotic female twins discordant for phenotype of Wilson's disease. *Mov Disord*. 2009;24:1066–1069.

22. Członkowska A, Litwin T, Dzieżyc K, Karliński M, Bring J, Bjartmar C. Characteristics of a newly diagnosed Polish cohort of patients with neurological manifestations of Wilson disease evaluated with the Unified Wilson's Disease Rating Scale. *BMC Neurol*. 2018;18:34.

23. Ferenci P, Caca K, Loudianos G, et al. Diagnosis and phenotypic classification of Wilson disease. *Liver Int*. 2003;23:139–142.

24. Kulecka M, Wierzbicka A, Paziewska A, et al. A heterozygous mutation in GOT1 is associated with familial macro-aspartate aminotransferase. *J Hepatol*. 2017;67:1026–1030.

25. Ala A, Schilsky M. Genetic modifiers of liver injury in hereditary liver disease. *Semin Liver Dis*. 2011;31:208–214.

26. Medici V, Weiss K-H. Genetic and environmental modifiers of Wilson disease. *Handb Clin Neurol*. 2017;142:35–41.

27. Stenson PD, Mort M, Ball EV, 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;136:665–677.

28. Lo C, Bandmann O. Epidemiology and introduction to the clinical presentation of Wilson disease. *Handb Clin Neurol*. 2017;142:7–17.

29. Ferenci P. Regional distribution of mutations of the ATP7B gene in patients with Wilson disease: impact on genetic testing. *Hum Genet*. 2006;120:151–159.

30. Czlonkowska A, Rodo M, Gajda J, Ploos van Amstel HK, Juyn J, Houwen RH. Very high frequency of the His1069Gln mutation in Polish Wilson disease patients. *J Neurol*. 1997;244:591–592.

31. Iacob R, Iacob S, Nastase A, et al. The His1069Gln mutation in the ATP7B gene in Romanian patients with Wilson's disease referred to a tertiary gastroenterology center. *J Gastrointestin Liver Dis*. 2012;21:181–185.

32. Wang C, Zhou W, Huang Y, et al. Presumed Missense and Synonymous Mutations in ATP7B Gene Cause Exon Skipping in Wilson Disease. *Liver Int*. 2018. doi:10.1111/liv.13754

33. Zare F, Dow M, Monteleone N, Hosny A, Nabavi S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinformatics*. 2017;18:286.

34. Wu M, Chen T, Jiang R. Global inference of disease-causing single nucleotide variants from exome sequencing data. *BMC Bioinformatics*. 2016;17:468.

35. Damiati E, Borsani G, Giacopuzzi E. Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum Genet*. 2016;135:499–511.

36. van der Velde KJ, Kuiper J, Thompson BA, et al. Evaluation of CADD Scores in Curated Mismatch Repair Gene Variants Yields a Model for Clinical Validation and Prioritization. *Hum Mutat*. 2015;36:712–719.

37. Chabik G, Litwin T, Członkowska A. Concordance rates of Wilson's disease phenotype among siblings. *J Inherit Metab Dis*. 2014;37:131–135.

38. Kegley KM, Sellers MA, Ferber MJ, Johnson MW, Joelson DW, Shrestha R. Fulminant Wilson's disease requiring liver transplantation in one monozygotic twin despite identical genetic mutation. *Am J Transplant*. 2010;10:1325–1329.

39. Loudianos G, Dessi V, Lovicu M, et al. Molecular characterization of wilson disease in the Sardinian population--evidence of a founder effect. *Hum Mutat*. 1999;14:294–303.

40. Coffey AJ, Durkie M, Hague S, et al. A genetic study of Wilson's disease in the United Kingdom. *Brain*. 2013;136:1476–1487.

41. Génin E, Feingold J, Clerget-Darpoux F. Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet*. 2008;124:357–368.

42. Genin E, Feingold J, Clerget-Darpoux F. Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet*. 2008;124:357–368.

43. Schiefermeier M, Kollegger H, Madl C, et al. The impact of apolipoprotein E genotypes on age at onset of symptoms and phenotypic expression in Wilson's disease. *Brain*. 2000;123 Pt 3:585–590.

44. Gu Y-H, Kodama H, Du S-L. Apolipoprotein E genotype analysis in Chinese Han ethnic children with Wilson's disease, with a concentration on those homozygous for R778L. *Brain Dev*. 2005;27:551–553.

45. Kocabay G, Tutuncu Y, Yilmaz H, Demir K. Impact of apolipoprotein E genotypes on phenotypic expression in Turkish patients with Wilson's disease. *Scand J Gastroenterol*. 2009;44:1270–1271.

46. Chuang LM, Tai TY, Wang TR, et al. Esterase D and retinoblastoma gene loci are tightly linked to Wilson's disease in Chinese pedigrees from Taiwan. *Hum Genet*. 1991;87:465–468.

47. Shen X, Mizuguchi G, Hamiche A, Wu C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature*. 2000;406:541–544.

48. Laget S, Joulie M, Le Masson F, et al. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS ONE*. 2010;5:e11982.

49. Jung JS, Jee MK, Cho HT, et al. MBD6 is a direct target of Oct4 and controls the stemness and differentiation of adipose tissue-derived stem cells. *Cell Mol Life Sci*. 2013;70:711–728.

50. Stapelbroek JM, Bollen CW, van Amstel JKP, et al. The H1069Q mutation in ATP7B is associated with late and neurologic presentation in Wilson disease: results of a meta-analysis. *J Hepatol*. 2004;41:758–763.

51. Merle U, Weiss KH, Eisenbach C, Tuma S, Ferenci P, Stremmel W. Truncating mutations in the Wilson disease gene ATP7B are associated with very low serum ceruloplasmin oxidase activity and an early onset of Wilson disease. *BMC Gastroenterol*. 2010;10:8.

52. Merle U, Schaefer M, Ferenci P, Stremmel W. Clinical presentation, diagnosis and long-term outcome of Wilson's disease: a cohort study. *Gut*. 2007;56:115–120.

53. Ferenci P. Phenotype-genotype correlations in patients with Wilson's disease. *Ann N Y Acad Sci*. 2014;1315:1–5.

54. Haaxma CA, Bloem BR, Borm GF, et al. Gender differences in Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2007;78:819–824.

55. Miller IN, Cronin-Golomb A. Gender differences in Parkinson's disease: clinical characteristics and cognition. *Mov Disord*. 2010;25:2695–2703.

56. Zhou X, Edmonson MN, Wilkinson MR, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nature Genetics*. 2016;48:4.

Tables

Table 1. Details of all mutations present in more than one WD patient.

Mutation	Frequency	Consequence	Protein position	Amino acid change
ENST00000242839.4:c.3207C>A	215	missense_variant	1069	H/Q
ENST00000242839.4:c.4051C>T	26	stop_gained	1351	Q/*
ENST00000242839.4:c.3402delC	18	frameshift_variant	1134	P/X
ENST00000242839.4:c.2336G>A	14	stop_gained	779	W/*
ENST00000242839.4:c.2962G>C	5	missense_variant	988	G/R
ENST00000242839.4:c.2752G>A	4	missense_variant	918	D/N
ENST00000242839.4:c.2906G>A	4	missense_variant	969	R/Q
ENST00000242839.4: c.2996_3011delCCGGGGTGGCCGCGCA	4	frameshift_variant	999-1004	TGVAAQ/X
ENST00000242839.4:c.3284A>C	4	missense_variant	1095	Q/P
ENST00000242839.4:c.2532delA	3	frameshift_variant	844	K/X
ENST00000242839.4:c.4022G>A	3	missense_variant& splice region_variant	1341	G/D
ENST00000242839.4:c.1063C>T	2	stop_gained	355	Q/*
ENST00000242839.4:c.1915C>T	2	missense_variant	639	H/Y
ENST00000242839.4:c.1958C>A	2	missense_variant	653	S/Y
ENST00000242839.4:c.2038C>T	2	stop_gained	680	Q/*
ENST00000242839.4:c.2304dupC	2	frameshift_variant	768-769	-/X
ENST00000242839.4:c.2930C>T	2	missense_variant	977	T/M
ENST00000242839.4:c.2975C>T	2	missense_variant	992	P/L
ENST00000242839.4:c.3419T>C	2	missense_variant	1140	V/A

ENST00000242839.4: c.3472_3482delGGTTTAACCAT	2	frameshift_variant	1158- 1161	GLTI/X
ENST00000242839.4:c.3556+2T>C	2	splice_donor_variant		

Mutation – base change; frequency – frequency in patients; consequence – consequence according to VEP (Variant Effect Predictor).

Table 2. Mutation type and frequency.

Mutation	HOM	HET	COMPOUND	OTHER	ALL
c.3207C>A	112	18	85	0	215
c.4051C>T	2	2	22	0	26
c.3402delC	2	0	16	0	18
c.2336G>A	0	1	13	0	14
c.2962G>C	0	3	2	0	5
c.2752G>A	0	0	4	0	4
c.2906G>A	0	0	4	0	4
c.2996_3011delCCGGGGTGGCCGCGCA	0	0	4	0	4
c.3284A>C	0	0	4	0	4
c.2532delA	0	1	2	0	3
c.4022G>A	0	0	3	0	3
c.1063C>T	0	0	2	0	2
c.1915C>T	0	0	2	0	2
c.1958C>A	0	0	2	0	2
c.2038C>T	0	0	2	0	2
c.2304dupC	0	0	2	0	2
c.2930C>T	0	0	2	0	2
c.2975C>T	0	0	2	0	2
c.3419T>C	0	0	0	2	2
c.3472_3482delGGTTTAACCAT	0	1	1	0	2
c.3556+2T>C	0	0	2	0	2
c.3914T>C	0	0	1	0	1
c.3128T>C	0	0	1	0	1
c.1770dupT	0	0	1	0	1
c.3140A>G	0	0	1	0	1

c.3620A>G	0	0	1	0	1
c.1481T>A	0	0	1	0	1
c.2332C>G	0	1	0	0	1
c.3809A>G	0	0	1	0	1
c.3079G>T	0	0	1	0	1
c.4058G>C	0	0	1	0	1
c.2597A>G	0	0	1	0	1
c.2954G>A	0	0	1	0	1
c.3762G>C	0	0	1	0	1
c.2668G>A	0	0	1	0	1
c.2525A>T	0	0	1	0	1
c.4009C>G	0	0	1	0	1
c.3115delG	0	0	1	0	1
c.503T>C	0	0	1	0	1
c.3800A>T	0	0	1	0	1
c.2998G>C	0	0	1	0	1
c.3818C>T	0	0	1	0	1
c.3760G>C	0	0	1	0	1
c.2720A>G	0	0	0	1	1
c.3305T>C	0	0	1	0	1
c.1772G>A	0	0	1	0	1
c.4005dupG	0	0	1	0	1
c.3649_3654delGTTCTG	0	0	1	0	1
c.2165dupT	0	0	1	0	1
c.1112C>T	0	0	0	1	1
c.3557-2A>G	1	0	0	0	1
c.712T>C	0	0	1	0	1

c.2129G>C	0	0	1	0	1
c.940C>T	0	0	1	0	1
c.3229A>G	0	0	1	0	1
c.1415C>G	0	0	0	1	1
c.3140delA	0	0	1	0	1

HOM – homozygous mutation; HET – single heterozygous mutation; COMPOUND – part of at least a two-mutation construct; OTHER – mutation present in addition to homozygous change; ALL – overall frequency.

Table 3. Novel variants in the ATP7B gene, with HIGH or MODERATE impact on coding, according to VEP.

Variant	Freq	Consequence	IMPACT	Condel score	CADD score
c.2996_3011delCCGGGGTGGCCGCGCA	4	frameshift_variant	HIGH		35
c.3556+2T>C	2	splice_donor_variant	HIGH		24.7
c.1112C>T	1	missense_variant	MODERATE	Neu (0.250)	12.74
c.1481T>A	1	missense_variant	MODERATE	D (0.871)	32
c.1770dupT	1	frameshift_variant	HIGH		35
c.2597A>G	1	missense_variant	MODERATE	D (0.869)	28.5
c.2720A>G	1	missense_variant	MODERATE	D (0.863)	29.5
c.3115delG	1	frameshift_variant	HIGH		35
c.3229A>G	1	missense_variant	MODERATE	D (0.695)	27.8
c.3557-2A>G	1	splice_acceptor_variant	HIGH		23.8
c.3760G>C	1	missense_variant	MODERATE	Neu (0.454)	22.4
c.3762G>C	1	missense_variant	MODERATE	D (0.847)	25.9
c.4005dupG	1	frameshift_variant	HIGH		34
c.4009C>G	1	missense_variant	MODERATE	D (0.935)	25.5
c.712T>C	1	missense_variant	MODERATE	Neu (0.027)	0.119
c.940C>T	1	missense_variant	MODERATE	Neu (0.414)	23.34

Freq - frequency; Condel score – score in Condel (as reported by VEP): Neu – neutral, D – deleterious; CADD – score in CADD (on PHRED scale).

Table 4 Variants in other disease-associated genes, present in Wilson's disease patients

Variant	Gene	Freq	Clinical significance
ENSP00000003084.6:p.Arg1162Leu	CFTR	4	benign&likely_benign
ENSP00000003084.6:p.Arg31Cys	CFTR	1	
ENSP00000003084.6:p.Arg668Cys	CFTR	2	uncertain_significance
ENSP00000003084.6:p.Asp891Gly	CFTR	1	
ENSP00000003084.6:p.Glu695Gly	CFTR	6	
ENSP00000003084.6:p.Ile148Thr	CFTR	1	not_provided&benign
ENSP00000003084.6:p.Leu997Phe	CFTR	1	benign&likely_benign&pathogenic&risk_factor
ENSP00000003084.6:p.Ser1235Arg	CFTR	3	benign
ENSP00000003084.6:p.Ser912Leu	CFTR	1	pathogenic
ENSP00000003084.6:p.Val754Met	CFTR	1	likely_benign
ENSP00000269228.4:p.Gln60His	NPC1	2	
ENSP00000269228.4:p.Gly1073Ser	NPC1	1	likely_benign
ENSP00000269228.4:p.Pro237Ser	NPC1	15	benign
ENSP00000269228.4:p.Pro255Ser	NPC1	1	
ENSP00000269228.4:p.Val494Met	NPC1	1	

ENSP00000416066.1:p.Ala308Ser	SERPINA1	2	
ENSP00000416066.1:p.Arg247Cys	SERPINA1	1	uncertain_significance&pathogenic&other
ENSP00000416066.1:p.Arg63Cys	SERPINA1	1	uncertain_significance&likely_pathogenic&pathogenic&other
ENSP00000416066.1:p.Glu366Lys	SERPINA1	4	pathogenic&other
ENSP00000416066.1:p.Gly172Trp	SERPINA1	1	other
ENSP00000417404.1:p.Asn130Ser	HFE	1	
ENSP00000417404.1:p.Ser65Cys	HFE	7	uncertain_significance
ENSP00000451112.1:p.Pro21Gln	NPC2	1	

Variant – HGVS protein sequence name, Gene – gene symbol, Freq – frequency, Clinical significance – clinical significance according to VEP ClinVar annotations

Supplementary Material Tables

Supplementary Tables S1-S7:

Supplementary Table S1. Results of CONTRA software analysis, with region length, adjusted p-value and mean of log-ratio in the region.

Supplementary Table S2. Results of logistic regression, modeling disease subtype on ATP7B mutation type, consequence, age and sex with odds ratio, 95% confidence interval for odds ratio and p-value.

Supplementary Table S3. Results of logistic regression, modeling age at diagnosis on ATP7B mutation type, consequence and sex with odds ratio, 95% confidence interval for odds ratio and p-value.

Supplementary Table S4. Results of logistic regression, modeling disease subtype on the presence of rare variants in genes, associated with Wilson's disease according do STRING

disease database. Odds ratio are presented with 95% confidence intervals and p-values. Please note that modeling was conducted only for patients with at least one mutation present in those genes.

Supplementary Table S5. Results of logistic regression, modeling age of onset on the presence of rare variants in genes, associated with Wilson's disease according do STRING disease database. Odds ratio are presented with 95% confidence intervals and p-values. Please note that modeling was conducted only for patients with at least one mutation present in those genes.

Supplementary Table S6. Tests conducted for WD patients.

Supplementary Table S7. Position and score of variants in "TP53 Regulates Metabolic Genes" pathway.

Supplementary Dataset S1:

This dataset contains all the information used for *ATP7B* gene influence on phenotype modeling, in sex: 1 - male, 2 - female. In symptoms 1 - symptoms absent, 2 - symptoms present. N - neurological symptoms, H - hepatological symptoms. Age is given at the moment of diagnosis.

Supplementary Dataset S2:

This dataset contains all the information, used for modelling of rare variants influence on phenotype modelling where: CHROM – chromosome number, POS – position in hg19 assembly, REF, ALT – reference and alternate base, Consequence – variant consequence, SYMBOL, Gene – gene identifiers in HGNC and Ensembl respectively, EXON, INTRON – exon or intron number, HGVSc, HGVSp – HGVS coding sequence and protein name, Existing_variation – variant identifiers in known databases, including dbSNP, AF – allele frequency in 1000 genomes. All annotations were retrieved with Variant Effect Predictor.

Figure legends

Figure 1. Mutations found in the *ATP7B* gene. Visualization was performed with ProteinPaint.[55]