JOURNAL OF CHEMICAL INFORMATION AND MODELING

Subscriber access provided by Bibliothèque de l'Université Paris-Sud

Computational Biochemistry

Computing the Pathogenicity of Wilson's Disease ATP7B Mutations: Implications for Disease Prevalence

Ning Tang, Thomas Damgaard Sandahl, Peter Ott, and Kasper P. Kepp

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.9b00852 • Publication Date (Web): 21 Nov 2019

Computing the Pathogenicity of Wilson's Disease ATP7B Mutations: Implications for Disease Prevalence

Ning Tang[†], Thomas D. Sandahl[§], Peter Ott[§], and Kasper P. Kepp[†]*

[†] *Technical University of Denmark, DTU Chemistry, Kemitorvet 206, 2800 Kongens Lyngby, Denmark*

[§] *The Danish Wilson Centre, Medical Department LMT, Hepatology, Aarhus University Hospital, Palle Juul Jensens Boulevard 99, 8200 Aarhus, Denmark.*

*Corresponding author. E-mail: kpj@kemi.dtu.dk. Phone: +045 45252409

Abstract

Genetic variations in the gene encoding the copper-transport protein ATP7B are the primary cause of Wilson's disease. Controversially, clinical prevalence seems much smaller than prevalence estimated by genetic screening tools, causing fear that many people are undiagnosed although early diagnosis and treatment is essential. To address this issue, we benchmarked 16 state-of-the-art computational disease-prediction methods against established data of missense *ATP7B* mutations. Our results show that the quality of the methods vary widely. We show the importance of optimizing the threshold of the methods used to distinguish pathogenic from non-pathogenic mutations against data of clinically confirmed pathogenic and non-pathogenic mutations. We find that most methods use thresholds that predict too many ATP7B mutations to be pathogenic. Thus, our findings explain the current controversy on Wilson's disease prevalence, because meta analysis and text search methods include many computational estimates that lead to higher disease prevalence than clinically observed. Since proteins and diseases differ widely, a one-size-fits-all threshold cannot distinguish efficiently pathogenic and non-pathogenic mutations, as shown here. We also show that amino acid changes with small evolutionary substitution probability, mainly due to amino acid volume, are more associated with disease, implying a pathological effect on the conformational state of the protein, which could affect copper transport or ATP recognition and hydrolysis. These findings may be a first step towards a more quantitative genotype-phenotype relationship of Wilson's disease.

Keywords: Wilson's disease; ATP7B mutations; pathogenic mutations; sequence conservation; amino acid volume; copper transport

Introduction

Wilson's disease (WD) is a rare autosomal recessive disorder of copper metabolism caused by pathogenic variants of the human *ATP7B* gene encoding the ATP7B protein, which is a copper-transporting P-type ATPase.[1,2] The approximately 160-kDa membrane protein contains a large N-terminal domain consisting of six metal-binding domains, eight transmembrane segments (TMs), an adenosine triphosphate (ATP) binding domain, and a soluble C-terminal tail.[3–8] In the hepatocyte, ATP7B transports copper from the cytosol into the Golgi apparatus and mediates either the incorporation of copper into ceruloplasmin or the excretion of excess copper into the bile.[9,10] To accomplish its copper transport function, the protein depends critically on its ability to use the energy gained by ATP hydrolysis.[3]

Pathogenic *ATP7B* mutations cause loss of copper transporting function resulting in accumulation of copper in multiple organs, most notably brain, liver, and kidney.[11–13] As a result, WD patients present with either hepatic, neurologic or psychiatric symptoms, often in combination.[14,15] If left untreated, WD with chronic presentation is fatal within a 5−10 year period from first symptom onset. However, apart from the rare acute fulminant hepatic presentation that requires liver transplantation, proper medical treatment can typically ensure a near-normal life expectancy, but this depends critically on early and accurate diagnosis.[16–19]

The clinical handling of patients with WD faces two major challenges; the uncertainty of the prevalence of the disease (the number of affected people within a given population) and the lack of a clear genotype-phenotype relation that enables the estimate of disease severity and manifestation. The worldwide prevalence has been estimated to be around 1 in 30,000[20], corresponding to a carrier frequency of approximately 1 in 90 with slightly higher numbers reported in areas where diagnostic awareness of WD is high, such as Austria[21] and France.[16,22–24] Due to the heterogeneity in the clinical presentation and the age of presentation, a substantial number of patients are undiagnosed.[25] Recent population-genetic studies based on the computer analysis of observed variants have led to estimates

of WD prevalence of 1:7,026[15], 1:7,200[26] or 1:4,000[27]. If true, these studies suggest that at least 75% of people affected by the disease are undiagnosed with potentially fatal consequences.

However, the difference between clinical observations and genetic predictions may also raise questions regarding the validity of current computational methods widely used to examine disease mechanisms and categorize pathogenic and non-pathogenic variants.[28–34] These predictive methods use the evolutionary "unlikeliness" of the amino acid substitution, the involved change in biochemical properties, and/or 3-dimensional protein structure to classify mutations. Loss of protein stability potentially leading to partial loss of function is a common feature of many inherited diseases. In such cases, structure-based computational methods can identify pathogenic mutations.[35] Alternatively, evolutionary conservation information may capture disruptive amino acid changes that are unlikely to occur during natural evolution as they typically impair protein conformational integrity or function.[36]

From a molecular evolution perspective, we expect that a majority of naturally occurring protein variants are nearly neutral in their functional effect, which is the basis for the so-called neutral theory of evolution and the widely supported use of molecular clocks in phylogeny[37]. In the clinical terminology[38], these probably tend to be the benign variants. This insight is further complicated by penetrance being modulated by non-genetic and genetic confounding factors. We hypothesize that, since proteins differ widely in size, shape, stability, location, and natural function, the impact of a typical human mutation will be very protein-dependent. For example, abundant proteins are known to evolve much more slowly than less abundant proteins due to selection pressures, and evolution rate is also very dependent on specific selection pressures of the protein[39–42]. Yet most methods suggest a default threshold to distinguish pathogenic from non-pathogenic mutations based on conservation patterns. From a clinical strategic perspective there is an urgent need to solve this issue and identify methods that correctly distinguish truly disease-causing mutations from benign (neutral) variants[38], and possibly also the severity and penetrance of the pathogenic mutations.

As ATP7B is the only identified gene known to cause WD, genetic screening for known pathogenic variants is a sensitive approach to diagnose WD. However, in cases where the functional impact of a variant is unclear, genetic testing only provides circumstantial evidence, as variants display diverse functional effects. Further complications such as life-style and environmental risk modifiers and the low frequency and unknown penetrance of the mutations complicate diagnosis even further.[43] Direct functional testing of disease-causing ATP7B mutations would ideally be the most sensitive method to diagnose WD, but such functional tests are time-consuming. Clarification of these issues may affect the use of genetic testing for diagnosing patients suspected of having WD[16] and may also have implications for conclusions based on genetic population screenings.

Genotype-phenotype relations would aid our understanding of the pathophysiology and the development of new diagnostic and therapeutic strategies. Such relations have so far met with little success.[17] More than 700 ATP7B natural variants have been identified, including mostly missense mutations, insertions/deletions, and some rare splice-site mutations as summarized in the WD database (http://www.wilsondisease.med.ualberta.ca).[44] Whereas truncating mutations tend to severely impair copper metabolism and cause early age of disease onset, some mutations do not relate to hepatic or neurologic presentations, supporting our assumption of a pool of disease-wise benign natural ATP7B variants.[45,46] For missense mutations the genotype-phenotype relationship is even weaker as they cause a wide variety of symptoms in WD patients implying that they may affect ATP7B function in different ways.[47,48] These missense mutations are distributed across the ATP7B gene, but tend to cluster in the ATP binding domain indicating its importance for the ATP-dependent copper transport function.[49] There is considerable phenotypic variation between individuals with the same mutation, even within the same families and in monozygotic twins[50,51] showing clearly the need for understanding the risk modulation effects of specific mutations on ATP7B functionality and clinical presentation.

In the present study we performed a detailed computational study of ATP7B protein variants using 16 widely used structure- and sequence-based methods with the specific aims i) to test the

application of state of the art computational screening methods to the problem of WD where diagnosis is challenged; ii) to identify amino acid properties that correlate with disease presentation. We show that several sequence-based methods can accurately classify pathogenic variants if the threshold is optimized before quantitative diagnosis of WD. However, outcomes are extremely dependent on the thresholds used, and default thresholds tend to overestimate disease prevalence. This finding largely explains the discrepancy between genetic-screening based and clinically observed WD prevalence. We also identify several important chemical features that determine whether a variant is disease-causing or not, which may aid the so far unsuccessful understanding of WD genotype-phenotype relations.

Computational Methods

Data for ATP7B genetic variants

We studied several data sets of ATB7B variants related to WD, but ultimately settled on using the mutations from the WD database[44] (http://www.wilsondisease.med.ualberta.ca) for reasons described below. As most of the studied methods can only deal with missense mutations, only these were selected for investigation, which substantially reduces the number of relevant data points and affects the data set choice. There are two main concerns: 1) the completeness of the dataset and 2) the confidence in the assignment of the clinical impact of each variant.

When comparing to the most recent 2019 database by Gao et al.[26] we found that the WD database includes almost all variants with high confidence of pathogenicity according to the more recent criteria by Richards et al.[38] We tested the sensitivity of our conclusions by including the most recent variants from 2019.[26] As shown below, this did not affect our conclusions, mainly because the confidently assigned variants have changed little compared to the major increase in the total number of inferred variants from genome screening, since functional and clinical testing has not experienced the same growth in capacity as sequencing and computational screening tools.

Many of the most confidently assigned loss-of-function variants are not missense mutations (e.g. deletions) and not studied by the applied methods; our analysis deals mainly with the difficult grey zone of missense mutations that are commonly nearly neutral (benign) and are the cause of the current controversy on disease prevalence. Gao et al.[26] use broad screening approaches (including text search and meta analysis) to maximize completeness at the expense of confidence in the assignment of pathogenicity. The new data thus contain many computational estimates of pathogenicity; these estimates emerged mainly during the last decade and in the case of WD, after the WD database was complete. We find that the WD database, by minimizing recent computational and low-confidence screening results, is optimal for the analysis that we conduct here, where we

specifically want to avoid pollution by computational screening estimates in the benchmark data. We discuss the insensitivity of our findings to reasonable variations in data set later in this paper.

The WD dataset includes 722 and 172 entries for pathogenic and non-pathogenic variants, respectively; of these, 291 variants are unique missense mutations studied in the present work. Clinical effects from loss of function studies (class 2) provide the best evidence for these mutations[44]. Since WD is a loss-of-function disease, functional studies provide a strong support for pathogenicity in particular in combination with control data for normal people (class 4). We estimate that, compared to the classification by Richards et al.[38], which was not available and thus not used in the WD database, the loss-of-function feature makes the confidence of pathogenicity approximately strong (PS), which is the best possible situation as statistical critical mass of data points is still required. Among the 291 missense mutations studied, there are 267 pathogenic mutations and 24 non-pathogenic mutations. In addition, 15 phenotypic mutations were reported as both pathogenic and non-pathogenic; these mutations were not included in our dataset as their pathogenicity is variable or debated. Details of the used mutation data set are shown in **Table S1**.

Studying mutations by computational mutagenesis using structure-based methods

Since the full ATP7B protein structure is not available, homology models will be unreliable, although additional (e.g. evolutionary) constraints can improve models, and work towards larger more reliable models of ATP7B is thus ongoing[52]. We used here the available NMR structures of several domains to perform structure-based mutation analysis where possible. A major issue is the quality of the protein structure input and its relevance to the real pathogenic process. As shown previously[29,31,53,54] the $\Delta\Delta G$ (the change in stability caused by point mutation) is very structure-dependent for some stability-prediction methods but less so for others, and this means that more methods and structures should be compared whenever possible. Five structures were used with the Protein Data Bank (PDB) IDs 2N7Y, 2LQB, 2ROP, 2EW9 and 2ARF corresponding to metal-

binding domains 1 (2N7Y), 2 (2LQB), 3 and 4 (2ROP), 5 and 6 (2EW9), and the ATP binding domain (2ARF), respectively.[55–59]

For FoldX (version 5),[60] the structures were first repaired using the RepairPDB function, and then the BuildModel function was applied to the repaired structures to obtain ΔΔG values from five independent runs. The final reported ΔΔG values were the averages of these five independent runs for each mutation. For Rosetta (2019.07.60616 weekly release version), the structures were relaxed to produce 20 optimized structures. The structure with the lowest score was then applied in the Cartesian version of the Rosetta protocol with three iterations.[61,62] The final ΔΔG values were calculated based on the difference in total scores averaged over three rounds for the mutant and wild type structures. For I-mutant (version 2.0),[63] the secondary structure of the different domains was calculated using the DSSP algorithm.[64] For mCSM,[65] SDM,[66] and DUET,[67] the web server versions of the programs were used with default settings for calculating the ΔΔG values using the original NMR ensemble as input. Four mutations (S291A, S291I, S291Q and M573H) could not be computed by SDM and DUET and were thus not included in the final SDM and DUET results. For POPMUSIC,[68] HOTMUSIC,[69] and SNPMUSIC,[70] the calculations were performed using the DEZYME (http://www.dezyme.com) platform using the original NMR ensemble as input.

As different ΔΔG sign conventions are used for labeling the stabilizing and destabilizing mutations, the sign was adjusted (ΔΔG < 0 stabilizing, ΔΔG > 0 destabilizing) in the present study in order to enable clear comparison. Only the values for the mutants relative to wild-type are of interest, as the absolute values are not very meaningful. A short summary of the used structure-based methods is given in **Table S2**.

Mutation analysis using sequence-based methods

The ATP7B protein sequence was obtained from Uniprot (ID P35670 ATP7B_HUMAN). The obtained sequence was then used to perform saturated mutagenesis with several state-of-the-art sequence-based disease-prediction methods, EASE-MM,[71] PolyPhen-2,[30] SIFT,[72] Envision,[73]

PROVEAN,[74] SNAP.2,[75] and FATHMM[32], using the corresponding default settings. For all the sequence-based methods, only the final scores were collected and used for analysis. A short summary of the used sequence-based methods is also given in **Table S2**.

Identifying disease causing mutations using ATP7B protein conservation analysis

The degree of evolutionary conservation of an amino acid in a protein reflects a balance between its natural tendency to mutate and the overall need to retain the structural integrity and function of the macromolecule. Conservation analysis was performed using the ConSurf server designed for estimating the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule based on comparison to homologous sequences.[76] The HMMER method was used for homology search with an E-value cutoff of 0.0001against the UNIREF-90 protein database.[77] The homologs were selected for analysis based on the ConSurf server default criteria, and the resulting allowed amino acid variation at each position was used in the mutation pathogenicity analysis. In addition, the residue classification (buried/exposed) was also calculated by ConSurf since the complete ATP7B structure is unknown and the available structures represent different smaller parts that interact in unknown ways, producing many sites that are not exposed in the full protein. Solvent exposure dependencies of the classifications may provide insight not available from the total set of mutations, the typical example being disruptive mutations being more hydrophilic inside the protein but more hydrophobic on the protein surface.

Co-variation analysis using GREMLIN

In order to analyze the functional effects of the mutations in ATP7B, we also built a global statistical model based on the ATP7B multiple sequence alignment to estimate the log-likelihood of any given ATP7B variant. The multiple sequence alignment was established using HHblits in HHsuite with an E-value cutoff of 10^{-10} and four iterations against the uniclust30_2018_08 database.[78] The obtained multiple sequence alignment was further filtered by removing the sequences (rows) that

have more than 50% gap, resulting in 767 homologs. The global statistical model was established based on the obtained ATP7B multiple sequence alignment using the GREMLIN (Beta version 2.1) algorithm implemented in Tensorflow kindly provided by Dr. Sergey Ovchinnikov (FAS Center for Systems Biology, Harvard University).[79–81] GREMLIN enables the production of statistical models based on both coevolution and conservation data of homologs, which reveals residue contacts and can thus be a powerful estimator of structural-function impacts of amino acid mutations. Please note that the pseudo likelihood optimization in Tensorflow was performed with Adam optimizer because the "LBFGS" optimizer in the original Matlab version of GREMLIN is slow in Tensorflow. The difference in log-likelihood between the wild type and mutant variant allows us to explore the variant pathogenicity considering both site conservation and pairwise co-varying positions, which in general has better accuracy than analyzing each site independently. Similar methods have been previously successfully applied to identify disease-causing mutations.[82]

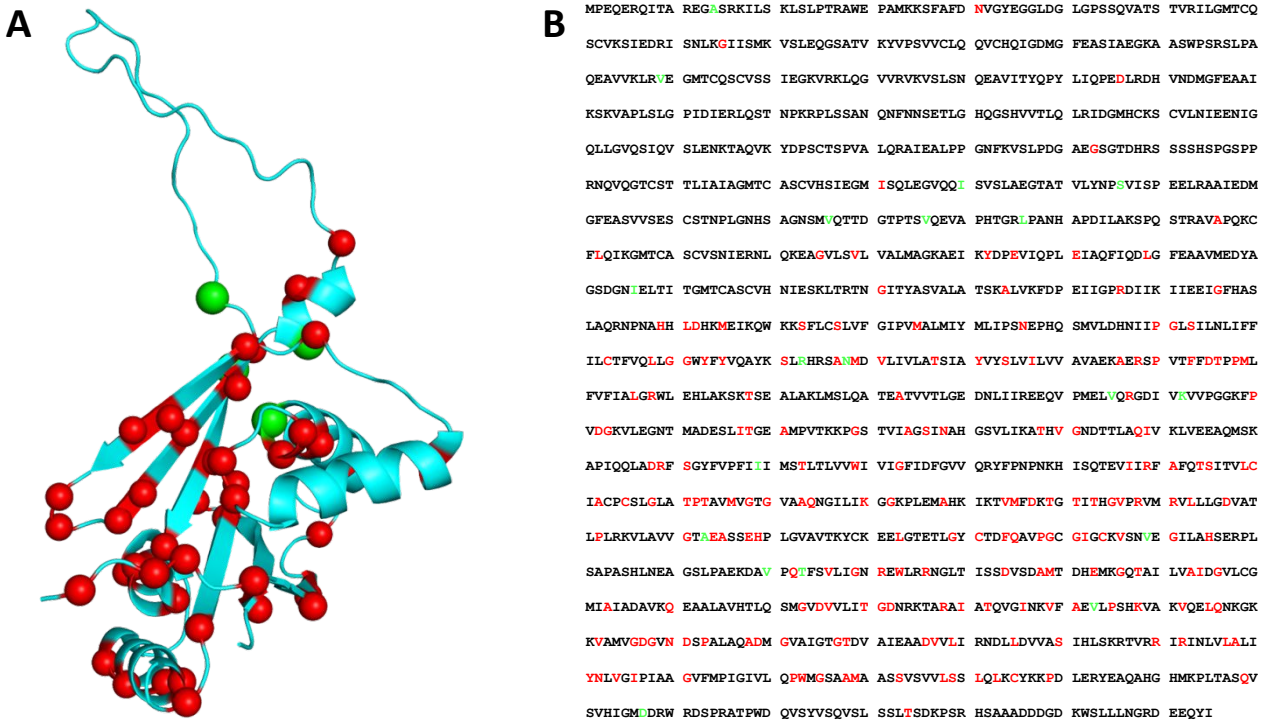


Figure 1. Example structure and full sequence used for in this study. (**A**) The NMR structural model of ATP binding domain (PDB ID 2ARF) as example of used NMR structures of different ATP7B protein domains with pathogenic mutations marked in red and non-pathogenic mutations marked in green. (**B**) The used canonical sequence of ATP7B obtained from Uniprot (ID P35670, ATP7B_HUMAN) with pathogenic mutations shown in red and non pathogenic mutations in green.

Results and Discussion

Structure- and sequence-based estimation of ATP7B variant pathogenicity

The missense ATP7B mutations could potentially exert pathogenicity in many ways and only in some relevant structural contexts. They may directly impair the ATP7B copper transport or the ATP binding or hydrolysis activity by mutation in the respective binding sites, but could also generally destabilize the membrane protein, partly dissociate it from the membrane or disrupt the trafficking of the protein to the membrane. Random missense mutations are more likely to reduce the thermochemical stability and typically impair the free energy of folding by 1 kcal/mol compared to the wild type, because the protein stability represents an optimized system.[31,83–86] Thus, loss of protein stability leading to excessive protein degradation and loss of functional copies of ATP7B protein available for copper transporting is a likely mechanism for WD.

To distinguish between different pathogenic mechanisms, we employed a wide range of both structure-based stability estimation methods and sequence-based disease prediction methods to estimate the consequence of mutations in ATP7B, and studied a broad range of chemical properties potentially affected by mutation. The relevant biological context where the mutations impair the protein may be very hard to model and as such the amino acid properties, which are context-independent, serve as important supplementary test cases. We used five available NMR structures of different ATP7B protein domains, with the partial structure of the important ATP binding domain shown for illustration in **Figure 1A.** To ensure a valid control test data set, we used the method of exhaustive control mutagenesis[87] by introducing all possible single-site amino acid substitutions into the wild type sequence, and using the distribution of scores as a control set in an analysis of variance (ANOVA), since properties of pathogenic mutations are meaningless by themselves if not compared to a random or non-pathogenic control set. Our exhaustive mutation control dataset for structure-based methods comprised 11,305 different ATP7B mutations after mapping to the original sequence numbering shown in **Figure 1B** (Uniprot P35670).

The computed values for all methods are shown in **Figure 2** with positive ΔΔG values indicating a destabilizing effect on the protein (signs were aligned for all ΔΔG methods to enable comparison). As seen from **Figure 2**, most of the pathogenic mutations destabilize the protein structure. However, the non-pathogenic mutations have similar effects, as expected for randomly introduced mutations,[88] clearly showing why conclusions cannot be drawn from a set of pathogenic mutations without a proper control. According to the ANOVA summarized in **Table S3**, the destabilization of the non-pathogenic and pathogenic mutations is not significantly different.

Current protein stability calculators are not expected to be as accurate for membrane proteins as for soluble proteins according to previous benchmarks, since the membrane environment contributes to the stability effect of the mutation;[89] the fact that ATP7B is a membrane protein could by itself affect the reliability of methods recently used to argue for very high WD prevalences[15,26,27]. Furthermore, some of these methods are structure-dependent and the structural input thus affects outcome substantially, with many snapshot structures required to generate an ensemble in agreement with experiment.[53,54,90] Accordingly, we also tested the sequence-based protein stability predictor EASE-MM, which does not rely on structure and thus is not impaired by potential weaknesses in the NMR structures. However, as seen from **Figure 2** and **Table S3**, it also produced insignificant differences in destabilization of pathogenic and control groups. We conclude that pathogenic ATP7B mutations destabilize the protein broadly, but the mutations are not more destabilizing than random mutations in the same protein. This does not rule out that destabilization can contribute to disease as it is a consistent feature of the mutations.

As an alternative to the hypothesis that thermodynamic destabilization of the folded state drives pathogenicity, we also tested whether sequence-based disease predictors using evolutionary conservation information (likelihood of substitution) can better describe pathogenicity. These tools capture important disruptive effects of mutations on the functional folded protein by considering the magnitude of the chemical perturbation and have the advantage of being applicable to all proteins[32,33] and several of them describe pathogenicity well in independent benchmarks.[29] Because

such computational methods are error-prone, using only one or two of them is not appropriate as conclusions may reflect the specific choice of methods.

We selected six methods (PolyPhen-2, SIFT, PROVEAN, Envision, SNAP.2 and FATHMM) to study the ATP7B variants. The resulting scores in **Figure 2** for non-pathogenic and pathogenic mutations are well-separated for all six methods except FATHMM. However, such conclusions can be deceptive due to biases in the thresholds, and thus one cannot conclude anything without statistical significance tests. If clinically confirmed non-pathogenic mutations are available at substantial count (> 10) they serve as a relevant test data set in a t-test or ANOVA. Since many data sets do not effectively separate non-pathogenic (benign) variants from confirmed pathogenic variants, we have advocated the use of exhaustive computational mutagenesis as a control data set for t-tests and ANOVA[87] to test whether a chemical property is significantly different for pathogenic and randomly occurring mutations. Such tests are easily performed using computational methods and provide a statistical quality that is hard to obtain experimentally, due to the cost of functional assaying of thousands of possible mutants. The ANOVA results (**Table S3)** show that the mean values of the obtained scores differ significantly between non-pathogenic and pathogenic categories at the 99% confidence level ($p < 0.01$) except for the FATHMM method.

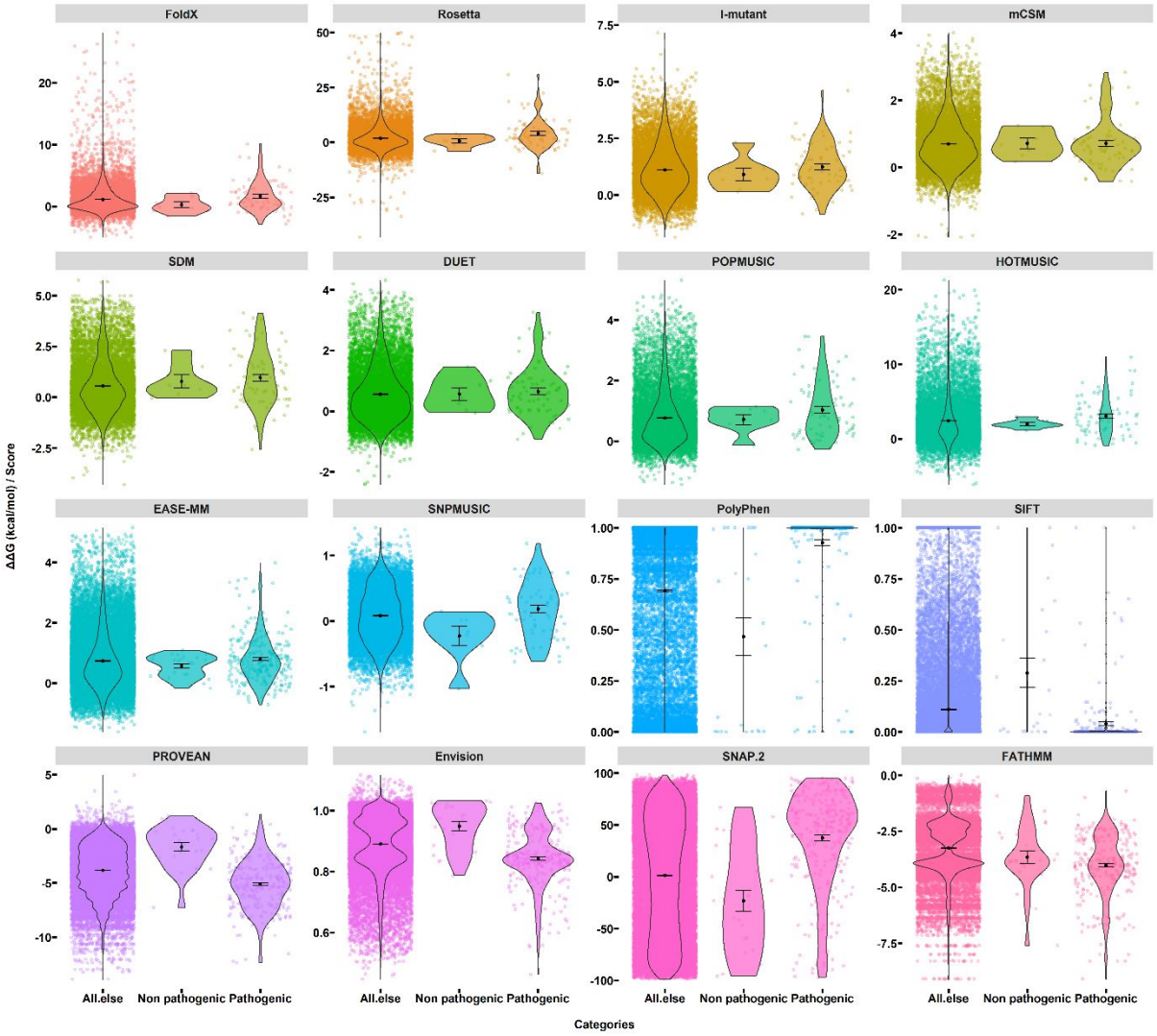


Figure 2. The ΔΔG values/scores for structure-based methods (FoldX, Rosetta, I-mutant, mCSM, SDM, DUET, POPMUSIC, HOTMUSIC, and SNPMUSIC) and sequence-based methods (EASE-MM, Polyphen-2, SIFT, PROVEAN, Envision, SNAP.2, and FATHMM) applied to 291 missense ATP7B mutations. The background dots with jitter function represent the obtained values. The black lines represent the distribution of values. Black dots represent the mean values. The stability methods have signs aligned, whereas the sequence-based methods show pathogenicity relative to their default ranges.

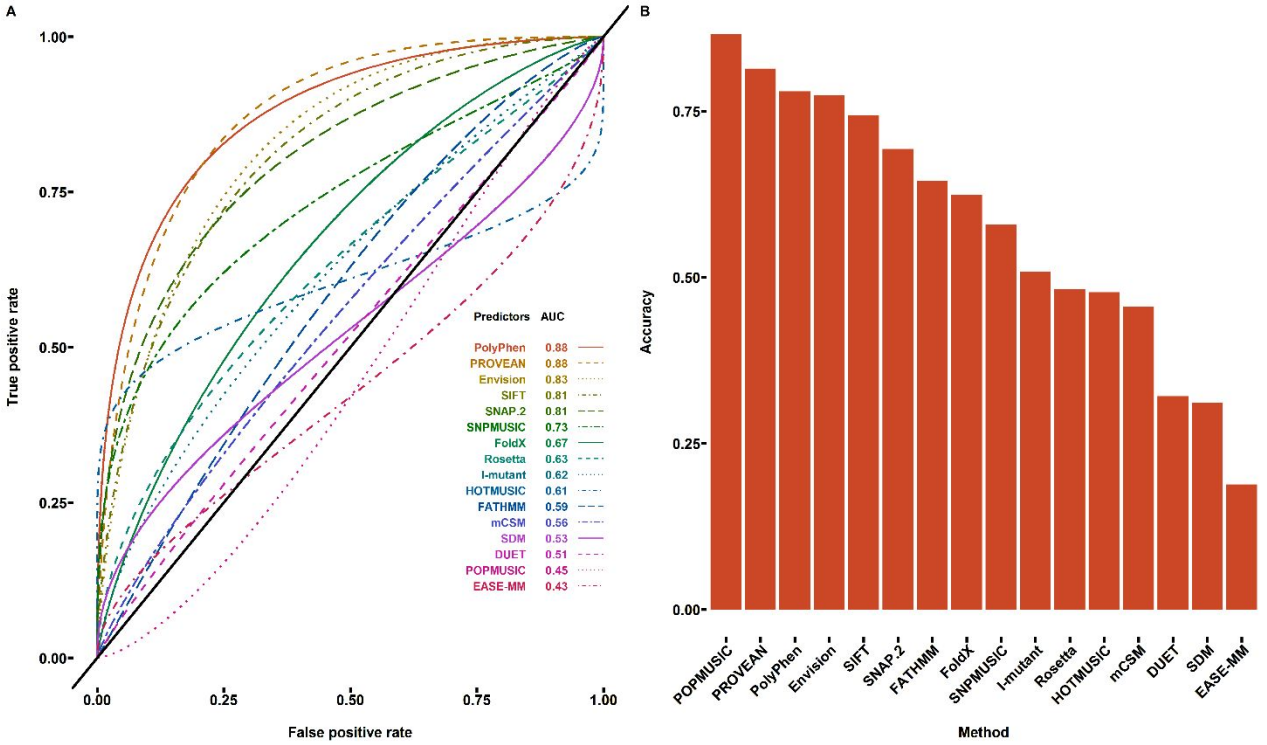


Figure 3. The ROC analysis of methods applied to 291 missense ATP7B mutations. (**A**) The ROC plot of the benchmarked methods for identifying the pathogenicity of the ATP7B protein mutations. (**B**) The identification accuracy of the used methods obtained from ROC analysis.

In order to test the ability of the methods to distinguish non-pathogenic and pathogenic *ATP7B* mutations, we performed a receiver operating characteristic analysis (**Figure 3A)**. In agreement with ANOVA results, the top-5 methods were all sequence-based methods. PolyPhen-2 and PROVEAN produced the highest area under curve (AUC) values of 0.88. Combined with the accuracy results in **Figure 3B**, PROVEAN was slightly (but insignificantly) more accurate than PolyPhen-2. Other properties of the ROC analysis are shown in **Table S4**. Notably, bootstrap analysis (95% confidence level, 2000 bootstrap replicates) shows that the optimized thresholds are generally robust to reasonable data variations, with the best-performing methods such as PROVEAN, Polyphen, and SIFT also well-determined optimized thresholds; whereas the thresholds of some other methods are very sensitive to data variations (**Table S4**).

The more accurate methods also predict well the pathogenicity of membrane protein mutations, indicating their value when structures are elusive.[29] One of the most important findings however is

that the optimized threshold values for distinguishing pathogenic and non-pathogenic mutations differ widely from the default values of the methods (**Table 1**). This shows that these widely used methods need to use thresholds optimized toward the specific protein of interest, or a class of proteins of related structures and diseases.

Correlations between clinical pathogenicity, amino acid properties, and allele frequency

In order to understand how ATP7B mutations confer pathogenicity and to identify valid semi-quantitative prediction tools of ATP7B variant pathogenicity, we analyzed the relationship between the clinically established pathogenicity and changes in 48 amino acid properties previously analyzed in similar work[91] (summarized in **Table S5**). We calculated the mutation-induced change in property using the equation $\Delta P = P_{mut} - P_{wt}$ for both non-pathogenic and pathogenic mutations. Representative results are shown in **Figure 4**, with remaining results shown in **Figure S1**. As illustrated in **Figure 4** and **Figure S1**, most of these properties show no significant separation between non-pathogenic and pathogenic mutations. Based on the ANOVA (**Table S6**) only changes in amino acid bulkiness (B1 in **Figure 4**), the ratio of the side chain volume to its length, differed significantly (95% confidence) for the two categories of mutations, with the pathogenic mutations displaying larger changes in amino acid bulkiness. Hydrophobicity changes, which relate to aggregation propensity and drive the pathogenicity of some other disease-causing mutations[87,92–96] were not important for the ATP7B variants.

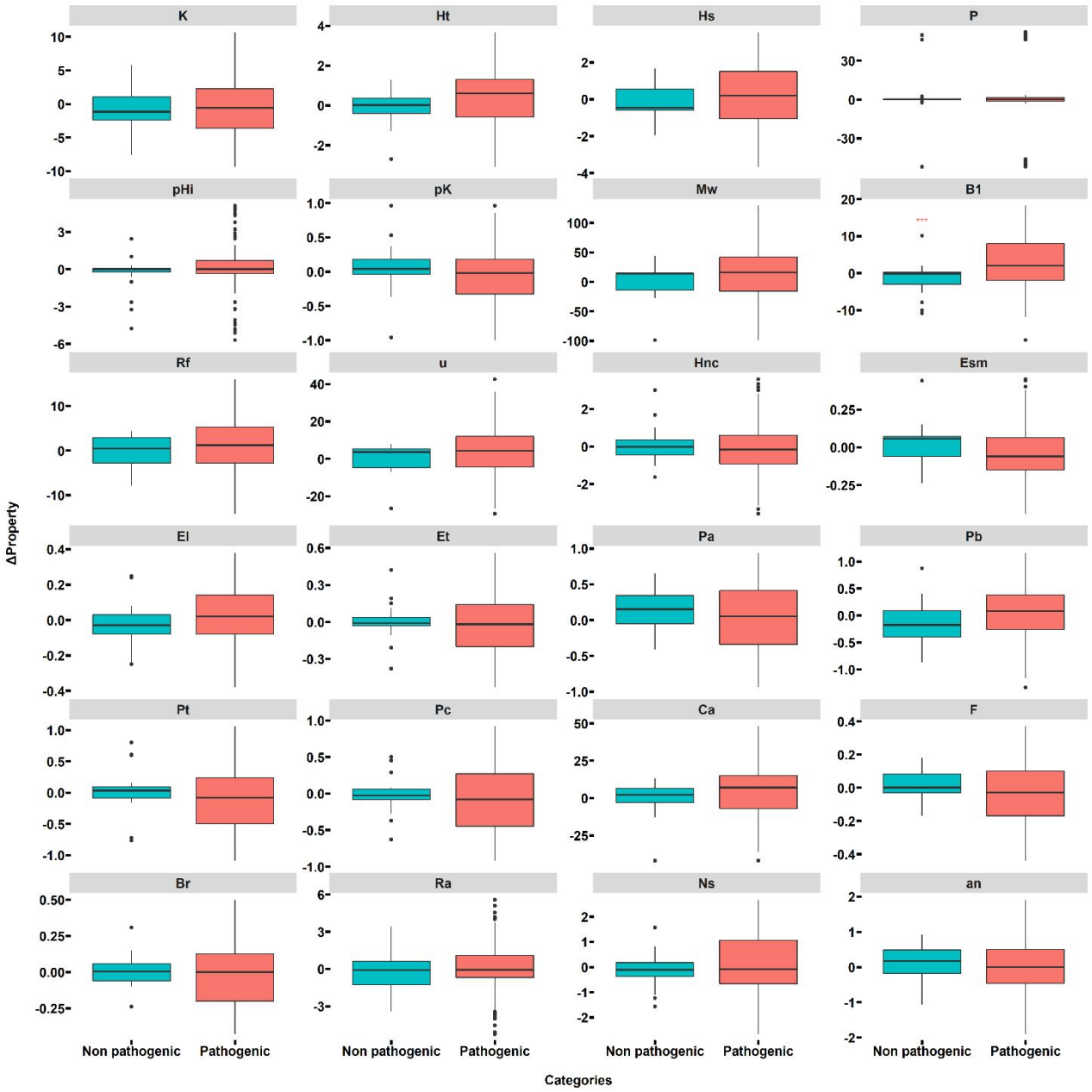


Figure 4. Amino acid property changes for 291 missense ATP7B non-pathogenic and pathogenic ATP7B mutations. Thick bars indicate the medians; the edges of the color-filled rectangles represent the 25th and 75th percentiles. The black dots represent outliers of the range covered by the black bars. *** for B1 indicates a p-value < 0.05. K: compressibility; Ht: thermodynamic transfer hydrophobicity; Hp: surrounding hydrophobicity; P: polarity; pHi: isoelectric point; pK: equilibrium constant for ionizing COOH; Mw: molecular weight; B1: bulkiness; Rf: chromatographic index; u: refractive index; Hnc: consensus hydrophobicity; Esm: short- and medium-range non-bonded energy; El: long-range non-bonded energy; Et: total non-bonded energy

(Esm + El); Pa, Pb, Pt, and Pc are α-helix, β-strand, turn, and coil tendencies; Ca: helical contact area; F: fluctuation displacement; Br: buriedness; Ra: solvent-accessible surface reduction; Ns: average number of surrounding residues; an: empirical tendency of the amino acid to be N-terminal.

These findings point towards two very general types of molecular pathogenicity; one that manifests in soluble proteins subject to aggregation toxicity, and one that manifests in membrane proteins and some soluble allosteric proteins by conformation changes affecting function. Previous work suggests that the amino acids bulkiness defines the local conformation and dynamics of natively folded proteins relevant to normal and pathological processes.[97] Mutation at position 653 of ATP7B indicated that bulky or charged amino acids mimic the phenotype of WD mutations, while small neutral substitutions do not, suggesting that the bulky substitutions distort the ATP7B protein conformation and thereby its function.[98] Our quantified changes in bulkiness significantly separate pathogenic and non-pathogenic ATP7B mutations for our large data set, suggesting that this hypothesis[98] is valid for the ATP7B mutations broadly and not just in single cases.

The database tool gnomAD[99] estimates the combined allele frequency of ATP7B variants in general populations and thus enables an analysis of the likely natural selection on missense ATP7B mutations across the allele frequency spectrum. The allele frequency is commonly used for clinical diagnostic filtration because disease-causing mutations are, all-else being equal, expected to be selected against. To understand the relationship between pathogenicity and allele frequency for WD mutations, the correlations between screening scores (PolyPhen-2, PROVEAN, Envision, and SNAP.2) and allele frequency for the ATP7B missense mutations found in gnomAD were analyzed. The results in **Figure 5** show that mutations with very high allele frequency are more likely to be benign as correctly predicted by these four methods. Moreover, some non-pathogenic mutations that are misidentified by these methods are relatively frequent, suggesting that incorporation of this information could improve the prediction of the pathogenicity of ATP7B variants.

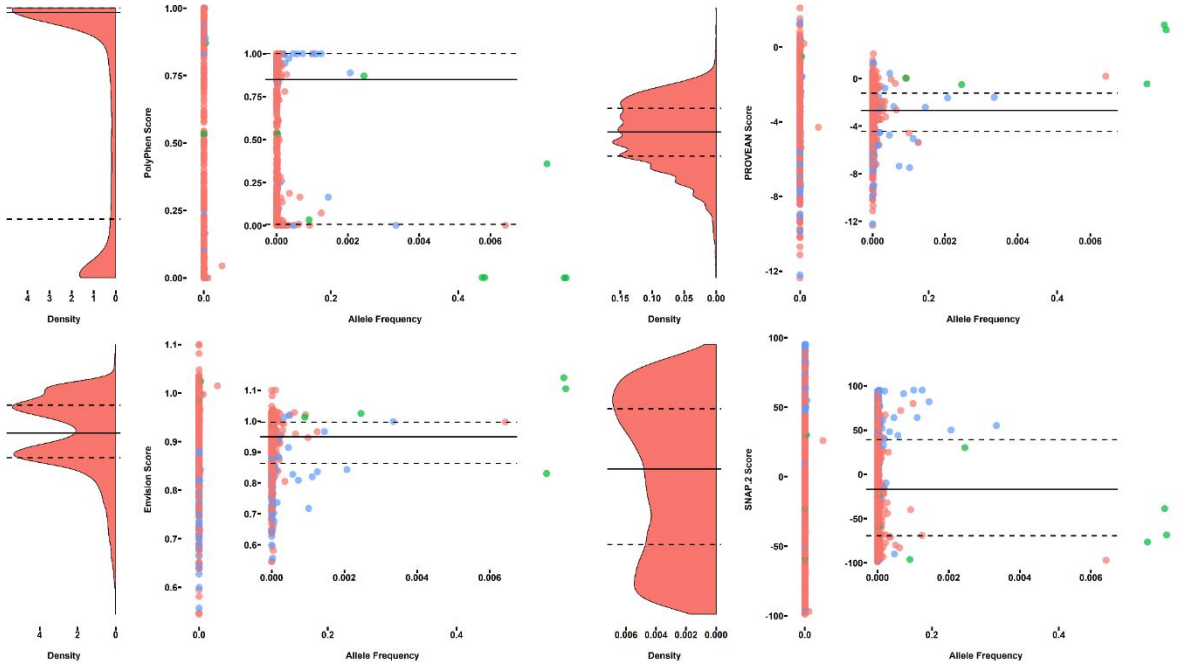


Figure 5. The correlation between PolyPhen, PROVEAN, Envision, and SNAP.2 scores and allele frequency. The plots on the left side in each category are the density plots of the scores. The inset plots on the right side are the zoom-in correlation plots with the allele frequency lower than 0.009. The red, green and blue dots represent the ATP7B protein mutations with unknown pathogenicity, non-pathogenic mutations, and pathogenic mutations, respectively.

Conservation and co-variation analysis

As we have shown above, several computational methods are capable of predicting the pathogenicity of ATP7B variants at a promising level of accuracy, but only after optimization of their thresholds against a clinically confirmed data set. These sequence-based methods all use evolutionary conservation information. To understand the contribution of this feature in more detail, we analyzed the evolutionary conservation of amino acid positions in ATP7B based on comparison to known ATP7B protein homologs. We extracted the allowed amino acid variations at each position and used this information as a predictor to identify pathogenicity. **Figure 6** shows the results. Simply using the evolutionary conservation information enables good performance with an AUC value of

0.86. This result clearly indicates that conservation information captures an important part of the pathogenic effect on ATP7B protein function.

Previous studies have indicated that most pathogenic mutations occur in buried sites of proteins or in surface-sites involved in molecular interactions.[100–105] It is thus of interest to know the performance of the computational methods for different types of sites. To explore this, we collected the residue classification information (buried/exposed) based on the evolutionary conservation analysis and divided the ATP7B protein residues into buried and exposed. We note that the structural assignment of exposure is uncertain because there is no complete ATP7B structure, and each NMR structure only reflects a small domain of the total membrane protein, and thus, many parts of the domains are exposed to other domains rather than water. We performed the ROC analysis for all the used methods for each residue category. As shown in **Figure 7**, it is very clear that most methods perform better for buried residues in agreement with our previous study.[29] Interestingly, we also found that most non-pathogenic mutations are variable and thus predicted to be exposed, consistent with a neutral effect on protein function.

We also analyzed the amino acid property changes for both residue categories using ANOVA (**Table S7**). The difference in the change of amino acid bulkiness for non-pathogenic and pathogenic mutations is only significant for exposed residues, but this may be due to the fact that most of the non-pathogenic mutations are exposed, making the test more assertive for this category. We also find that the β-strand tendency change is significantly different for non-pathogenic and pathogenic mutations but less clearly so than the amino acids bulkiness. Changes in β-strand tendency upon mutation has been shown to correlate with protein stability change in some proteins[91] and contribute to the aggregation propensity of some proteins.[106]

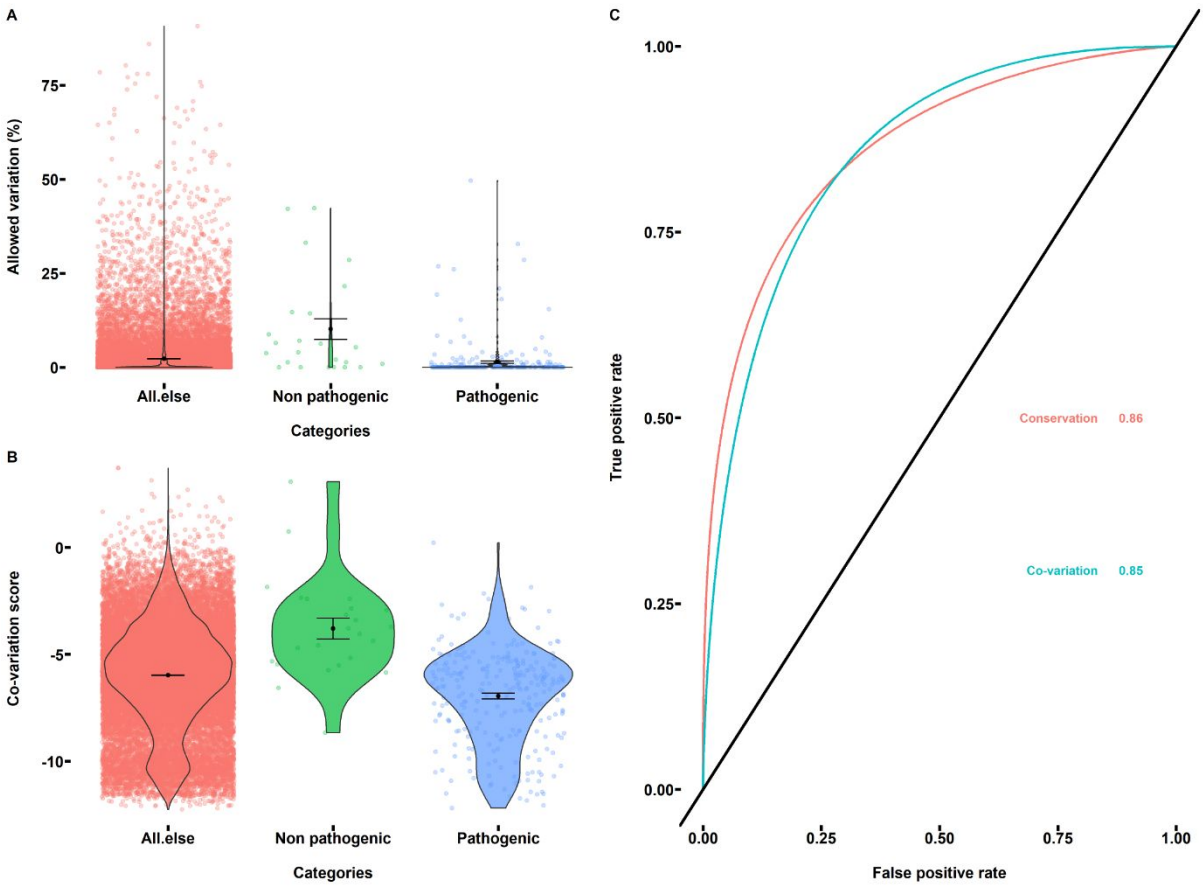


Figure 6. Conservation analysis and co-variation analysis used to identify pathogenic ATP7B mutations. (**A**) The allowed amino acid variation at each position and variation distributions for different categories. The dots in the background represent the obtained allowed variation; the black lines represent the variation distribution. The black dots represent the mean values in each category, and the error bars represent the standard errors in each category. (**B**) The co-variation scores obtained from GREMLIN and score distributions for different categories. (**C**) The ROC plot for conservation analysis and co-variation analysis with the final AUC value labeled.

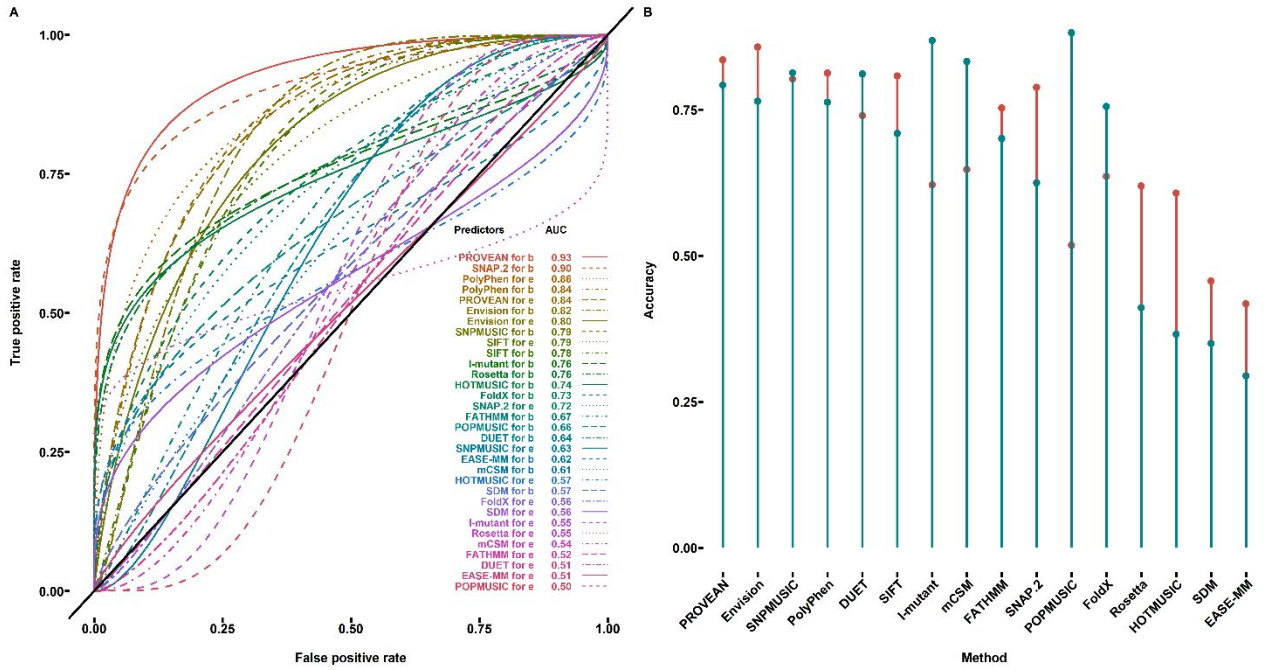


Figure 7. ROC analysis for buried and exposed residues. (**A**) ROC plot of the used structure-based and sequence-based methods applied to predict the pathogenicity of mutations in buried (b) and exposed (e) sites. (**B**) Identification accuracy of the used methods obtained from ROC analysis. The red and green colors represent the accuracy for buried and exposed residues, respectively.

Since PROVEAN displayed good ability to distinguish the non-pathogenic and pathogenic ATP7B mutations, we investigated this method further. We divided the 20 amino acids into nine groups: Hydrophobic amino acids (AVILM), polar amino acids (SCTQN), aromatic amino acids (FWY), negatively charged (DE), positively charged (HKR), phosphorylatable (STY), small (AGST), proline (P), and glycine (G). Then PROVEAN scores were calculated for each group and residue category (buried/exposed) as shown in **Figure S2**. For buried residues, the mutations related to aromatic amino acids were more likely to be pathogenic. According to the mutation data set shown in **Table S1**, most of the identified pathogenic mutations related to aromatic amino acids substitution were buried and there were no non-pathogenic mutations involving aromatic amino acid substitutions. This result largely supports the bulkiness change identified in the ANOVA analysis. In addition, a previous study indicated that some aromatic amino acids are highly conserved and

these positions had to be aromatic amino acids and of a precise size to maintain copper transport function.[107] For exposed residues, we did not find any systematic tendencies.

So far, our analysis has considered each site independently, neglecting the potentially important interactions between residues during substitution. Higher-order statistical models which consider both conservation at individual sites and pairwise coevolution positions may be appropriate to handle this.[28,80,108,109] These approaches require larger numbers of homologous sequences to build global sequence models.[79,81] We applied this method to investigate if such models improve the identification of pathogenic ATP7B variants, as shown in **Figure 6**. We did not see a markedly better performance as indicated by an AUC value of 0.85, suggesting that the interactions between pairs of residues is not critical in driving pathogenicity of ATP7B mutations. When analyzing the conservation and co-variation together, the two methods agree indicating that conservation is the major factor rather than co-variation in the global statistical model.

As shown in **Figure S3**, the ATP binding domain harbors most of the non-pathogenic and pathogenic mutations. The 3D structure of this domain is well defined (PDB ID 2ARF). The relatively small importance of co-variation led us to analyze the structure-based methods again only using the ATP binding domain (**Figure S4)**. As seen from the resulting ROC analysis, the AUC value for the best predictor, FoldX, increased to 0.8 indicating that loss of the stability is a relevant driver of disease for the mutations within the ATP binding domain, and thus the failure to identify stability above may be due to poor structural information. The two pathogenic mutations E1064A and H1069Q lower the stability of the ATP binding domain rather than impair ATP binding directly, suggesting a stability rather than functional effect in some of these mutations.[86]

To use our identified drivers in a best-possible combination, we combined the methods to create a two-dimensional representation for the mutations in the ATP binding domain, as shown in **Figure S5.** The individual threshold obtained from ROC analysis is shown as the vertical and horizontal dash lines. Using a two-dimensional representation improves the accuracy of identifying pathogenic mutations but is still not perfect, and assumptions on the pathogenicity of new variants

will clearly be error-prone. Thus, we recommend using the combination of the structure-based approach and conservation analysis to identify the pathogenic ATP7B variants as this two-dimensional representation could greatly improve the accuracy, by e.g. plotting the structure-based output against the sequence-based co-variation data, and optimizing the threshold.

Sensitivity of findings to data set

Our main finding in this work is that there are major variations in optimal thresholds of the many computational methods, which greatly affect pathogenicity estimates. As argued in the Method section, the WD data is ideal for our purpose as it avoids pollution from computational estimates of the last decade while being nearly complete in confidently assigned mutations ("PS"-type using the Richards et al. classification[38]). To test whether our findings are affected by the inclusion of recently identified variants, we extracted the missense mutations from the Gao et al. data set from 2019[26] with likely pathogenicity and performed all the analysis done above also for this data set.

The results in **Figures S6-S12** and **Tables S8-S9** show that all main findings are unaffected by using the newer data, as also supported by the intra-data bootstrapping analysis (**Table S4**). In particular, our finding that the computational methods overestimate pathogenicity of ATP7B mutations is unaffected mainly because the confidently assigned missense mutations are similar. In contrast, the total number of reported variants with computationally estimated pathogenicity have increased dramatically the last decade, but were not included in order to avoid computational self-reference in the benchmarking. The sequence-based methods again showed better performance for identifying the pathogenic variants. Furthermore, according to **Figure S8** and **Table S9**, the only significant changes in amino acid properties affecting pathogenicity were again related to amino acid volume or size. For the best-performing sequence-based method PolyPhen and SIFT (those with the highest AUC values), substitutions involving aromatic amino acids tend to confer pathogenicity, consistent with the importance of volume changes inside the membrane protein.

Implications for estimates of disease prevalence

The true prevalence of WD has been a longstanding matter of debate: The long-accepted estimate of 1:30,000 rests on a 30-year old publication, whereas recent genetic studies in large populations[15,26,27] have suggested that the true prevalence is in fact four times higher. These new estimates would imply severe consequences for the large number of undiagnosed patients. The new estimates were based on computer evaluations of likely pathogenicity of the discovered variants in genetic samples. Thus missense mutations were analyzed by Polyphen-2, PhyloP, CADD and MutationTaster in the French study[27], SIFT and PolyPhen2 in the British study[15], and SIFT and Polyphen in the global study[26]. Since these authors did not have access to our present evaluation of these methods, we think that they may have overestimated the prevalence by including benign variants. Specifically, the thresholds required to accurately discriminate pathogenic from benign variants is likely to be very protein-dependent, and unless corrections for this weakness are developed by e.g. using protein-class specific threshold categories in the methods, each protein and disease case requires a specific optimization against known clinical data to set the threshold properly in order to estimate disease prevalence. Our study shows how to select the best possible model for predicting mutation pathogenicity and use an optimized threshold (**Table 1)**, which will change the genetically estimated prevalence of WD. For most methods in **Table 1**, including Polyphen and SIFT used in the studies mentioned above, the number of estimated pathogenic variants will be substantially smaller when using optimized rather than default thresholds, thereby lowering the inferred prevalence WD, although the actual error made depends on the specific variant frequencies. Using these accurate thresholds may also enable a first step towards quantifying variants with low penetrance, which is crucial for a correct prediction of the number of undiagnosed patients with WD.

Better knowledge of the performance of the computer prediction will also affect the diagnostic work up in patients with suspected WD. According to the Leipzig Criteria,[16] two disease causing mutations is sufficient for diagnosis which was earlier based on clinical criteria. In such cases it is

absolutely important to avoid false diagnoses caused by erroneous computer evaluation of a given variant, a problem that easily arises if thresholds are not optimized as we show here. Considering the likely prevalence of many nearly neutral natural variants, which will probably be clinically benign[38], we question whether the identification of two mutations is sufficient for diagnosis of WD without assessment of disturbances in copper metabolism. A similar concern relates to the use of genetic testing as mean for neonatal screening for WD, the value of which will heavily depend on the validity of the computer prediction. Our findings regarding the accuracy and thresholds of the applied methods should be important in all of these contexts.

Conclusions

We have benchmarked state-of-the-art available computational disease prediction methods against a well-known and clinically confirmed data set of pathogenic and non-pathogenic *ATP7B* variants. The data presented suggest that structure-based analysis of the variants does not effectively separate non-pathogenic from pathogenic WD variants whereas sequence-based methods that account for the evolutionary conservation are more effective, but only if their thresholds for distinction are optimized.

Our findings are consistent with a view that proteins (and diseases) differ much more than a default generic threshold a single method can reasonably represent. Thus, different proteins, because of their diversity, have very different thresholds for pathogenicity of an arising mutation, and thus, each method applied should be optimized against real clinical data if possible. As discussed above, this can affect both diagnosis and the estimation of the real prevalence of diseases. Our results show that prevalence estimates based on these methods are not reliable as they tend to overestimate the pathogenicity of ATP7B mutations. Our finding explains why meta analysis and text search methods[26,27], which include many computational estimates, have concluded higher prevalence of WD than clinically observed.

Interestingly, once optimized, the best methods were more effective for buried rather than exposed sites and pointing towards an important role of the bulkiness of the specific amino acid change. The ATP7B protein includes a large transmembrane domain that mediates the transport of copper across the membrane, and buried sites may be those who most likely to affect the transport of copper; our finding that the side-chain volume of buried residues is an important correlator of pathogenicity may be the first structural-functional clue to future more quantitative genotype-phenotype relationships and more accurate prevalence estimates of WD.

Acknowledgements

The Danish Council for Independent Research | Natural Sciences (DFF), grant case 7016-00079B, and The Memorial Foundation of Manufacturer Vilhelm Petersen & Wife are gratefully acknowledged for supporting this work. The authors are particularly grateful to Dr. Sergey Ovchinnikov for providing the Tensorflow version of GREMLIN and useful suggestions for co-variation analysis.

Supporting Information Available.

The supporting information file contains details on the analysis including raw data, all p-values from ANOVA, and supplementary figures and analysis. The Supporting Information is available free of charge on the ACS Publications website.

References

(1) Ala, A.; Walker, A. P.; Ashkan, K.; Dooley, J. S.; Schilsky, M. L. Wilson's Disease. *Lancet* **2007**, *369*, 397–408.

(2) Thomas, G. R.; Forbes, J. R.; Roberts, E. A.; Walshe, J. M.; Cox, D. W. The Wilson Disease Gene: Spectrum of Mutations and Their Consequences. *Nat. Genet.* **1995**, *9*, 210–217.

(3) Gupta, A.; Das, S.; Ray, K. A Glimpse into the Regulation of the Wilson Disease Protein, ATP7B, Sheds Light on the Complexity of Mammalian Apical Trafficking Pathways. *Metallomics* **2018**, *10*, 378–387.

(4) Lutsenko, S.; Jayakanthan, S.; Dmitriev, O. Y. Molecular Architecture of the Copper-Transporting ATPase ATP7B. In *Clinical and Translational Perspectives on Wilson's disease*; Academic Press, 2019; pp 33–43.

(5) DiDonato, M.; Narindrasorasak, S.; Forbes, J. R.; Cox, D. W.; Sarkar, B. Expression, Purification, and Metal Binding Properties of the N-Terminal Domain from the Wilson Disease Putative Copper-Transporting ATPase (ATP7B). *J. Biol. Chem.* **1997**, *272*, 33279–33282.

(6) Payne, A. S.; Kelly, E. J.; Gitlin, J. D. Functional Expression of the Wilson Disease Protein Reveals Mislocalization and Impaired Copper-Dependent Trafficking of the Common H1069Q Mutation. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 10854–10859.

(7) Huster, D.; Lutsenko, S. The Distinct Roles of the N-Terminal Copper-Binding Sites in Regulation of Catalytic Activity of the Wilson's Disease Protein. *J. Biol. Chem.* **2003**, *278*, 32212–32218.

(8) Gourdon, P.; Liu, X.-Y.; Skjørringe, T.; Morth, J. P.; Møller, L. B.; Pedersen, B. P.; Nissen,

P. Crystal Structure of a Copper-Transporting PIB-Type ATPase. *Nature* **2011**, *475*, 59–64.

(9) Squitti, R.; Ghidoni, R.; Simonelli, I.; Ivanova, I. D.; Colabufo, N. A.; Zuin, M.; Benussi, L.; Binetti, G.; Cassetta, E.; Rongioletti, M. Copper Dyshomeostasis in Wilson Disease and Alzheimer's Disease as Shown by Serum and Urine Copper Indicators. *J. Trace Elem. Med. Biol.* **2018**, *45*, 181–188.

(10) Polishchuk, R. S. Cellular Function of ATP7B (Wilson ATPase). In *Clinical and Translational Perspectives on Wilson's disease*; Elsevier, 2019; pp 45–56.

(11) Tao, T.; Gitlin, J. D. Hepatic Copper Metabolism: Insights from Genetic Disease. *Hepatology* **2003**, *37*, 1241–1247.

(12) De Bie, P.; Muller, P.; Wijmenga, C.; Klomp, L. W. J. Molecular Pathogenesis of Wilson and Menkes Disease: Correlation of Mutations with Molecular Defects and Disease Phenotypes. *J. Med. Genet.* **2007**, *44*, 673–688.

(13) McCann, C. J.; Jayakanthan, S.; Siotto, M.; Yang, N.; Osipova, M.; Squitti, R.; Lutsenko, S. Single Nucleotide Polymorphisms in the Human ATP7B Gene Modify the Properties of the ATP7B Protein. *Metallomics* **2019**, *11*, 1128–1139.

(14) Lv, T.; Li, X.; Zhang, W.; Zhao, X.; Ou, X.; Huang, J. Recent Advance in the Molecular Genetics of Wilson Disease and Hereditary Hemochromatosis. *Eur. J. Med. Genet.* **2016**, *59*, 532–539.

(15) Coffey, A. J.; Durkie, M.; Hague, S.; McLay, K.; Emmerson, J.; Lo, C.; Klaffke, S.; Joyce, C. J.; Dhawan, A.; Hadzic, N. A Genetic Study of Wilson's Disease in the United Kingdom. *Brain* **2013**, *136*, 1476–1487.

(16) "European Association for the Study of the Liver." EASL Clinical Practice Guidelines: Wilson's Disease. *J. Hepatol.* **2012**, *56*, 671–685.

(17) Ferenci, P.; Stremmel, W.; Członkowska, A.; Szalay, F.; Viveiros, A.; Stättermayer, A. F.; Bruha, R.; Houwen, R.; Pop, T. L.; Stauber, R. Age and Sex but Not ATP7B Genotype Effectively Influence the Clinical Phenotype of Wilson Disease. *Hepatology* **2019**, *69*, 1464–1476.

(18) Schilsky, M. L. Long-Term Outcome for Wilson Disease: 85% Good. *Clin. Gastroenterol. Hepatol.* **2014**, *12*, 690–691.

(19) Harada, M. Pathogenesis and Management of Wilson Disease. *Hepatol. Res.* **2014**, *44*, 395–402.

(20) Scheinberg, I.; Sternlieb, I. Wilson Disease. In *Major Problems in Internal Medicine*; Lloyd, H., Smith, J., Eds.; Saunders: Philadelphia, 1984; p 23.

(21) Beinhardt, S.; Leiss, W.; Stättermayer, A. F.; Graziadei, I.; Zoller, H.; Stauber, R.; Maieron, A.; Datz, C.; Steindl-Munda, P.; Hofer, H. Long-Term Outcomes of Patients with Wilson Disease in a Large Austrian Cohort. *Clin. Gastroenterol. Hepatol.* **2014**, *12*, 683–689.

(22) Rodriguez-Castro, K. I.; Hevia-Urrutia, F. J.; Sturniolo, G. C. Wilson's Disease: A Review of What We Have Learned. *World J. Hepatol.* **2015**, *7*, 2859–2870.

(23) Bandmann, O.; Weiss, K. H.; Kaler, S. G. Wilson's Disease and Other Neurological Copper Disorders. *Lancet Neurol.* **2015**, *14*, 103–113.

(24) 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.

(25) Lin, L.-J.; Wang, D.-X.; Ding, N.-N.; Lin, Y.; Jin, Y.; Zheng, C.-Q. Comprehensive Analysis on Clinical Features of Wilson's Disease: An Experience over 28 Years with 133 Cases. *Neurol. Res.* **2014**, *36*, 157–163.

(26) Gao, J.; Brackley, S.; Mann, J. P. The Global Prevalence of Wilson Disease from Next-Generation Sequencing Data. *Genet. Med.* **2019**, *21*, 1155–1163.

(27) Collet, C.; Laplanche, J.-L.; 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.

(28) Stein, A.; Fowler, D. M.; Hartmann-Petersen, R.; Lindorff-Larsen, K. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends Biochem. Sci.* **2019**, *44*, 575–588.

(29) Tang, N.; Dehury, B.; Kepp, K. P. Computing the Pathogenicity of Alzheimer's Disease Presenilin 1 Mutations. *J. Chem. Inf. Model.* **2019**, *59*, 858–870.

(30) Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* **2010**, *7*, 248–249.

(31) Kepp, K. P. Computing Stability Effects of Mutations in Human Superoxide Dismutase 1. *J. Phys. Chem. B* **2014**, *118*, 1799–1812.

(32) Shihab, H. A.; Gough, J.; Cooper, D. N.; Stenson, P. D.; Barker, G. L. A.; Edwards, K. J.; Day, I. N. M.; Gaunt, T. R. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions Using Hidden Markov Models. *Hum. Mutat.* **2013**, *34*, 57–65.

(33) Flanagan, S. E.; Patch, A.-M.; Ellard, S. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. *Genet. Test. Mol. Biomarkers* **2010**, *14*, 533–537.

(34) Mehra, R.; Kepp, K. P. Computational Analysis of Alzheimer-Causing Mutations in Amyloid Precursor Protein and Presenilin 1. *Arch. Biochem. Biophys.* **2019**, *678*, 108168.

(35) Pey, A. L.; Stricher, F.; Serrano, L.; Martinez, A. Predicted Effects of Missense Mutations on Native-State Stability Account for Phenotypic Outcome in Phenylketonuria, a Paradigm of Misfolding Diseases. *Am. J. Hum. Genet.* **2007**, *81*, 1006–1024.

(36) Niroula, A.; Vihinen, M. Variation Interpretation Predictors: Principles, Types, Performance, and Choice. *Hum. Mutat.* **2016**, *37*, 579–597.

(37) Ohta, T. The Nearly Neutral Theory of Molecular Evolution. *Ann. Rev. Ecol. System*. **1992**, *23*, 263–286.

(38) Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W. W.; Hegde, M.; Lyon, E.; Spector, E. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17* (5), 405.

(39) Kepp, K. P.; Dasmeh, P. A Model of Proteostatic Energy Cost and Its Use in Analysis of Proteome Trends and Sequence Evolution. *PLoS One* **2014**, *9*, e90504.

(40) Drummond, D. A.; Bloom, J. D.; Adami, C.; Wilke, C. O.; Arnold, F. H. Why Highly Expressed Proteins Evolve Slowly. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14338–14343.

(41) Pollock, D. D.; Thiltgen, G.; Goldstein, R. A. Amino Acid Coevolution Induces an Evolutionary Stokes Shift. *Proc. Natl. Acad. Sci.* **2012**, *109*, E1352–E1359.

(42) Dasmeh, P.; Kepp, K. P. Superoxide Dismutase 1 Is Positively Selected to Minimize Protein Aggregation in Great Apes. *Cell. Mol. Life Sci.* **2017**, *74*, 3023–3037.

(43) 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.; Ostrowski, J. Whole-Exome Sequencing Identifies Novel Pathogenic Variants across the ATP7B Gene and Some Modifiers of Wilson's Disease Phenotype. *Liver Int.* **2019**, *39*,

177–186.

(44) Kenney, S. M.; Cox, D. W. Sequence Variation Database for the Wilson Disease Copper Transporter ATP7B. *Hum. Mutat.* **2007**, *28*, 1171–1177.

(45) Merle, U.; Weiss, K. H.; 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.

(46) Gromadzka, G.; Schmidt, H. H.-J.; Genschel, J.; Bochow, B.; Rodo, M.; Tarnacka, B.; Litwin, T.; Chabik, G.; Członkowska, A. 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.

(47) Zhu, M.; Dong, Y.; Ni, W.; Wu, Z.-Y. Defective Roles of ATP7B Missense Mutations in Cellular Copper Tolerance and Copper Excretion. *Mol. Cell. Neurosci.* **2015**, *67*, 31–36.

(48) Wu, Z.-Y.; Wang, N.; Lin, M.-T.; Fang, L.; Murong, S.-X.; Yu, L. Mutation Analysis and the Correlation Between Genotype and Phenotype of Arg778Leu Mutation in Chinese Patients With Wilson Disease. *Arch. Neurol.* **2001**, *58*, 971–976.

(49) Forbes, J. R.; Cox, D. W. Functional Characterization of Missense Mutations in ATP7B: Wilson Disease Mutation or Normal Variant? *Am. J. Hum. Genet.* **1998**, *63*, 1663–1674.

(50) Członkowska, A.; Rodo, M.; Gromadzka, G. Late Onset Wilson's Disease: Therapeutic Implications. *Mov. Disord.* **2008**, *23*, 896–898.

(51) Członkowska, A.; Gromadzka, G.; Chabik, G. Monozygotic Female Twins Discordant for Phenotype of Wilson's Disease. *Mov. Disord.* **2009**, *24*, 1066–1069.

(52) Schushan, M.; Bhattacharjee, A.; Ben-Tal, N.; Lutsenko, S. A Structural Model of the

Copper ATPase ATP7B to Facilitate Analysis of Wilson Disease-Causing Mutations and Studies of the Transport Mechanism. *Metallomics* **2012**, *4*, 669–678.

(53) Kepp, K. P. Towards a "Golden Standard" for Computing Globin Stability: Stability and Structure Sensitivity of Myoglobin Mutants. *Biochim. Biophys. Acta - Proteins Proteomics* **2015**, *1854*, 1239–1248.

(54) Christensen, N. J.; Kepp, K. P. Stability Mechanisms of Laccase Isoforms Using a Modified FoldX Protocol Applicable to Widely Different Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3210–3223.

(55) Dmitriev, O.; Tsivkovskii, R.; Abildgaard, F.; Morgan, C. T.; Markley, J. L.; Lutsenko, S. Solution Structure of the N-Domain of Wilson Disease Protein: Distinct Nucleotide-Binding Environment and Effects of Disease Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5302–5307.

(56) Yu, C. H.; Lee, W.; Nokhrin, S.; Dmitriev, O. Y. The Structure of Metal Binding Domain 1 of the Copper Transporter ATP7B Reveals Mechanism of a Singular Wilson Disease Mutation. *Sci. Rep.* **2018**, *8*, 581.

(57) Dolgova, N. V.; Nokhrin, S.; Yu, C. H.; George, G. N.; Dmitriev, O. Y. Copper Chaperone Atox1 Interacts with the Metal-Binding Domain of Wilson's Disease Protein in Cisplatin Detoxification. *Biochem. J.* **2013**, *454*, 147–156.

(58) Banci, L.; Bertini, I.; Cantini, F.; Rosenzweig, A. C.; Yatsunyk, L. A. Metal Binding Domains 3 and 4 of the Wilson Disease Protein: Solution Structure and Interaction with the Copper(I) Chaperone HAH1. *Biochemistry* **2008**, *47*, 7423–7429.

(59) Yuan, D. S.; Stearman, R.; Dancis, A.; Dunn, T.; Beeler, T.; Klausner, R. D. The Menkes/Wilson Disease Gene Homologue in Yeast Provides Copper to a Ceruloplasmin-like Oxidase Required for Iron Uptake. *Proc. Natl. Acad. Sci.* **1995**, *92*, 2632–2636.

(60) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382–W388.

(61) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 830–838.

(62) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048.

(63) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306–W310.

(64) Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G. A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Res.* **2015**, *43*, D364–D368.

(65) Pires, D. E.; Ascher, D. B.; Blundell, T. L. MCSM: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. *Bioinformatics* **2013**, *30*, 335–342.

(66) Pandurangan, A. P.; Ochoa-Montaño, B.; Ascher, D. B.; Blundell, T. L. SDM: A Server for Predicting Effects of Mutations on Protein Stability. *Nucleic Acids Res.* **2017**, *45*, W229–W235.

(67) Pires, D. E. V; Ascher, D. B.; Blundell, T. L. DUET: A Server for Predicting Effects of Mutations on Protein Stability Using an Integrated Computational Approach. *Nucleic Acids Res.* **2014**, *42*, W314–W319.

(68) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. PoPMuSiC 2.1: A Web Server for the Estimation of Protein Stability Changes upon Mutation and Sequence Optimality. *BMC Bioinformatics* **2011**, *12*, 151.

(69) Pucci, F.; Bourgeas, R.; Rooman, M. Predicting Protein Thermal Stability Changes upon Point Mutations Using Statistical Potentials: Introducing HoTMuSiC. *Sci. Rep.* **2016**, *6*, 23257.

(70) Ancien, F.; Pucci, F.; Godfroid, M.; Rooman, M. Prediction and Interpretation of Deleterious Coding Variants in Terms of Protein Structural Stability. *Sci. Rep.* **2018**, *8*, 4480.

(71) Folkman, L.; Stantic, B.; Sattar, A.; Zhou, Y. EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. *J. Mol. Biol.* **2016**, *428*, 1394–1405.

(72) Ng, P. C.; Henikoff, S. SIFT: Predicting Amino Acid Changes That Affect Protein Function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814.

(73) Gray, V. E.; Hause, R. J.; Luebeck, J.; Shendure, J.; Fowler, D. M. Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. *Cell Syst.* **2018**, *6*, 116–124.

(74) Choi, Y.; Chan, A. P. PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels. *Bioinformatics* **2015**, *31*, 2745–2747.

(75) Bromberg, Y.; Yachdav, G.; Rost, B. SNAP Predicts Effect of Mutations on Protein Function. *Bioinformatics* **2008**, *24*, 2397–2398.

(76) Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules. *Nucleic Acids Res.* **2016**, *44*, W344–W350.

(77) Suzek, B. E.; Huang, H.; McGarvey, P.; Mazumder, R.; Wu, C. H. UniRef: Comprehensive and Non-Redundant UniProt Reference Clusters. *Bioinformatics* **2007**, *23*, 1282–1288.

(78) Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment. *Nat. Methods* **2012**, *9*, 173–175.

(79) Ovchinnikov, S.; Kamisetty, H.; Baker, D. Robust and Accurate Prediction of Residue-Residue Interactions across Protein Interfaces Using Evolutionary Information. *Elife* **2014**, *3*, e02030.

(80) Balakrishnan, S.; Kamisetty, H.; Carbonell, J. G.; Lee, S. I.; Langmead, C. J. Learning Generative Models for Protein Fold Families. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 1061–1078.

(81) Kamisetty, H.; Ovchinnikov, S.; Baker, D. Assessing the Utility of Coevolution-Based Residue-Residue Contact Predictions in a Sequence- and Structure-Rich Era. *Proc. Natl. Acad. Sci.* **2013**, *110*, 15674–15679.

(82) Hopf, T. A.; Ingraham, J. B.; Poelwijk, F. J.; Schärfe, C. P. I.; Springer, M.; Sander, C.; Marks, D. S. Mutation Effects Predicted from Sequence Co-Variation. *Nat. Biotechnol.* **2017**, *35*, 128–135.

(83) Huster, D.; Khne, A.; Bhattacharjee, A.; Raines, L.; Jantsch, V.; Noe, J.; Schirrmeister, W.; Sommerer, I.; Sabri, O.; Berr, F.; Mssner, J.; Stieger, B.; Caca, K.; Lutsenko, S. Diverse Functional Properties of Wilson Disease ATP7B Variants. *Gastroenterology* **2012**, *142*, 947–956.

(84) de Bie, P.; van de Sluis, B.; Burstein, E.; van de Berghe, P. V. E.; Muller, P.; Berger, R.; Gitlin, J. D.; Wijmenga, C.; Klomp, L. W. J. Distinct Wilson's Disease Mutations in ATP7B Are Associated With Enhanced Binding to COMMD1 and Reduced Stability of ATP7B. *Gastroenterology* **2007**, *133*, 1316–1326.

(85) Parisi, S.; Polishchuk, E. V.; Allocca, S.; Ciano, M.; Musto, A.; Gallo, M.; Perone, L.; Ranucci, G.; Iorio, R.; Polishchuk, R. S.; Bonatti, S. Characterization of the Most Frequent ATP7B Mutation Causing Wilson Disease in Hepatocytes from Patient Induced Pluripotent Stem Cells. *Sci. Rep.* **2018**, *8*, 6247.

(86) Dmitriev, O. Y.; Bhattacharjee, A.; Nokhrin, S.; Uhlemann, E. M. E.; Lutsenko, S. Difference in Stability of the N-Domain Underlies Distinct Intracellular Properties of the E1064A and H1069Q Mutants of Copper-Transporting ATPase ATP7B. *J. Biol. Chem.* **2011**, *286*, 16355–16362.

(87) Kepp, K. P. Genotype-Property Patient-Phenotype Relations Suggest That Proteome Exhaustion Can Cause Amyotrophic Lateral Sclerosis. *PLoS One* **2015**, *10*, e0118649.

(88) Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D. S. The Stability Effects of Protein Mutations Appear to Be Universally Distributed. *J. Mol. Biol.* **2007**, *369*, 1318–1332.

(89) Kroncke, B. M.; Duran, A. M.; Mendenhall, J. L.; Meiler, J.; Blume, J. D.; Sanders, C. R. Documentation of an Imperative to Improve Methods for Predicting Membrane Protein Stability. *Biochemistry* **2016**, *55*, 5002–5009.

(90) Christensen, N. J.; Kepp, K. P. Accurate Stabilities of Laccase Mutants Predicted with a Modified FoldX Protocol. *J. Chem. Inf. Model.* **2012**, *52*, 3028–3042.

(91) Gromiha, M. M.; Oobatake, M.; Kono, H.; Uedaira, H.; Sarai, A. Relationship Between Amino Acid Properties and Protein Stability: Buried Mutations. *J. Protein Chem.* **1999**, *18*, 565–578.

(92) Somavarapu, A. K.; Kepp, K. P. Loss of Stability and Hydrophobicity of Presenilin 1 Mutations Causing Alzheimer's Disease. *J. Neurochem.* **2016**, *137*, 101–111.

(93) Tiwari, M. K.; Kepp, K. P. Modeling the Aggregation Propensity and Toxicity of Amyloid-

β Variants. *J. Alzheimer's Dis.* **2015**, *47*, 215–229.

(94) Somavarapu, A. K.; Kepp, K. P. Direct Correlation of Cell Toxicity to Conformational Ensembles of Genetic Aβ Variants. *ACS Chem. Neurosci.* **2015**, *6*, 1990–1996.

(95) Stefani, M.; Dobson, C. M. Protein Aggregation and Aggregate Toxicity: New Insights into Protein Folding, Misfolding Diseases and Biological Evolution. *J. Mol. Med.* **2003**, *81*, 678–699.

(96) Münch, C.; Bertolotti, A. Exposure of Hydrophobic Surfaces Initiates Aggregation of Diverse ALS-Causing Superoxide Dismutase-1 Mutants. *J. Mol. Biol.* **2010**, *399*, 512–525.

(97) Cho, M.-K.; Kim, H.-Y.; Bernado, P.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. Amino Acid Bulkiness Defines the Local Conformations and Dynamics of Natively Unfolded α-Synuclein and Tau. *J. Am. Chem. Soc.* **2007**, *129*, 3032–3033.

(98) Braiterman, L. T.; Murthy, A.; Jayakanthan, S.; Nyasae, L.; Tzeng, E.; Gromadzka, G.; Woolf, T. B.; Lutsenko, S.; Hubbard, A. L. Distinct Phenotype of a Wilson Disease Mutation Reveals a Novel Trafficking Determinant in the Copper Transporter ATP7B. *Proc. Natl. Acad. Sci.* **2014**, *111*, E1364–E1373.

(99) Weisburd, B.; Ruano-Rubio, V.; Daly, M. J.; Moonshine, A. L.; Rivas, M. A.; Kiezun, A.; Flannick, J.; Ardissino, D.; MacArthur, D. G.; Donnelly, S.; McGovern, D.; Cummings, B. B.; Hultman, C. M.; Orozco, L.; Tukiainen, T.; Danesh, J.; Boehnke, M.; Duncan, L. E.; Yu, D.; Fromer, M.; Rose, S. A.; Cooper, D. N.; Minikel, E. V.; Hill, A. J.; McPherson, R.; Tusie-Luna, M. T.; Gupta, N.; Gauthier, L.; Florez, J. C.; Shakir, K.; Kosmicki, J. A.; Karczewski, K. J.; Neale, B. M.; Tsuang, M. T.; Thomas, B. P.; Watkins, H. C.; Banks, E.; Natarajan, P.; Stenson, P. D.; Gabriel, S. B.; Zou, J.; Glatt, S. J.; Deflaux, N.; O'Donnell-Luria, A. H.; Howrigan, D.; Birnbaum, D. P.; Ware, J. S.; Saleheen, D.; Tiao, G.; Palotie, A.; Kathiresan, S.; Laakso, M.; Peloso, G. M.; Zhao, F.; McCarroll, S.; McCarthy, M. I.;

Poplin, R.; Sklar, P.; DePristo, M.; Berghout, J.; Lek, M.; Ruderfer, D. M.; Sullivan, P. F.; Goldstein, J.; Pierce-Hoffman, E.; Elosua, R.; Altshuler, D. M.; Wilson, J. G.; Scharf, J. M.; Samocha, K. E.; Purcell, S. M.; Stevens, C.; Fennell, T.; Estrada, K.; Do, R.; Getz, G.; Kurki, M. I.; Tuomilehto, J.; Won, H.-H. Analysis of Protein-Coding Genetic Variation in 60,706 Humans. *Nature* **2016**, *536*, 285–291.

(100) Kucukkal, T. G.; Petukh, M.; Li, L.; Alexov, E. Structural and Physico-Chemical Effects of Disease and Non-Disease NsSNPs on Proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 18–24.

(101) Petukh, M.; Kucukkal, T. G.; Alexov, E. On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns. *Hum Mutat* **2015**, *36*, 524–534.

(102) Wang, Z.; Moult, J. SNPs, Protein Structure, and Disease. *Hum. Mutat.* **2001**, *17*, 263–270.

(103) Ferrer-Costa, C.; Orozco, M.; de la Cruz, X. Characterization of Disease-Associated Single Amino Acid Polymorphisms in Terms of Sequence and Structure Properties. *J. Mol. Biol.* **2002**, *315*, 771–786.

(104) Sunyaev, S.; Ramensky, V.; Bork, P. Towards a Structural Basis of Human Non-Synonymous Single Nucleotide Polymorphisms. *Trends Genet.* **2000**, *16*, 198–200.

(105) Teng, S.; Madej, T.; Panchenko, A.; Alexov, E. Modeling Effects of Human Single Nucleotide Polymorphisms on Protein-Protein Interactions. *Biophys. J.* **2009**, *96*, 2178–2188.

(106) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. Inherent Toxicity of Aggregates Implies a Common Mechanism for Protein Misfolding Diseases. *Nature* **2002**, *416*, 507–511.

(107) Braiterman, L.; Nyasae, L.; Guo, Y.; Bustos, R.; Lutsenko, S.; Hubbard, A. Apical Targeting and Golgi Retention Signals Reside within a 9-Amino Acid Sequence in the

Copper-ATPase, ATP7B. *Am. J. Physiol. Liver Physiol.* **2009**, *296*, G433–G444.

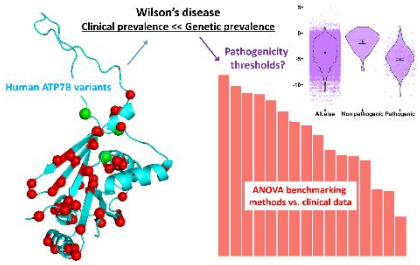
(108) Weigt, M.; White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. Identification of Direct Residue Contacts in Protein-Protein Interaction by Message Passing. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 67–72.

(109) Marks, D. S.; Colwell, L. J.; Sheridan, R.; Hopf, T. A.; Pagnani, A.; Zecchina, R.; Sander, C. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS One* **2011**, *6*, e28766.

Table 1. Optimized and default threshold for pathogenicity assignment and the number of the pathogenic and non-pathogenic mutations based on the applied threshold.

Method	Threshold (ROC)	Non Pathogenic	Pathogenic	Threshold (Default)	Non Pathogenic	Pathogenic
DUET	1.465	9596	1705	0	3238	8063
EASE-MM	0.115	6703	21132	0	5093	22742
Envision	0.864	16415	11420	1	2952	24883
FATHMM	-3.945	19980	7855	-1.5	1567	26268
FoldX	-0.154	2885	8420	0	3570	7735
HOTMUSIC	2.965	7869	3436	0	1721	9584
I-mutant	1.180	6607	4698	0	1447	9858
mCSM	0.163	2553	8752	0	1632	9673
PolyPhen	0.872	10827	17008	0.5	8430	19405
POPMUSIC	1.155	8444	2861	0	1920	9385
PROVEAN	-2.795	10151	17684	-2.5	8782	19053
Rosetta	3.976	8488	2790	0	4368	6910
SDM	0.650	6900	4401	0	4034	7267
SIFT	0.0245	11312	16523	0.05	9227	18608
SNAP.2	5.50	13757	14078	0	12949	14886
SNPMUSIC	0.185	6745	4560	0	4988	6317

TOC graphic


Wilson's disease
Clinical prevalence << Genetic prevalence
Pathogenicity thresholds?
Human ATP7B variants
ANOVA benchmarking methods vs. clinical data

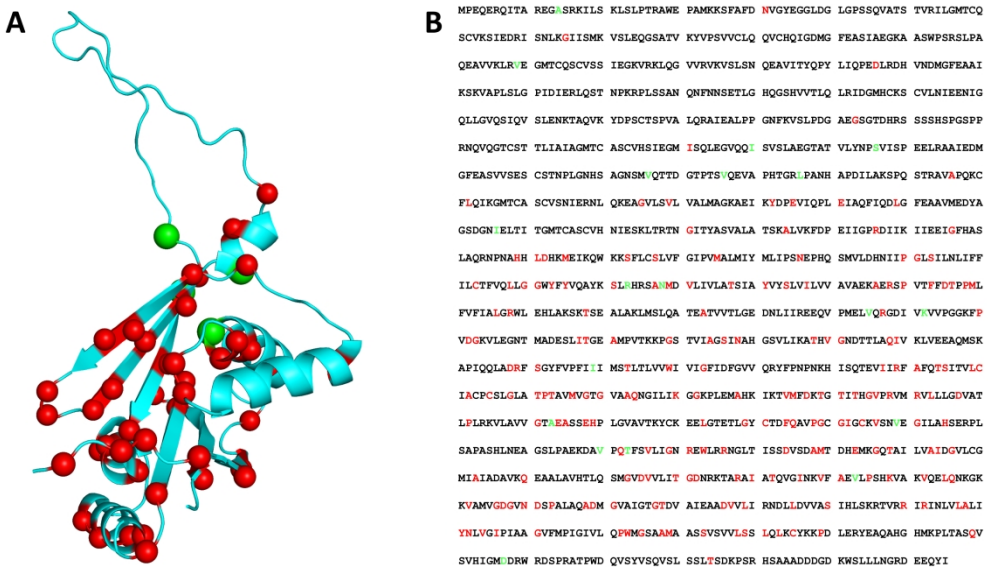


Figure 1

338x190mm (300 x 300 DPI)

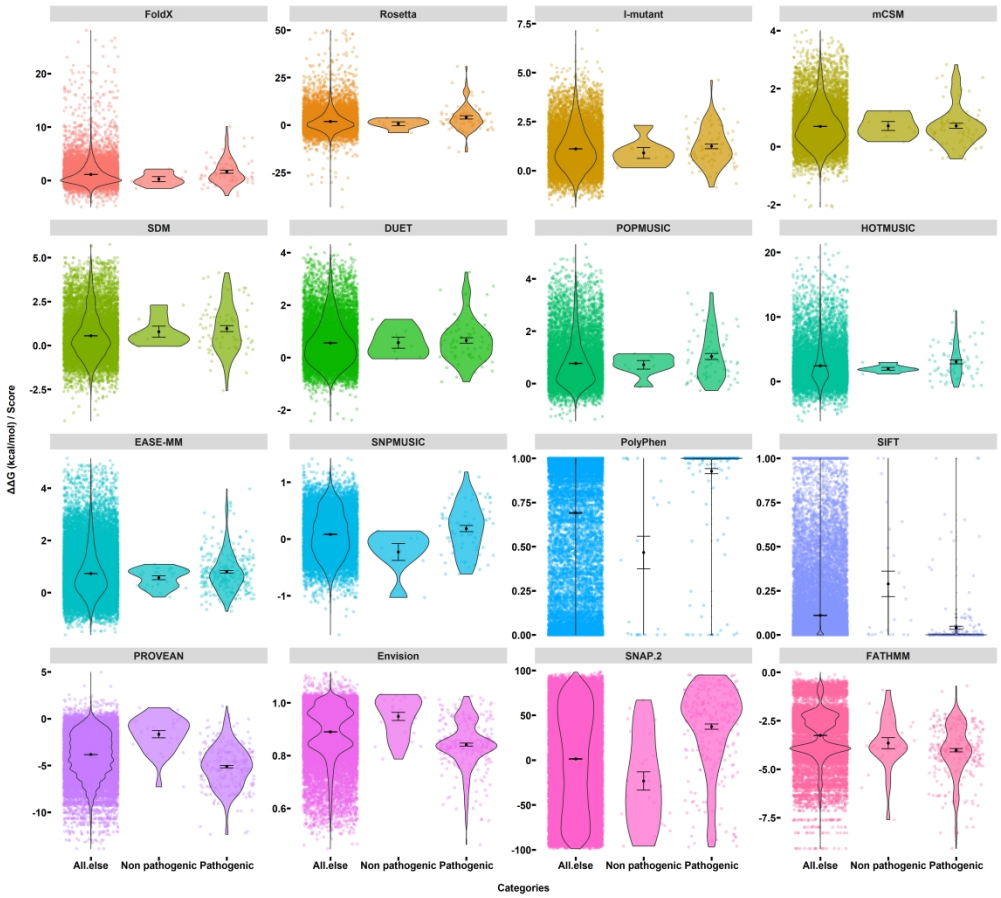


Figure 2

482x431mm (300 x 300 DPI)

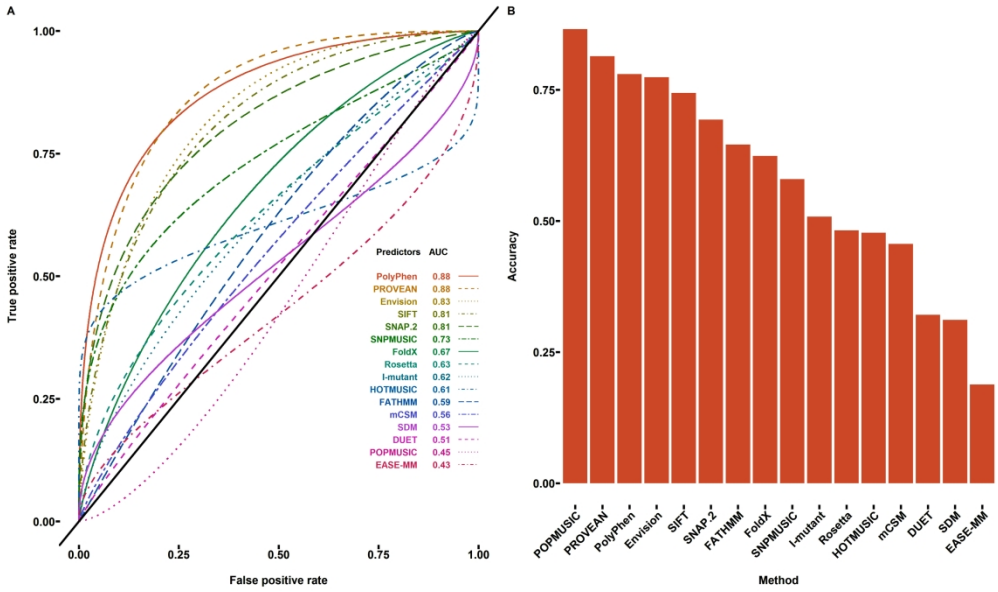


Figure 3

259x152mm (300 x 300 DPI)

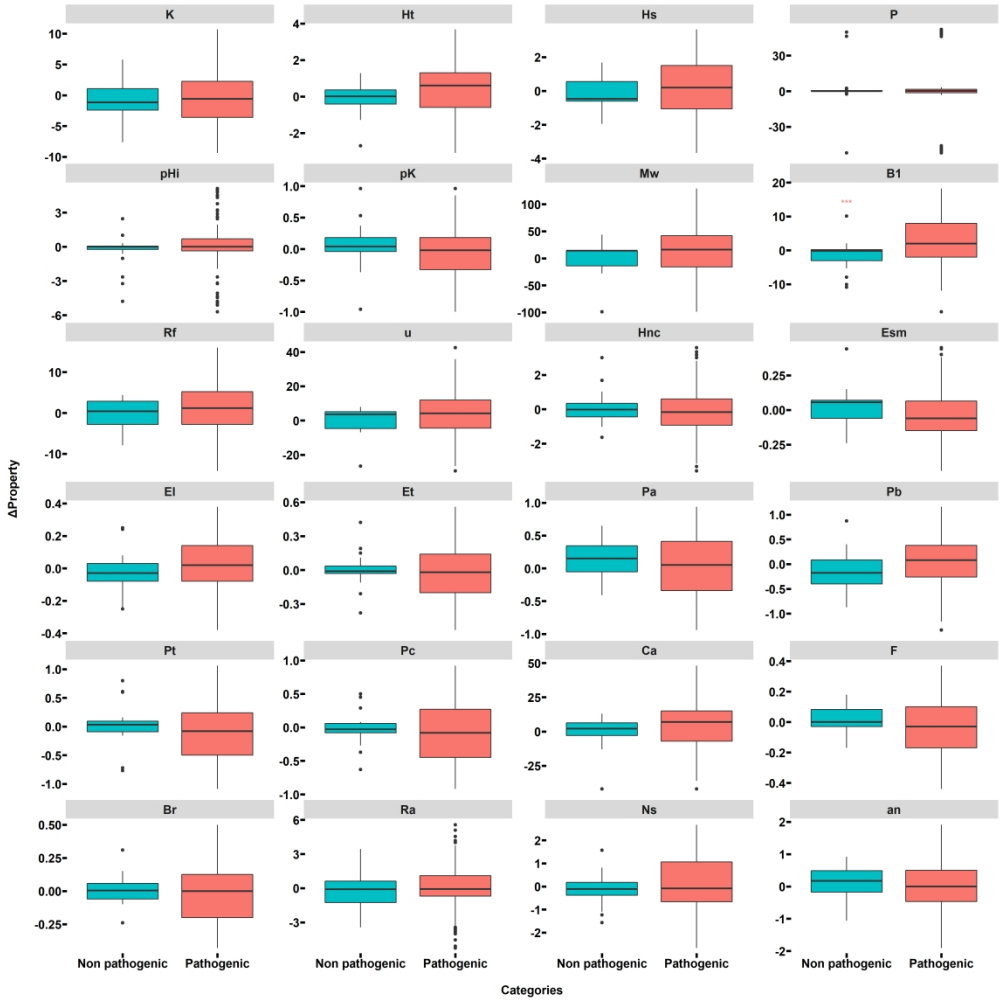


Figure 4

406x406mm (300 x 300 DPI)

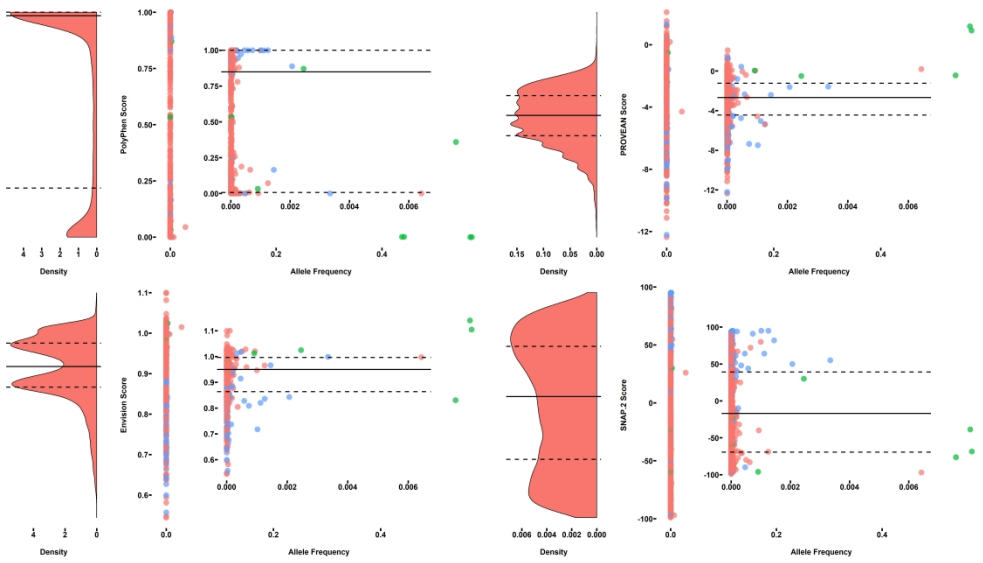


Figure 5

635x355mm (300 x 300 DPI)

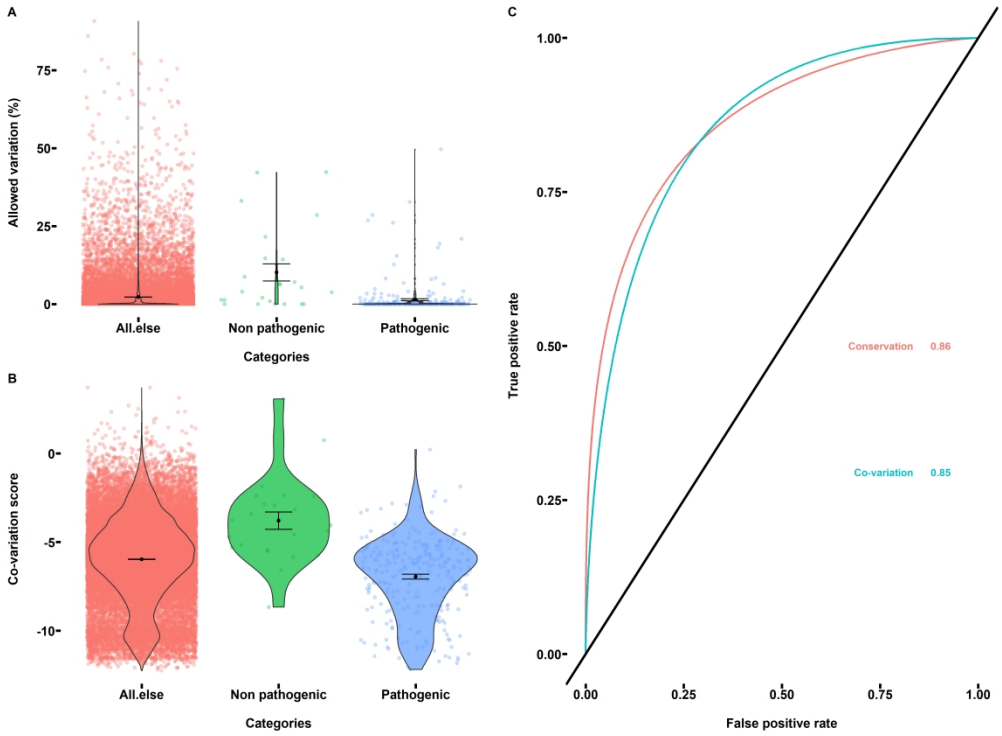


Figure 6

381x279mm (300 x 300 DPI)

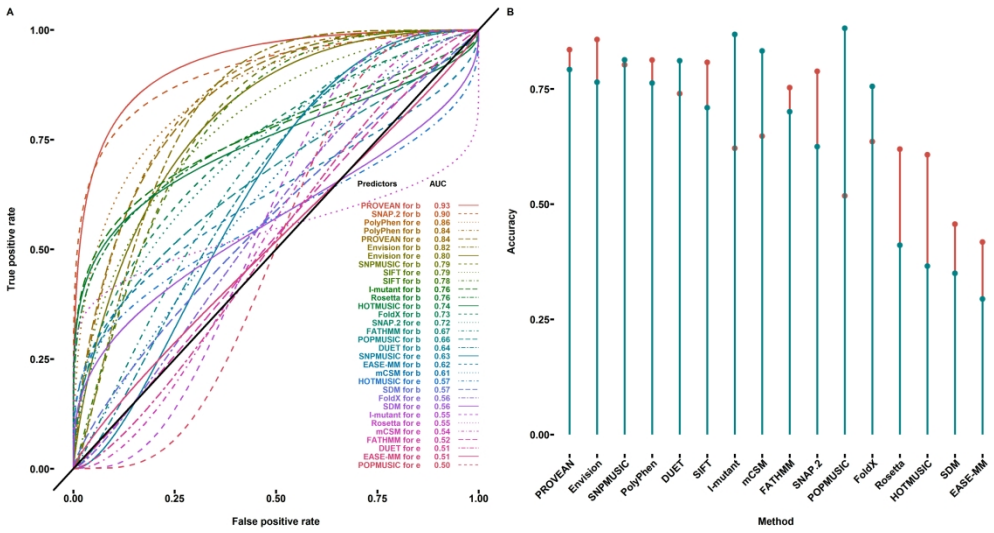


Figure 7

289x152mm (300 x 300 DPI)