Published in final edited form as:

Circ Cardiovasc Genet. 2017 October ; 10(5): . doi:10.1161/CIRCGENETICS.117.001754.

Predicting the Functional Impact of KCNQ1 Variants of Unknown Significance

Bian Li, MSc[1,2], **Jeffrey L. Mendenhall, MSc**[1,2], **Brett M. Kroncke, PhD**[2,3,4], **Keenan C. Taylor, PhD**[2,3], **Hui Huang, PhD**[2,3], **Derek K. Smith, DDS, PhD**[6], **Carlos G. Vanoye, PhD**[5], **Jeffrey D. Blume, PhD**[6], **Alfred L. George Jr., MD**[5], **Charles R. Sanders, PhD**[2,3,4], and **Jens Meiler, PhD**[1,2]

[1]Department of Chemistry, Vanderbilt University, Nashville, TN

[2]Center for Structural Biology, Vanderbilt University, Nashville, TN

[3]Department of Biochemistry, Vanderbilt University, Nashville, TN

[4]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

[5]Department of Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, IL

[6]Department of Biostatistics, Vanderbilt University, Nashville, TN

Abstract

Background—An emerging standard-of-care for long QT syndrome (LQTS) employs clinical genetic testing to identify genetic variants of the KCNQ1 potassium channel. However, interpreting results from genetic testing is confounded by the presence of "variants of unknown significance" (VUS) for which there is inadequate evidence of pathogenicity.

Methods and Results—In this study, we curated from the literature a "high-quality" set of 107 functionally characterized KCNQ1 variants. Based on this dataset, we completed a detailed quantitative analysis on the sequence conservation patterns of subdomains of KCNQ1 and the distribution of pathogenic variants therein. We found that conserved subdomains generally are critical for channel function and are enriched with dysfunctional variants. Using this experimentally validated dataset, we trained a neural network, designated Q1VarPred, specifically for predicting the functional impact of KCNQ1 VUS. The estimated predictive performance of Q1VarPred in terms of Matthew's correlation coefficient and area under the receiver operating characteristic curve were 0.581 and 0.884, respectively, superior to the performance of eight previous methods tested in parallel. Q1VarPred is publicly available as a web server at http://meilerlab.org/q1varpred.

Correspondence: Jens Meiler, PhD, Departments of Chemistry and Pharmacology, Center for Structural Biology, Vanderbilt University, 465 21st Ave South, BIOSCI/MRBIII, Room 5144B, Nashville, TN 37232-8725, Tel: +1 (615) 936-5662, jens.meiler@vanderbilt.edu.

Data and Software Availability: The curated data set is included in Data Supplement and designated as Table S1. The dataset for training Q1VarPred is provided as a spreadsheet in Supplemental Materials. Q1VarPred was developed under the framework of the Biochemical Library (available at http://www.meilerlab.org/bclcommons) and is made publicly available as a web server at http://meilerlab.org/q1varpred.

Disclosures: None

Conclusions—Although a plethora of tools are available for making pathogenicity predictions over a genome-wide scale, previous tools fail to perform in a robust manner when applied to KCNQ1. The contrasting and favorable results for Q1VarPred suggests a promising approach, where a machine learning algorithm is tailored to a specific protein target and trained with a functionally validated dataset to calibrate informatics tools.

Keywords

long QT syndrome; KCNQ1; variants of unknown significance; neural networks

Introduction

Congenital long QT syndrome (LQTS) is a heart rhythm disorder that affects ~ 1 in 2,500 births.[1] It predisposes children and young adults to a type of ventricular tachycardia (torsades de pointes) and sudden cardiac death.[2] LQTS is associated with pathogenic variants in several genes that lead to dysfunctional cardiac ion channels. Among the 16 known LQTS-associated genes, *KCNQ1* variants account for ~ 30–35% of all LQTS cases. *KCNQ1* encodes the α-subunit of the voltage-gated K^+ channel KCNQ1 (also known as $K_V7.1$) that regulates the slow delayed rectifier current (I_{Ks}), a major driver of cardiac repolarization.[3] Loss of KCNQ1 function leads to diminished or dysfunctional I_{Ks}, impaired myocardial repolarization and LQTS.[4]

An emerging standard-of-care for LQTS employs clinical genetic testing to identify LQTS-associated variants.[4] Established genotype-phenotype relations should be factored into the assessment of the risk of sudden cardiac death and the selection of appropriate therapeutic interventions.[5] However, variants of unknown significance (VUS) for which there is inadequate evidence to classify as being pathogenic are common findings.[6] This issue is further confounded by the presence of background genetic “noise” (the frequency of genetic variations of a particular gene in a healthy population) and variants with incomplete penetrance.[5–7] Variant interpretation is bound to present an increasingly daunting challenge in the era of next-generation sequencing.[7–9]

Ideally, except for certain well-established disease-causing variants, positive LQTS genetic testing results should be evaluated by physiologically relevant experimental functional assays, but experimental characterization remains labor-intensive and costly to scale.[9, 10] Under such constraints, computational methods, which are usually machine learning-based, represent a common predictive approach.[8, 11, 12] However, hardly any computational methods are sufficiently accurate for clinical use related to channelopathies or other genetic disorders.[13, 14] Most existing computational methods have been trained on datasets pulled from online databases that have not been subjected to rigorous functional validation.[12] These datasets may be significantly contaminated with erroneous annotations and thereby provide machine-learning algorithms with misleading information.[12, 15] Further, a potentially even more crucial issue is that current methods intermingle two related but separate questions: whether a given variant causes functional impact at the molecular level and, if so, whether that functional effect will be manifested at the organismal level. Making such distinctions is

important when delivering predictions because dysfunction at molecular level does not necessarily equate to organismal deleteriousness.[7, 8]

In this study, we sought to develop a protein-specific algorithm capable of accurately predicting functional consequences of KCNQ1 variants. We first curated a set of functionally validated KCNQ1 variants. We then trained a neural network-based, KCNQ1-specific genotype–channel function relationship predictor Q1VarPred. In contrast to genome-wide methods, whose performances have suffered from dataset contamination and heterogeneity and do not differentiate between functional impact and organismal deleteriousness when delivering predictions, Q1VarPred was trained on the functionally validated dataset to predict molecular functional impact.

Materials and Methods

Dataset and criteria for annotating functional impact

KCNQ1 variants and their associated electrophysiological (EP) effects in the dataset for this study were collected from the literature (Table S1 in Data Supplement). We only considered data from experiments where the auxiliary subunit KCNE1 was also expressed. Each variant was annotated in terms of functional impact based on two experimental parameters (peak current relative to the wild-type and change in voltage of half-maximal activation V½). Specifically, a variant was defined as "Normal" if 1) 75% ≤ *peak current* ≤ 125% and 2) there was ≤10 mV depolarization or hyperpolarization shift in V½. "Mild Loss of Function" was defined as 1) 25% < *peak current* < 75% or 2) 10–20 mV depolarization shift in V½. "Severe Loss of Function" was defined as 1) *peak current* < 25% or 2) >20 mV depolarization shift in V½. "Severe Gain of Function" was defined as 1) >150% peak current or 2) 120 to 150 % peak current and >15 mV hyperpolarization shift in V½. Clinical classification (case variant versus control) of each variant was sourced from previous large-scale clinical studies,[16, 17] or EP studies that reported such information. Case variants were identified in patient cohort whereas control variants were found in healthy cohort. In addition, in accordance to the recent ACMG/AMP standards and guidelines for the interpretation of sequence variants,[12] variants with a minor allele frequency of > 1 / 2500 (LQTS prevalence) in the general population were removed. For training the binary classification model Q1VarPred, loss-of-function and gain-of-function variants were grouped together as dysfunctional and a mild loss-of-function variant was either labeled as dysfunctional if its peak current was < 50%, or normal otherwise. The common variant G643S was classified as having normal function.[18]

Neural network architecture and training

The neural network in the present study was a fully connected three-layer feed-forward network with a sigmoid transfer function. The input layer consists of two nodes, one for each predictive feature. The output layer consists of a single neuron, which outputs a numerical prediction of the functional impact of a given variant on the scale of 0 to 1 with 1 being complete dysfunction. A hidden layer with 3 neurons was chosen considering the fact that the "dropout" technique[19] was adopted to prevent the neural network from overfitting, a phenomenon in which the learned model is excessively complex (e.g. too many model

parameters relative to the number of observations for training) and is poorly generalizable. However, we also tested hidden layers with up to 8 neurons, the results of which showed that the size of the hidden layer did not affect the performance of the neural network in a significant way (Table S2 in Data Supplement). The neural network was trained on numeric encoding of variant functional labels (1 for complete dysfunction 0 for normal), with back-propagation of errors. The learning rate was set to 0.05 and momentum was set to 0.8. Weights were updated after each presentation of a variant to the network and a constant weight decay of 0.02 was applied to reduce model flexibility.

Predictive features

We used two features to characterize an amino acid substitution, namely rate of evolution, which quantifies the conservation of the sequence position where the substitution has occurred, and position-specific scoring matrix (PSSM)-based perturbation, which measures the radicalness of the substitution itself. These two features were chosen, before the dataset was inspected, based on the rationale that a very conserved position may tolerate less radical substitutions while a variable position may not tolerate more radical substitutions, as—for example—observed in a systematic mutation study of bacteriophage T4 lysozyme.[20] We confirmed that these features are critical among a limited number of features tested (Table S3 in Data Supplement). Details on how these two features were computed can be found in Methods in Data Supplement.

Performance metrics

The performance of the learned neural network model and other evaluated methods were quantified using the following metrics: true positive rate (TPR), true negative rate (TNR), positive predictive value (PPV), negative predictive value (NPV), accuracy, Matthew's correlation coefficient (MCC),[21] and area under the receiver operating characteristic (ROC) curve (AUC). Note that the first six metrics can be computed only after all variants are classified at a specific threshold. Using the notation of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), these metrics are defined as:

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\mathrm{TNR}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$$

$$\mathrm{PPV}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

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

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

$$\text{MCC}=\frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TN}+\text{FN})(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})}}$$

respectively. A TP is a dysfunctional variant classified as dysfunctional and TN is a normal variant classified as normal. MCC measures the correlation between predicted and observed binary classifications, with a value between −1 and 1. A MCC of 1 means perfect prediction, a value of 0 means no better than random prediction, and −1 indicates a completely reversed prediction. As MCC is unaffected by class size, it is a particularly useful measure of classification quality when the two classes are of very different sizes.[21] Computation of all performance metrics was accomplished using the ROCR package[22] implemented in the R programming environment.[23]

Estimating generalization ability

The generalization ability of a learned model is defined as its performance in predicting new variants that are not used for training. A model with higher generalization ability is favored over ones with lower generalization ability. A common practice to estimate a model's generalization ability is through a procedure called "*k*-fold cross validation" where the dataset is randomly divided into *k* equally-sized mutually exclusive subsets. The model is trained on $k-1$ subsets (collectively known as the training set) and its generalization ability is estimated on the remaining one subset (test set). Specifically, after the model is trained, a threshold is determined which maximizes the MCC on the training set, the same threshold is then used for computing the performance metrics on the test set. This process is repeated *k* times each using a different one of the *k* subsets as the test set and the remaining $k-1$ subsets as the training set. Every time a model is trained, its performance metrics are computed on the test set. In a *k*-fold cross-validation, the generalization ability is estimated as the average of performance metrics over *k* test sets. Because the number of ways a dataset can be split into *k* subsets is enormous, it is desirable to repeat the random splitting *k* times to reduce artifacts. In the current study, we chose $k=3$ and $p=200$, similar to a previous study.[24] The splitting was stratified such that the class proportions of the training set and the test set are as close to that of the whole dataset as possible. To ensure the consistency of comparison, the performance metrics of all evaluated methods were estimated using the exact same data. This means that every time the dataset was randomly split into 3 subsets, these subsets were used for calculating the performance metrics of all methods. The variability in performance metrics associated with random splitting of dataset is presented in Table S4 in Data Supplement.

Results

Functional studies do not always agree with clinical testing

We compiled a total of 107 functionally characterized KCNQ1 variants (Table S1 in Data Supplement). Two important observations were made on this dataset. First, a handful of case variants (variants identified in LQTS patient cohort, a total of 99 in our dataset) are functionally normal. Per our scheme of functional annotation (see Dataset and criteria for annotating functional impact), 6 out of 99 case variants are functionally normal and 8 out of 99 cause only mild loss of function. Interestingly, these two fractions roughly agree with the previous estimate that ~10% case variants may be false positives.[16] On the other hand, a few variants identified in presumed healthy controls are severely dysfunctional (for example, V110I and A300T). A300T, which occurs within the pore helix of the channel was shown to cause a massive reduction of I_{Ks} and hyperpolarization of the voltage of half-activation ($V_{1/2}$) of the channel both with and without the presence of the wild-type subunit.[25] The V110I variant showed significant reduction in I_{Ks} and depolarization of voltage of half-maximal activation when expressed in the absence of the wild-type subunit.[26] This analysis reinforces the argument that translating protein dysfunction at the molecular level to clinical manifestation and also attributing clinical manifestation to protein dysfunction both need to be carried out with caution.[5]

Position-specific rate of evolution reflects functionally-critical subdomains

The importance of a sequence site for protein structure or function can often be inferred from its conservation over evolution. We computed the position-specific rate of evolution for the entire sequence as well as the mean rate of evolution for each of the 24 subdomains of KCNQ1 (Methods in Data Supplement). A lower rate of evolution indicates higher conservation.

Overall, the N-terminal domain (NTD) and C-terminal domain (CTD) are generally less conserved than subdomains within the channel domain (CD), as shown in Figure 1. The average rates of evolution for the NTD and CTD are 3.2 and 2.5, respectively, whereas the average rate of evolution in the CD is 1.0. Within the CD, six subdomains have a mean rate of evolution below 1.0 (S4, S4–S5, S5, pore-helix, pore-loop, and S6). As expected, the pore-helix (residues 299–312) and pore-loop (residues 313–322) of the channel are the most conserved subdomains, with mean rates of evolution of only 0.38 and 0.41, respectively. This correlates with the critical role played by these components in achieving high ion selectivity for K^+ over Na^+ ions.[27] The S4 segment of the CD, which harbors basic residues for sensing and responding to changes in membrane potential,[28] has a mean rate of evolution of 0.61. The S4–S5 linker, which is believed to be responsible for transferring the conformational changes in the voltage sensor domain to the pore[29] and serve as binding sites for phosphatidylinositol-4,5-bisphosphate (PIP2) to modulate the deactivation rate of the channel,[30] has a mean rate of evolution of 0.92. The S2–S3 linker, proposed in a recent study to also bind PIP2,[31] is only moderately conserved. Interestingly, although most subdomains of the CD exhibit a low mean rate of evolution, two subdomains namely the S1–S2 linker and the S5-Pore linker, show substantially higher mean rates of evolution (2.5 and 1.9, respectively) than the rest of the CD.

As the CTD has been shown to have four helices designated A–D,[32] we computed the mean rate of evolution of each of these helices and their linkers to see if any of these subdomains are conserved. Our analysis shows that only helices A, B, and C have a mean rate of evolution < 1.0, whereas the mean rate of evolution of helix D is substantially higher (1.9). This observation agrees with the functional role of helices A and B in binding calmodulin (CaM) and the critical role of helix C in tetramerization of the intracellular C-terminal domain.[32, 33] The juxtramembrane subdomain S6-A, with a mean rate of evolution of 0.9, as well as the B-C linker, considered extremely conserved according to its mean rate of evolution (0.23), have yet to be shown to play any particular functional role.

Dysfunctional variants are enriched in selected subdomains

Results from a recent study suggested that the probability of pathogenicity of a KCNQ1 variant depends in part on the topological location of the variant.[17] However, in the previous study the protein was only divided into three topological domains namely NTD, CD, and CTD. We mapped all variants in our dataset onto the curve of position-specific rates of evolution (Figure 1A). We observed that dysfunctional variants preferentially occur at positions with low rate of evolution, especially within a selected set of subdomains.

In fact, 95.7% (90/94) dysfunctional variants occur at positions where the rate of evolution is under 2. In contrast, 61.5% (8/13) of normal variants occur at positions with rates of evolution above 2. The five normal variants that occur at positions with a rate of evolution under 2 are: Q147R, G179S, T391I, R533W, and G643S. Interestingly, Q147R, G179S, T391I, and G643S are chemically conserved, as judged by their Grantham distances:[34] Q→R (68), G→S (56), T→I (89). Nevertheless, this clear segregation of the functional impact of variants with respect to position-specific rate of evolution indicates that the rate of evolution of a sequence site pre-selected as one of the predictive features is indeed a strong predictor on whether variants occurring at the site will be dysfunctional or not.

In addition, we also computed the enrichment of dysfunctional variants for each subdomain, to confirm that such variants are indeed localized within a selected set of subdomains (Methods and Table S5 in Data Supplement). An enrichment of > 1.0 indicates that the corresponding subdomain has higher than random chance of harboring dysfunctional variants. As shown in Figure 2, subdomains with higher than random chance for dysfunctional variants are: S0, S2–S3 linker, S3, S4, S4–S5, S5, pore-helix, pore-loop, S6, S6-A, B–C, and C. In particular, S0, S3, S4–S5 linker, S5, pore-loop, and S6-A each have an enrichment ≥ 3. As discussed in the previous section, these subdomains are highly conserved.

Q1VarPred: A KCNQ1-specific predictor

A schematic representation of the architecture of Q1VarPred is shown in Figure 3A. Figure 3B shows a visualization of the Q1VarPred model of the relationship between predictive features (rate of evolution and PSSM-based perturbation) and the prediction about functional impact (impact score 0 − most likely normal, 1 − most likely dysfunctional). The contour surface indicates that the impact score has a sharper dependence on the rate of evolution than it does on PSSM-based perturbation. In particular, variants at conserved positions (rate of

evolution close to 0) are very likely to be dysfunctional (impact score > 0.5) even if the perturbation is very small. An example of such variants is the dysfunctional V307L whose impact was predicted to be 0.68. The estimated rate of evolution of this position is 0.52, whereas the perturbation introduced by substituting Val for Leu at this position is considerably small (3.7). Similarly, variants at evolutionarily tolerated positions (rate of evolution > 3.0 for example) tend to be normal even if the perturbation is very large (for example, R583H). However, the impact score does rise along with increasing magnitude of perturbation, which is particularly important for predicting the impact at positions exhibiting intermediate rates of evolution.

Comparing Q1VarPred with other methods

We employed a procedure called “repeated cross-validation ”[24] to estimate the generalization ability of Q1VarPred and other methods (see Estimating generalization ability). Seven commonly used genome-wide methods: PhD-SNP, Polyphen-2, PredictSNP, PROVEAN, SIFT, SNAP, and SNPs&GO and one potassium channel-specific method called KvSNP were examined (Methods and Table S6 in Data Supplement). Table 1 shows that all performance metrics rank Q1VarPred the best, except for NPV and TPR. In general, AUC and MCC are considered the most robust metrics for evaluating classifiers. AUC is independent of user-chosen and therefore possibly biased thresholds. MCC has the advantage to consider all four numbers (TP, TN, FP, FN) and provides a much more balanced evaluation than TPR or TNR individually.[35] In terms of AUC, Q1VarPred > PROVEAN > PhD-SNP > SNPs&GO > SIFT > KvSNP > PredictSNP > PolyPhen-2 > SNAP. This is similar to the findings of Leong et al.[36] except that PolyPhen-2 was shown to rank between PROVEAN and SNP&GO, and PhD-SNP and KvSNP were not evaluated in Leong et al. Methods that perform better than Q1VarPred in TPR, do so at a cost of a very low TNR, i.e. the threshold is chosen to minimize the loss of true positives at the cost of predicting many false positives. In some disease conditions, a high fraction of false positives might be acceptable. However, in LQTS and related channelopathies, the cost of false positives is as drastic as that of false negatives.[6] It is also worth noting that while KvSNP is gene-specific, our evaluation shows that its performance is worse than most genome-wide methods on this dataset. The primary cause of the inflation in KvSNP’s claimed performance is probably its convolution of dataset preparation and feature selection, where 85.5% of “neutral variants” were generated from variable sequence positions and later several sequence conservation-based features were selected as predictive features.[37]

Discussion

From functional impact to clinical disease diagnosis

The goal of our study was to create a highly tailored computational method to predict functional impact. However, translating evidence on functional impact to clinical disease diagnosis is far from trivial. First, every computational method has a certain degree of accuracy and reliability, and those of genome-wide methods are particularly limited. In fact, this is one of the primary motivations of the present study. Second, variants that are dysfunctional at the molecular level may not have clinical manifestation. For example, the A300T variant, which was confirmed experimentally to be severely dysfunctional,[25] was

later identified in a cohort considered to be clinically normal.[16] Such dysfunctional variants may have been rescued by compensating genetic variations. Third, trying to predict the clinical outcome without considering the mode of inheritance of LQTS may be problematic. The mode of inheritance is a key factor when determining the clinical relevance of a genotype for LQTS. For example, four variants in our dataset (R231H, W305S, A525T, and R594Q) were functionally normal when expressed in combination with the wild-type channel, but were severely dysfunctional in the absence of the wild-type. W305S was identified in members of two consanguineous families with the recessive JLN syndrome[38] and A525T was suspected to cause the recessive form of RW syndrome.[39] Moreover, a functionally normal variant may have compound genetic variations within the same gene or other genes that may obviate or, alternatively, contribute to the clinical phenotype.[40] In light of these considerations, Q1VarPred was intended for judicious use by researchers or clinicians in conjunction with complementary clinical and genetic evidence to assess the disease susceptibility caused by KCNQ1 variants.

Unexpected conserved subdomains in the C-terminal domain

Figure 4 shows the topological distributions of position-specific rate of evolution and subdomain-specific enrichment of dysfunctional variants. In our analysis of the rate of evolution in the CTD, we found a few topological subdomains with conserved mean rate of evolution (Figure 1B), predicting important functional or structural roles. Two subdomains, the S6-A linker and the B–C linker, were shown to have a surprisingly low mean rate of evolution (0.88 and 0.24, respectively). While S6-A has an estimated enrichment of dysfunctional variants of 3.0, that of the B–C linker is unexpectedly low (1.0) (Figure 2 and Table S5 in Data Supplement). The low enrichment of the B–C linker is likely biased because of the sparsity of functionally validated variants (e.g. only three functionally validated variants are located in the B–C linker). In fact, another six variants (Table S7 in Data Supplement) found in this subdomain have been deposited in ClinVar.[41] However, they were not included into our dataset as we were not able to find literature describing their functional validation. The enrichment of the B–C linker is likely to increase when larger datasets of functionally validated variants become available for estimating enrichments. More importantly, there seems to be a lack of study documenting the functional roles the S6-A linker and the B–C linker. Nevertheless, based on their low rate of evolution, we alert investigators about the potential high functional impact of variants found in these two subdomains.

The machine learning model

Ideally, a machine learning algorithm should produce a learned model that is accurate at predicting new observations and, at the same time, simple enough to allow straightforward interpretation. In general, linear models are easier to interpret, while nonlinear models are more powerful in cases where classes are not linearly separable. We chose a neural network, which generally is considered to be a nonlinear model, for the present study to leverage our extensive experience with neural networks and an established library for feature engineering and model building.[42–46] Admittedly, a logistic regression model performed only slightly worse (AUC = 0.855) and a linear discriminant classifier performed comparably (AUC = 0.870). However, given the complexity in the mechanisms behind KCNQ1 dysfunction, we

expect that the “true” decision boundary between normal and dysfunctional variants is complex. As additional experimental data become available, the advantage of neural networks for prediction over linear models is likely to become more substantial.

Factors contributing to the improved performance of Q1VarPred

Q1VarPred offers improved overall performance in predicting functional impact of variants on a KCNQ1-specific basis compared to the other evaluated tools (Table 1). Although most tools allow for predictions for a wide range of proteins, the fact that each method applies a single threshold to classify variants on all proteins may be partially responsible for their weaker overall performance on KCNQ1 variants. Additionally, recent work has shown that contemporary variant–phenotype and variant–stability prediction algorithms are substantially worse at predicting outcomes for membrane proteins, such as KCNQ1, than for water soluble proteins.[47]

The observed higher performance of Q1VarPred may also be attributed to better predictive features. Many methods use MSA-derived position-specific conservation scores as predictive feature, presumably based on the assumptions that the functional importance of a given position dictates how conserved this position is and, conversely, that the degree of conservation indicates the functional importance of this position. While this latter assumption is often valid, position-specific conservation scores computed directly from MSA without considering the evolutionary history of the aligned protein family may be biased because of unevenly sampled sequence space. Numerous position-specific quantitative conservation scores have been proposed over the years[48] and all evaluated methods except the meta-predictor PredictSNP use as position-specific conservation measures of some sort derived from MSA as predictive features. However, none of these methods consider the topology and branch lengths of phylogenetic trees as the method used in the current study does (Methods in Data Supplement). Thus, these conservation measures may lead to less accurate estimations of rate of evolution.

The other predictive feature used in Q1VarPred is the perturbation derived in the context of a PSSM. This feature measures how much less likely it is for the variant to occur at a sequence position relative to the wild-type. The higher the perturbation the less likely for the variant to replace the wild type residue at a specific position. While the position-specific rate of evolution presumably is a strong predictor of functional impact, it only indicates how likely it is that the wild-type amino acid at this sequence position changes. It does not, however, tell how likely it is that the wild-type amino acid is changed to one particular amino acid type over the others. In other words, the perturbation adds additional information by complementing position-specific rates of evolution with what the actual variants are.

Limitations and future direction

The primary limitation of the current study is the size of the dataset. Although a substantial amount of effort was spent by many labs to experimentally characterize the 107 variants treated in this study, the dataset used in this study is still very small, relative to that used to train other contemporary variant-effect predictors. As a result, we were limited from selecting a set of most relevant features in a systematic, algorithmic manner. Thus, it is very

likely that we missed some very informative sequence-based features. When larger datasets become available, Q1VarPred can be re-trained and new predictive features can be tested. In addition, our estimation of enrichment of dysfunctional variants for each subdomain is also likely to be biased due to this data sparsity. Even though the enrichment values correlate well with average rates of evolution and our analysis shows that functionally important subdomains tend to be more enriched with dysfunctional variants, there is currently not enough data available to demonstrate that such relationship for KCNQ1 is statistically significant.

Recent investigations into machine learning have shown that training neural networks on multiple traits/outcomes per training example can improve performance.[49, 50] Specifically, the advantages of simultaneously training a neural network to predict multiple outcome variables (disease severity, electrophysiological parameters, etc.) may enable a more accurate prediction of phenotype traits as well. Previous work aimed at predicting secondary structure and membrane burial for residues has suggested that neural networks trained to predict multiple outcomes are particularly beneficial when the dataset size is especially small.[43] This suggests that such neural networks may be particularly suitable to leverage the diverse experimental parameters available for LQTS variants and phenotypes.

The method developed in this study is modular in the sense that one possible future direction is to combine this method with other predictors—such as estimation of the impact of genetic variations on protein stability, to come up with predictions that are both more reliable and that also suggest mechanisms underlying variation-induced gain or loss of function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding: This project was supported by National Institutes of Health (NIH) Grant R01 HL122010. B.L. was supported by American Heart Association Pre-doctoral Fellowship Award 16PRE27260211. B.M.K. was supported by NIH Grant F32 GM113355. K.C.T. was supported by F32 GM11777 and T32 NS00749.

References

1. Schwartz PJ, Stramba-Badiale M, Crotti L, Pedrazzini M, Besana A, Bosi G, et al. Prevalence of the congenital long-QT syndrome. Circulation. 2009; 120:1761–7. [PubMed: 19841298]
2. Goldenberg I, Moss AJ. Long QT syndrome. J Am Coll Cardiol. 2008; 51:2291–300. [PubMed: 18549912]
3. Barhanin J, Lesage F, Guillemare E, Fink M, Lazdunski M, Romey G. K(V)LQT1 and lsK (minK) proteins associate to form the I(Ks) cardiac potassium current. Nature. 1996; 384:78–80. [PubMed: 8900282]
4. Schwartz PJ, Ackerman MJ, George AL Jr, Wilde AA. Impact of genetics on the clinical management of channelopathies. J Am Coll Cardiol. 2013; 62:169–80. [PubMed: 23684683]
5. Giudicessi JR, Ackerman MJ. Genetic testing in heritable cardiac arrhythmia syndromes: differentiating pathogenic mutations from background genetic noise. Curr Opin Cardiol. 2013; 28:63–71. [PubMed: 23128497]
6. Ackerman MJ. Genetic purgatory and the cardiac channelopathies: Exposing the variants of uncertain/unknown significance issue. Heart Rhythm. 2015; 12:2325–31. [PubMed: 26144349]

7. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum Mol Genet. 2010; 19:R125–30. [PubMed: 20805107]
8. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011; 12:628–40. [PubMed: 21850043]
9. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet. 2013; 14:415–26. [PubMed: 23681062]
10. Bhuiyan ZA. Silent mutation in long QT syndrome: Pathogenicity prediction by computer simulation. Heart Rhythm. 2012; 9:283–284. [PubMed: 22001705]
11. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet. 2006; 7:61–80. [PubMed: 16824020]
12. 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. Genet Med. 2015; 17:405–24. [PubMed: 25741868]
13. Ohanian M, Otway R, Fatkin D. Heuristic Methods for Finding Pathogenic Variants in Gene Coding Sequences. J Am Heart Assoc. 2012; 1
14. Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: Proceed with caution in clinical genetics. Clin Chem. 2004; 50:1974–1978. [PubMed: 15502081]
15. Care MA, Needham CJ, Bulpitt AJ, Westhead DR. Deleterious SNP prediction: be mindful of your training data! Bioinformatics. 2007; 23:664–72. [PubMed: 17234639]
16. Kapa S, Tester DJ, Salisbury BA, Harris-Kerr C, Pungliya MS, Alders M, et al. Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. Circulation. 2009; 120:1752–60. [PubMed: 19841300]
17. Giudicessi JR, Kapplinger JD, Tester DJ, Alders M, Salisbury BA, Wilde AA, et al. Phylogenetic and physicochemical analyses enhance the classification of rare nonsynonymous single nucleotide variants in type 1 and 2 long-QT syndrome. Circ Cardiovasc Genet. 2012; 5:519–28. [PubMed: 22949429]
18. Modell SM, Lehmann MH. The long QT syndrome family of cardiac ion channelopathies: a HuGE review. Genet Med. 2006; 8:143–55. [PubMed: 16540748]
19. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research. 2014; 15:1929–1958.
20. Rennell D, Bouvier SE, Hardy LW, Poteete AR. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol. 1991; 222:67–88. [PubMed: 1942069]
21. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et biophysica acta. 1975; 405:442–51. [PubMed: 1180967]
22. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005; 21:3940–1. [PubMed: 16096348]
23. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
24. Smith GC, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. Am J Epidemiol. 2014; 180:318–24. [PubMed: 24966219]
25. Bianchi L, Priori SG, Napolitano C, Surewicz KA, Dennis AT, Memmi M, et al. Mechanisms of I(Ks) suppression in LQT1 mutants. Am J Physiol Heart Circ Physiol. 2000; 279:H3003–11. [PubMed: 11087258]
26. Cordeiro JM, Perez GJ, Schmitt N, Pfeiffer R, Nesterenko VV, Burashnikov E, et al. Overlapping LQT1 and LQT2 phenotype in a patient with long QT syndrome associated with loss-of-function variations in KCNQ1 and KCNH2. Can J Physiol Pharmacol. 2010; 88:1181–90. [PubMed: 21164565]
27. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, et al. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science. 1998; 280:69–77. [PubMed: 9525859]
28. Choe S. Potassium channel structures. Nat Rev Neurosci. 2002; 3:115–21. [PubMed: 11836519]

29. Labro AJ, Boulet IR, Choveau FS, Mayeur E, Bruyns T, Loussouarn G, et al. The S4–S5 linker of KCNQ1 channels forms a structural scaffold with the S6 segment controlling gate closure. J Biol Chem. 2011; 286:717–25. [PubMed: 21059661]
30. Taylor KC, Sanders CR. Regulation of KCNQ/Kv7 family voltage-gated K+ channels by lipids. Biochimica et biophysica acta. 2016
31. Chen L, Zhang Q, Qiu Y, Li Z, Chen Z, Jiang H, et al. Migration of PIP2 lipids on voltage-gated potassium channel surface influences channel deactivation. Sci Rep. 2015; 5:15079. [PubMed: 26469389]
32. Wiener R, Haitin Y, Shamgar L, Fernández-Alonso MC, Martos A, Chomsky-Hecht O, et al. The KCNQ1 (Kv7.1) COOH terminus, a multitiered scaffold for subunit assembly and protein interaction. J Biol Chem. 2008; 283:5815–30. [PubMed: 18165683]
33. Sachyani D, Dvir M, Strulovich R, Tria G, Tobelaim W, Peretz A, et al. Structural basis of a Kv7.1 potassium channel gating module: studies of the intracellular c-terminal domain in complex with calmodulin. Structure. 2014; 22:1582–94. [PubMed: 25441029]
34. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974; 185:862–4. [PubMed: 4843792]
35. Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics. 2000; 16:412–424. [PubMed: 10871264]
36. Leong IUS, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. Bmc Med Genet. 2015; 16
37. Stead LF, Wood IC, Westhead DR. KvSNP: accurately predicting the effect of genetic variants in voltage-gated potassium channels. Bioinformatics. 2011; 27:2181–2186. [PubMed: 21685056]
38. Neyroud N, Denjoy I, Donger C, Gary F, Villain E, Leenhardt A, et al. Heterozygous mutation in the pore of potassium channel gene KvLQT1 causes an apparently normal phenotype in long QT syndrome. Eur J Hum Genet. 1998; 6:129–33. [PubMed: 9781056]
39. Larsen LA, Fosdal I, Andersen PS, Kanters JK, Vuust J, Wettrell G, et al. Recessive Romano-Ward syndrome associated with compound heterozygosity for two mutations in the KVLQT1 gene. Eur J Hum Genet. 1999; 7:724–8. [PubMed: 10482963]
40. Westenskow P, Splawski I, Timothy KW, Keating MT, Sanguinetti MC. Compound mutations: a common cause of severe long-QT syndrome. Circulation. 2004; 109:1834–41. [PubMed: 15051636]
41. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42:D980–5. [PubMed: 24234437]
42. Butkiewicz M, Lowe EW Jr, Mueller R, Mendenhall JL, Teixeira PL, Weaver CD, et al. Benchmarking ligand-based virtual High-Throughput Screening with the PubChem database. Molecules. 2013; 18:735–56. [PubMed: 23299552]
43. Leman JK, Mueller R, Karakas M, Woetzel N, Meiler J. Simultaneous prediction of protein secondary structure and transmembrane spans. Proteins. 2013; 81:1127–40. [PubMed: 23349002]
44. Mendenhall J, Meiler J. Improving quantitative structure-activity relationship models using Artificial Neural Networks trained with dropout. J Comput Aided Mol Des. 2016; 30:177–89. [PubMed: 26830599]
45. Li B, Mendenhall J, Nguyen ED, Weiner BE, Fischer AW, Meiler J. Accurate Prediction of Contact Numbers for Multi-Spanning Helical Membrane Proteins. J Chem Inf Model. 2016; 56:423–34. [PubMed: 26804342]
46. Li B, Mendenhall J, Nguyen ED, Weiner BE, Fischer AW, Meiler J. Improving Prediction of Helix–Helix Packing in Membrane Proteins Using Predicted Contact Numbers as Restraints. Proteins. 2017; 85:1212–21. [PubMed: 28263405]
47. Kroncke BM, Duran AM, Mendenhall JL, Meiler J, Blume JD, Sanders CR. Documentation of an Imperative To Improve Methods for Predicting Membrane Protein Stability. Biochemistry. 2016; 55:5002–9. [PubMed: 27564391]
48. Valdar WS. Scoring residue conservation. Proteins. 2002; 48:227–41. [PubMed: 12112692]

49. Qi Y, Oja M, Weston J, Noble WS. A unified multitask architecture for predicting local protein properties. PLoS One. 2012; 7:e32235. [PubMed: 22461885]

50. Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, et al. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci Rep. 2015; 5:11476. [PubMed: 26098304]

Clinical Perspective

Congenital LQTS is a heart rhythm disorder that affects ~ 1 in 2,500 births. It predisposes children and young adults to a type of ventricular tachycardia (torsades de pointes) and sudden cardiac death. An emerging standard-of-care for LQTS employs clinical genetic testing to identify LQTS-associated variants in the KCNQ1 potassium channel. However, variants of unknown significance for which there is inadequate evidence to classify as being pathogenic are common findings. While computational methods, mostly developed for genome-scale predictions, have been a common predictive approach to suggest genotype-phenotype relations for variants of unknown significance, hardly any is sufficiently accurate for clinical use related to channelopathies. This study presents a KCNQ1-specific genotype–channel function relationship predictor Q1VarPred, which was trained on a dataset of KCNQ1 variants whose functional impact have been experimentally validated. Q1VarPred offers substantially improved overall performance in predicting functional impact of variants on a KCNQ1-specific basis compared to the other eight methods evaluated in the study. It is publicly available as a web server at http://meilerlab.org/q1varpred to ease its access by researchers and clinicians. Along with developing this method, a detailed analysis on the conservation of the amino acid sequence of KCNQ1 showed that dysfunctional variants are enriched in a selected set of highly conserved subdomains. This finding together with the functional impact predicted by Q1VarPred may be considered as supplementary information to the interpretation of variants of known significance.

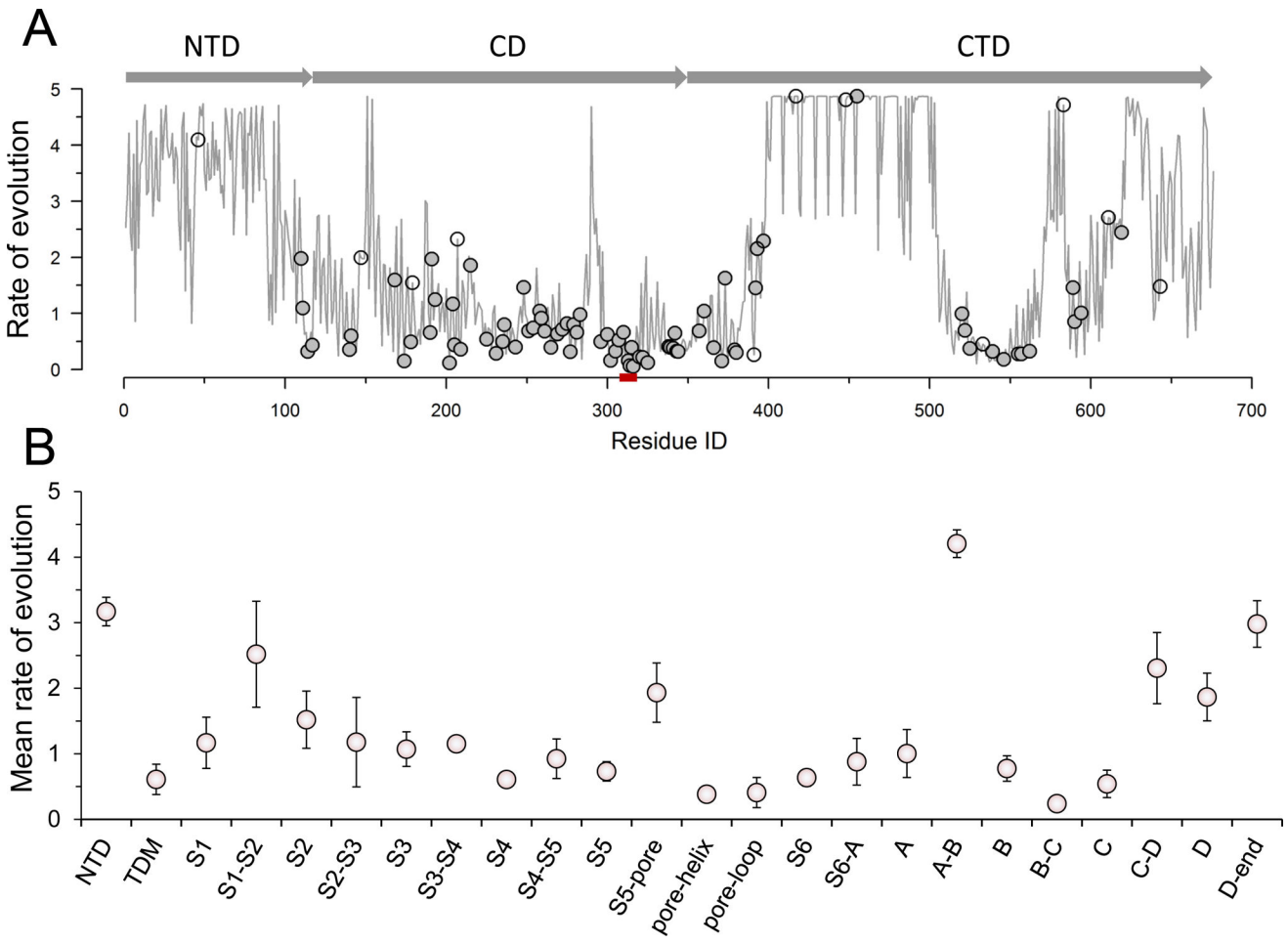

Figure 1.

Analysis on the evolutionary variability of the KCNQ1 sequence. A) Position-specific rate of evolution. Shaded arrow bars on the top indicate the sequence range of NTD, CD, and CTD respectively. The small red bar on the horizontal axis highlights the "selectivity filter" TIGYG. Closed circles represent dysfunctional variants and open circles represent normal variants. B) mean rates of evolution for structurally distinct subdomains of NTD, CD, and CTD. Note that the trafficking determinant motif (TDM), which resides within the NTD, is singled out for its distinct functional role. Error bars indicate the 95% confidence intervals (under Student-t distribution) for the mean rate of evolution.

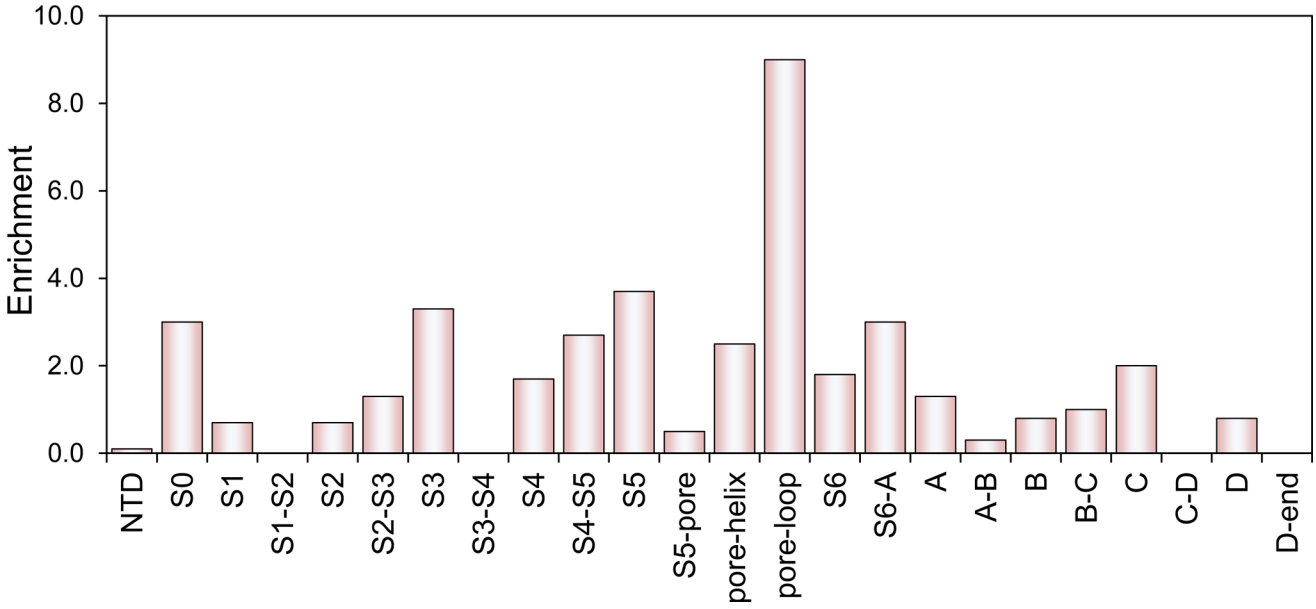

Figure 2.

This bar graph plots subdomain-specific enrichment of dysfunctional variants, showing that dysfunctional variants are enriched in a selected set of subdomains (S0, S3, S4–S5, S5, pore-helix, pore-loop, S6-A, see Table S5 for the residue ranges these subdomains correspond to). One needs to keep in mind that due to the sparsity of functionally characterized variants, the estimates of enrichments are likely to be biased.

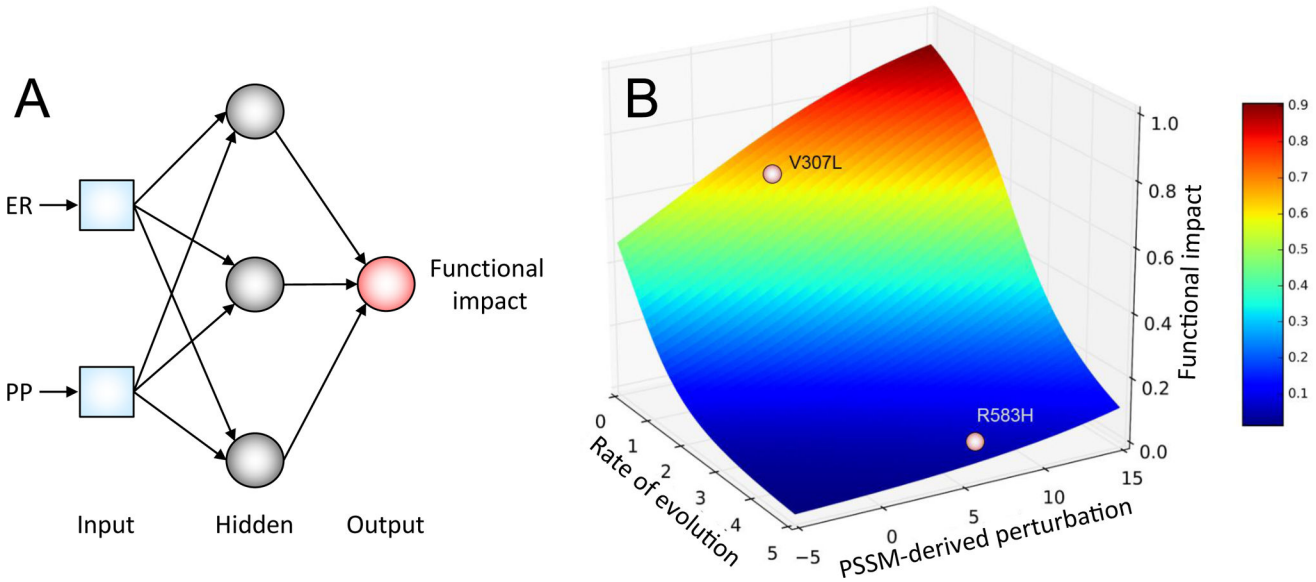

Figure 3.

A) A schematic representation of the architecture of Q1VarPred. The input layer is composed of two predictive features: rate of evolution (ER) and perturbation derived from PSSM (PP). The hidden layer has three neurons and the output layer has one neuron that computes the final predicted functional impact. B) A visualization of the Q1VarPred-mapped mathematical relationship between predictive features (rate of evolution and perturbation) and functional impact. The vertical axis is functional impact on the scale of 0 to 1 with 1 being complete dysfunction.

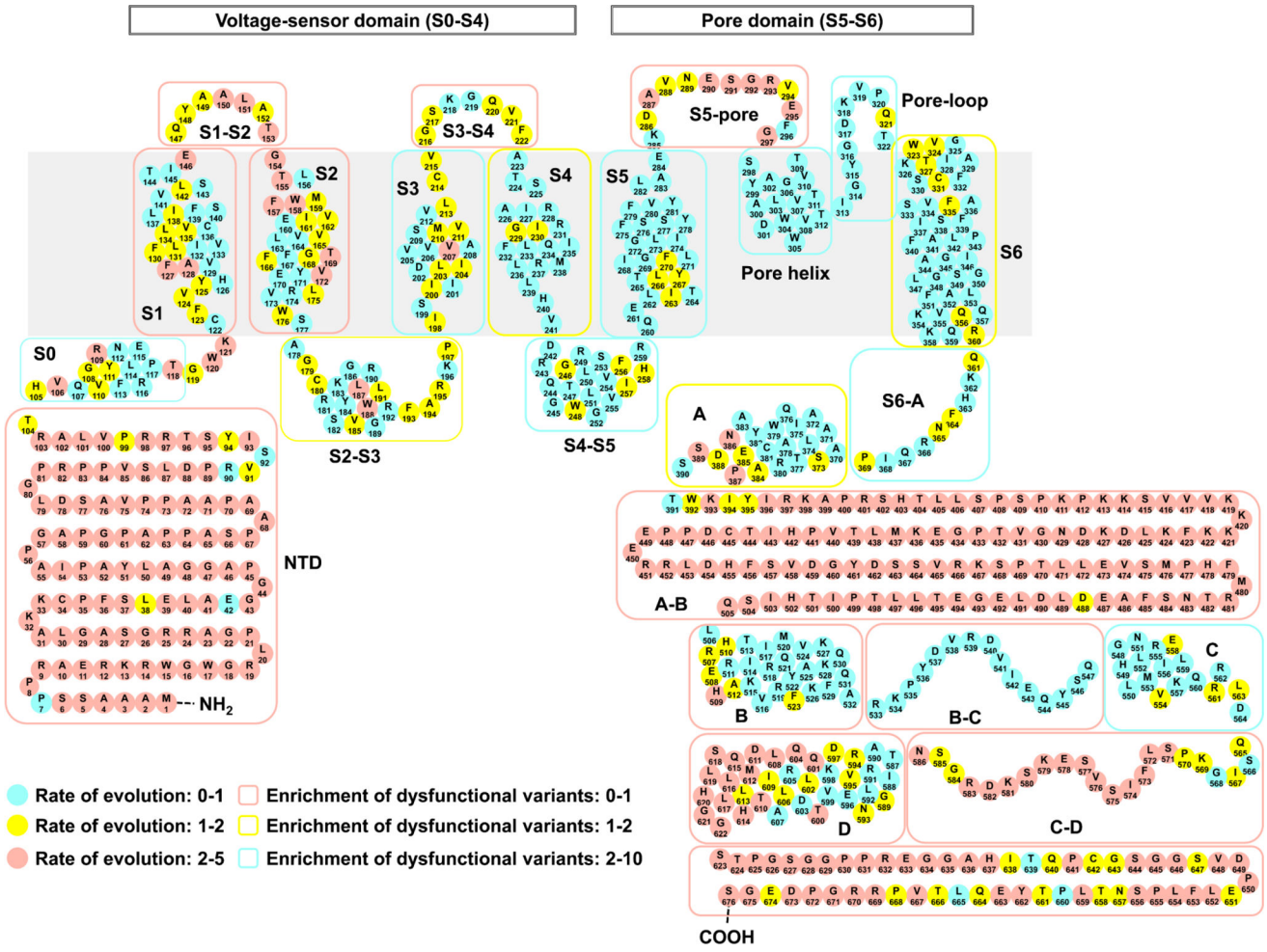

Figure 4.

A "global" view of the topological distribution of rate of evolution and enrichment of dysfunctional variants.

Table 1

Comparison of Q1VarPred with other methods.

Method	Mean performance metric							
	AUC	MCC	PPV	NPV	Accuracy	TPR+TNR	TPR	TNR
Q1VarPred	0.884	0.581	0.968	0.537	0.881	1.680	0.895	0.785
KvSNP	0.662	0.313	0.922	0.344	0.832	1.255	0.887	0.438
PhD-SNP	0.727	0.386	0.941	0.390	0.820	1.453	0.850	0.603
PolyPhen-2	0.636	0.340	0.912	0.547	0.866	1.272	0.939	0.333
PredictSNP	0.652	0.355	0.918	0.459	0.850	1.303	0.912	0.391
PROVEAN	0.770	0.510	0.949	0.537	0.869	1.536	0.902	0.634
SIFT	0.680	0.360	0.927	0.503	0.861	1.364	0.921	0.443
SNAP	0.542	0.101	0.895	0.158	0.771	1.085	0.844	0.241
SNPs&GO	0.697	0.307	0.939	0.296	0.767	1.384	0.792	0.592