Vasor: Accurate prediction of variant effects for amino acid substitutions in MDR3

Annika Behrendt¹, Pegah Golchin², Filip König¹, Daniel Mulnaes¹, Amelie Stalke^{3,4}, Carola Dröge^{5,6}, Verena Keitel^{5,6}, Holger Gohlke^{1,7,*}

¹Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Germany

²Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt

³Department of Human Genetics, Hannover Medical School, Hannover, Germany

⁴Department of Pediatric Gastroenterology and Hepatology, Division of Kidney, Liver and Metabolic Diseases, Hannover Medical School, Hannover, Germany

⁵Department for Gastroenterology, Hepatology and Infectious Diseases, Medical Faculty, Otto von Guericke University, Magdeburg, Germany

⁶Department for Gastroenterology, Hepatology and Infectious Diseases, University Hospital, Medical Faculty, Heinrich Heine University Düsseldorf, Germany

⁷John-von-Neumann-Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), Institute of Biological Information Processing (IBI-7: Structural Biochemistry), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52428 Jülich

Keywords: ABCB4, predictor, machine learning, benign, pathogenic, structure-based

Running title: Vasor: Accurate prediction of MDR3 variant effects

Funding: This study was supported by the BMBF through HiChol (01GM1904A to H.G., V.K., C.D., 01GM1904B to A.S.) and, in part, InCelluloProtStruct (031L0182 to H.G.).

Address correspondence to:

Prof. Dr. Holger Gohlke

Institute for Pharmaceutical and Medicinal Chemistry

Heinrich-Heine-Universität Düsseldorf

Universitätsstr. 1

40225 Düsseldorf

Germany

Tel. +49 211 81 13662

Email: gohlke@uni-duesseldorf.de

Abstract

Background / Rationale: The phosphatidylcholine floppase MDR3 is an essential hepatobiliary transport protein. MDR3 dysfunction is associated with various liver diseases, ranging from severe progressive familial intrahepatic cholestasis to transient forms of intrahepatic cholestasis of pregnancy and familial gallstone disease. Single amino acid substitutions are often found as causative of dysfunction, but identifying the substitution effect in *in vitro* studies is time- and cost-intensive.

Main results: We developed Vazor (**V**ariant **a**ssessor of MDR3), a machine learning-based model to classify novel MDR3 missense variants into the categories benign or pathogenic. Vazor was trained on the, to date, largest dataset specific for MDR3 of benign and pathogenic variants and uses general predictors, namely EVE, EVmutation, PolyPhen-2, I-Mutant2.0, MUpro, MAESTRO, PON-P2, and other variant properties such as half-sphere exposure, PTM site, and secondary structure disruption as input. Vazor consistently outperformed the integrated general predictors and the external prediction tool MutPred2, leading to the current best prediction performance for MDR3 single-site missense variants (on an external test set: F1-score: 0.90, MCC: 0.80). Furthermore, Vazor predictions cover the entire sequence space of MDR3. Vazor is accessible as a webserver at https://cpclab.uni-duesseldorf.de/mdr3_predictor/ for users to rapidly obtain prediction results and a visualization of the substitution site within the MDR3 structure.

Conclusion: The MDR3-specific prediction tool Vazor can provide reliable predictions of single site amino acid substitutions, giving users a fast way to assess initially whether a variant is benign or pathogenic.

1. Introduction

Bile formation is a carefully regulated system, from bile acid synthesis to secretion of bile acids across the canalicular membrane. Adenosine triphosphate binding cassette (ABC) transporters present on the canalicular membrane of hepatocytes are responsible for the transport of primary bile components, namely bile acids via the bile salt export pump (BSEP, *ABCB11*), cholesterol via the ABC sub-family G members 5 and 8 (*ABCG5/ABCG8*), as well as phospholipids via the multidrug resistance protein 3 (MDR3). MDR3 encoded by the *ABCB4* gene acts as a floppase, translocating substrates such as phosphatidylcholine from the inner to the outer membrane leaflet (1,2) and exposing the substrate for extraction by bile acids into primary bile (3). Recent studies have suggested different transport pathways, following either an alternating two-site access model through the protein's inner cavity (4) or a credit-card swipe mechanism along transmembrane helix 7 (5), indicating a need for further research on the exact molecular mechanism. MDR3 dysfunction has been linked to various liver-associated diseases, including intrahepatic cholestasis of pregnancy, low phospholipid-associated cholelithiasis, drug-induced liver injury, progressive familial intrahepatic cholestasis type 3, liver fibrosis/cirrhosis as well as hepatobiliary malignancy (6–12).

It is estimated that at least 70 % of disease-causing *ABCB4* variants are amino acid substitutions, whereas variants leading to premature stop codons and protein truncations are in the minority (13). However, while the advancement of sequencing allows rapid testing of patients, it remains challenging for clinicians and researchers to assess the potential impact of novel missense variants.

Evaluation of newly found MDR3 amino acid substitutions by *in vitro* cellular assays remains time-consuming. Machine learning-based prediction tools instead offer rapid analysis and have led in recent years to a plethora of predictors (14,15). Nonetheless, general predictors do not consistently perform well on all proteins, necessitating the development of protein-specific prediction tools. To date, there is no MDR3-specific predictor available for classifying amino acid substitutions despite MDR3's vital role in bile homeostasis. An initial evaluation of general predictor performances on MDR3 variants suggested MutPred as a well-performing tool (16); however, generalization is difficult due to only 21 tested variants with established cellular effects. Additionally, the tested variants presented a clear bias towards pathogenic effects.

Here, we created an MDR3-specific variant dataset and trained a machine learning algorithm using several established general prediction tools, namely EVE, EVmutation, PolyPhen-2, I-Mutant2.0, MUpro, MAESTRO, and PON-P2 (17–23), as well as half-sphere exposure, post-translational modification (PTM) site influence, and secondary structure disruption as features to obtain an MDR3-specific prediction tool for help in classifying variants as benign or pathogenic (see Fig. 1 for a graphical overview). Our predictor, Vamor (**V**ariant **a**ssessment **o**f MDR3), performed better than each integrated general predictor. Additionally, Vamor outperformed MutPred2, a general predictor we chose for comparison based on the suggested high performance of its predecessor MutPred on MDR3 (16). We provide easy access to Vamor via a webserver, where users can enter a missense variant of interest and obtain a prediction if it is benign or pathogenic together with an estimate of the prediction probability. Additionally, the mutation site is displayed on the structure of MDR3, giving the user a comprehensive view of the local site and the overall position of the assessed variant.

2. Method and Implementation

2.1. MDR3 missense variants

MDR3 variants were obtained from a literature search for variants causative of MDR3 dysfunction or known variants with no effect in any MDR3-associated disease (4,6,10,12,13,16,24–57). We excluded variants with unclear information on disease association (i.e., no *in vitro* verification analysis and no information on clinical indications for disease association) to eliminate False Positives or False Negatives. As studied benign variants for MDR3 are rare (13,16), further missense variants were obtained from gnomAD v2.1.1 (58) to increase the number of benign variants. During the generation of the gnomAD database, individuals with severe pediatric diseases are removed; however, it is possible that pathogenic variants exist in the gnomAD dataset. Accordingly, we employed a selection step to exclude false-negative cases of MDR3 variants. Using the platform VarSome (59), variants were pre-classified following the guidelines of The American College of Medical Genetics and Association for Molecular Pathology (ACMG-AMP) (60) rules, and variants with a likely pathogenic or pathogenic effect were removed, whereas variants with uncertain significance, likely benign, or benign classification by VarSome were integrated into the dataset. These

steps were included to create a high-quality dataset to keep the number of misclassified variants low but at the same time retain a sufficiently high number of variants. The final list of variants contained 85 pathogenic and 279 benign variants. Every variant was mapped to the longest MDR3 isoform, corresponding to Uniprot (61) entry P21439-1.

2.2. Dataset and features

The list of MDR3 variants was subjected to established general predictors for missense mutations (EVE, PolyPhen-2, I-Mutant2.0, MUpro, MAESTRO, PON-P2, and EVmutation), and additional features (half-sphere exposure, secondary structure, PTM site, and relative solvent accessibility) were computed, creating an MDR3-specific feature set.

EVE (Evolutionary models of Variant Effects) is a recently developed unsupervised computational method, which trained Bayesian variational autoencoders on multiple sequence alignments to classify variant effects based on a variant-specific, computed evolutionary index followed by a fitted global-local mixture of Gaussian Mixture Models (17). PolyPhen-2 employs a Naïve Bayes classifier for predicting variant effects using sequence-based features and structure-based features, if available (18). I-Mutant2.0 predicts protein stability changes, using a Support Vector Machine-based tool trained on either sequence or structural information (19). MUpro predicts stability changes upon single-site mutations, using sequence and structural information if available, using a Support Vector Machine Approach (20). Both I-Mutant2.0 and MUpro predict the direction of stability change and the energy difference. MAESTRO employs a combination of machine learning approaches (neural network, support vector machine, and multiple linear regression) to predict the energy difference introduced by missense mutations based on consensus, along with predicting a confidence score (21). PON-P2 applies selected features from evolutionary conservation and biochemical properties of amino acids to develop a random forest classifier that classifies mutations into benign or pathogenic cases, or those with unknown significance (22). EVmutation explicitly considers interdependencies between residues or nucleotide bases in their unsupervised statistical method to include epistasis (23).

EVE and EVmutation predictions for the MDR3 protein were accessed using the pre-computed dataset available from the creators of EVE (<https://evemodel.org/>) and EVmutation

(https://marks.hms.harvard.edu/evmutation/human_proteins.html), respectively. I-Mutant2.0, Mupro, and MAESTRO predictions were generated using their standalone downloadable versions. PolyPhen-2 predictions were accessed using the batch query of the webserver (<http://genetics.bwh.harvard.edu/pph2/bgi.shtml>) with the default values. PON-P2 predictions were generated using the sequence submission feature for variants of the webserver (<http://structure.bmc.lu.se/PON-P2/>).

Additional features were added to explicitly integrate effects on PTM sites, variant location in α -helical or β -sheet secondary structure, and effects on residue solvent accessibility. Known PTM sites from the literature were supplemented by potential PTM sites, which were predicted using PhosphoMotif (62), PhosphoSitePlus (63), NetPhos (61), and the ELM database (65). Secondary structure was extracted from the MDR3 structure (PDB ID 6S7P) using DSSP (66,67). Relative solvent accessibility was computed based on residue exposure calculated with DSSP divided by the maximal residue solvent accessibility calculated by Thien et al. (68). Half-sphere exposure was introduced by Hamelryk (69) to measure residue solvent exposure and surpass limitations of relative solvent accessibility, which does not differentiate between residues closely buried beneath the surface and residues deeply buried. It was implemented using values from the Biopython HSExposure module calculated according to the half-sphere corresponding to the direction of the sidechain of the residue as measured from the C α atom.

2.3. Machine learning

The obtained dataset was cleaned from non-numerical values. In the case of binary features, such as classification features of general predictors, -1 was set if no prediction was available to distinguish from benign (indicated by value 0) or pathogenic (indicated by value 1) predictions. Additionally, relative solvent accessibility and half-sphere exposure were set to -1 if no prediction value was obtained, to distinguish from prediction values of 0. Other numerical features were replaced by 0 if no prediction for the respective feature was available. The correlation between features within the dataset was assessed by the Spearman *R* correlation coefficient.

A test set was generated by selecting 20 benign and 20 pathogenic variants from the overall dataset. To avoid a bias towards specific amino acids, we minimized the root-mean-square

deviation (RMSD)-based difference between the amino acid distribution of the variants within the test set compared to the overall dataset (Suppl. Fig. 1): After randomly drawing ten variants into the test set, the RMSD-based difference between the amino acid distribution of the general dataset and current test set was computed; further variants were only transferred into the test set if they met one of the following conditions: (a) the RMSD between reference sequence and substituted amino acid distributions decreased by addition of the new variant, (b) the RMSD between reference sequence amino acid distributions decreased while the RMSD between substituted amino acid distributions did not increase more than 0.1, or (c) the RMSD between substituted amino acid distributions decreased while the RMSD between reference sequence amino acid distributions did not increase more than 0.1. Due to the limited size of the dataset, otherwise, it might not be possible to draw a variant for the test set. The test set was withheld from the machine learning training step and used for final validation.

To handle the imbalance between the pathogenic (85 variants) and benign (279 variants) class, we used the synthetic minority oversampling technique (SMOTE) (70). This method generates new synthetic data points by using existing minority data points within the N -dimensional dataset space, drawing lines to the five nearest minority class neighbors, and randomly selecting synthetic data points along these lines to balance out the classes.

On the training dataset, the XGBoost algorithm (71) (as implemented in the python library) was trained using the default gradient-boosted tree (gbtree); the maximum depth of a tree (max_depth) was set to 3, subsample to 0.6, and the step size (learning_rate) to 0.02. The training was evaluated using repeated k -fold cross-validation, with k set to 3 and the value of repeats (n_repeats) to 5. Using this procedure, the training dataset was randomly split into three equally sized folds, where each fold is used as an internal test dataset with the remaining two folds as training datasets, respectively. The performance results were measured and visualized in receiver-operator curves (ROC) for comparison to the final test set. These steps were repeated five times.

To reduce features and estimate feature importance, we analyzed the tree-based feature importance and the permutation importance, leading to the removal of the four least-informative features shared in both feature importance measures: relative solvent accessibility, I-Mutant2.0 stability sign, I-Mutant2.0 deltaG value, and PON-P2 probability

value. Tree-based feature importance was computed using the XGBoost algorithm built-in feature and the 'gain' (average gain across all splits where a feature is used). Permutation-based feature importance was computed by random shuffling each feature consecutively, followed by a performance test, denoting performance alterations upon feature permutation. The performance of the model without feature selection is shown in Suppl. Fig. 2.

The trained model, termed Vador (**V**ariant **a**ssessment of **M**DR3), predicts a probability, ranging from 0 to 1, for a given variant to belong to the pathogenic class. Predictions above (below) 0.5 are classified as pathogenic (benign).

2.4. Comparison to established predictors

To assess the general performance of Vador, we compared it to the general predictors EVE, PolyPhen-2, PON-P2, and MutPred2. MutPred2 predictions were used to compare our prediction tool to an 'external' general predictor, as MutPred2 was not used as an input feature for Vador. The standalone version of MutPred2 was used to classify each variant within the entire dataset. The performance of Vador and the other predictors was evaluated on the entire dataset and the test set. This ensured increased fairness for the performance comparison, as Vador may have an advantage over other predictors based on its training on the training dataset. ROC and precision-recall-curves were adjusted to the availability of variants each predictor was able to classify over the entire dataset (i.e., if general predictors did not classify a variant into the category "benign" or "pathogenic", the respective variant could not be assessed, and curves were shown only on assessable variants). To account for this, the coverage of each predictor of the MDR3 dataset was computed.

2.5. Performance evaluation

The performance of Vador and the other prediction tools was evaluated using recommended measures for binary classifiers (72), including additionally the F1-Score as well as visualization in ROC and precision-recall-curves. The measures are based on the values of correctly classified variants, indicated by True Positives (TP) for correctly predicted pathogenic variants and True Negatives (TN) for correctly predicted benign variants, as well as incorrectly classified variants, indicated by False Positives (FP) for variants predicted as pathogenic albeit being benign and False Negatives (FN) for variants predicted as benign albeit being

pathogenic. The analyzed measures of recall, specificity, precision, negative predictive value (NPV), accuracy, F1-score, and Matthews correlation coefficient were calculated as follows:

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2 TP}{2 TP + FP + FN}$$

$$Matthews correlation coefficient = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

2.6. Webserver tool

Vasor can be accessed online via https://cpclab.uni-duesseldorf.de/mdr3_predictor/. Users can enter a single-site amino acid missense MDR3 variant of interest, keeping in mind that the tool will only recognize MDR3 variants corresponding to the largest protein isoform, UniProt ID P21439-1. Further, the entry needs to be in the format of the standard IUPAC code for amino acids, entering first the one-letter code of the amino acid of the reference sequence, followed by the position and the amino acid substitution of interest. On the results page, users can see the predicted classification (either benign or pathogenic) and the probability of the given variant being pathogenic. This probability ranges from 0 (highest probability for the variant to be benign) to 1 (highest probability for the variant to be pathogenic). Probability values close to the cut-off value of 0.5 indicate less confidence in the prediction.

Additionally, the results page displays the structure of the MDR3 protein (PDB ID: 6S7P) with the NGL Viewer (73,74), including the membrane localization obtained from the OPM database (75) as a red and blue plane. The substituted residue is colored according to the predicted effect either in red (pathogenic) or green (benign). The user can download a zip archive containing a high-resolution image of the complete protein, PDB files of the reference sequence and the variant protein, and high-resolution images of the position with the reference sequence residue or the substituted one.

2.7. Code availability

The code for Vador was written in Python 3.9 and will be provided for download at <https://cpclab.uni-duesseldorf.de/index.php/Software>.

3. Results

3.1. Generation of a dataset with informative features and good overall coverage of the MDR3 protein

To establish an MDR3-specific prediction tool, we first prepared a dataset of benign and pathogenic MDR3 variants. Relevant literature on MDR3-associated diseases was screened. Variants with unclear association to effects were omitted to avoid misclassified variants within the dataset. In addition, the gnomAD database (58) was screened for MDR3 variants, and the results were subjected to filtering by VarSome (59) using ACMG-AMP rules (60) to remove variants with a high potential for a pathogenic effect. This step was necessary as pathogenic MDR3 variants on a single allele with a potential late-onset or mild phenotype might have been included in the gnomAD database. Next, we used well-established general predictors (EVE (17), EVmutation (23), PolyPhen-2 (18), I-Mutant2.0 (19), MUpro (20), MAESTRO (56) and PON-P2 (22)) and descriptors of the variant site, namely the disruption of secondary structure, possible PTM site disturbance, and changes in the relative solvent accessibility and half-sphere exposure of the position in question, as features in the dataset. Projecting the variant locations from the dataset onto the known cryo-EM structure of MDR3 (PDB ID 6S7P) (4) revealed a broad coverage of the structure with benign and pathogenic variants (Fig. 2A). No functional domain is devoid of variants, and we do not observe large

clusters of benign or pathogenic variants, which may be indicative of a potential bias within the dataset. Such a bias might prevent applying the tool to areas of low coverage. Hence, we expect that our tool can generalize predictions to every position of MDR3.

To further probe for domains of low applicability, we mapped variants misclassified by Vamor to the MDR3 structure. Misclassified variants from the dataset tend to occur on the solvent-exposed surface of the protein rather than within buried regions of the protein (Suppl. Fig. 3). As solvent-exposed residues are less evolutionary conserved than buried residues (76), the obtained trend might visualize the underlying increased uncertainty of those integrated general predictors that are based on evolutionary sequence conservation. Overall, also given the small number of misclassifications, we do not see indications of domains of increased uncertainty for MDR3 predictions. The correlation coefficients between input features range from -0.64 to 0.76 over the 18 features (Fig. 2 B), indicating that each feature adds information that does not overlap with information from another feature.

3.2. Generating Vamor: training the XGBoost algorithm on the dataset

For Machine Learning models to function reliably, it is vital to estimate potential over- or under-fitting of the trained model. One of the most important techniques in that respect is the hold-out method, where a subsection of the entire dataset is split off as an external test set. Ideally, the test set has a similar probability distribution as the entire dataset (77); however, this is not certain if a test set is randomly drawn. Therefore, we paid attention to drawing our test set with a similar distribution of amino acids, as to both reference sequence and variant amino acid distributions, by minimizing the RMSD-based difference in amino acid distributions to the overall dataset. Besides, the test set contained an equal amount of benign and pathogenic variants, 20 each (Suppl. Fig. 1).

Next, for the remaining dataset, SMOTE (70) was used to create synthetic examples of the minority class (pathogenic variants) to balance the classes as binary classifiers otherwise tend to favor the majority class (benign variants) for prediction outcomes. The final training dataset consisted of 259 data points for each class, benign and pathogenic, upon which an XGBoost algorithm was trained. To evaluate the most important features, we measured and visualized feature importance (Suppl. Fig. 4) and removed the four consistently least-important features (Suppl. Fig. 5) without reducing performance. Of note, EVE is highly important for the

prediction outcome of the model, indicating that Vador primarily relies on EVE's predictions compared to other features.

Performance estimates were visualized within a repeated *k*-fold cross-validation and compared to the performance against the held-out test set (Fig. 3 A). The trained model performs on the test set with an accuracy of 90 %, with 18 out of 20 variants being predicted correctly, both for the benign and the pathogenic class (Fig. 3 B). Notably, the performance based on the *k*-fold cross-validation does not differ from that on the independent test set, indicating a well-fit model without over- or under-fitting.

3.3. Vador outperforms integrated general predictors and the external general predictor MutPred2

We compared the performance of Vador with general predictors on the entire dataset. We compared Vador to EVE, PolyPhen-2, and PON-P2, which were integrated as features into the dataset on which Vador was trained. As such, Vador should outperform each predictor due to the additional information gathered from the other features. Additionally, we compared Vador to MutPred2 as an external prediction tool; the predecessor tool MutPred was indicated to perform well on MDR3 classification problems (16). Vador outperformed EVE, PolyPhen-2, PON-P2, and MutPred2 according to ROC (Fig. 4 A) and precision-recall-curves (Fig. 4 C), with an area under the curve (AUC) of 0.98 for Vador against 0.90 for EVE, 0.89 for MutPred2, 0.87 for PolyPhen2, and 0.80 for PON-P2 for the ROC and an AUC of 0.94 for Vador against an AUC of 0.86 for EVE, 0.74 for MutPred2, 0.72 for PolyPhen2, and 0.51 for PON-P2 for the precision-recall-curves. Precision-recall-curves have been shown to be more robust and accurate for binary classifiers on imbalanced datasets (77), like in our case.

Noteworthy, the second-best performing predictor, EVE, was the most important feature for Vador, suggesting that the machine learning model recognized the information contained within this feature as highly correlated with the true output and its value in predicting the output correctly. However, EVE could only predict 85.7 % of the variants in the dataset, whereas Vador, by design, predicted an outcome for every possible missense variant of MDR3 (Fig. 4 B and Table 1).

Additional performance measures are summarized in Table 1, indicating that Vador outperforms existing prediction tools according to the weighted measures F1-Score (0.85)

and Matthews correlation coefficient (MCC) (0.81). Specifically, Vazor achieved a low number of False Negatives. Comparable low values in False Negatives are achieved by PolyPhen2, but at the cost of an increased number of False Positives, and PON-P2, but only at coverage of 45.1 % of the variants in the MDR3 protein and an increased number of False Positives.

When comparing the performance of the missense predictors on the test set only (Suppl. Table 1), our tool reached the best scores in F1-Score and MCC (0.90 and 0.80, respectively) compared to other predictors with full coverage of the test set. EVE showed F1-Score and MCC values of 0.91 and 0.83, respectively, on a subset (82.5 %) of variants where it reached a prediction. By contrast, MutPred2 was able to predict every pathogenic variant as pathogenic, albeit at the cost of predicting almost half of the benign variants as pathogenic, resulting in a high number of False Positives.

Overall, Vazor outperformed other predictors consistently according to ROC and precision-recall-curves, revealing a well-balanced prediction with few False Negatives and False Positives both on the entire dataset and the test set.

3.4. Vazor classifies the majority of variants with high certainty

Additionally, we investigated the distribution of Vazor's output, the probability of pathogenicity values. Vazor assigns the majority of benign cases low probability values (75% of benign variants < 0.23 probability of pathogenicity), whereas the majority of pathogenic cases are assigned a high probability value (75% of pathogenic variants > 0.76 probability of pathogenicity) (Fig. 5). Furthermore, Vazor showed no misclassifications of variants in the dataset for values below 0.24 and above 0.84, indicating high certainty for benign variant predictions in the range 0 to 0.24 (75 % of the benign variants) and pathogenic variant predictions in the range 0.84 to 1 (60 % of the pathogenic variants).

We further investigated the usage of SMOTE to generate data points for the minority class (i.e., pathogenic variants). Due to the method underlying SMOTE, SMOTE-generated data points are expected to follow the distribution of pathogenic variants within the probability of pathogenicity curve. Accordingly, no SMOTE data point was predicted with a lower value of probability of pathogenicity than 0.29, and data points mainly clustered within the high certainty zone (Suppl. Fig. 6).

Overall, Vazor showed a robust separation of probability of pathogenicity values of both variant classes, indicating that Vazor classified most variants within the dataset with high certainty.

3.5. Easy accessibility of Vazor as a webserver tool

Using Vazor, we precalculated the effect of every possible amino acid substitution for MDR3, resulting in a heatmap of 1286 * 20 probabilities of pathogenicity (Fig. 6, SI). We mapped the average probability of pathogenicity of each position onto the MDR3 protein structure to visualize positions that are functionally more sensitive to substitutions (Fig. 7). As expected, areas near the ATP-binding site within the nucleotide-binding domain displayed a high average probability of pathogenicity. Similarly, buried residues within the helices forming the transmembrane part showed high sensitivity as several missense mutations may lead to a disruption of the helical structure. More exposed residues located on the outsides of helices or in flexible regions, such as the small extracellular loops, displayed less sensitivity. However, this trend does not exclude that specific variants at seemingly less sensitive sites can be pathogenic, and vice versa.

We also used the pre-computed heatmap for rapid lookup and output generation of the webserver tool, thus, eliminating waiting time for users needing a prediction for a specific MDR3 variant. The webserver can be accessed at https://cpclab.uni-duesseldorf.de/mdr3_predictor/. It requires as input an MDR3 variant (with the amino acid of the reference sequence in the one-letter format, its position within the canonical sequence of Uniprot ID P21439-1, and the substituted amino acid in the one-letter format) and yields the predicted effect of the entered variant, either benign or pathogenic, together with the probability of pathogenicity. Additionally, the variant position is depicted in the 3D structure of MDR3, and high-quality images of reference sequence amino acid, variant, and the overall MDR3 structure can be downloaded. The heatmap is also downloadable from the webserver for implementation in other applications.

4. Discussion

Although recent years have resulted in a plethora of general predictors for protein properties, their performance on specific proteins of interest can differ greatly (79). While existing state-of-the-art tools to predict substitution effects perform admirably on the MDR3 protein, especially EVE (17), the potential for improvement is given both for the performance on and coverage of the MDR3 dataset, since not every general predictor can classify each MDR3 variant.

To improve predictions, we created the, to our knowledge, largest dataset specific for pathogenic and benign variants of MDR3, obtained from the literature and gnomAD database and comprising 85 pathogenic and 279 benign variants. As the generation of a high-quality dataset is a critical first step for any machine learning approach (71,78), we carefully screened the literature specifically for MDR3 variants, filtering out variants with unclear disease associations. To counteract the bias that mainly pathogenic variants are chosen for detailed *in vitro* or *in vivo* analysis, we obtained variants from the gnomAD database (58). Since there may be potentially disease-associated variants in the database, we implemented an additional filtering step of removing variants categorized as likely pathogenic or pathogenic as evaluated by VarSome (59) to exclude false-negative variants. The dataset resulting from this strategy was then kept as is, i.e., no variants were added or removed, thus eliminating the potential to introduce bias from the researcher. Using established general predictors and variant site properties, we trained an MDR3-specific machine-learning model, termed Vador, to classify protein missense variants into benign or pathogenic. Vador outperforms general predictors: Over the entire dataset, Vador shows F1-Score and MCC of 0.85 and 0.81, respectively, while the second-best method, EVE, follows with scores of 0.81 and 0.77, respectively, but coverage of only 85.7 %. By contrast, Vador ensures high-quality predictions for all MDR3 missense variants. As machine learning models trained on a specific dataset exhibit a bias towards overperformance on this dataset, Vador has an inherent advantage when evaluated on the entire dataset over other predictors. Notably, the superior performance of Vador is also present on the independent test set, where Vador only misclassified two (5 %) benign and two (5 %) pathogenic variants, leading to the highest performance compared to other predictors as indicated by F1-Score and MCC of 0.9 and 0.8, respectively. Although EVE and PON-P2 achieve similar performances for the test set, they

only cover a fraction of the variants (82.5 % and 37.5 %, respectively). Overall, no other analyzed predictor provided a similarly good balance of consistently low False Negative and False Positive predictions. Both measures have important implications for using Vazor within a clinical setting: Predictors with a high number of False Negatives will lead to variants found within patients being falsely given no attention, whereas a high number of False Positives will result in a predictor raising too often a false alarm for an actually benign variant.

We established an easily accessible webserver for reliable and fast predictions of novel MDR3 variants based on Vazor. It can serve as an important step for deciding which variants to study and to get the first indication of a variant effect. It does not eliminate the need for classical *in vitro* studies for mutational impact, however, and in a clinical setting, the ACMG-AMP guidelines (60) should be followed. The webserver classifies single-site amino acid substitutions into the categories benign or pathogenic. Truncation, insertion, and deletion variants of MDR3 cannot be assessed. However, the probability of pathogenicity for such variants is often more definite (81,82). Of note, the effect of a single missense variant within the biological context might not always be a clear-cut pathogenic or benign effect. Therefore, the probability of pathogenicity provided by the webserver can act as an indicator of prediction reliability.

As a limitation, the exact mechanism underlying a pathogenic variant cannot be inferred from the current tool. MDR3 missense variants may impact protein folding and maturation, activity, or stability (13), and several of these categories can be influenced. Information on mechanistic dysfunction may aid in targeted therapy. In terms of machine learning, such a multi-class classification problem might be solved – with the premise of a sizeable dataset of quality-assured variants. Unfortunately, we are unaware of such a dataset for MDR3. The currently employed dataset strived for such quality-assured variants; however, especially lacking large-scale functional studies of benign variants, variants indicated by VarSome as of unclear significance were included. Thus, we encourage the scientific community to submit novel MDR3 variants with a proven effect on folding, maturation, activity, and stability to the authors for addition into the dataset to improve and develop Vazor further.

5. Acknowledgments

We are grateful for computational support and infrastructure provided by the “Zentrum für Informations- und Medientechnologie” (ZIM) at the Heinrich Heine University Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) to HG on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) (user ID: HKF7, VSK33).

Fig. 1: Graphical overview of dataset generation and machine learning approach. For details see text.

Fig. 2: Coverage of MDR3 by the dataset and correlation analysis of features. [A] Mapping of dataset variants onto the MDR3 structure. Benign variants are marked in green and pathogenic variants in magenta. [B] Spearman rank correlation matrix of features computed for the dataset. Abb.: prob. – probability, conf. – confidence, st. sign – stability sign, RI – reliability index, sec. structure – secondary structure, NBD - nucleotide-binding domain, PTM – posttranslational modification, RSA – relative solvent accessibility, HSE – half-sphere exposure, epi. – epistatic, ind. – independent.

Fig. 3: Performance of Vasor on the test set. [A] ROC curve of the Vasor performance on the test set (green line) compared to performance estimates from repeated *k*-fold cross-validation (black lines). [B] Confusion matrix of Vasor performance on the test set.

Fig. 4: Performance of Vasor in comparison to established general predictors. [A] ROC curve of the performance of Vasor, EVE, PolyPhen-2, PON-P2, and MutPred2 on the variants of the entire dataset. Note that the performance was determined for those variants each predictor was able to make a prediction for (see [B]). [B] Coverage of dataset variants by the predictors. [C] Precision-recall-curves of the predictors. Performance was determined for those variants each predictor was able to make a prediction for.

Fig. 5: Distribution of probability of pathogenicity values over the entire dataset. Distribution of Vazor's probability of pathogenicity output for benign (blue) and pathogenic (red) variants. Vazor classified 75 % of benign variants into the benign category with values below 0.23, which is below the lowest probability value of any pathogenic variant (0.24) within the dataset. 60 % of pathogenic variants were classified into the pathogenic category with values above 0.84, which is greater than the highest probability value of any benign variant (0.84) within the dataset. 75 % of pathogenic variants were classified with probability values greater than 0.76.

Fig. 6: Heatmap of predictions for every possible amino acid substitution in MDR3. [A] Color-coded predictions for every position (displayed on the y-axis) within the MDR3 protein and every possible amino acid substitution (x-axis). Prediction values range from likely benign (blue) to likely pathogenic (red). [B] Secondary structure of MDR3. α -helical stretches are depicted as green zig-zag curves, β -sheet stretches as orange arrows. [C] Domains, secondary structure elements, and characteristic motives are indicated on the right. TM H: Transmembrane Helix, NBD: Nucleotide-binding domain.

Fig. 7: Mapping the average pathogenicity onto the structure of MDR3. Prediction values for each position were averaged over all possible substitutions. Values closer to 0 (most likely benign) correspond to blue, values closer to 1 (most likely pathogenic) correspond to red residues.

Table 1: Detailed performance measurements of Vazor in comparison to EVE, PolyPhen-2, PON-P2, and MutPred2 on the entire dataset.

	Vazor	EVE	PolyPhen-2	PON-P2	MutPred2
Recall	0.84	0.73	0.84	0.74	0.52
Specificity	0.96	0.98	0.74	0.89	0.95
Precision	0.87	0.91	0.49	0.52	0.77
NPR	0.95	0.93	0.94	0.95	0.87
Accuracy	0.93	0.92	0.76	0.87	0.85
F1-Score	0.85	0.81	0.62	0.61	0.62
MCC	0.81	0.77	0.50	0.54	0.55
TP	71	52	71	17	44
FN	14	19	14	6	41
TN	268	236	206	125	266
FP	11	5	73	16	13
Coverage [%]	100	85.7	100	45.1	100

References

1. Smith AJ, Timmermans-Hereijgers JLPM, Roelofsen B, Wirtz KWA, van Blitterswijk WJ, Smit JJM, et al. The human MDR3 P-glycoprotein promotes translocation of phosphatidylcholine through the plasma membrane of fibroblasts from transgenic mice. *FEBS Letters*. 1994 Nov 14;354(3):263–6.
2. van Helvoort A, Smith AJ, Sprong H, Fritzsche I, Schinkel AH, Borst P, et al. MDR1 P-Glycoprotein Is a Lipid Translocase of Broad Specificity, While MDR3 P-Glycoprotein Specifically Translocates Phosphatidylcholine. *Cell*. 1996 Nov;87(3):507–17.
3. Oude Elferink RPJ, Paulusma CC. Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein). Vol. 453, *Pflugers Archiv European Journal of Physiology*. 2007. p. 601–10.
4. Olsen JA, Alam A, Kowal J, Stieger B, Locher KP. Structure of the human lipid exporter ABCB4 in a lipid environment. *Nature Structural and Molecular Biology*. 2020 Jan 1;27(1):62–70.
5. Prescher M, Bonus M, Stindt J, Keitel-Anselmino V, Smits SHJ, Gohlke H, et al. Evidence for a credit-card-swipe mechanism in the human PC floppase ABCB4. *Structure*. 2021 Oct;29(10):1144-1155.e5.
6. Rosmorduc O, Hermelin B, Poupon R. MDR3 gene defect in adults with symptomatic intrahepatic and gallbladder cholesterol cholelithiasis. *Gastroenterology*. 2001;120(6):1459–67.
7. Deleuze J, Jacquemin E, Dubuisson C, Cresteil D, Dumont M, Erlinger S, et al. Defect of multidrug-resistance 3 gene expression in a subtype of progressive familial intrahepatic cholestasis. *Hepatology*. 1996 Apr;23(4):904–8.
8. Lang C, Meier Y, Stieger B, Beuers U, Lang T, Kerb R, et al. Mutations and polymorphisms in the bile salt export pump and the multidrug resistance protein 3 associated with drug-induced liver injury. *Pharmacogenetics and Genomics*. 2007 Jan;17(1):47–60.
9. Dröge C, Bonus M, Baumann U, Klindt C, Lainka E, Kathemann S, et al. Sequencing of FIC1, BSEP and MDR3 in a large cohort of patients with cholestasis revealed a high number of different genetic variants. *Journal of Hepatology*. 2017;67(6):1253–64.
10. Pauli-Magnus C, Lang T, Meier Y, Zodan-Marin T, Jung D, Breymann C, et al. Sequence analysis of bile salt export pump (ABCB11) and multidrug resistance p-glycoprotein 3 (ABCB4, MDR3) in patients with intrahepatic cholestasis of pregnancy. *Lippincott Williams & Wilkins Pharmacogenetics*. 2004;14:91–102.
11. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics*. 2015 May 25;47(5):435–44.
12. Dong C, Condat B, Picon-Coste M, Chretien Y, Potier P, Noblinski B, et al. Low phospholipid-associated cholelithiasis syndrome: prevalence, clinical features, and comorbidities. *JHEP Reports [Internet]*. 2020;100201. Available from: <https://doi.org/10.1016/j.jhepr.2020.100201>

13. Delaunay JL, Durand-Schneider AM, Dossier C, Falguières T, Gautherot J, Davit-Spraul A, et al. A functional classification of ABCB4 variations causing progressive familial intrahepatic cholestasis type 3. *Hepatology*. 2016;63(5):1620–31.
14. Hassan MS, Shaalan AA, Dessouky MI, Abdelnaïem AE, ElHefnawi M. A review study: Computational techniques for expecting the impact of non-synonymous single nucleotide variants in human diseases. *Gene*. 2019 Jan;680:20–33.
15. Niroula A, Vihinen M. Variation Interpretation Predictors: Principles, Types, Performance, and Choice. Vol. 37, *Human Mutation*. John Wiley and Sons Inc.; 2016. p. 579–97.
16. Khabou B, Durand-Schneider AM, Delaunay JL, Aït-Slimane T, Barbu V, Fakhfakh F, et al. Comparison of in silico prediction and experimental assessment of ABCB4 variants identified in patients with biliary diseases. *International Journal of Biochemistry and Cell Biology*. 2017 Aug 1;89:101–9.
17. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature*. 2021 Nov 4;599(7883):91–5.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Vol. 7, *Nature Methods*. 2010. p. 248–9.
19. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*. 2005 Jul;33(SUPPL. 2).
20. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function and Genetics*. 2006 Mar 1;62(4):1125–32.
21. Laimer J, Hofer H, Fritz M, Wegenkittl S, Lackner P. MAESTRO - multi agent stability prediction upon point mutations. *BMC Bioinformatics*. 2015 Dec;16(1).
22. Niroula A, Urolagin S, Vihinen M. PON-P2: Prediction method for fast and reliable identification of harmful variants. *PLoS ONE*. 2015 Feb 3;10(2).
23. Hopf TA, Ingraham JB, Poelwijk FJ, Schärfe CPI, Springer M, Sander C, et al. Mutation effects predicted from sequence co-variation. *Nature Biotechnology*. 2017 Feb 1;35(2):128–35.
24. Kubitz R, Bode J, Erhardt A, Graf D, Kircheis G, Müller-Stöver I, et al. Cholestatic liver diseases from child to adult: The diversity of MDR3 disease. *Zeitschrift für Gastroenterologie*. 2011;49(6):728–36.
25. Wendum D, Barbu V, Rosmorduc O, Arrivé L. Aspects of liver pathology in adult patients with MDR3 / ABCB4 gene mutations. 2012;291–8.
26. Gautherot J, Delautier D, Maubert MA, Aït-Slimane T, Bolbach G, Delaunay JL, et al. Phosphorylation of ABCB4 impacts its function: Insights from disease-causing mutations. *Hepatology*. 2014;60(2):610–21.
27. Poupon R, Rosmorduc O, Boëlle PY, Chrétien Y, Corpechot C, Chazouillères O, et al. Genotype-phenotype relationships in the low-phospholipid-associated cholelithiasis syndrome: A study of 156 consecutive patients. *Hepatology*. 2013 Sep;58(3):1105–10.

28. Gordo-Gilart R, Andueza S, Hierro L, Martínez-Fernández P, D'Agostino D, Jara P, et al. Functional analysis of ABCB4 mutations relates clinical outcomes of progressive familial intrahepatic cholestasis type 3 to the degree of MDR3 floppase activity. *Gut*. 2015 Jan 1;64(1):147–55.
29. Colombo C, Vajro P, Degiorgio D, Coviello DA, Costantino L, Tornillo L, et al. Clinical features and genotype-phenotype correlations in children with progressive familial intrahepatic cholestasis type 3 related to ABCB4 mutations. *Journal of Pediatric Gastroenterology and Nutrition*. 2011 Jan;52(1):73–83.
30. Degiorgio D, Colombo C, Seia M, Porcaro L, Costantino L, Zazzeron L, et al. Molecular characterization and structural implications of 25 new ABCB4 mutations in progressive familial intrahepatic cholestasis type 3 (PFIC3). *European Journal of Human Genetics*. 2007 Dec;15(12):1230–8.
31. Pauli-Magnus C, Kerb R, Fattinger K, Lang T, Anwald B, Kullak-Ublick GA, et al. BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. *Hepatology*. 2004 Mar;39(3):779–91.
32. Gordo-Gilart R, Hierro L, Andueza S, Muñoz-Bartolo G, López C, Díaz C, et al. Heterozygous ABCB4 mutations in children with cholestatic liver disease. *Liver International*. 2016 Feb 1;36(2):258–67.
33. Andress EJ, Nicolaou M, McGeoghan F, Linton KJ. ABCB4 missense mutations D243A, K435T, G535D, I490T, R545C, and S978P significantly impair the lipid floppase and likely predispose to secondary pathologies in the human population. *Cellular and Molecular Life Sciences*. 2017 Jul 1;74(13):2513–24.
34. Rosmorduc O, Hermelin B, Boelle PY, Parc R, Taboury J, Poupon R. ABCB4 gene mutation-associated cholelithiasis in adults. *Gastroenterology*. 2003 Aug 1;125(2):452–9.
35. Andress EJ, Nicolaou M, Romero MR, Naik S, Dixon PH, Williamson C, et al. Molecular mechanistic explanation for the spectrum of cholestatic disease caused by the S320F variant of ABCB4. *Hepatology*. 2014;59(5):1921–31.
36. Poupon R, Barbu V, Chamouard P, Wendum D, Rosmorduc O, Housset C. Combined features of low phospholipid-associated cholelithiasis and progressive familial intrahepatic cholestasis 3. *Liver International*. 2010 Feb;30(2):327–31.
37. Bacq Y, Gendrot C, Perrotin F, Lefrou L, Chrétien S, Vie-Buret V, et al. ABCB4 gene mutations and single-nucleotide polymorphisms in women with intrahepatic cholestasis of pregnancy. *Journal of Medical Genetics*. 2009 Oct;46(10):711–5.
38. Keitel V, Vogt C, Häussinger D, Kubitz R. Combined Mutations of Canalicular Transporter Proteins Cause Severe Intrahepatic Cholestasis of Pregnancy. *Gastroenterology*. 2006;131(2):624–9.
39. Jacquemin E, DeVree JML, Cresteil D, Sokal EM, Sturm E, Dumont M, et al. The wide spectrum of multidrug resistance 3 deficiency: From neonatal cholestasis to cirrhosis of adulthood. *Gastroenterology*. 2001;120(6):1448–58.

40. Saleem K, Cui Q, Zaib T, Zhu S, Qin Q, Wang Y, et al. Evaluation of a Novel Missense Mutation in ABCB4 Gene Causing Progressive Familial Intrahepatic Cholestasis Type 3. *Disease Markers*. 2020;2020.
41. Degiorgio D, Corsetto PA, Rizzo AM, Colombo C, Seia M, Costantino L, et al. Two ABCB4 point mutations of strategic NBD-motifs do not prevent protein targeting to the plasma membrane but promote MDR3 dysfunction. *European Journal of Human Genetics*. 2014;22(5):633–9.
42. Fang LJ, Wang XH, Knisely AS, Yu H, Lu Y, Liu LY, et al. Chinese children with chronic intrahepatic cholestasis and high γ -glutamyl transpeptidase: Clinical features and association with ABCB4 mutations. *Journal of Pediatric Gastroenterology and Nutrition*. 2012 Aug;55(2):150–6.
43. Tougeron D, Fotsing G, Barbu V, Beauchant M. ABCB4/MDR3 gene mutations and cholangiocarcinomas. Vol. 57, *Journal of Hepatology*. 2012. p. 467–8.
44. Ziol M, Barbu V, Rosmorduc O, Frassati-Biaggi A, Barget N, Hermelin B, et al. ABCB4 Heterozygous Gene Mutations Associated With Fibrosing Cholestatic Liver Disease in Adults. *Gastroenterology*. 2008;135(1):131–41.
45. Floreani A, Carderi I, Paternoster D, Soardo G, Azzaroli F, Esposito W, et al. Intrahepatic cholestasis of pregnancy: Three novel MDR3 gene mutations. *Alimentary Pharmacology and Therapeutics*. 2006 Jun;23(11):1649–53.
46. Delaunay J-L, Bruneau A, Hoffmann B, Durand-Schneider A-M, Eronique Barbu V, Jacquemin E, et al. Functional Defect of Variants in the Adenosine Triphosphate-Binding Sites of ABCB4 and Their Rescue by the Cystic Fibrosis Transmembrane Conductance Regulator Potentiator, Ivacaftor (VX-770). 2016;
47. Lucena JF, Herrero JI, Quiroga J, Sangro B, Garcia-Foncillas J, Zabalegui N, et al. A multidrug resistance 3 gene mutation causing cholelithiasis, cholestasis of pregnancy, and adulthood biliary cirrhosis. *Gastroenterology*. 2003 Apr 1;124(4):1037–42.
48. Floreani A, Carderi I, Paternoster D, Soardo G, Azzaroli F, Esposito W, et al. Hepatobiliary phospholipid transporter ABCB4, MDR3 gene variants in a large cohort of Italian women with intrahepatic cholestasis of pregnancy. *Digestive and Liver Disease*. 2008 May;40(5):366–70.
49. Degiorgio D, Crosignani A, Colombo C, Bordo D, Zuin M, Vassallo E, et al. ABCB4 mutations in adult patients with cholestatic liver disease: impact and phenotypic expression. *Journal of Gastroenterology*. 2016 Mar 1;51(3):271–80.
50. Dixon PH, Weerasekera N, Linton KJ, Donaldson O, Chambers J, Egginton E, et al. Heterozygous MDR3 missense mutation associated with intrahepatic cholestasis of pregnancy: evidence for a defect in protein trafficking. Vol. 9, *Human Molecular Genetics*. 2000.
51. Gotthardt D, Runz H, Keitel V, Fischer C, Flechtenmacher C, Wirtenberger M, et al. A mutation in the canalicular phospholipid transporter gene, ABCB4, is associated with cholestasis, ductopenia, and cirrhosis in adults. *Hepatology*. 2008 Oct;48(4):1157–66.
52. Keitel V, Burdelski M, Warskulat U, Kühlkamp T, Keppler D, Häussinger D, et al. Expression and localization of hepatobiliary transport proteins in progressive familial intrahepatic cholestasis. *Hepatology*. 2005 May;41(5):1160–72.

53. Keitel V, Dröge C, Stepanow S, Fehm T, Mayatepek E, Köhrer K, et al. Intrahepatic cholestasis of pregnancy (ICP): case report and review of the literature. *Zeitschrift für Gastroenterologie*. 2016 Dec 1;54(12):1327–33.
54. Denk GU, Bikker H, Lekanne dit Deprez RH, Terpstra V, van der Loos C, Beuers U, et al. ABCB4 deficiency: A family saga of early onset cholelithiasis, sclerosing cholangitis and cirrhosis and a novel mutation in the ABCB4 gene. *Hepatology Research*. 2010 Sep;40(9):937–41.
55. Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E. The spectrum of liver diseases related to ABCB4 gene mutations: Pathophysiology and clinical aspects. *Seminars in Liver Disease*. 2010;30(2):134–46.
56. Kluth M, Stindt J, Dröge C, Linnemann D, Kubitz R, Schmitt L. A mutation within the extended X loop abolished substrate-induced ATPase activity of the human liver ATP-binding cassette (ABC) transporter MDR3. *Journal of Biological Chemistry*. 2015 Feb 20;290(8):4896–907.
57. Dzaganian T, Engelmann G, Häussinger D, Schmitt L, Flechtenmacher C, Rtskhiladze I, et al. The histidin-loop is essential for transport activity of human MDR3. A novel mutation of MDR3 in a patient with progressive familial intrahepatic cholestasis type 3. *Gene*. 2012 Sep 10;506(1):141–5.
58. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020 May 28;581(7809):434–43.
59. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. *Bioinformatics*. 2019 Jun 1;35(11):1978–80.
60. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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. *Genetics in Medicine*. 2015 May;17(5):405–24.
61. Bateman A, Martin M-J, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D480–9.
62. Amanchy R, Periaswamy B, Mathivanan S, Reddy R, Tattikota SG, Pandey A. A curated compendium of phosphorylation motifs. *Nature Biotechnology*. 2007 Mar;25(3):285–6.
63. Hornbeck P v., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*. 2015 Jan 28;43(D1):D512–20.
64. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*. 1999 Dec;294(5):1351–62.
65. Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, et al. ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Research*. 2019 Nov 4;
66. Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, et al. A series of PDB related databases for everyday needs. *Nucleic Acids Research*. 2011 Jan 1;39(Database):D411–9.

67. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983 Dec;22(12):2577–637.
68. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum allowed solvent accessibilities of residues in proteins. *PLoS ONE*. 2013 Nov 21;8(11).
69. Hamelryck T. An amino acid has two sides: A new 2D measure provides a different view of solvent exposure. *Proteins: Structure, Function and Genetics*. 2005 Apr 1;59(1):38–48.
70. Chawla N v, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. Vol. 16, *Journal of Artificial Intelligence Research*. 2002.
71. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2016. p. 785–94.
72. Vihinen M. How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC Genomics*. 2012;13(Suppl 4):S2.
73. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlić A, Rose PW. Web-based molecular graphics for large complexes. In: *Proceedings of the 21st International Conference on Web3D Technology*. New York, NY, USA: ACM; 2016. p. 185–6.
74. Rose AS, Hildebrand PW. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Research*. 2015 Jul 1;43(W1):W576–9.
75. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research*. 2012 Jan;40(D1):D370–6.
76. Echave J, Spielman SJ, Wilke CO. Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics*. 2016 Feb 19;17(2):109–21.
77. Raschka S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. 2018 Nov 13; Available from: <http://arxiv.org/abs/1811.12808>
78. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*. 2015 Mar 4;10(3):e0118432.
79. Riera C, Padilla N, de la Cruz X. The Complementarity Between Protein-Specific and General Pathogenicity Predictors for Amino Acid Substitutions. *Human Mutation*. 2016 Oct 1;37(10):1013–24.
80. Walsh I, Pollastri G, Tosatto SCE. Correct machine learning on protein sequences: A peer-reviewing perspective. *Briefings in Bioinformatics*. 2016 Sep 1;17(5):831–40.
81. Balasubramanian S, Fu Y, Pawashe M, McGillivray P, Jin M, Liu J, et al. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nature Communications*. 2017 Dec 29;8(1):382.
82. Lek M, Karczewski KJ, Minikel E v., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016 Aug 17;536(7616):285–91.