# Continuous Bayesian Variant Interpretation Accounts for Incomplete Penetrance among Mendelian Cardiac Channelopathies

## Supplemental Material

**Supplemental Methods**

## Supplemental Methods

*Appraisal of Literature Data.* The presence of absence of a clinical diagnosis of LQTS was defined by a Schwartz score[1] ≥3.5 or a single QT interval ≥480 ms[3]. Affected status of LQT1 was annotated in compliance with the diagnostic guidelines published by the HRS/EHRA/APHRS Expert Consensus Statement on the Management of inherited Arrhythmias[2], withholding point contributions from genotype-positive status. In literature reports where patient level data was not available for manual review, the status was assigned in accordance with that reported by the authors. The *KCNQ1* transcript ENST0000155840 was used to map all variants. PubMed was searched for the term "KCNQ1" and identified a total of 2216 manuscripts, of which 326 provided sufficient information to interpret meaningful clinical phenotypes. Each paper was manually reviewed for LQT1 status in affected and unaffected individuals (described below). The full set of heterozygotes for variants in *KCNQ1* from gnomAD was added to this curation. All individuals from gnomAD were assumed not to have LQT1, an assumption previously shown not to alter the model performance[3,4]. From the literature and gnomAD X heterozygotes for XXX unique missense or in-frame insertion/deletion (indel) variants were identified. The *SCN5A* transcript ENST00000333535 was used to annotate all variants. For *SCN5A*-LQT3, a previously reported dataset was queried for association with LQTS[4]. This comprised 86118 heterozygotes with 1667 unique missense or in-frame insertion/deletions.

*Collation of Cohort Data*. Three international centers with arrhythmia genetics expertise contributed clinical phenotypes and genotypes for *KCNQ1* probands and family members when available; Italy (600 heterozygotes harboring 76 unique missense variants), France (438 heterozygotes with 88 unique missense variants), and Japan (350 heterozygotes with 82 unique missense variants). The literature data was reviewed to remove potential overlap of patients

between the cohort and literature set. The non-overlapping literature dataset and cohort set were combined and optimism was estimated by 10-fold cross validation.

KCNQ1 *Variant Specific Features: Functional Data, Structural Data, and* in silico *Covariates.* Functional data were collected from a previously assembled dataset of *KCNQ1* variant functional data[5], and from recent high-throughput automated patch clamp experiments[6,7]. When multiple functional measurements were available for a single variant, high-throughput SyncroPatch measurements were given priority. A published cryo-EM model of *KCNQ1* enabled high-resolution structural studies[8]. We applied our previously developed penetrance density metric as a covariate in the *KCNQ1*-LQT1 model (see previous manuscript for full description)[5]. We obtained pathogenicity predictions using the *in silico* servers CardioBoost[9], PolyPhen-2[10], and PROVEAN[11]. Conservation scores were obtained by the basic local alignment search tool position-specific scoring matrix (BLAST-PSSM) and point accepted mutation score. The meta-predictor REVEL was additionally included given its previous impressive performance in similar applications[12].

*Expectation-Maximization (EM) Algorithm.* The present EM algorithm comprises 2 steps: 1) calculate the expected penetrance from an empirical Bayes penetrance model; 2) maximum likelihood regression of estimated penetrance on variant-specific features. The EM algorithm was performed using the covariates LQT1 probability density, CardioBoost, Provean, REVEL, and homozygous peak current. The covariates were regressed to the fraction of heterozygous carriers diagnosed with LQT1 using a logistic regression model. Missing data was handled using a pattern mixture model, with a model fit for each missing data pattern. Full details of the approach are described in the Supplementary Data (full code available on GitHub).

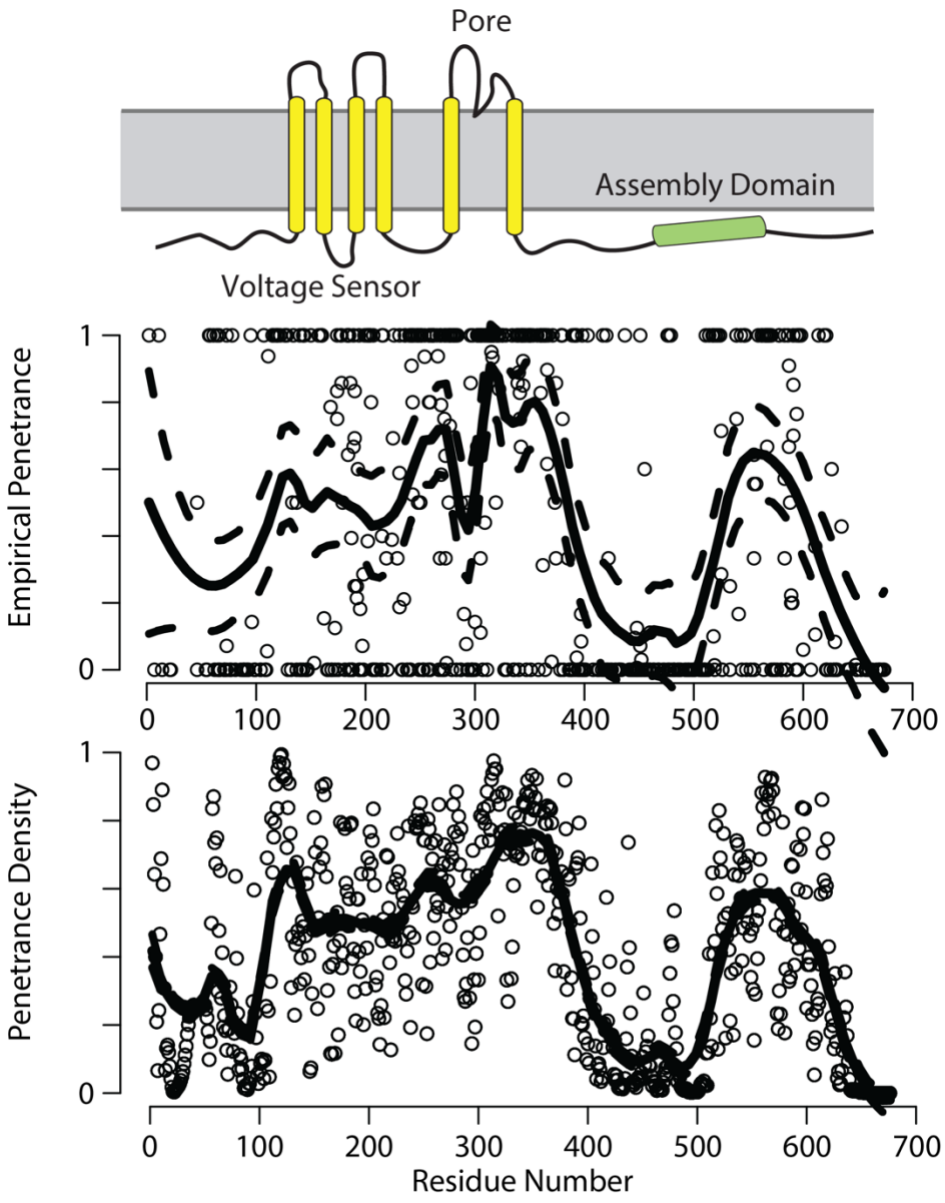
*Statistical evaluation.* All reported correlation coefficients, Spearman ρ, Pearson's $R^2$, and coefficient of determination were weighted by the function $1 - \frac{1}{0.01+total\ heterozygotes}$, as previously described[3,4], unless otherwise noted, to ensure variants with higher total heterozygote counts had greater influence in the resulting correlation coefficient estimate. All scripts and data used are available at the Kroncke Lab GitHub page. Additionally, a compiled and curated form of the data presented here are available in the *KCNQ1* variant browser. We performed 10-fold cross-validation using the caret package in R to create folds. All variants with heterozygous carriers were included in the model. We calculated Spearman and Pearson correlations between the Bayesian prior and Bayesian penetrance for each fold. We report summary statistics for the average correlation coefficients and variances observed for 10 independent implementations of the 10-fold cross-fold validation. Brier scores were used as a probabilistic classification metric as defined below:

$$\text{Brier Score} = \frac{1}{N}\sum_{t=1}^{n}(f_t - o_t)^2$$

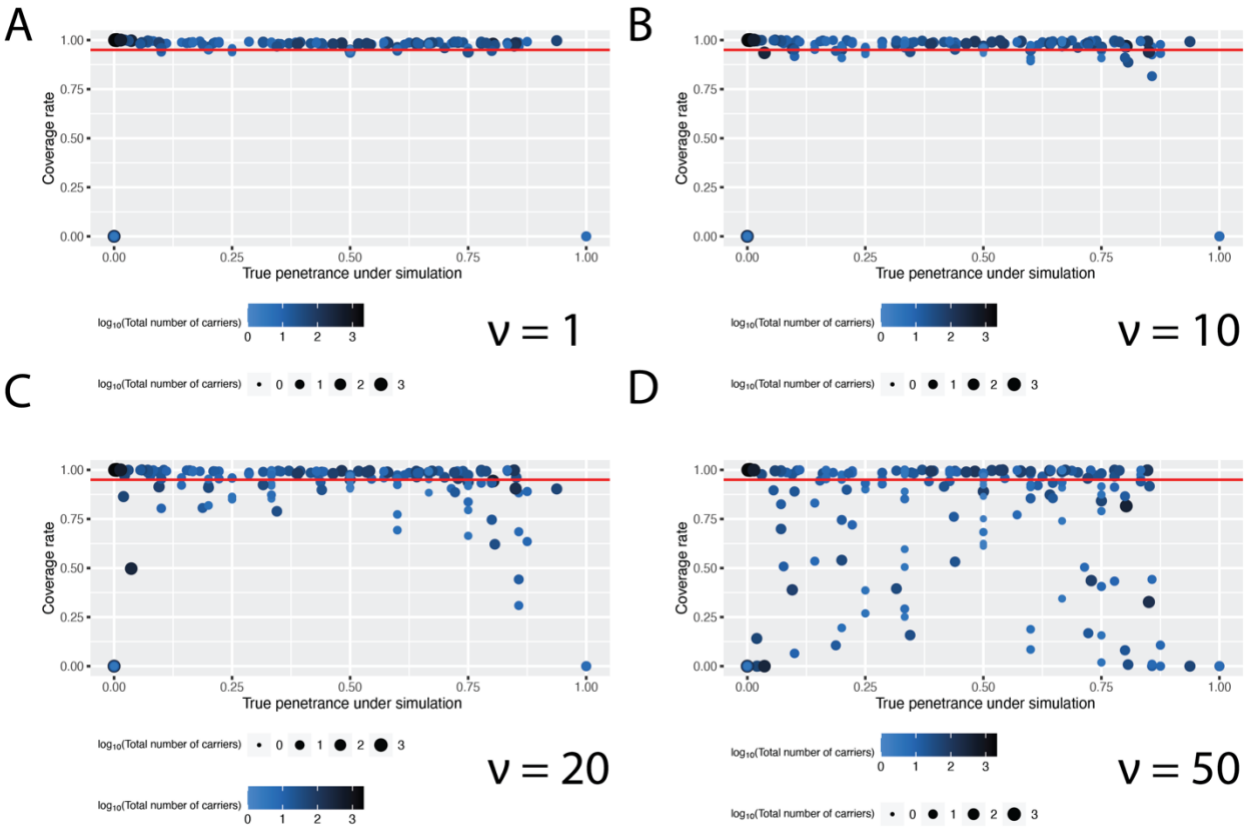
Where N is the number of observations for a covariate, $f_t$ is the probability of the covariate for the $t^{th}$ variant, and $o_t$ is the observed penetrance for the $t^{th}$ covariate.

**Supplemental Table 1.** Weighted Spearman and Pearson correlation coefficients and p-values. Homozygous (hm); Heterozygous (ht).

| Covariate | Weighted Spearman | Spearman p-Value | Weighted Pearson | Pearson p-Value | n |
|---|---|---|---|---|---|
| lqt1_dist | 0.54 | 0 | 0.55 | 0 | 586 |
| cardiacboost | 0.65 | 0 | 0.59 | 0 | 582 |
| hm_peak | -0.56 | 0 | -0.53 | 0 | 143 |
| hm_Vhalfact | 0.43 | 0 | 0.38 | 0 | 102 |
| hm_tauact | 0.014 | 0.94 | 0.063 | 0.67 | 45 |
| hm_taudeact | -0.24 | 0.076 | -0.13 | 0.40 | 73 |
| ht_peak | -0.55 | 0 | -0.59 | 0 | 79 |
| ht_Vhalfact | 0.27 | 0.047 | 0.26 | 0.054 | 68 |
| ht_tauact | 0.044 | 0.81 | -0.075 | 0.69 | 39 |
| ht_taudeact | -0.27 | 0.13 | -0.076 | 0.74 | 45 |
| pph2_prob | 0.54 | 0 | 0.42 | 0 | 586 |
| provean_score | -0.62 | 0 | -0.59 | 0 | 586 |
| revel_score | 0.64 | 0 | 0.62 | 0 | 586 |
| blast_pssm | -0.26 | 0 | -0.25 | 0 | 586 |
| Bayes prior | 0.71 | 0 | 0.70 | 0 | 586 |
| Bayes posterior | 0.95 | 0 | 0.95 | 0 | 586 |




**Supplemental Figure 1.** Observed Penetrance and Penetrance Density by Residue. Cartoon schematic of the structure of KCNQ1 followed by empirical penetrance by residue (weighted across all affected and total heterozygotes at each residue) and penetrance density as described for the LQT1-dist covariate. Note smoothing of extremes in the LQT1 dist analysis. Solid bars indicate 95% confidence intervals.



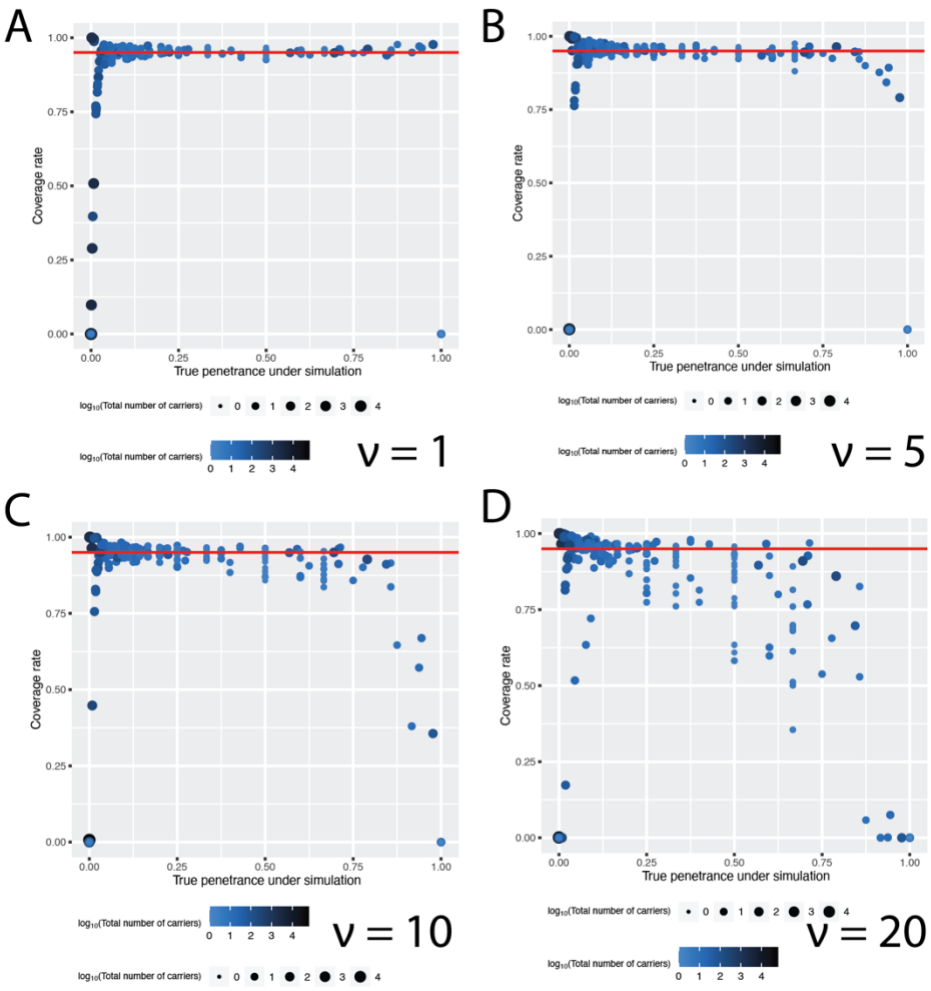

**Supplementary Figure 2.** Coverage plots for *KCNQ1*-LQT1 model.

A) Tuning parameter of 1 is too stringent – more than 95% of values are above the 0.95 coverage rate.

B) Calibrated tuning parameter of 10 – approximately 95% of variants have a 0.95 coverage rate.

C) Overly permissible tuning parameter of 20 – less than 95% of variants have a coverage rate less than 0.95.

D) Overly permissible tuning parameter of 50 – approximately half of variants have a coverage rate less than 0.95.




**Supplementary Figure 3.** Coverage plots for *SCN5A*-LQT3 model.

A) Tuning parameter of 1 is too stringent – more than 95% of values are above the 0.95 coverage rate.

B) Calibrated tuning parameter of 5 – approximately 95% of variants have a 0.95 coverage rate.

C) Overly permissible tuning parameter of 10 – less than 95% of variants have a coverage rate less than 0.95.

D) Overly permissible tuning parameter of 20 – approximately half of variants have a coverage rate less than 0.95.

## Supplemental References


1 Schwartz, P. J., Crotti, L. & Insolia, R. Long-QT syndrome: from genetics to management. *Circ Arrhythm Electrophysiol* **5**, 868-877 (2012). https://doi.org:10.1161/circep.111.962019

2 Priori, S. G. *et al.* HRS/EHRA/APHRS expert consensus statement on the diagnosis and management of patients with inherited primary arrhythmia syndromes: document endorsed by HRS, EHRA, and APHRS in May 2013 and by ACCF, AHA, PACES, and AEPC in June 2013. *Heart Rhythm* **10**, 1932-1963 (2013). https://doi.org:10.1016/j.hrthm.2013.05.014

3 Kozek, K. *et al.* Estimating the Post-Test Probability of Long QT Syndrome Diagnosis for Rare KCNH2 Variants. *Circ Genom Precis Med* (2021). https://doi.org:10.1161/circgen.120.003289

4 Kroncke, B. M. *et al.* A Bayesian method to estimate variant-induced disease penetrance. *PLoS Genet* **16**, e1008862 (2020). https://doi.org:10.1371/journal.pgen.1008862

5 Kroncke, B. M. *et al.* Protein structure aids predicting functional perturbation of missense variants in SCN5A and KCNQ1. *Comput Struct Biotechnol J* **17**, 206-214 (2019). https://doi.org:10.1016/j.csbj.2019.01.008

6 Glazer, A. M. *et al.* Arrhythmia variant associations and reclassifications in the eMERGE-III sequencing study. *medRxiv*, 2021.2003.2030.21254549 (2021). https://doi.org:10.1101/2021.03.30.21254549

7 Vanoye, C. G. *et al.* High-Throughput Functional Evaluation of KCNQ1 Decrypts Variants of Unknown Significance. *Circ Genom Precis Med* **11**, e002345 (2018). https://doi.org:10.1161/circgen.118.002345

8 Sun, J. & MacKinnon, R. Cryo-EM Structure of a KCNQ1/CaM Complex Reveals Insights into Congenital Long QT Syndrome. *Cell* **169**, 1042-1050.e1049 (2017). https://doi.org:10.1016/j.cell.2017.05.019

9 Zhang, X. *et al.* Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. *Genet Med* **23**, 69-79 (2021). https://doi.org:10.1038/s41436-020-00972-3

10 Adzhubei, I. A. *et al.* in *Nat Methods* Vol. 7 248-249 (2010).

11 Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745-2747 (2015). https://doi.org:10.1093/bioinformatics/btv195

12 Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877-885 (2016). https://doi.org:10.1016/j.ajhg.2016.08.016