

## Protein structure aids predicting functional perturbation of missense variants in *SCN5A* and *KCNQ1*

Brett M. Kroncke<sup>a,\*</sup>, Jeffrey Mendenhall<sup>b,c</sup>, Derek K. Smith<sup>d</sup>, Charles R. Sanders<sup>c,e</sup>, John A. Capra<sup>f,g</sup>, Alfred L. George<sup>h</sup>, Jeffrey D. Blume<sup>d</sup>, Jens Meiler<sup>b,c,i</sup>, Dan M. Roden<sup>a,g,i</sup>

<sup>a</sup> Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA

<sup>b</sup> Department of Chemistry, Vanderbilt University, Nashville, TN 37232, USA

<sup>c</sup> Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA

<sup>d</sup> Department of Biostatistics, Vanderbilt University, Nashville, TN 37240, USA

<sup>e</sup> Department of Biochemistry, Vanderbilt University, Nashville, TN, 37232, USA

<sup>f</sup> Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

<sup>g</sup> Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37235, USA

<sup>h</sup> Department of Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

<sup>i</sup> Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA

### ARTICLE INFO

#### Article history:

Received 17 November 2018

Received in revised form 21 January 2019

Accepted 23 January 2019

Available online 1 February 2019

#### Keywords:

*SCN5A*

*KCNQ1*

Function prediction

Protein structure

And protein function

### ABSTRACT

Rare variants in the cardiac potassium channel *K<sub>v</sub>7.1* (*KCNQ1*) and sodium channel *Na<sub>v</sub>1.5* (*SCN5A*) are implicated in genetic disorders of heart rhythm, including congenital long QT and Brugada syndromes (LQTS, BrS), but also occur in reference populations. We previously reported two sets of *Na<sub>v</sub>1.5* ( $n = 356$ ) and *K<sub>v</sub>7.1* ( $n = 144$ ) variants with in vitro characterized channel currents gathered from the literature. Here we investigated the ability to predict commonly reported *Na<sub>v</sub>1.5* and *K<sub>v</sub>7.1* variant functional perturbations by leveraging diverse features including variant classifiers PROVEAN, PolyPhen-2, and SIFT; evolutionary rate and BLAST position specific scoring matrices (PSSM); and structure-based features including “functional densities” which is a measure of the density of pathogenic variants near the residue of interest. Structure-based functional densities were the most significant features for predicting *Na<sub>v</sub>1.5* peak current (adj.  $R^2 = 0.27$ ) and *K<sub>v</sub>7.1* + KCNE1 half-maximal voltage of activation (adj.  $R^2 = 0.29$ ). Additionally, use of structure-based functional density values improves loss-of-function classification of *SCN5A* variants with an ROC-AUC of 0.78 compared with other predictive classifiers (AUC = 0.69; two-sided DeLong test  $p = .01$ ). These results suggest structural data can inform predictions of the effect of uncharacterized *SCN5A* and *KCNQ1* variants to provide a deeper understanding of their burden on carriers.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1. Introduction

Of an estimated 20,000 nonsynonymous single nucleotide polymorphisms (nsSNPs) in each individual's protein-coding genome, approximately 10 are presently predicted to be clinically actionable [26]. nsSNPs in *KCNQ1* (*K<sub>v</sub>7.1* channel protein, which complexes with the protein KCNE1 to generate the slow cardiac potassium repolarization current,  $I_{Ks}$ ) and *SCN5A* (*Na<sub>v</sub>1.5* channel protein, which generates the cardiac depolarizing sodium current,  $I_{Na}$ ), are associated with heritable diseases of the heart [4,36,37,38,49] including dilated cardiomyopathy [14,28], cardiac conduction disease [6,29], short QT syndrome [13], sick sinus syndrome [15], types 1 and 3 congenital long QT syndromes

(LQTS) [7,18,30,38], and Brugada syndrome (BrS) [5]. However, in aggregate, rare nsSNPs in *SCN5A* and *KCNQ1* also appear at ~2% in the population, being more common than the rare arrhythmia disorders associated with these genes, suggesting only limited roles in disease. Determining the significance and effect size of these nsSNPs will be of increasing importance as more people undergo genome or exome sequencing [3,27].

Models used to predict the effect of these nsSNPs are most commonly trained on the information-poor inputs of binary disease-inducing/benign classification. Binary classification reduces information. Moreover, the disease-inducing vs. benign distinction ignores penetrance and the underlying molecular phenotype—or potentially multiple overlapping molecular phenotypes—that may be most informative for therapy. A striking example involves patients presenting with type 3 long QT syndrome due to a gain-of-function *SCN5A* variant that also impairs trafficking of the encoded channel *Na<sub>v</sub>1.5*. Therapeutic

\* Corresponding author at: Vanderbilt University Medical Center, 2215B Garland Ave, 1285 MRBIV, Nashville, TN 37232, USA.

E-mail address: [brett.m.kroncke.1@vumc.org](mailto:brett.m.kroncke.1@vumc.org) (B.M. Kroncke).

targeting of this gain-of-function with the antiarrhythmic drug mexiletine can increase cell surface expression of the mutant channel, leading to the unintended consequence of exaggerating the long QT phenotype [34,35,46].

Using literature datasets we have recently curated for both  $I_{Ks}$  [25,47] and  $I_{Na}$ , [22] we test the hypothesis that incorporating variant-specific functional features from *KCNQ1* and *SCN5A* nsSNPs and structure-based features into prediction models will improve our ability to predict if previously uncharacterized nsSNPs will result in altered currents. Secondary structural elements are independent predictors of deleterious variants in *SCN5A* and can improve current prediction models [20], suggesting the potential utility of structure-based approaches. In fact, the highest densities of disease-associated variants across the entire spectrum of proteins fall largely in structured, functional segments: the structure/function of these molecules are compromised in the disease state [23,50]. Here, we generated a set of models able to predict  $I_{Na}$  and  $I_{Ks}$  variant-specific current phenotypes. Identifying the variant-specific functional perturbation will provide an additional tool to geneticists and physicians to determine if variants are likely disease-causing and to more accurately stratify the degree of risk that carriers who present without a phenotype will eventually develop channelopathy-based heart disease.

## 2. Methods

### 2.1. Quantified functional parameters of *KCNQ1* and *SCN5A* chosen for analysis

For  $I_{Na}$ , we analyzed peak current, steady state  $V_{1/2}$  activation and inactivation, late/persistent current, and recovery from inactivation [49]. For  $I_{Ks}$ , we analyzed peak current,  $V_{1/2}$  activation, and activation and deactivation time constants [19]. We selected these functional features because these parameters are most consistently reported in the literature. We only included functional data from  $K_v7.1$  variants when functional protocols involved homotetrameric mutated  $K_v7.1$  coexpressed with KCNE1, since this protocol was most commonly reported in the literature. Details about how each dataset was collected is contained in the original papers. ([22,25]; C. G. [48]) Briefly, all variants were normalized to WT measurements included in the same publication, i.e. peak current mutant/peak current WT, or  $V_{1/2}$  activation (mutant) –  $V_{1/2}$  activation (WT), etc.

Most functionally characterized variants in *SCN5A* were characterized by heterologous expression in human embryonic kidney cells (291 of 356 total), so we used only patch-clamp data derived in human embryonic kidney cells when available. For *KCNQ1*-KCNE1, most variants were characterized in CHO cells (79 of 165 total). We averaged the individual parameters in cases where multiple articles reported functional characterization of the same variant in the same cell system.

### 2.2. Generating structural models of $K_v7.1$ (*KCNQ1*)

No experimental structure of transmembrane domains of human  $K_v7.1$  exists, so we generated models using the recently released *Xenopus* structure of a closed pore and open voltage sensor and the human sequence NP\_000209.2 with 91% identity [45]. We used comparative modeling within the Rosetta scripts utility in Rosetta 3.8 to build  $K_v7.1$  [44]. We rebuilt loops on  $K_v7.1$  monomers, followed by rebuilding the functional homotetramer with symmetry for 1000 models. Most best-scoring structures had reasonable  $C_\alpha$  RMSDs between 1 and 3. We selected the best scoring model for subsequent analysis. We built models both with, and without, human calmodulin (CaM) bound; however no significant differences were observed in structure-based features, therefore, we selected  $K_v7.1$  with CaM bound for the analysis presented here.

### 2.3. Generating structural models of $Na_v1.5$

We generated two human  $Na_v1.5$  structural models using the human sequence NP\_000326.2 with the American cockroach sodium channel  $Na_vPaS$  structure [41] (45% identity), and electric eel  $Na_v1.4$  structure [52] (67% identity). Models of  $Na_v1.5$  were refined with small, unstructured segments rebuilt using established protocols as for  $K_v7.1$ , generating 1000 models. Most best-scoring structures had reasonable  $C_\alpha$  RMSDs between 2 and 4. We selected the best scoring model for subsequent analysis. We tested the performance of structure-based features using both models, with very similar results. Because models based on the  $Na_vPaS$  structure allow the inclusion of more variants in the analysis, we report here features calculated using those structural models.

### 2.4. Summary of predictive features

Our objective was to predict variant-specific functional perturbations for the cardiac ion channels  $K_v7.1$  + KCNE1 ( $I_{Ks}$ ) and  $Na_v1.5$  ( $I_{Na}$ ). We used the variant classifier models PROVEAN [9], PolyPhen-2 [1], and SIFT [24]; sequence alignment-based rate of evolution [32], and mutation rates derived from BLAST position specific scoring matrices (PSSM), and Point Accepted Mutation (PAM) matrix score [39]; and several structure-based features including burial propensities (how often certain residues are in the interior of the protein), neighbor counts (number of neighboring residues), neighbor identities (propensity of neighboring residues to be close in space) and what we term functional density (k-nearest neighbors-inspired metric to estimate functional perturbation). These predictive features are described below and summarized in Table S1. As can be seen in the higher off-diagonal  $R^2$ s, predictive classifiers were modestly degenerate; functional density weight only, i.e. the local enrichment for variants that had been functionally characterized, were more degenerate (described below, Figs. S1 and S2).

### 2.5. Calculating structure-derived features

NeighborCount is derived from the number of nearest neighbors weighted by distance and within 11.4 Å of the residue of interest, a cutoff found to be optimized to predict protein structure [12]. NeighborVector is a variation of neighbor density, scaled by how evenly distributed the nearest neighbor residues are to the residue of interest. Amino acid neighbor count (aaneigh) and amino acid neighbor vector (aaneighvector) are analogous to NeighborCount and NeighborVector, respectively, modified to account for amino acid-specific propensities for a given degree of burial [12,51]. NeighborCount, NeighborVector, aaneigh, aaneighvector predictive features were generated using the BioChemical Library (BCL) and the structures described above (for more detail see [12,51]).

### 2.6. Generating a structure-based functional density predictor

In addition to the structure-based features described above, we leveraged both the structural models and variant-specific functional datasets for  $I_{Ks}$  and  $I_{Na}$  in estimating the “functional density”. Using an approach akin to k-nearest neighbors, we calculated functional density by averaging functional perturbations of variants near the variant of interest weighted by the inverse of their distance from the variant of interest. This calculated feature therefore depends on how many functionally perturbed variants are near the variant of interest, with regions in three-dimensional space dense with functionally perturbed variants—“hotspots”—yielding a more perturbed prediction. However, all functionally characterized variants contribute to this parameter. We did not use a cutoff to determine whether or not to include a variant

in this analysis. Functional density is calculated as follows:

$$\rho_{j,x} = \frac{\sum_{i=0}^n \Delta \text{function}_{x,i,\text{mutation}(i)} \cdot \frac{1}{1 + e^{\left(\frac{d_{i,j}}{r}\right)}}}{\sum_{i=0}^n \frac{1}{1 + e^{\left(\frac{d_{i,j}}{r}\right)}}$$

where  $\rho_j$  is functional density of the  $j^{\text{th}}$  residue and  $x^{\text{th}}$  functional parameter,  $\Delta \text{function}_{x,i}$  is the change in functional parameter  $x$  for the  $i^{\text{th}}$  variant, and  $d_{i,j}$  is the distance between the center of mass of residues  $i$  and  $j$ .  $i$  does include residue  $j$ , but only if the identity of the amino-acid mutation is changed, i.e.  $\text{mutation}(i) \neq \text{mutation}(j)$ . A graphical representation is shown in Fig. S3. The distribution of neighboring residues is similar between  $K_v7.1$  and  $Na_v1.5$ , with a first shell of contacting residues at  $\sim 6 \text{ \AA}$  and a second shell at  $\sim 11 \text{ \AA}$  (Fig. S4). Additionally, we calculated the functional density weights alone (same equation as above, but with  $\Delta \text{function} = 1$ ) to test whether signal derived from functional densities could be attributed to protein region bias in the variants that have been functionally characterized.

### 2.7. Variant-specific $I_{Na}$ and $I_{Ks}$ functional perturbation predictive models

Because the number of features in our dataset was large relative to the number of variants, regularization was used to fit predictive models. We used a fully relaxed LASSO penalty, which has good predictive performance overall [16]. Prediction models were 10-fold cross-validated. After feature selection, the relaxed generalized linear model was bootstrapped (1000 times) to obtain bootstrapped percentile intervals for quantities of interest. We report the adjusted coefficient of determination,  $\text{adj. } R^2$ , with 95% confidence intervals as a measure of overall

prediction of the relaxed LASSO model. We focused on models where LASSO shrinkage yielded at least one significant predictive feature and the lower bound of the naive 95% confidence interval for the  $\text{adj. } R^2$  was  $>0.10$ . Relatively few models were able to meet these minimum criteria. Note that since the functional density features were calculated from the data, we additionally subjected the fully relaxed LASSO to higher-level 10-fold cross validation procedure which included a functional density construction step. This accounts for any variability or overfitting that might result from using data-determined functional covariates.

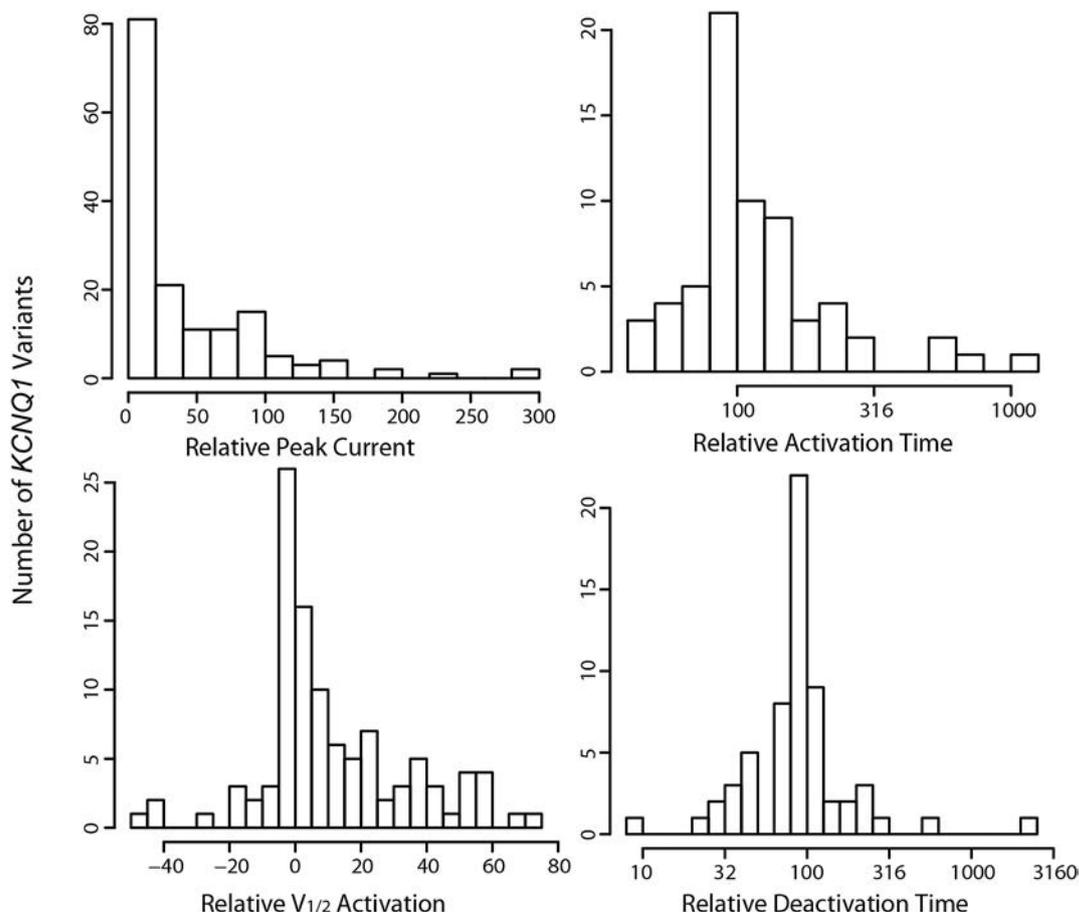
### 2.8. Loss-of-function classification of $I_{Na}$ and $I_{Ks}$ with and without structure-based features

We further classified loss-of-function variants by degree of functional perturbation, for  $I_{Na}$  defined as  $<50\%$  peak current [22] and for  $I_{Ks}$   $<50\%$  peak current or  $>10 \text{ mV}$  positive shift in  $V_{1/2}$  activation [25], to estimate the impact of functional densities on this task. We used commonly available variant sequence-based classifiers PolyPhen2, PROVEAN, BLAST-PSSM, and rate of evolution individually, all combined, and all combined with peak current functional density in a logistic regression model. We generated 95% confidence intervals on AUCs from the candidate models using bootstrap with 2000 replicates and used a two-sided DeLong test to evaluate ROC difference significance.

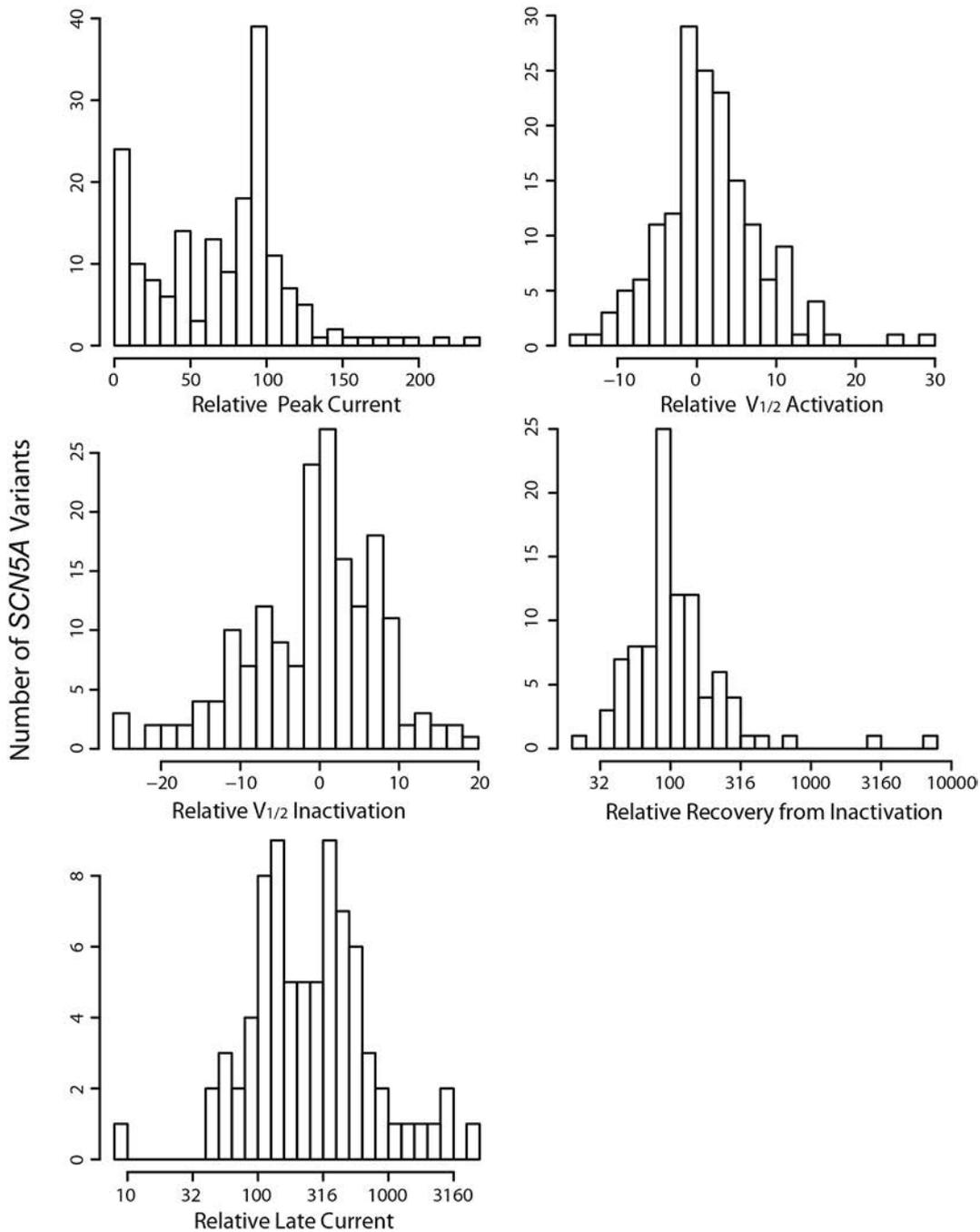
## 3. Results

### 3.1. Ion channel missense variants have diverse effects on current

Histograms of all functional parameters analyzed are shown in Figs. 1 and 2 and Table 1. For homotetrameric  $K_v7.1$  variants, the



**Fig. 1.** Histogram distributions of all functional parameters for  $K_v7.1 + KCNE1 (I_{Ks})$  analyzed in this paper. All values are referenced to WT which is either 100% or 0 mV.



**Fig. 2.** Histogram distributions of all functional parameters for Nav1.5 ( $I_{Na}$ ) analyzed in this paper. All values are referenced to WT which is either 100% or 0 mV.

**Table 1**  
Summary statistics of functional parameters.

Nav1.5	# of variants	Median [1st Q, 3rd Q]	WT
Peak Current	162	82 [36, 100] (%WT)	100%
Late Current	61	253 [122, 474] (%WT)	100%
$V_{1/2}$ Activation	163	0.00 [−1.63, 3.09] (mV)	0 mV
$V_{1/2}$ Inactivation	141	0.00 [−4.00, 3.44] (mV)	0 mV
Inactivation Recovery	85	98 [76, 138] (%WT)	100%
Kv7.1	# of variants	Median [1st Q, 3rd Q]	WT
$I_{Ks}$ peak	142	17 [0, 59] (%WT)	100%
$V_{1/2}$ Act	93	6.40 [0.00, 23.80] (mV)	0 mV
tau_act	58	106 [94, 150] (%WT)	100%
tau_deact	57	87 [70, 115] (%WT)	100%

distribution of  $I_{Ks}$  current maxima is skewed towards 0% current compared to WT function, likely a reflection of literature bias. The distribution of  $I_{Na}$  variant current maxima is bimodal with centers at 0% (complete LOF) and 100% (WT).  $I_{Ks}$   $V_{1/2}$  activation is also skewed towards more positive values, whereas  $I_{Na}$   $V_{1/2}$  activation is more evenly distributed about 0 mV.  $I_{Na}$  late current is skewed towards higher values. Time constants for  $I_{Ks}$  activation and inactivation and  $I_{Na}$  recovery from inactivation are clustered around WT with very wide ranges, populated with few points at extremely long characteristic times.

**3.2. Models can significantly predict  $I_{Na}$  and  $I_{Ks}$  peak current but rely on different predictive features**

Using a linear model, we could predict peak current, a proxy for overall channel function, for both  $I_{Ks}$  and  $I_{Na}$  (lower bound 95% CI adj.  $R^2$  of

0.14 and 0.18 respectively; Table 2 and Fig. 3). Interestingly, sequence-based predictors, especially BLAST-PSSM, had the most significant association with  $I_{Ks}$  peak current (Table S2, Fig. S5) but were not as integral to predicting  $I_{Na}$  peak current (Table S3, Fig. S6). Conversely, functional density for peak current provided most of the signal for  $I_{Na}$  but did not contribute meaningfully to  $I_{Ks}$  peak current prediction. This suggests a spatial dependence of peak current for  $I_{Na}$  not recapitulated by other published predictive models, contrary to  $I_{Ks}$ . This difference may be due in part to the comparatively large fraction of reported *SCN5A* variants that do not perturb peak current yet are still associated with cardiac diseases compared to *KCNQ1*, such as LQT3 variants with increased late current but no change in peak current; BLAST-PSSM is sensitive to evolutionary fitness of residue changes which may be more homogeneously dependent on peak current for *KCNQ1* and more heterogeneous for *SCN5A*. Alternatively, the spatial distribution of  $I_{Ks}$  peak current may be more heterogeneous than for  $I_{Na}$ . The functional density weight, a measure of the number of functionally characterized variants proximal to a residue of interest, was selected out of the  $I_{Ks}$  peak current model, but not for  $I_{Na}$  suggesting a modest sampling bias in regions of  $Na_v1.5$  sensitive to peak current perturbation.

### 3.3. Models can predict steady-state $I_{Ks}$ $V_{1/2}$ activation but not $I_{Na}$ $V_{1/2}$ activation or inactivation

We were able to significantly model  $I_{Ks}$   $V_{1/2}$  activation. However, no models could reliably predict  $I_{Na}$   $V_{1/2}$  activation or inactivation. The  $I_{Ks}$   $V_{1/2}$  activation variance explained is relatively high, 0.29 with a 95% confidence interval lower bound of 0.12 (Table 2). The functional density feature had a significant  $p$ -value, suggesting a three-dimensional localization of regions that influence  $V_{1/2}$  activation (Table S2, Fig. S7).

### 3.4. Most $I_{Na}$ and $I_{Ks}$ functional parameters cannot be reliably predicted

Most  $I_{Ks}$  and  $I_{Na}$  functional parameters assessed could not be predicted with stable fully relaxed LASSO-regularized linear models and a lower bound of the 95% confidence interval in  $\text{adj. } R^2 > 0.10$ . In many cases for these functional parameters, at least one of the 10 folds in the cross validation resulted in only an intercept, i.e.  $\beta$  coefficients for all inputted features shrunk to 0. For some functional parameters, such as time constants for  $I_{Ks}$  activation and inactivation and  $I_{Na}$  late current and recovery from inactivation times, lower numbers of characterized variants and relatively low dispersion of values (Table 1, Figs. 1 and 2) mean the data themselves are limiting prediction. Alternatively, or in addition, our chosen feature set may contain little information relevant to the prediction of these values, likely the case for  $I_{Na}$   $V_{1/2}$  activation and inactivation, which may be under sampled for the functional density analysis.

### 3.5. Structural features improve $I_{Na}$ but not $I_{Ks}$ loss-of-function classification

For comparison with published variant classifiers predicting binary functional perturbation of these two channels [22,25], we calculated receiver operating characteristic curves for models trained using only published models as features and models trained additionally with structure-based features. We generated binary classifications of loss-of-function *SCN5A* and *KCNQ1* variants using criteria described above

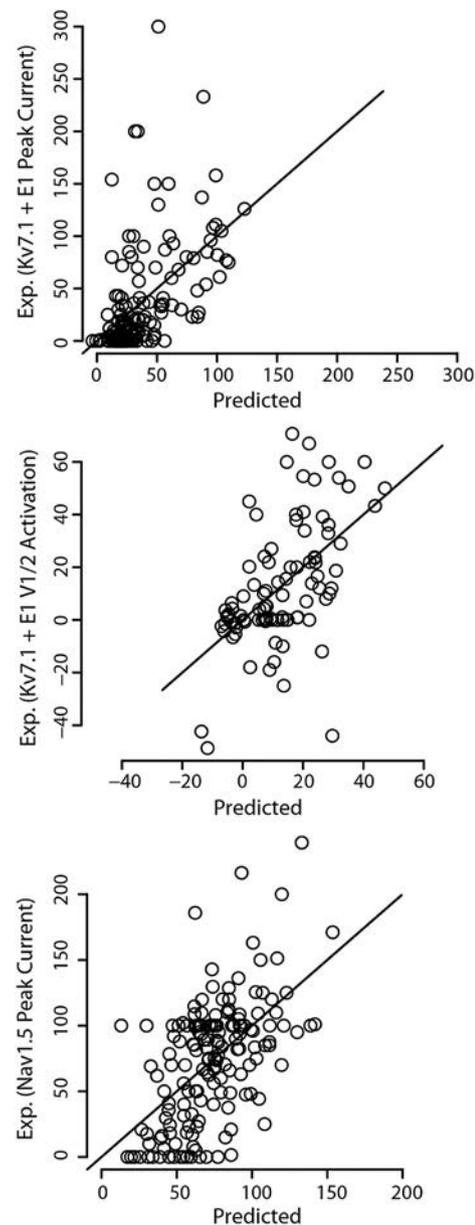
**Table 2**

Summary statistics of predictive model.

Functional parameter	Adj. $R^{2\dagger}$ [95% CI <sup>†</sup> ; CV <sup>‡</sup> ]
$I_{Ks}$ Peak Current	0.24 [0.14–0.46; 0.24]
$I_{Ks}$ $V_{1/2}$ Activation	0.29 [0.12–0.48; 0.23]
$Na_v1.5$ Peak Current	0.27 [0.18–0.45; 0.23]
$Na_v1.5$ $V_{1/2}$ Inact.	0.16 [0.08–0.34; 0.05]

<sup>†</sup> CI (confidence interval)

<sup>‡</sup> CV (10 fold cross-validation)



**Fig. 3.** Experimental vs. predicted functional parameters for the subset of functional features with significant predictive models (Table 2). Plot of experimental  $I_{Ks}$  peak current,  $I_{Ks}$   $V_{1/2}$  activation, and  $I_{Na}$   $Na_v1.5$  peak current vs. predictions from a linear regression. The resulting models explain 0.24, 0.29, and 0.27 of the variance in  $I_{Ks}$  peak current,  $I_{Ks}$   $V_{1/2}$  activation, and  $I_{Na}$  peak current, respectively.

in the methods section. We calculated the ability of several variant classifiers to correctly classify LOF variants. The resulting areas under the curve (AUCs) from logistic models trained to predict *KCNQ1* LOF were as follows (AUC; [95% CI]): PolyPhen-2 (0.81; [0.74–0.92]), rate of evolution (0.77; [0.67–0.87]), BLAST-PSSM (0.84; [0.76–0.92]), PROVEAN (0.83; [0.75–0.91]), all published predictive models (0.86; [0.78–0.94]), all published predictive models with functional density for peak current (0.87; [0.79–0.94]). Most variant classifiers performed reasonably well and the addition of structural information did not meaningfully improve classification for this task. However, the resulting AUCs from logistic models trained to predict *SCN5A* LOF were as follows: PolyPhen-2 (0.60; [0.51–0.68]), rate of evolution (0.51; [0.42–0.60]), BLAST-PSSM (0.61; [0.52–0.69]), PROVEAN (0.66; [0.57–0.75]), SIFT (0.53; [0.48–0.58]), all published variant classifiers (0.69; [0.60–0.77]), all published variant classifiers with functional density for peak current (0.78; [0.70–0.85]). This improvement in classification ability for LOF

variants in *SCN5A* when adding functional density for peak current (0.69 without vs. 0.78 with,  $p = .01$ ) suggests structure-based features contribute information not contained in other predictive features (Fig. S8) an observation gaining appreciation elsewhere [42,43].

## 4. Discussion

### 4.1. A limited number of $I_{Ks}$ and $I_{Na}$ functional parameters can be predicted reliably

Most  $I_{Ks}$  and  $I_{Na}$  functional parameters analyzed could not be predicted reliably:  $I_{Ks}$  time constants of activation and inactivation; and  $I_{Na}$   $V_{1/2}$  activation/inactivation, recovery from inactivation, and late current. However, three important functional parameters could be predicted:  $I_{Ks}$  peak current and  $V_{1/2}$  activation and  $I_{Na}$  peak current. In two of these models,  $I_{Ks}$   $V_{1/2}$  activation and  $I_{Na}$  peak current, the functional density features have the greatest predictive value, indicating three-dimensional enrichment of regions of the proteins that influence these functional parameters (Table S2 and S3, Figs. S6 and S7).

### 4.2. Functional density suggests regions in three-dimensional space are enriched for influence on $I_{Ks}$ $V_{1/2}$ activation and $I_{Na}$ peak current

“Functional densities” are measure of how dense pathogenic variants are near the residue of interest, i.e. are they near “hotspots” that influence a particular function. Given the influence of the functional density calculation in predicting  $I_{Ks}$   $V_{1/2}$  activation and  $I_{Na}$  peak current, there is likely a spatial influence over both of these parameters. As can be seen in Figs. 4 and 5, there are regions (black circles) where variants that have a large influence on  $I_{Ks}$   $V_{1/2}$  activation and  $I_{Na}$  peak current are localized. Not surprisingly, the greatest perturbations in  $I_{Ks}$   $V_{1/2}$  activation are in the regions of the channel known to be functionally critical: the selectivity filter, voltage-sensing helix in the voltage sensing domain, and in the constriction point in the middle of the pore, as we have seen previously. [25] The S6 helix in  $Kv7.1$  influences activation in part through its intrinsic flexibility, a necessary property for activation. [40] S0 helix has been found to provide stabilization to the voltage sensing domain. [17] S4 helix is canonically responsible for voltage-dependent activation

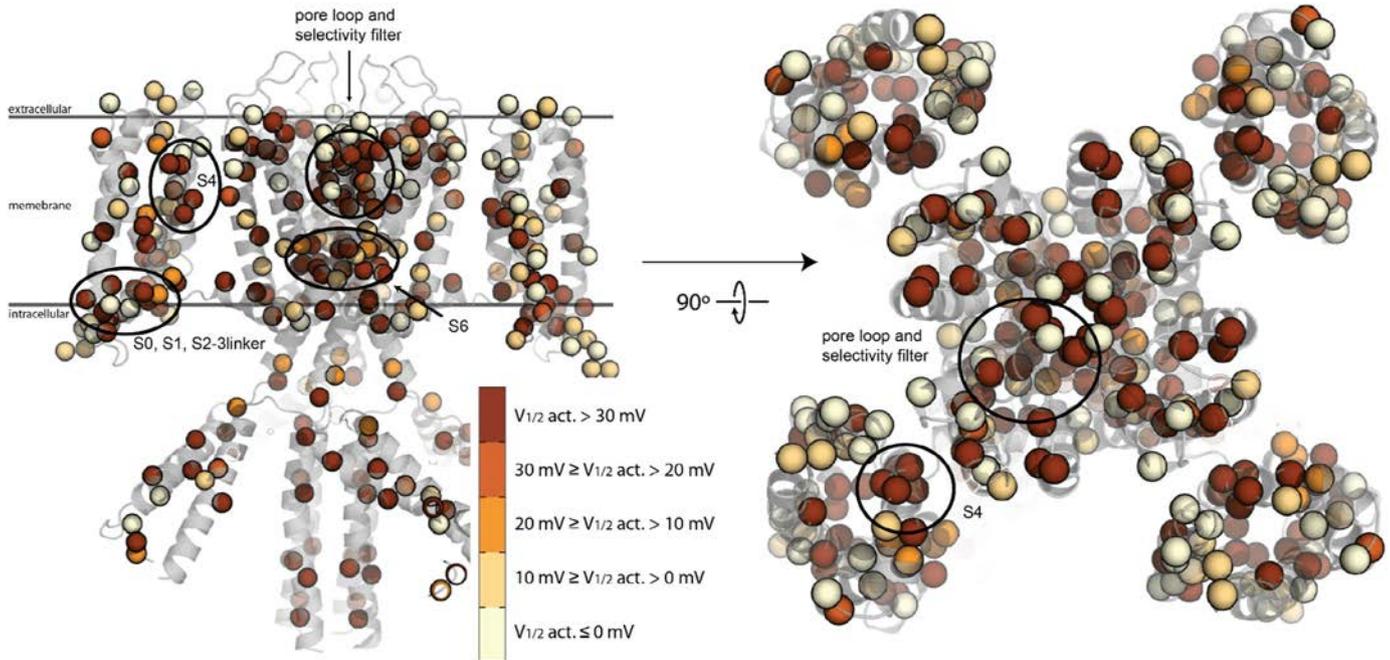
[8,11,31]. Interestingly, the variants most disruptive to  $I_{Na}$  peak current are located in the extracellular region of the channel, mostly near the selectivity filter. The pore region of voltage-gated sodium channels is canonically responsible for  $Na^+$  conduction [2] and is also enriched BrS1 variants, an  $Na_v1.5$  loss-of-function disorder [21,22]. These data suggest the utility in leveraging combined structural and previously determined functional perturbation datasets to predict functional disruption of previously uncharacterized channel variants.

### 4.3. Challenging regions to predict

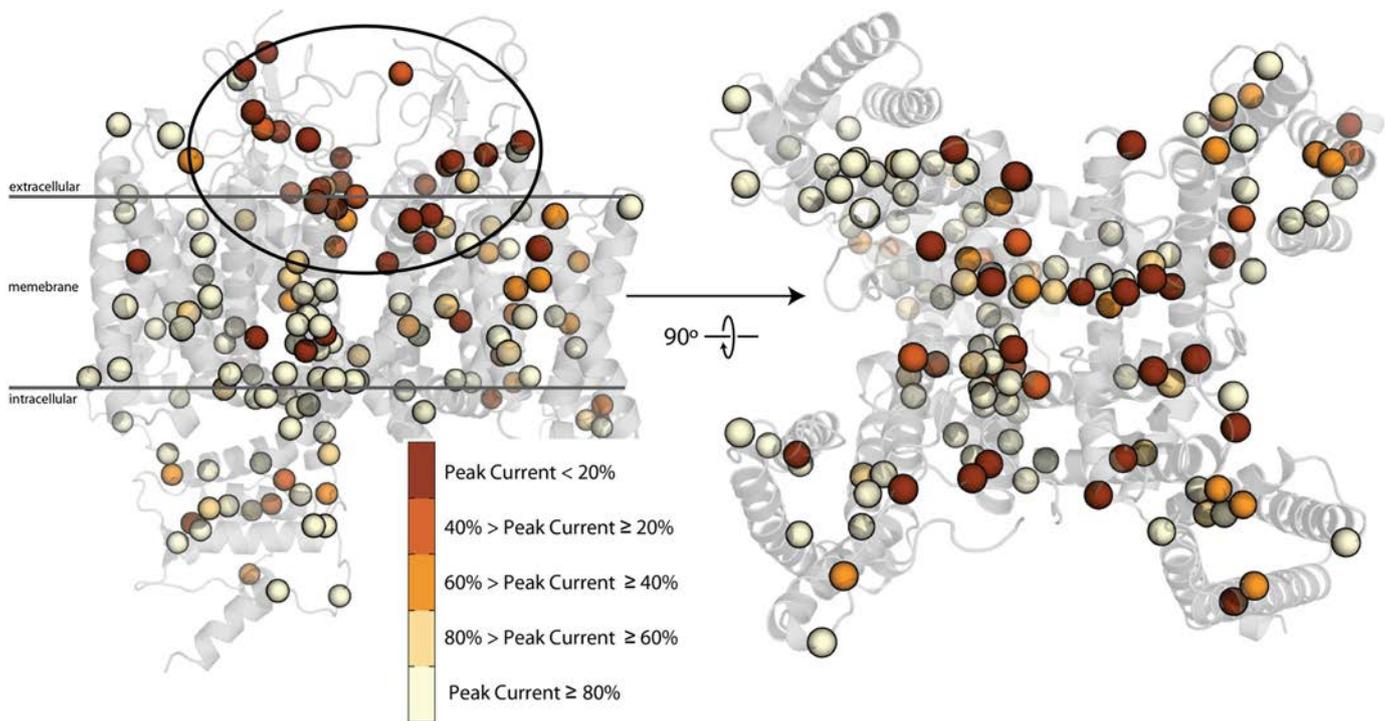
To identify potential commonalities among the most challenging variants to predict, we identified the five least congruent predictions, at extremes both greater and less than experiment, for  $I_{Ks}$  peak current,  $I_{Ks}$   $V_{1/2}$  activation, and  $I_{Na}$  peak current (Fig. S8–10). All variants, with one exception, occur in the transmembrane region and on structured segments, not flexible loops or linkers. Some commonalities for challenges in predicting  $I_{Ks}$  peak current and  $V_{1/2}$  activation prediction are the extracellular half of the voltage sensing domain, especially S3 and S4 helices, and the interface between the pore loop helix and helices S5 and S6. The S3 and S4 helices of the voltage sensing domain undergo large conformational changes in response to voltage [8,11,31] which are not captured by the static structure we used in this analysis. However, the distribution of predictions both greater than and less than experiment within these two segments suggests changes in function in these regions are heterogeneous possibly due to individual residues in these regions having special roles in voltage-gated activation. Interestingly, several of the challenging  $I_{Ks}$  peak current variants are located on the S0 helix in  $Kv7.1$ . We previously observed an anomalous sensitivity to expression level in the S0 helix and suggest the protein is stabilized by intramolecular interactions between the S0 helix and the rest of the voltage sensing domain. [17] Challenging variants for  $I_{Na}$  peak current are more evenly distributed though the protein molecule (Fig. S10).

### 4.4. Classification of loss-of-function *KCNQ1* and *SCN5A* variants

Classification of variants inherently reduces the richness of available data, in our case the continuous functional perturbation induced by



**Fig. 4.** Structural model of  $Kv7.1$  with colored spheres at  $C_{\alpha}$  positions where variants have  $V_{1/2}$  activation data available. Colors indicate the degree of perturbation from WT  $Kv7.1$ , with the darker color displaying variants with more positive shifts in  $V_{1/2}$  activation. Selection criteria are displayed in the inset. Several regions of apparent enrichment are highlighted by circles. The tetrameric structure gives the appearance of a greater number of functionally characterized variants.



**Fig. 5.** Structural model of Na<sub>v</sub>1.5 with colored spheres at C<sub>α</sub> positions where variants have peak current available. Colors indicate the degree of perturbation from WT Na<sub>v</sub>1.5, with the darker color displaying variants with less peak current. Selection criteria are displayed in the inset. A single extracellular region shows apparent enrichment and is circled. Even though there are a greater number of variants functionally characterized for Na<sub>v</sub>1.5, K<sub>v</sub>7.1 appears to have a greater number due to its homotetrameric structure.

variants in *SCN5A* and *KCNQ1*. However, to assess how well structure-based features contribute to predicting variant loss-of-function classification, we built logistic models trained on variants classified as loss-of-function or not loss-of-function. For I<sub>Na</sub>, structure-based features improve the AUC (Fig. S11); for I<sub>Ks</sub> there is no significant improvement. This is consistent with our previous *KCNQ1* work suggesting sequence and evolutionary-based features, BLAST-PSSM and residue rate of evolution, yield a competent classification model and suggests alternative features will be needed to further improve prediction of *KCNQ1* variants [25]. For *SCN5A*, structure-based features improve the classification of loss-of-function variants from an AUC of 0.69 to 0.78 ( $p = .01$ ).

#### 4.5. Recent interest in predicting functional perturbation

Recently, Clerx et al. attempted to predict classification of functionally compromised I<sub>Na</sub> for many of the functional parameters we report here [10]. The authors report modest classification ability for I<sub>Na</sub> late current and V<sub>1/2</sub> activation/inactivation with better performance predicting complete loss of function. We too find limited ability to predict most functional perturbations; however, we found significant and quantitative correlations between predicted and experimental I<sub>Na</sub> peak current and challenge the use of functional classification in favor of quantitative perturbation prediction. Interestingly, the authors also noted difficulty in predicting late current which we recapitulate here suggesting this feature is a more challenging target to predict. Furthermore, here we put forward a feature based on knowledge of the three-dimensional structure, functional density, and demonstrate its utility in predicting variant phenotype.

#### 4.6. Application to variant annotation

The field is still evolving on how to include *in silico* predictions and experimental functional data quantitatively [33]. We suggest the model presented here could be useful in a pipeline whose first-pass filter aims to detect pathogenic variants. Our previous

publication suggested the degree to which a loss-of-function variant produces non-negligible penetrance was an I<sub>Na</sub> peak current 50% or less than that of WT. We suggest this implies the need to have a variance explained of experimental data from our predictions >50% such that the probability a variant predicted to be WT actually has <50% peak current is very low. Predicting around 0.2 of the variance in relevant I<sub>Ks</sub> and I<sub>Na</sub> functional parameters we show here is significant; however, further improvement is needed before the predictive models will be useful in classifying variants for clinical use.

#### 4.7. Limitations

The dataset used was limited by those variants available in the literature, which are biased towards functionally perturbed variants. We chose to analyze I<sub>Ks</sub> generated with homozygous K<sub>v</sub>7.1 variants (co-expressed with KCNE1) because this configuration is reported most consistently in the literature. In a majority of cases, K<sub>v</sub>7.1 variants are heterozygous in individuals. Furthermore, we have begun to investigate the influence of variant-specific functional perturbation on clinical presentation [22], but the exact relationship is complicated (notably including β-adrenergic regulation for I<sub>Ks</sub>) and warrants further investigation. Another limitation is that the structural models are imperfect estimates of the functional state they represent and are also only representative of a single functional state in channels known to have at least two functional states. Models reflecting greater conformational diversity may be another source for improved features.

#### 4.8. Conclusions

We have derived predictive features from three-dimensional structures of Na<sub>v</sub>1.5 and K<sub>v</sub>7.1 and have demonstrated these features improve our ability to predict variant-induced functional perturbations in each channel. These predictive features are based on recognizing that residue positions for pathogenic variants are likely to be clustered in three-dimensional space in proximity to other pathogenic residues. Based on

this recognition, we can account for approximately 0.2 of the variance in  $I_{Ks}$  peak current,  $I_{Ks} V_{1/2}$  activation, and  $I_{Na}$  peak current. For  $I_{Ks} V_{1/2}$  activation and  $I_{Na}$  peak current, structure-based features contribute meaningfully to the predictive model and in a way not recapitulated by commonly used sequence, evolutionary features, or genetic variant classifiers methods. For predicting variant-induced loss-of-function, structure-based features contribute meaningfully to  $I_{Na}$  but not  $I_{Ks}$ .

## Funding

This work was supported by the National Institutes of Health K99HL135442 to B.M.K.; R35GM127087 to J.A.C.; HL122010 to A.L.G., C.R.S., and J.M.; and P50GM115305 to D.M.R.

## Acknowledgements

The authors would like to thank Bian Li and Mike Sivley for useful discussions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2019.01.008>.

## References

- [1] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7(4):248–9. <https://doi.org/10.1038/nmeth0410-248>.
- [2] Ahern CA, Payandeh J, Bosmans F, Chanda B. The hitchhiker's guide to the voltage-gated sodium channel galaxy. *J Gen Physiol* 2016;147(1):1–24. <https://doi.org/10.1085/jgp.201511492>.
- [3] Andrews TD, Sjollem G, Goodnow CC. Understanding the immunological impact of the human mutation explosion. *Trends Immunol* 2013;34(3):99–106. <https://doi.org/10.1016/j.it.2012.12.001>.
- [4] Barhanin J, Lesage F, Guillemare E, Fink M, Lazdunski M, Romey G. K(V)LQT1 and Isk (minK) proteins associate to form the I(Ks) cardiac potassium current. *Nature* 1996;384(6604):78–80. <https://doi.org/10.1038/384078a0>.
- [5] Bezzina C, Veldkamp MW, van Den Berg MP, Postma AV, Rook MB, Viersma JW, et al. A single Na(+) channel mutation causing both long-QT and Brugada syndromes. *Circ Res* 1999;85(12):1206–13 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10590249>.
- [6] Bezzina CR, Rook MB, Groenewegen WA, Herfst LJ, van der Wal AC, Lam J, et al. Compound heterozygosity for mutations (W156X and R225W) in SCN5A associated with severe cardiac conduction disturbances and degenerative changes in the conduction system. *Circ Res* 2003;92(2):159–68 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12574143>.
- [7] Charpentier F, Bourge A, Merot J. Mouse models of SCN5A-related cardiac arrhythmias. *Prog Biophys Mol Biol* 2008;98(2–3):230–7. <https://doi.org/10.1016/j.pbiomolbio.2008.10.012>.
- [8] Choe S. Potassium channel structures. *Nat Rev Neurosci* 2002;3(2):115–21. <https://doi.org/10.1038/nrn727>.
- [9] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7(10):e46688. <https://doi.org/10.1371/journal.pone.0046688>.
- [10] Clerx M, Heijman J, Collins P, Volders PGA. Predicting changes to INa from missense mutations in human SCN5A. *Sci Rep* 2018;8(1):12797. <https://doi.org/10.1038/s41598-018-30577-5>.
- [11] Cui J. Voltage-Dependent Gating: Novel Insights from KCNQ1 Channels. *Biophys J* 2016;110(1):14–25. <https://doi.org/10.1016/j.bpj.2015.11.023>.
- [12] Durham E, Dorr B, Woetzel N, Staritzbichler R, Meiler J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J Mol Model* 2009;15(9):1093–108. <https://doi.org/10.1007/s00894-009-0454-9>.
- [13] Gaita F, Giustetto C, Bianchi F, Wolpert C, Schimpf R, Riccardi R, et al. Short QT Syndrome: a familial cause of sudden death. *Circulation* 2003;108(8):965–70. <https://doi.org/10.1161/01.CIR.0000085071.28695.C4>.
- [14] Ge J, Sun A, Paajanen V, Wang S, Su C, Yang Z, et al. Molecular and clinical characterization of a novel SCN5A mutation associated with atrioventricular block and dilated cardiomyopathy. *Circ Arrhythm Electrophysiol* 2008;1(2):83–92. <https://doi.org/10.1161/CIRCEP.107.750752>.
- [15] Gui J, Wang T, Jones RP, Trump D, Zimmer T, Lei M. Multiple loss-of-function mechanisms contribute to SCN5A-related familial sick sinus syndrome. *PLoS One* 2010;5(6):e10985. <https://doi.org/10.1371/journal.pone.0010985>.
- [16] Hastie T, Tibshirani R, Tibshirani RJ. Extended Comparisons of best Subset selection, Forward Stepwise selection, and the Lasso arXiv, 1707 ; 2017; 08692.
- [17] Huang H, Kuenze G, Smith JA, Taylor KC, Duran AM, Hadziselimovic A, et al. Mechanisms of KCNQ1 channel dysfunction in long QT syndrome involving voltage sensor domain mutations. *Sci Adv* 2018;4(3):eaar2631. <https://doi.org/10.1126/sciadv.aar2631>.
- [18] Huang H, Priori SG, Napolitano C, O'Leary ME, Chahine M. Y1767C, a novel SCN5A mutation, induces a persistent Na+ current and potentiates ranolazine inhibition of Nav1.5 channels. *Am J Physiol Heart Circ Physiol* 2011;300(1):H288–99. <https://doi.org/10.1152/ajpheart.00539.2010>.
- [19] Jespersen T, Grunnet M, Olesen SP. The KCNQ1 potassium channel: from gene to physiological function. *Physiology (Bethesda)* 2005;20:408–16. <https://doi.org/10.1152/physiol.00031.2005>.
- [20] Juang JM, Lu TP, Lai LC, Hsueh CH, Liu YB, Tsai CT, et al. Utilizing multiple in silico analyses to identify putative causal scn5a variants in brugada syndrome. *Sci Rep* 2014;4 doi: ARTN 385010.1038/srep03850.
- [21] Kapplinger JD, Tester DJ, Alders M, Benito B, Berthet M, Brugada J, et al. An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart Rhythm* 2010;7(1):33–46. <https://doi.org/10.1016/j.hrthm.2009.09.069>.
- [22] Kroncke BM, Glazer AM, Smith DK, Blume JD, Roden DM. SCN5A (NaV1.5) variant functional perturbation and clinical presentation: variants of a certain significance. *Circ Genom Precis Med* 2018;11(5):e002095. <https://doi.org/10.1161/CIRCGEN.118.002095>.
- [23] Kroncke BM, Vanoye CG, Meiler J, George AL, Sanders CR. Personalized Biochemistry and Biophysics. *Biochemistry* 2015;54(16):2551–9. <https://doi.org/10.1021/acs.biochem.5b00189>.
- [24] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4(7):1073–81. <https://doi.org/10.1038/nprot.2009.86>.
- [25] Li B, Mendenhall JL, Kroncke BM, Taylor KC, Huang H, Smith DK, et al. Predicting the functional impact of KCNQ1 variants of unknown significance. *Circ Cardiovasc Genet* 2017;10(5). <https://doi.org/10.1161/CIRCGENETICS.117.001754>.
- [26] Marian AJ. Challenges in medical applications of whole exome/genome sequencing discoveries. *Trends Cardiovasc Med* 2012;22(8):219–23. <https://doi.org/10.1016/j.tcm.2012.08.001>.
- [27] Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, et al. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci U S A* 2015;112(37):E5189–98. <https://doi.org/10.1073/pnas.1511585112>.
- [28] Moreau A, Gosselin-Badaroudine P, Delemotte L, Klein ML, Chahine M. Gating pore currents are defects in common with two Nav1.5 mutations in patients with mixed arrhythmias and dilated cardiomyopathy. *J Gen Physiol* 2015;145(2):93–106. <https://doi.org/10.1085/jgp.201411304>.
- [29] Neu A, Eisel M, Paul M, Sauter K, Stallmeyer B, Isbrandt D, et al. A homozygous SCN5A mutation in a severe, recessive type of cardiac conduction disease. *Hum Mutat* 2010;31(8):E1609–21. <https://doi.org/10.1002/humu.21302>.
- [30] Nuyens D, Stengl M, Dugarmaa S, Rossenbacher T, Compennolle V, Rudy Y, et al. Abrupt rate accelerations or premature beats cause life-threatening arrhythmias in mice with long-QT3 syndrome. *Nat Med* 2001;7(9):1021–7. <https://doi.org/10.1038/nm0901-1021>.
- [31] Panaghie G, Abbott GW. The role of S4 charges in voltage-dependent and voltage-independent KCNQ1 potassium channel complexes. *J Gen Physiol* 2007;129(2):121–33. <https://doi.org/10.1085/jgp.200609612>.
- [32] Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* 2002;18(Suppl. 1):S71–7 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12169533>.
- [33] 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(5):405–24. <https://doi.org/10.1038/gim.2015.30>.
- [34] Ruan Y, Denegri M, Liu N, Banchetti T, Seregni M, Morotti S, et al. Trafficking defects and gating abnormalities of a novel SCN5A mutation question gene-specific therapy in long QT syndrome type 3. *Circ Res* 2010;106(8):1374–83. <https://doi.org/10.1161/CIRCRESAHA.110.218891>.
- [35] Ruan Y, Liu N, Bloise R, Napolitano C, Priori SG. Gating properties of SCN5A mutations and the response to mexiletine in long-QT syndrome type 3 patients. *Circulation* 2007;116(10):1137–44. <https://doi.org/10.1161/CIRCULATIONAHA.107.707877>.
- [36] Ruan Y, Liu N, Priori SG. Sodium channel mutations and arrhythmias. *Nat Rev Cardiol* 2009;6(5):337–48. <https://doi.org/10.1038/nrcardio.2009.44>.
- [37] Sanguinetti MC, Curran ME, Zou A, Shen J, Spector PS, Atkinson DL, et al. Coassembly of K(V)LQT1 and minK (IsK) proteins to form cardiac I(Ks) potassium channel. *Nature* 1996;384(6604):80–3. <https://doi.org/10.1038/384080a0>.
- [38] Schwartz PJ, Ackerman MJ, George Jr AL, Wilde AAM. Impact of genetics on the clinical management of channelopathies. *J Am Coll Cardiol* 2013;62(3):169–80. <https://doi.org/10.1016/j.jacc.2013.04.044>.
- [39] Schwarz R, Dayhoff M. Matrices for detecting distant relationships. In: Dayhoff M, editor. *Atlas of protein sequences*. National Biomedical Research Foundation; 1979. p. 353–8.
- [40] Seeböhm G, Strutz-Seeböhm N, Ureche ON, Baltaev R, Lampert A, Kornichuk G, et al. Differential roles of S6 domain hinges in the gating of KCNQ potassium channels. *Biophys J* 2006;90(6):2235–44. <https://doi.org/10.1529/biophysj.105.067165>.
- [41] Shen H, Zhou Q, Pan X, Li Z, Wu J, Yan N. Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science* 2017;355(6328). <https://doi.org/10.1126/science.aal4326>.
- [42] Sivley RM, Dou X, Meiler J, Bush WS, Capra JA. Comprehensive Analysis of Constraint on the Spatial Distribution of Missense Variants in Human Protein Structures. *Am J Hum Genet* 2018;102(3):415–26. <https://doi.org/10.1016/j.ajhg.2018.01.017>.

- [43] Sivley RM, Sheehan JH, Kropski JA, Cogan J, Blackwell TS, Phillips JA, et al. Three-dimensional spatial analysis of missense variants in RTEL1 identifies pathogenic variants in patients with Familial Interstitial Pneumonia. *BMC Bioinformatics* 2018;19(1):18. <https://doi.org/10.1186/s12859-018-2010-z>.
- [44] Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, et al. High-resolution comparative modeling with RosettaCM. *Structure* 2013;21(10):1735–42. <https://doi.org/10.1016/j.str.2013.08.005>.
- [45] Sun J, MacKinnon R. Cryo-EM Structure of a KCNQ1/CaM complex reveals Insights into Congenital Long QT Syndrome. *Cell* 2017;169(6). <https://doi.org/10.1016/j.cell.2017.05.019> 1042–1050 e1049.
- [46] Valdivia CR, Ackerman MJ, Tester DJ, Wada T, McCormack J, Ye B, et al. A novel SCN5A arrhythmia mutation, M1766L, with expression defect rescued by mexiletine. *Cardiovasc Res* 2002;55(2):279–89 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12123767>.
- [47] Vanoye C, Desai R, Fabre K, Potet F, DeKeyser J-M, Macaya D, et al. High throughput functional evaluation of KCNQ1 decrypts variants of unknown significance bioRxiv ; 2017. <https://doi.org/10.1101/223206>.
- [48] Vanoye CG, Desai RR, Fabre KL, Gallagher SL, Potet F, DeKeyser JM, et al. High-throughput functional evaluation of KCNQ1 decrypts variants of unknown significance. *Circ Genom Precis Med* 2018;11(11):e002345. <https://doi.org/10.1161/CIRCGEN.118.002345>.
- [49] Veerman CC, Wilde AA, Lodder EM. The cardiac sodium channel gene SCN5A and its gene product Nav1.5: Role in physiology and pathophysiology. *Gene* 2015;573(2):177–87. <https://doi.org/10.1016/j.gene.2015.08.062>.
- [50] Wang Z, Moulton J. SNPs, protein structure, and disease. *Hum Mutat* 2001;17(4):263–70. <https://doi.org/10.1002/humu.22>.
- [51] Woetzel N, Karakas M, Staritzbichler R, Muller R, Weiner BE, Meiler J. BCL::Score-knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PLoS One* 2012;7(11):e49242. <https://doi.org/10.1371/journal.pone.0049242>.
- [52] Yan Z, Zhou Q, Wang L, Wu J, Zhao Y, Huang G, et al. Structure of the Nav1.4-beta1 complex from Electric Eel. *Cell* 2017;170(3). <https://doi.org/10.1016/j.cell.2017.06.039> 470–482 e411.