Published in final edited form as:

J Cardiovasc Transl Res. 2015 April ; 8(3): 187–197. doi:10.1007/s12265-015-9622-8.

Enhancing the Predictive Power of Mutations in the C-Terminus of the *KCNQ1*-Encoded Kv7.1 Voltage-Gated Potassium Channel

Jamie D. Kapplinger, BA^{1,2,*}, Andrew S. Tseng, BA/BS^{1,*}, Benjamin A. Salisbury, PhD³, David J. Tester, BS^{2,4}, Thomas E. Callis, PhD⁵, Marielle Alders, PhD⁶, Arthur A.M. Wilde, MD, PhD^{7,8}, and Michael J. Ackerman, MD, PhD^{1,2,4,9}

¹Mayo Medical School, Mayo Clinic, Rochester, Minnesota, United States ²Department of Molecular Pharmacology & Experimental Therapeutics; Windland Smith Rice Sudden Death Genomics Laboratory, Mayo Clinic, Rochester, Minnesota, United States ³Knome, Inc., Cambridge, Massachusetts, United States ⁴Department of Medicine, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota, United States ⁵Transgenomic Inc., New Haven, Connecticut, United States ⁶Department of Clinical Genetics, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands ⁷Heart Centre, Department of Clinical and Experimental Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands ⁸Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders, Jeddah, Kingdom of Saudi Arabia ⁹Department of Pediatrics, Division of Pediatric Cardiology, Mayo Clinic, Rochester, Minnesota, United States

Abstract

Despite the overrepresentation of Kv7.1 mutations among patients with a robust diagnosis of LQTS, a background rate of innocuous Kv7.1 missense variants observed in healthy controls creates ambiguity in the interpretation of LQTS genetic test results. A recent study showed the probability of pathogenicity for rare missense mutations depends in part on the topological location of the variant in Kv7.1's various structure-function domains. Since the Kv7.1 C-terminus accounts for nearly 50% of the overall protein and nearly 50% of the overall background rate of rare variants falls within the C-terminus, further enhancement in mutation calling may provide guidance in distinguishing pathogenic LQT1-causing mutations from non-disease causing rare variants in Kv7.1's C-terminus. Therefore, we have used conservation analysis and a large case/control study to generate topology-based estimative predictive values to aid in interpretation; identifying three regions of high conservation within the Kv7.1 C-terminus which have a high probability of LQT1 pathogenicity.

Keywords

conservation analysis; estimated predictive value; KCNQ1 (Kv7.1); long QT syndrome

Address for correspondence: Michael J. Ackerman, MD, PhD., Mayo Clinic Windland Smith Rice Sudden Death Genomics Laboratory, Guggenheim 501, Mayo Clinic, Rochester, MN 55905, 507-284-0101 (phone), 507-284-3757 (fax), ackerman.michael@mayo.edu.

*Contributed equally and are co-equal first authors

INTRODUCTION

Congenital long QT syndrome (LQTS) stems from heritable defects in the repolarization of the myocardium leading to a prolonged QT interval. LQTS has an estimated prevalence of 1 in 2000 and carriers an increased risk for potentially fatal arrhythmias. Since the sentinel discovery of *KCNQ1* mutations in the first LQTS-susceptibility locus (long QT syndrome type 1; LQT1),[1] hundreds of mutations have been identified in *KCNQ1* and account for 35-40% of LQTS cases (LQT1 prevalence = ~1:5000).[2] *KCNQ1* encodes for the Kv7.1 voltage-gated potassium channel alpha subunit responsible for the slow activating late repolarizing potassium current in the human heart.

Despite the overrepresentation of Kv7.1 mutations among patients with a clinically robust diagnosis of LQTS, a background rate of likely innocuous rare Kv7.1 missense variants observed in ostensibly healthy controls creates ambiguity in the interpretation of LQTS genetic test results.[3,4] A recent study has shown that the probability of pathogenicity for rare missense variants depends in part on the topological location of the variant in Kv7.1's various structure-function domains, with high estimated predictive values (EPVs) assigned for variants localizing to Kv7.1's transmembrane and pore-forming regions.[4] Importantly, since the C-terminus (defined here as the entirety of the protein beyond the end of the 6th transmembrane segment) accounts for nearly 50% of the overall Kv7.1 protein (> 300 amino acids) and since nearly 50% of the identified overall background rate of rare variants falls within the C-terminus of Kv7.1,^{4, 5} further enhancement in mutation calling efforts may provide guidance in distinguishing pathogenic LQT1-causing mutations from non-disease causing rare variants localizing to Kv7.1's C-terminus.

Given the ever increasing use of clinical-based whole exome sequencing and the recent mutation reporting guidelines from the American College of Medical Genetics[5], there will be an increasing number of incidental findings of rare *KCNQ1* genetic variants that are reported, which could lead subsequently to an overzealous increase in incorrect LQTS diagnosis and unwarranted prophylactic treatment with beta-blocker therapy and/or internal cardioverter defibrillator implantation. Therefore, it is paramount to provide physicians with enhanced tools to assist in better mutation calling efforts so the best genetic testing-based decisions in medical care can be made.

Previously, *in silico* tools, which largely rely on conservation comparisons, have been utilized to improve the EPV (i.e. estimated predictive value or probability of pathogenicity) of genetic variants in LQTS. However, it was identified that topology largely superseded the *in silico* predictions, suggesting that the tools were simply identifying the mutations that fell into functional regions.[6] As recent studies have identified four helical regions (A-D, Figure 1) within the C-terminus that play a critical role in Kv7.1's tetramerization, autonomic regulation of the channel, and subunit binding, the goal of the present study was to incorporate the analyses of these newly elucidated C-terminal functional regions and phylogenetic amino acid conservation in an effort to increase the positive and negative predictive power for rare genetic variants that localize to Kv7.1's C-terminus.

METHODS

Study Populations

In order to identify the amino acid regions of the Kv7.1 C-terminus that may be critical in the pathogenesis of LQTS, two compendia of *KCNQ1* single nucleotide variants resulting in missense changes (the exchange of one amino acid for another) were compiled from publically available databases or the literature (Supplemental Table 1). Presumably benign control variants were defined as missense variants that were identified among two publically available exome sequencing databases (the 1000 Genome Project [1kG, n=1092 individuals, <http://www.1000genomes.org/>][7] and the National Heart, Lung, and Blood Institute Exome Sequencing Project [ESP, n=6503, <http://evs.gs.washington.edu/EVS/>]) and 1344 *KCNQ1* (Kv7.1) Sanger-sequenced “in-house” controls. In contrast, putative LQT1-cases associated mutations were defined as missense variants that were identified in LQTS cases from the literature and were completely absent in the two publically available exome databases and our 1344 “in-house” controls. This strict definition of a putative LQT1-associated missense mutation was used in order to polarize the two groups of missense variants for our case-control comparative analysis. All missense variants were named at the nucleotide level using NM_000218.2 and at the protein level using NP_000209.2 according to standard Human Genome Variation Society (HGVS) nomenclature.

Both control variants and LQT1 mutations were mapped onto the Kv7.1 C-terminus sequence obtained from Uniprot (P51787). Additionally, four predicted functional domains corresponding to known and predicted alpha helical domains based on previous studies[8] (Helix A: 370-389, Helix B: 506-532, Helix C: 548-562, Helix D [which also corresponds with the Kv7.1's C-terminal subunit assembly domain (SAD)]: 588-622) were analyzed to determine whether control variants or LQT1-associated mutations differentially clustered to any of these particular domains.

Kv7.1 Ortholog and Paralog Conservation Analysis

For inter-species (ortholog) conservation analysis, the University of California Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu/>) alignment of 43 species to the primary human Kv7.1 sequences was used. Additionally, the sequences of the five human paralogs from the Kv7 family were obtained from UniProt and aligned using ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). See the Supplemental Methods for further alignment details.

A moving window was used to plot the average percent conservation separately among the 43 orthologs and the five paralogs in 12 amino acid increments for the 328 amino acids of Kv7.1's C-terminus. Distinct 12 amino acid windows hosting an average conservation greater than the average conservation across the entire 328 amino acid segment of the Kv7.1 C-terminus for both the orthologs and paralogs were identified.

In Silico Pathogenicity Predictions

In order to assess whether *in silico* tools could enhance mutation interpretation within the C-terminus, ortholog conservation, Grantham values, SIFT, PolyPhen2, MutPred, and KvSNP

were utilized using default settings. For each *in silico* tool, the sensitivity was calculated as the number of case mutations predicted pathogenic over the total number of case mutations and the specificity was calculated as the number of control variants predicted benign over the total number of control variants. In order to compare the *in silico* tools, the Matthews Correlation Coefficient (MCC) was calculated for each tool according to the following equation:

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

where TP = number of case mutations predicted pathogenic, TN = number of control variants predicted benign, FP = number of control variants predicted pathogenic, and FN = number of case mutations predicted benign.

Estimated Predictive Values

In order to provide an estimate of the disease likelihood for an identified missense variant, estimated predictive values (EPVs) were calculated as previously described[4] and further detailed in the Supplemental Methods.

As LQT1 has an estimated prevalence of ~1:5000, of which, ~45% is C-terminus-mediated. Therefore, the prevalence of C-terminus-mediated LQT1 would be ~1:10,000. In an effort to account for the potential for reduced penetrance, the EPV calculations were based on only those rare control variants identified with a frequency less than 0.02% (1:5000) across all 8939 unrelated control subjects examined. The rare control variant status is not meant to imply pathogenicity nor non-pathogenicity but rather to reflect that had these missense variants been identified during a clinical LQTS genetic test, they would have been considered a variant of uncertain significance (VUS) or possibly pathogenic. As the EPV was calculated based on variant frequency among cases or controls, only case variants identified in cohorts with a known number of cases were used [LQTS case series from Tester *et al.*[9] (n = 541) and Kapplinger *et al.*[10] (n = 2500)].

Statistical Analysis

A Fisher's exact test was used to identify regions with an overrepresentation of LQT1 mutations by comparing the number of case mutations inside and outside a given region vs. control variants inside and outside the same region. Additionally, the Fisher's exact test was used to determine whether case or control variants were more likely to affect the amino acids inside or outside of these regions by comparing the number of amino acids in a given region affected by case or control variants over all the amino acids in the region vs. the amino acids affected by a case or control variant outside the given region over all of the amino acids in that region. The average conservation was reported as average $\pm$ standard error. A student t-test was used to compare the average percent conservation between regions.

RESULTS

Localization of LQT1-Associated Missense Mutations and Control Missense Variants in the Helical Domains

To determine whether the previously identified helical regions may enhance interpretation of variants within the Kv7.1 C-terminus, 82 LQT1-associated mutations from the literature and 31 control variants residing within the C-terminus were identified (Supplemental Table 1). The amino acid residue boundaries for known and predicted functional domains were previously identified as Helix A (amino acids 370-389), Helix B (507-532), Helix C (548-562) and Helix D (588-622), where Helix D corresponds to the known subunit assembly domain (SAD).[8]

Highlighting the importance of these four helices, the amino acids within these helices were statistically more likely to host a case mutation, with 27.8% (27/97) of the amino acids in the helices hosting a case mutation, while only 36 of the 231 (15.6%) amino acids outside the helices hosted a case mutation ($p = 0.01$). Despite this, only 35/82 (43%) LQT1 mutations were identified within the helical domains, while these helices also hosted 8/31 control variants (26%, $p = 0.13$). The identification of a large number of LQT1 mutations outside of the helical regions would suggest the existence of additional critical functional domains outside of these established helical domains.

Identification of Highly Conserved Regions within the Kv7.1 C-terminus

Conservation has been well known to correlate with regions critical to a protein's function. Therefore, conservation of the Kv7.1 C-terminus was assessed, using a 12-amino acid moving window analysis, in order to identify potential critical domains residing outside of the four known helices. The moving window analysis showed that the conservation level was not homogenous throughout the C-terminus (Figure 2), but rather showed regions of high conservation and regions of strikingly low conservation.

In order to identify regions critical to the overall channel function, regions with high conservation were identified as the amino acids within moving windows with an average conservation (solid lines, Figure 2) greater than the average conservation across the entire C-terminus (Ortholog: 85.4%, Paralog: 22.5%, dashed lines, Figure 2) in both paralogs and orthologs. Using this definition, three regions (amino acid 349-391, 509-575, and 585-607) in the Kv7.1 C-terminus were identified. As expected, the average conservation for the 133 amino acids within these three windows was substantially higher than outside of these regions, especially among the paralogs (Orthologs: $94.4\% \pm 0.9\%$ compared to $79.3\% \pm 1.7\%$; Paralogs: $44.0\% \pm 4.1\%$ compared to $7.8\% \pm 1.4\%$). Interestingly, the average percent conservation for amino acids within the three identified regions was indistinguishable statistically from the average across the transmembrane region (Orthologs: $95.6\% \pm 0.6\%$ [$p = \text{ns}$]; Paralogs: $54.4\% \pm 3.0\%$ [$p = \text{ns}$], Figure 3 and Table 1), suggesting that these regions may be as vital to the overall functional of the channel as the transmembrane-spanning regions.

Not unexpectedly, the identified highly conserved regions closely correlated with the previously reported helical domains. All of helix A and C and the majority of helix B fall

within the identified conserved regions (Figure 4). Interestingly, only ~50% of helix D (SAD) was within the highly conserved region. Despite the majority of the known helical domains being within the highly conserved region, over 40% of the highly conserved regions fell outside of these helical domains, further supporting that other yet defined critical functional domains within these conserved regions exist.

Finally, case mutations had a clear preponderance for the identified regions, with 50 of the 133 amino acids (37.6%) in the highly conserved regions hosting a case mutation, while only 6.7% (13/195) of the amino acids outside the highly conserved regions hosted a case mutation ($p = 5.0 \times 10^{-12}$, Table 1).

Given the lack of phenotypic data for a large subset of the cases examined, we performed a subset analysis examining the location of the variants identified in 388 “clinically definite” LQTS cases.[4] Among the rare variants identified in these cases, 90% (18/20) were identified within the conserved regions. Interestingly, the only variant among these cases identified within the controls, found in 5 controls, resided outside of the conserved regions and was identified in a case hosting a likely pathogenic mutation in *KCNH2*.

Furthermore, when examining 45 *KCNQ1* missense variants from the literature with functional characterization, the highly conserved region correlated with the functional status of the variants; as 73% (29/40) of the functionally abnormal variants resided within these highly conserved regions while only 1 of the 5 variants with wildtype functional properties localized to this region ($p = 0.035$).

While we cannot definitely label the control variants as benign, there is a paucity of control variants in these identified regions as only 3.8% (5/133) of the amino acids within the conserved regions hosted a control variant whereas 12.3% (24/195, $p = 0.009$) of the amino acids outside the highly conserved region hosted a control variant. This contrast is much greater than was observed for the four helical domains, suggesting the helical domains do not adequately represent the true critical regions within the C-terminus with respect to potential for LQT1 pathogenicity. Further highlighting the conserved regions’ role in the pathogenesis of LQTS, 68/82 (83%) of the LQT1 mutations resided within these highly conserved regions while only 5/31 control variants (16%, $p = 6.3 \times 10^{-11}$) resided within these regions (Figure 4 and Table 1).

Given the relatively low number of control variants identified in our control cohort, we additionally examined the 109 *KCNQ1* C-terminus variants identified from the recently released Exome Aggregation Consortium (Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: <http://exac.broadinstitute.org>) [date (March, 2015) accessed]. With only 33% (36/109) of the ExAC variants residing within one of these highly conserved regions, the localization of the ExAC variants was indistinguishable from that of the control variants from this study ($p = 0.08$), however it is important to note that the ESP and 1kG samples are included in the ExAC samples. Even with this larger dataset, ExAC-derived control variants were far less likely to localize to one of these highly conserved regions than the case-derived variants ($p = 4.7 \times 10^{-12}$).

Comparison of *In Silico* Tools for their Ability to Distinguish LQT1 Mutations from Control Variants

Pathogenicity predictions using the identified highly conserved regions from this study were compared to five highly utilized *in silico* predictive algorithms, one of which was designed specifically for the Kv channels (KvSNP). All the *in silico* algorithms were able to statistically distinguish case mutations from control variants (Figure 5). For example, orthologs conservation predicted 61% (50/82) of case mutation as pathogenic, but only predicted 13% (4/27) control variants as pathogenic ($p = 4.8 \times 10^{-7}$). The pathogenic predictions for case mutations and control variants from each tool are shown in Figure 5.

While each prediction (except the use of the C-terminus helices) was able to statistically distinguish case mutations from control variants, the use of the highly conserved regions was superior when comparing the sensitivity and specificity of each tool as compared by the Matthews Correlation Coefficient (MCC). For example, the use of the highly conserved regions to establish predictive pathogenicity resulted in a sensitivity of 82.9% and a specificity of 83.9%, giving the highest MCC of 62.3%. The pathogenic predictions, sensitivity, specificity, and MCCs for each prediction algorithm can be seen in Table 2.

Mutations in Identified Conserved Regions Have Greater Estimated Predictive Values (EPVs) than do Mutations Localizing Outside of These Regions

In an effort to provide physicians with a framework to interpret the next missense variant identified within the C-terminus of Kv7.1, EPVs were calculated. In this study, the overall EPV for missense variants within the entire C-terminus was calculated to be 97% (95-98%). This would suggest a high probability of pathogenicity for all missense variants in the C-terminus. However, we have shown a non-random clustering of LQT1 mutations to highly conserved regions within the C-terminus. Therefore, we compared the EPV for variants residing within one of these highly conserved regions to those localizing outside of these regions. The EPV for the highly conserved regions was 99% [98-100%].

Conversely, the EPV for missense variants outside the highly conserved regions fell to 66% (22-85%). This would suggest a high probability of pathogenicity for missense variants identified within the conserved region while a great deal of caution must be applied to those missense variants falling outside these three identified conserved regions. Interestingly, when assessing variants based upon the previously identified helical domains within the C-terminus, the EPVs are not polarized: the EPV within the helical regions was calculated to be 98% (95-99%) and outside the helices it was 96% (93-98%). This highlights the fundamental importance of the newly identified highly conserved regions from this study in the interpretation of missense variants within the C-terminus of Kv7.1.

Additionally, as previous studies have suggested enhanced predictions can be achieved when synergistically using predictions from multiple *in silico* algorithms, the ability of *in silico* algorithms to enhance even further the interpretation inside or outside the identified highly conserved regions was assessed. Despite all seven *in silico* tools being able to distinguish LQT1 mutations from control variants when assessing the entire C-terminus, when the variants were assessed based on their location inside or outside the identified highly

conserved regions, none of the tools were able to statistically distinguish case mutations from control variants in both regions. Looking at PolyPhen2 as an example, 94.1% (64/68) case mutations within the highly conserved regions were predicted pathogenic, however 60% of the control variants were also predicted pathogenic (3/5) ($p > 0.05$) within this region. Additionally, outside of the highly conserved regions, PolyPhen2 only predicted 28.6% (4/14) LQTS mutations as pathogenic, while predicting 46.2% (12/26) control variants as pathogenic ($p > 0.05$). The pathogenicity predictions for each tool inside and outside the identified conserved regions are provided (Table 3). While no tool reached statistical significance in both regions, all the tools scored more cases mutations as pathogenic than control mutations in both regions (except PolyPhen2 outside the highly conserved regions) suggesting that there potentially could be added benefit, but this study may be limited by the small number when examining the subset regions.

DISCUSSION

Issues of Interpretation in LQTS Genetic Testing

With the expanding use of clinical genetic testing and the introduction of clinical whole exome/genome sequencing, the clinical interpretation of genetic tests is becoming more and more complex. The ever-increasing amount of genetic data has highlighted that genetic test interpretation is not simply completed with the identification of a rare variant in an established disease-susceptibility gene, especially given recent studies identifying a background rate of rare variation in apparently healthy individuals leading to a level of ambiguity in genetic test interpretation.

Highlighting this issue, a recent study identified a high prevalence of genetic variants previously associated with LQTS within exome data from large population studies, casting doubt on their previous annotation as LQT1-causative mutations.[11] Because LQTS genetic testing provides diagnostic, prognostic, and therapeutic implications, the correct interpretation of an identified variant is critical.[12] However, the level of ambiguity presents a challenge to the referring physician, which has led to a number of studies identifying methods to distinguish pathogenic mutations from benign “genetic noise” within genes associated with LQTS.[4,6,13]

While the “gold-standard” for mutation interpretation continues to be functional studies examining mutant’s electrophysiology properties, these studies are expensive and time consuming requiring extensive expertise. Therefore, the utilization of interpretive algorithms, harnessing the identification of critical functional regions or *in silico* predictions, has gained favor due to their ease of use. For *KCNQ1*, topological and *in silico* analyses have been used previously in an effort to improve EPVs for variants in *KCNQ1*. [6,4,13] While previous topology studies have suggested the importance of the entire C-terminus of Kv7.1 providing an EPV of 95% (0.95CI: 89-98%)[4], the C-terminus of the Kv7.1 channel spans roughly half of the entire protein and also hosts nearly half of the genetic noise within the gene. Attempts to further polarize the EPVs within the C-terminus with *in silico* and physiochemical analyses have been attempted. However, it was identified that topology superseded the *in silico* predictions as the *in silico* tools were unable to distinguish case from

control variants when assessing variants within topological regions.[6] Therefore, improvements need to be made to increase the analytical resolution of this region.

Identification of Highly Conserved Regions

As topology has been shown previously to be critical in the interpretation of variants, we assessed known functional regions within the C-terminus of Kv7.1. Studies have identified four helical regions involved in protein trafficking, subunit assembly, as well as interactions with the β -subunit KCNE1 and calmodulin.[14-16] Despite the critical role of these regions for the overall channel function, a large fraction (57%) of case variants fell outside of these four helices suggesting there may be additional regions within the C-terminus playing a key role in the function of the channel.

Therefore, through an analysis of conservation among orthologs and paralogs, we identified three highly conserved regions in Kv7.1's C-terminus. These regions had a conservation level comparable to the critical transmembrane-spanning regions, hosted 83% of the case variants identified in this study, and were associated closely with the four previously identified helices. Although there was a close association with the known helical regions, there were some interesting differences. Regions showing high level of conservation and hosting a number of case variants were identified outside the helical regions, including the regions between the 6th transmembrane segment and helix A, between helix B and C, and a large portion of the region between helix C and D (Figure 4). This would suggest a critical role for these regions in the overall assembly or function of the channel. Interestingly, only the proximal portion of the SAD was included in the highly conserved region, suggesting either a limitation to the conservation analysis or a less critical requirement for the distal portion of the SAD, more likely the latter given that LQT1 mutations were much more common in the highly conserved portion of the SAD. Overall, the identification of the identified highly conserved regions can guide future structural/functional studies.

Impact on Interpretation of Variants of Uncertain Significance within the Kv7.1 C-Terminus

In an effort to improve interpretation within the C-terminus, we assessed the EPVs inside and outside of the identified conserved regions. The EPV was increased to 99% (98-100%) within the three identified conserved regions, while the EPV for variants localizing outside of these three regions has fallen significantly (66% [22% - 85%]). These EPVs mean that the next novel rare variant identified within the Kv7.1 C-terminus localizing in the highly conserved regions can be more reliably predicted to be pathogenic, while similarly rare variants identified outside of these regions need to be interpreted more cautiously as they remain stuck in that nebulous VUS category, which we have coined as genetic purgatory.

While this discovery is important for the interpretation of cases suspected of having LQTS, it is equally important and perhaps even more so for the emerging issue of incidental findings from clinical whole genome/exome sequencing. In fact, recently the American College of Medical Genetics has recommended the reporting of incidental findings of *KCNQ1* mutations even when in cases without any previous indication of LQTS. The reporting of such incidental findings certainly may identify the 1 in 2000 individuals who truly have undiagnosed LQTS and will benefit from potentially life-saving prophylactic

treatment and avoidance of QT-prolonging medication. However, there is a stronger possibility that incidental findings may result in not only an overzealous misdiagnosis of LQTS but also the possible initiation of unnecessary and excessive therapies, including ICD implantation.

Prior to this study, any VUS identified in the Kv7.1's C-terminus could have been interpreted as carrying a high likelihood of pathogenicity (EPV>90) when found in a suspected LQTS patient.[4] However, this study now shows the majority of the genetic noise within the C-terminus occurs outside of the identified conserved regions (84% of control variants fall outside the three conserved regions). The low pathogenicity prediction for this region (EPV = 66%) suggests that one must be more critical of the role a variant in this region may play in the disease, and that without a clinical evaluation supporting the diagnosis, the variant may not be disease causing.

In support of the lower pathogenic prediction outside of the identified conserved regions, some case variants within the large loop between helices A and B were reviewed and were identified to have low evidence for pathogenicity. For example, KCNQ1 p.V417M was identified in a case that also hosted the KCNQ1 mutation p.V254M. Following functional characterization, it was identified that the helical A to helical B loop where the p.V417M variant resided had negligible changes in the electrophysiological properties of the channel while the transmembrane localizing p.V254M variant resulted in a significant loss of current.[17] Additionally, three other case variants (p.D446E, p.P448L, and p.R451W) found within the A to B loop were identified in cases hosting a second mutation that was either within the transmembrane region or previously shown to have an abnormal electrophysiological properties. This suggests that these four A to B loop variants, previously published as possible LQT1 mutations, are likely false positives or at most modifiers. It is important to note that while we highlight these four variants outside the critical region, we did not do a systematic analysis of all case variants inside and outside the highly conserved region due to limitation in the published data. Therefore, there is a possibility that cases hosting a variant that localizes to one of these highly conserved regions may also host other pathogenic variants outside the Kv7.1 C-terminus or outside Kv7.1 altogether.

***In Silico* Predictions for *KCNQ1* Interpretation**

The use of *in silico* tools to enhance variant interpretation has gained traction as these tools require limited expertise and provide immediate predictions. Despite these advantages, very few of the tools have been vetted for particular diseases, but rather have been shown to distinguish case from control variants using large datasets of variants across a broad spectrum of disorders. This broad characterization may have limitations for particular genes or regions despite showing a statistical ability across the broad spectrum of mutations. This has been demonstrated for disease-susceptibility genes for hypertrophic cardiomyopathy (HCM) where *in silico* algorithms were only advantageous in the two major genes, *MYBPC3* and *MYH7*. [18] Additionally, *in silico* tools have shown limited utility when assessing sub-regions within *KCNH2* and *KCNQ1* for LQTS.⁶

While we have shown that *in silico* tools are able to distinguish case mutations from control variants within the Kv7.1 C-terminus, the use of these newly identified, highly conserved regions (amino acid 349-391, 509-575, and 585-607) to assign pathogenic predictions provided a stronger predictive output when comparing sensitivity and specificity. Additionally, we highlight the inability of the *in silico* tools to enhance predictions when assessing variants inside and outside of these three highly conserved regions.

Importantly, the interpretation of a genetic test is never a black and white, binary interpretation. Instead, the genetic test must be interpreted in the context of the clinical picture, the mutation involved, and the pathogenicity predictions. Recently, an algorithm has been proposed which uses a Bayesian model to synergistically use multiple predictive methods to enhance interpretation.[19] To that end, we have identified three highly conserved regions (amino acid 349-391, 509-575, and 585-607) within the C-terminus of Kv7.1, which strongly correspond with putative case mutations and previously functionally characterized variants, and could be added quickly and easily to such a model to further improve variant interpretation.

LIMITATIONS

Conservation has been used widely to identify regions critical to the overall function of a particular protein. However, there are limitations to the use of conservation to identify all critical regions. While conservation across orthologs or paralogs can be indicative of a region's necessity for the function of a protein, absence of conservation does not necessarily indicate a region has no role in the protein's function. Nevertheless, our approach has identified regions with likely critical function that can enhance interpretation of VUSs within Kv7.1's C-terminus.

While we classify any variant identified in the controls as benign, there may be control variants that are pathogenic. However, given an estimated prevalence of Kv7.1 C-terminus-mediated LQT1 at ~1:10,000, the vast majority of the Kv7.1 C-terminus variants identified in the controls can be assumed to have little effect individually on the LQTS phenotype. This assumption is further supported by the decreased pathogenic predictions of the *in silico* tools, as well as the clear separation between case and control variant location with the C-terminus. While many LQTS-causative mutations are associated with decreased penetrance, given the background variation in the major genes of LQTS, the penetrance of the control variants would have to be less than 5%, making the interpretation nearly impossible.[11]

Further, if some of the control variants are in fact pathogenic, the EPVs would be underestimated, making our resultant EPVs a conservative assumption. Conversely, it is also possible that given the inability to adjudicate each case with a putative LQT1-associated mutation as a robust, definite case of LQTS, it is conceivable that some of the case variants may in fact be benign, as we have highlighted in the discussion. By definition, the *KCNQ1* background rate defines that there will be a certain presence of benign genetic noise in any cohort of samples tested for *KCNQ1*. While the presence of benign variants within the case samples cannot be avoided, the EPV calculation attempts to account for the background genetic noise and the inclusion of benign variants in the cases should make the EPVs

conservative estimates. As the inherent issue of potential pathogenic variants in controls and benign variants result in conservative estimates, our identification of regions that provide such a strong separation would only be enhanced with the removal of these confounding variants.

CONCLUSION

We have identified three highly conserved regions in the C-terminus of *KCNQ1*-encoded Kv7.1 channel (amino acid 349-391, 509-575, and 585-607) with significantly greater EPVs. This finding is further supported by functional and structural analyses, showing a significant association among conservation, structure, and pathogenicity. The methods utilized in this study can be readily applied to other genetic syndromes to aide in the interpretation of genetic tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

DISCLOSURES

J.D.K. is supported by the NIH grant GM72474-08 and thanks the Mayo Clinic MSTP for fostering an outstanding environment for physician-scientist training. This project was supported by the Mayo Clinic Windland Smith Rice Comprehensive Sudden Cardiac Death Program (M.J.A.). We acknowledge the support from the Netherlands CardioVascular Research Initiative (CVON-PREDICT project): the Dutch Heart Foundation, Dutch Federation of University Medical Centres, the Netherlands Organisation for Health Research and Development and the Royal Netherlands Academy of Sciences (A.A.M.W.). T.E.C. is an employee of Transgenomic Inc.. B.A.S. is an employee of Knome, Inc.. M.J.A. is a consultant for Boston Scientific, Gilead Sciences, Medtronic, St. Jude Medical, Inc., and Transgenomic. Intellectual property derived from M.J.A.'s research program resulted in license agreements in 2004 between Mayo Clinic Ventures (formerly Mayo Medical Ventures) and Genaissance Pharmaceuticals (now Transgenomic) with respect to their FAMILION-LQTS and FAMILION-CPVT genetic tests.

References

1. Wang Q, Curran ME, Splawski I, Burn TC, Millholland JM, VanRaay TJ, Shen J, Timothy KW, Vincent GM, de Jager T, Schwartz PJ, Toubin JA, Moss AJ, Atkinson DL, Landes GM, Connors TD, Keating MT. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat Genet.* 1996; 12(1):17–23. [PubMed: 8528244]
2. Perrin MJ, Gollob MH. Genetics of cardiac electrical disease. *The Canadian journal of cardiology.* 2013; 29(1):89–99. DOI: 10.1016/j.cjca.2012.07.847 [PubMed: 23062665]
3. Ackerman MJ. Cardiac channelopathies: it's in the genes. *Nature medicine.* 2004; 10(5):463–464. DOI: 10.1038/nm0504-463
4. Kapa S, Tester DJ, Salisbury BA, Harris-Kerr C, Pungliya MS, Alders M, Wilde AA, Ackerman MJ. Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. *Circulation.* 2009; 120(18):1752–1760. [PubMed: 19841300]
5. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2013; 15(7): 565–574. DOI: 10.1038/gim.2013.73 [PubMed: 23788249]
6. Giudicessi JR, Kapplinger JD, Tester DJ, Alders M, Salisbury BA, Wilde AAM, Ackerman MJ. Phylogenetic and physicochemical analyses enhance the classification of rare nonsynonymous

single nucleotide variants in type 1 and 2 long-QT syndrome. *Circulationcardiovascular Genetics*. 2012; 5(5):519–528.

7. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. DOI: 10.1038/nature11632 [PubMed: 23128226]
8. Wiener R, Haitin Y, Shamgar L, Fernandez-Alonso MC, Martos A, Chomsky-Hecht O, Rivas G, Attali B, Hirsch JA. The KCNQ1 (Kv7.1) COOH terminus, a multitiered scaffold for subunit assembly and protein interaction. *J Biol Chem*. 2008; 283(9):5815–5830. DOI: 10.1074/jbc.M707541200 [PubMed: 18165683]
9. Tester DJ, Will ML, Haglund CM, Ackerman MJ. Compendium of cardiac channel mutations in 541 consecutive unrelated patients referred for long QT syndrome genetic testing. *Heart Rhythm*. 2005; 2(5):507–517. [PubMed: 15840476]
10. Kapplinger JD, Tester DJ, Salisbury BA, Carr JL, Harris-Kerr C, Pollevick GD, Wilde AA, Ackerman MJ. Spectrum and prevalence of mutations from the first 2,500 consecutive unrelated patients referred for the FAMILION long QT syndrome genetic test. *Heart Rhythm*. 2009; 6(9): 1297–1303. [PubMed: 19716085]
11. Refsgaard L, Holst AG, Sadjadieh G, Haunso S, Nielsen JB, Olesen MS. High prevalence of genetic variants previously associated with LQT syndrome in new exome data. *European journal of human genetics* : EJHG. 2012; 20(8):905–908. DOI: 10.1038/ejhg.2012.23 [PubMed: 22378279]
12. Giudicessi JR, Ackerman MJ. Genotype- and phenotype-guided management of congenital long QT syndrome. *Curr Probl Cardiol*. 2013; 38(10):417–455. DOI: 10.1016/j.cpcardiol.2013.08.001 [PubMed: 24093767]
13. Ware JS, Walsh R, Cunningham F, Birney E, Cook SA. Paralogous annotation of disease-causing variants in long QT syndrome genes. *Hum Mutat*. 2012; 33(8):1188–1191. DOI: 10.1002/humu.22114 [PubMed: 22581653]
14. Zheng R, Thompson K, Obeng-Gyimah E, Alessi D, Chen J, Cheng H, McDonald TV. Analysis of the interactions between the C-terminal cytoplasmic domains of KCNQ1 and KCNE1 channel subunits. *Biochem J*. 2010; 428(1):75–84. DOI: 10.1042/BJ20090977 [PubMed: 20196769]
15. Schmitt N, Calloe K, Nielsen NH, Buschmann M, Speckmann EJ, Schulze-Bahr E, Schwarz M. The novel C-terminal KCNQ1 mutation M520R alters protein trafficking. *Biochemical and Biophysical Research Communications*. 2007; 358(1):304–310. DOI: 10.1016/j.bbrc.2007.04.127 [PubMed: 17482572]
16. Sato A, Arimura T, Makita N, Ishikawa T, Aizawa Y, Ushinohama H, Kimura A. Novel mechanisms of trafficking defect caused by KCNQ1 mutations found in long QT syndrome. *J Biol Chem*. 2009; 284(50):35122–35133. DOI: 10.1074/jbc.M109.017293 [PubMed: 19825999]
17. Wedekind H, Schwarz M, Hauenschild S, Djonlagic H, Haverkamp W, Breithardt G, Wulffing T, Pongs O, Isbrandt D, Schulze-Bahr E. Effective long-term control of cardiac events with beta-blockers in a family with a common LQT1 mutation. *Clin Genet*. 2004; 65(3):233–241. [PubMed: 14756674]
18. Kapplinger JD, Landstrom AP, Bos JM, Salisbury BA, Callis TE, Ackerman MJ. Distinguishing hypertrophic cardiomyopathy-associated mutations from background genetic noise. *J Cardiovasc Transl Res*. 2014; 7(3):347–361. DOI: 10.1007/s12265-014-9542-z [PubMed: 24510615]
19. Ruklisa D, Ware JS, Walsh R, Balding DJ, Cook SA. Bayesian models for syndrome- and gene-specific probabilities of novel variant pathogenicity. *Genome Med*. 2015; 7(1):5.doi: 10.1186/s13073-014-0120-4 [PubMed: 25649125]

Figure 1.
Topology of the *KCNQ1*-encoded Kv7.1 voltage-gated potassium channel alpha subunit highlighting the amino acid location of the previously reported C-terminus helices (AD).

Figure 2.

Moving window analysis for the Kv7.1 C-terminus orthologs and paralogs conservation. A window size of 12 amino acids was used with a step size of 1 amino acid. The solid lines represent the moving averages for the percent conservation of orthologs (black) and paralogs (orange). The dashed lines represent the average percent conservation across the entire C-terminus among orthologs (black) and paralogs (orange). Regions of high conservation (highlighted in blue) were identified by windows with an average percent conservation greater than the average percent conservation across the entire C-terminus in both orthologs and paralogs.

Figure 3. Comparison of percent amino acid residue conservation within (Blue) and outside (White) the regions of high conservation to the percent conservation of Kv7.1's transmembrane-spanning regions. Averages for both orthologs and paralogs are shown. Error bars represent standard error of the mean. Student t-test p-values are provided above the bars. ns = p-value not significant.

Figure 4. Linear topology of case and control mutations within the C-terminus of Kv7.1. The location of case mutations (red) and control variants (Black; all controls represent all variants identified in controls, rare control represent control variants identified in a single control and used to calculate the EPV) are shown. The previously identified helices (A-B) are shown as gray cylinders. The identified highly conserved region is shown in blue.

Figure 5. Percent of case mutations (red) and control variants (black) predicted pathogenic by each *in silico* algorithm. * indicates $0.05 > p \geq 0.001$. ** indicates $p < 0.001$

Table 1

Table summarizing the data for the highly conserved regions (Conserved) and remaining regions of the Kv7.1 C-terminus

Region	Conserved Regions	Outside Conserved Regions	P-value
% Conservation (Ortholog)	94.4% $\pm$ 0.9%	79.3% $\pm$ 1.7%	p = ns, vs Transmembrane conservation (95.6% $\pm$ 0.6%)
% Conservation (Paralog)	44.0% $\pm$ 4.1%	7.8% $\pm$ 1.4%	p = ns, vs Transmembrane conservation (54.4% $\pm$ 3.0%)
% of Amino Acids hosting Case variants	37.6% (50/133)	6.7% (13/195)	p = 5.0×10^{-12} , Inside vs. Outside for Case variants
% of Amino Acids hosting Control variants	3.8% (5/133)	12.3% (24/195)	p = 0.009, Inside vs. Outside for Control variants
% Distribution of Case Variants	82.9% (68/82)	15.7% (14/82)	p = 6.3×10^{-11} , Case distribution vs. Control distribution
% Distribution of Control Variants	16.1% (5/31)	83.9% (26/31)	

Table 2

Performance of the *in silico* algorithms predictions for case mutations and control variants. The sensitivity and specificity for each tool is provided as well as the resulting Matthew Correlation Coefficient (MCC)

Prediction Algorithm	Case % Pathogenic	Control % Pathogenic	Sens.	Spec.	MCC
Highly Conserved Regions	68/82 (82.9%)	5/31 (16.1%)	82.9%	83.9%	62.3%
C-terminus Helices	35/82 (42.7%)	8/31 (25.8%)	42.7%	74.2%	15.5%
Paralog Conservation	25/82 (30.5%)	1/31 (3.2%)	30.5%	96.8%	28.9%
Ortholog Conservation	50/82 (61.0%)	4/31 (12.9%)	61.0%	87.1%	42.9%
Grantham	34/82 (41.5%)	4/31 (12.9%)	41.5%	87.1%	27.0%
SIFT	64/82 (78.0%)	6/31 (19.4%)	78.0%	80.6%	53.9%
PolyPhen2	68/82 (82.9%)	15/31 (48.4%)	82.9%	51.6%	34.9%
KsSNP	76/82 (92.7%)	15/31 (48.4%)	92.7%	51.6%	49.9%
MutPred	68/82 (82.9%)	7/31 (22.6%)	82.9%	77.4%	57.0%

Table 3

Percent of case mutations and control variants predicted pathogenic by each tool inside or outside of the highly conserved region

<i>In silico</i> Tool	Inside highly conserved regions		Outside highly conserved regions	
	Case Mutations	Control Variants	Case Mutations	Control Variants
Paralog Conservation	36.8%	20.0%	0.0%	0.0%
Ortholog Conservation	73.5%	40.0%	0.0%	7.7%
Grantham Values	42.6%	20.0%	35.7%	11.5%
SIFT	88.2%	20.0%	28.6%	19.2%
PolyPhen2	94.1%	60.0%	28.6%	46.2%
KvSNP	97.0%	60.0%	71.4%	46.2%
MutPred	83.6%	20.0%	78.6%	23.1%