Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

# HHS Public Access

Author manuscript

*J Cardiovasc Electrophysiol*. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

*J Cardiovasc Electrophysiol*. 2018 January ; 29(1): 98–104. doi:10.1111/jce.13355.

# Re-evaluating pathogenicity of variants associated with the long-QT syndrome

**Jonathan R. Kaltman, MD**, **Frank Evans, PhD**, and **Yi-Ping Fu, PhD**

Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD

## Abstract

**Introduction—**Genetic testing for congenital long-QT syndrome (LQTS) has become common. Recent studies have shown that some variants labelled as pathogenic might be misclassified due to sparse case reports and relatively common allele frequencies in the general population. This study aims to evaluate the presence of LQTS-associated variants in the Genome Aggregation Database (gnomAD) population, and assess the functional impact of these variants.

**Methods and Results—**Variants associated with LQTS from the Human Gene Mutation Database were extracted and matched to the gnomAD to evaluate population-based allele frequencies (AF). We used MetaSVM to predict the function of LQTS variants. Allele distribution by protein topology in *KCNQ1*, *KCNH2*, and *SCN5A* was compared between gnomAD (n=123,136) and a cohort of LQTS patients aggregated from 8 published studies (n = 2,683). Among the 1,415 LQTS-associated single nucleotide variants in 30 genes, 347 (25%) are present in gnomAD; 24% of the 347 variants were predicted as functionally tolerated compared with 4% of variants not present in gnomAD ($p < 0.001$). Of the 347 pathogenic variants in gnomAD, 7 (2%) had an AF of  0.001 and 65 (19%) variants had an AF of  0.0001. In *KCNQ1*, *KCNH2*, and *SCN5A*, allele distribution by protein functional region was significantly different with gnomAD alleles appearing less frequently in highly pathogenic domains than case alleles.

**Conclusion—**A significant number of LQTS variants have insufficient evidence for pathogenicity and relatively common allele frequencies in the general population. Caution should be used when ascribing pathogenicity to these variants.

### Keywords

long QT syndrome; clinical genetics; exome; genetic variation; Exome Aggregation Consortium

## Introduction

The congenital long QT syndrome (LQTS) is a genetic arrhythmia syndrome characterized by delayed repolarization, manifest as a prolonged QTc on electrocardiogram, torsades de pointes, and sudden cardiac arrest. The prevalence of LQTS is estimated at 1 per 2,500

Correspondence: Jonathan Kaltman, MD, Division of Cardiovascular Sciences, NHLBI, 6701 Rockledge Drive, Bethesda, MD 20187, Tel: 301-435-0510, Fax: 301-480-2858, kaltmanj@nhlbi.nih.gov.

Disclosures: None. The views expressed are those of the authors and not necessarily that of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

persons of European and other ancestries.[1,2] There are 15 defined sub-types of the LQTS, but the vast majority of genetically confirmed LQTS is explained by mutations in three genes, *KCNQ1*, *KCNH2*, and *SCN5A*.[3] Beta-blockers are the primary therapy for LQTS with sodium channel blockers sometimes used in patients with *SCN5A* mutations. Implantable defibrillator-cardioverters are reserved for patients who have experienced a prior cardiac arrest or who are at highest risk.

Approximately 72% of those with high probability of LQTS carry a mutation in a known LQTS gene,[4] including pathogenic variants identified in genes associated with cardiac potassium and sodium channels. Recent studies have expanded the type of genes associated with LQTS to non-channel proteins,[5,6] increasing the catalog of potential susceptibility genes to over thirty.[7]

Clinical genetic testing for the LQTS has become common with the goal of confirming disease, refining treatment strategies, and facilitating cascade screening to identify at-risk relatives. Several commercial companies currently offer gene panel testing that evaluates pathogenic variants in 12 to 17 LQTS-associated genes.[8,9]

Pathogenic variants of genetic diseases have historically been determined by their presence in one or more clinical cases and their absence in healthy populations. However, assessing the rarity of a variant in a control population has been constrained by the relatively limited number of available sequenced genomes. The recent release of data from the Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD), including 60,706 and 123,136 exomes sequences from unrelated individuals respectively, has allowed for a deeper understanding of genetic diversity within the general population.[10,11] With the much improved catalog of exome-wide variation, ExAC has demonstrated that rare genetic variation is, collectively, not so rare in the general population. Thus, this type of collective database can serve as a powerful filter to rule out candidate pathogenic variants in Mendelian disease.[10,11] The recent American College of Medical Genetics and Genomics guidelines for interpretation of sequence variants state that an allele frequency for a candidate variant in a control population, such as ExAC and gnomAD, that is greater than expected for the disorder is strong evidence for a benign interpretation.[12] Studies in cardiomyopathy [13] and catecholaminergic polymorphic ventricular tachycardia[14] have shown that some variants labelled as pathogenic might be misclassified due, in part, to relatively common allele frequencies in the ExAC population.

Re-evaluating the pathogenicity of LQTS variants is critical since genetic testing results increasingly guide treatment options and screening decisions, and false positive results might have adverse clinical implications. Therefore, the purpose of this study is to examine the presence of LQTS-associated genetic variants in the Genome Aggregation Database (gnomAD) population, and assess the impact of these variants in different functional domains in the most commonly mutated genes, *KCNQ1*, *KCNH2*, and *SCN5A*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

## Methods

### Variant identification

Genetic variants associated with congenital LQTS were identified from the Human Gene Mutation Database (HGMD® Professional 2016.3; HGMD)[7] by searching the term "long QT syndrome " for the HGMD phenotype with results manually reviewed (Supplementary Table 1). The literature supporting the HGMD designation of pathogenicity was manually curated,[7] and only variants labelled as disease causing (DM) and likely disease causing (DM?) were extracted. Variants from HGMD were subsequently identified in gnomAD (data release 2.0.1, 2017 February)[11] by matching on chromosomal position and reference/alternative alleles (genome build GRCh37/hg19). gnomAD includes high quality exome sequence data on 123,136 individuals of diverse geographic ancestry. Most individuals were sequenced because of prevalent adult diseases with exclusion of any severe pediatric diseases as well as their first-degree relatives.[10,11] Only single nucleotide variants that passed the high-quality filter in gnomAD (PASS filter) were included for further analysis. We used dbNSFP (v3.2) to conveniently annotate these LQTS variants with popular functional prediction algorithms[15] and the ensemble prediction score MetaSVM to assess whether a variant was computationally predicted to be damaging to protein function or tolerated.[16]

### Defining allele frequency threshold

Allele counts (AC) and population-based allele frequencies (AF) were determined for each variant from the combined gnomAD cohort across all ethnic groups. We selected AF thresholds of 0.001 and 0.0001 to define potentially pathogenic variants for LQTS since a variant with an allele frequency greater than these thresholds would be too common in the population to be a highly penetrant, monogenic cause of disease. An AF of 0.001 is well established as a rare variant threshold.[10] The AF of 0.0001 is estimated by the method proposed by Walsh et al.[13] and the LQTS prevalence of 1 in 2500 in the general population.[1] The most frequent allele in the clinical cohort (see description below) is *KCNQ1* c.760G>A with 26 alleles observed in 2683 individuals. This corresponds to a population allele frequency of 1.9E-06 ($26/2683 \times 1/2500 \times ½ = 1.9$E-06). The upper bound for the true frequency would be 2.7E-06 based on a binomial distribution. Therefore, we consider an AF of 0.0001 as a conservative threshold which takes into account incomplete penetrance and other biases.

### Clinical cohort

An aggregated clinical cohort of 2683 LQTS patients was synthetically assembled from the results of eight published studies (Supplementary Table 2). Only studies that provided allele counts for all variants identified were included. Three studies sequenced 3 genes (*KCNQ1, KCNH2, SCN5A*)[17–19] and four studies sequenced 5 genes (*KCNQ1, KCNH2, SCN5A, KCNE1, KCNE2*)[20–23] in patients with clinically diagnosed LQTS. In the eighth study, 12 genes (*KCNQ1, KCNH2, SCN5A, ANK2, KCNE1, KCNE2, CACNA1C, KCNJ2, CAV3, SCN4B, AKAP9, SNTA1*)[24] were sequenced in patients referred for genetic testing for the LQTS. Clinical information was available for most of the patients in this study and those

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

with positive genetic tests had clinical features suggestive of the LQTS (longer QTc interval, positive family history).

### Functional domain mapping

Variants identified in the gnomAD and clinical cohorts were mapped to functional protein domains based on the designated amino acid ranges in the three most commonly mutated LQTS genes, *KCNQ1, KCNH2, and SCN5A* (Supplementary Table 3).[17,25] Kapa et al. and Kapplinger et al. also calculated estimated predictive values (EPV) for pathogenicity to identify regions that were considered likely to be disease causing. Seven domains (TMLP and CT-c in *KCNQ1*; NT-PAS, TMLP, and CT-cNBD in *KCNH2*; TML and CT in *SCN5A*) had EPVs of greater than or equal to 75% and were considered pathogenic (Supplementary Table 3).

### Statistical analysis

Fisher's exact and Chi-square tests were applied to compare LQTS variants in different sub-groups. A two-sided p value of < 0.05 was considered significant for comparisons of pathogenicity for groups of variants based on MetaSVM prediction. For protein topology analysis, allele distribution by functional domain was compared between gnomAD and clinical cohorts. A Bonferroni-adjusted p value of < 3.85E-03 was used for significance.

## Results

From HGMD, 1415 single nucleotide, disease-causing or likely disease-causing variants were identified from 31 genes (Table 1). After quality filtering and position/allele matching, 347 of the 1415 (25%) variants from 22 genes were identified in gnomAD, including 303 disease-causing and 44 likely disease-causing variants (Table 2). The ensemble algorithm MetaSVM predicted that 24% of the 347 LQTS variants present in gnomAD would be functionally tolerated, compared with 4% of variants not present in gnomAD ($p < 0.001$, Table 2). These 347 variants were found in 9525 individuals of the gnomAD cohort representing 9652 alleles (9398 individuals were heterozygotes and 127 individuals were homozygous for the variant). The average number of alleles sequenced per these LQTS variants was 236,063, or 118,031 individuals of the gnomAD cohort. Assuming that each variant is pathogenic and an individual only carries one variant, the calculated prevalence of LQTS in the gnomAD cohort would be 1 in 12.4. Using the previously reported LQTS prevalence of 1 in 2500 in general population, the expected number of LQTS cases in gnomAD would be 47.

Among the 347 pathogenic variants in gnomAD, 7 (2%) had an AF of 0.001 and 65 (19%) variants had an AF of 0.0001 (Supplementary Table 4). MetaSVM predicted that 39% of the 65 LQTS variants with AF > 0.0001 would be tolerated compared with 20% of variants with AF < 0.0001 ($p = 0.001$, Table 2). Literature review demonstrated that pathogenicity of the 65 variants with AF 0.0001 was initially based on comparison to 100 to 3000 control alleles and only 5 (7%) demonstrated segregation in family-based studies (Supplementary Table 4). Only 9 variants (13%) had experimental data investigating their effects on ion channel function. Of these, one variant (*KCNE3* p.T4A) had no significant effect on channel

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

function, two variants (*SCN5A* p.S1904L; *KCNE1* p.V109I) had modest effects on channel function and the remainder had significant functional effects (Supplementary Table 4).

To examine the variant distribution in the LQTS patient population, we synthetically aggregated a clinical cohort from eight published studies. In a total of 2683 subjects from this clinical cohort, 1950 pathogenic alleles were identified in 1647 subjects. We limited our analysis to the 1430 alleles that matched to the 1415 single nucleotide variants identified from HGMD, which includes 580 variants from 10 genes (Table 1). The majority (96%) of these 580 variants was predicted as deleterious by MetaSVM, while 128 (22%) of them were also detected in the gnomAD cohort with a wide range of allele frequencies in both datasets (Figure 1). All sequenced alleles in *KCNQ1*, *KCNH2*, and *SCN5A* genes from gnomAD and the clinical cohort were mapped to the protein function regions. Allele distribution by functional domain was significantly different between gnomAD and the clinical cohort (Table 3). In 6 of 7 pathogenic domains, gnomAD alleles appeared less frequently than clinical cohort alleles, meeting Bonferroni corrected significance level ($p<0.0038$) in 5 of these pathogenic domains (Table 3). There were 45 variants in *KCNQ1*, *KCNH2*, and *SCN5A* that had an $AF > 0.0001$ in gnomAD. Of these, 18 (40%) mapped to the pathogenic domains. An additional comparison between the gnomAD and clinical cohorts shows that deleterious variants in *KCNQ1*, *KCNH2*, and *SCN5A* as predicted by MetaSVM are more prevalent in the clinical cohort than in gnomAD ($p < 0.001$) (Supplementary Table 5).

## Discussion

This study demonstrates that a significant number of variants associated with the LQTS designated as disease-causing or likely disease-causing from the HGMD are probably mislabeled. This conclusion is based on several intersecting lines of evidence. First, many variants have an allele frequency greater than 0.0001 that might be too common in the general population to be the monogenic cause of a rare disease. Since the gnomAD cohort is biased against rare disease, many of the variants found in the gnomAD cohort with an $AF < 0.0001$ are also suspect. Second, the ensemble algorithm MetaSVM predicted a quarter of the variants present in gnomAD are functionally tolerated. Furthermore, the variants present in gnomAD are less likely to localize to the highly pathogenic protein regions of the 3 most commonly mutated LQTS genes. Taken together, these data suggest that many of the variants associated with LQTS present in gnomAD are either bystanders, modifiers of disease, or associated with reduced penetrant form of this disease.

Prior to the release of ExAC and gnomAD data, understanding of the range of genetic diversity of rare variants in the general population was limited. Thus, the absence of a disease associated allele in a control set of 3000 alleles was thought sufficient to conclude that a variant was pathogenic. However, large scale sequencing data such as ExAC has shown that small comparator groups do not sufficiently reflect the background rate of rare variation within the population. Using ExAC as a benchmark, recent studies have cast doubt on the pathogenicity of some disease-associated variants. Walsh et al. analyzed and compared sequence data from a large cohort of cardiomyopathy patients (hypertrophic, dilated, and arrhythmogenic right ventricular cardiomyopathy) to the ExAC reference cohort.[13] They found that in some genes, especially those more recently identified through

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

candidate gene studies, there was no enrichment of rare variants in the cardiomyopathy cohort compared to ExAC, suggesting that these genes are likely not causative of highly penetrant disease. Looking at a variant level, the authors identified a significant number of putatively pathogenic variants (6.5% of HCM variants, 11.9% of DCM, and 13.5% of ARVC) that had a minor allele frequency that was too common to be associated with a highly penetrant cause of cardiomyopathy. Paludan-Muller et al. determined that 15% of 246 catecholaminergic polymorphic VT (CPVT)-associated variants (catalogued from HGMD) were identified in ExAC suggesting a significant over-representation.[14] When compared to variants not found in ExAC, CPVT variants found in ExAC were less likely to have a disease-causing effect on the expressed protein based on *in silico* predictions. The CPVT variants identified in ExAC are less likely compatible with a penetrant monogenic form of the disease.

Our study builds on the work of others who found an over-representation of LQTS-associated variants in the general population. Ackerman et al. demonstrated the presence of a non-trivial amount of rare variation in *KCNQ1*, *KCNH2*, and *SCN5A* in healthy subjects.[17,26] Focusing on LQTS-associated pathologic variants catalogued by HGMD, Refsgaard et al. and Ghouse et al. identified a high prevalence of these variants in the NHLBI Exome Sequencing Project (ESP; n = 5400 subjects) and in a Danish population-based sample (n = 7030 subjects), respectively.[27,28]

The findings from this study extend the prior work in important ways. Compared to ESP, the sample size of gnomAD provided greater power to filter rare variants.[10,11] More variants overall were identified in gnomAD than in ESP, and a higher percentage of these variants were identified in more than one individual. We also used two methods to show that many of the LQTS-associated variants in gnomAD, especially those with higher allele frequencies, are expected to result in non-pathogenic effects. One method involved the support vector machine (SVM) based ensemble prediction score which incorporated ten commonly-used annotation tools and the allele frequency information from the 1000 Genomes Project to evaluate the functionality of genetic variants.[16] The other leveraged the work of Ackerman et al. who identified pathogenic topological regions in the three most commonly mutated LQTS genes.[17,25] They reported that variants within these functional regions were more likely to be associated with true disease. By mapping the variants from gnomAD onto these functional domains, we were able to show that the gnomAD alleles were less likely to be located in one of the pathogenic regions. Additionally, we confirmed that many of the variants with an allele frequency above 0.0001 have insufficient evidence (small control groups, lack of functional or segregation data) to support the designation of pathogenicity.

It is important to explain two apparent paradoxes in the results of the study. In particular, 8 variants had functional data supporting a pathogenic effect of the variant but were also found at high allele frequency within the general population of gnomAD. Also, 75% of LQTS-associated variants found in gnomAD were predicted to be damaging by MetaSVM. How could these variants with presumably pathogenic effects be found in the general population of gnomAD (not enriched for LQTS)? These paradoxes may be reconciled by several explanations. *In vitro* experiments, by necessity, oversimplify and do not adequately represent complex *in vivo* interactions. Without the modulating effects of various factors

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

found in the cardiomyocyte, functional studies may not reflect true effects on ion channel function.[28] Also, *in silico* prediction models have generally performed better at identifying truly pathogenic variants (few false negatives) but at the expense of mislabeling tolerated variants as damaging (many false positives).[29] Finally, it is possible that some of the variants, likely a small minority, are in fact truly pathogenic (see Limitations). Both *in vitro* studies and *in silico* prediction models should be used as supportive evidence for pathogenicity rather than confirmation.[28]

This study has important clinical implications. With its expanding use in the clinic, genetic testing results help guide diagnosis, prognosis, therapy refinement, and cascade screening. Correct interpretation of an identified variant is essential for appropriate treatment and counseling for a patient and their family. Classifying a variant as pathogenic requires distinguishing its signal above the background noise within the human genome. In this study, the gnomAD dataset served as a noise filter. Allele frequencies greater than 0.0001 in the gnomAD cohort for candidate LQTS variants weakens the evidence for pathogenicity of that variant. Clinicians should not assume that a genetic testing result is a binary assessment – yes, the patient's disease is due to this variant or no, it is not. Rather a probabilistic framework should be used in evaluating genetic testing results.[30,31] Prior probabilities are established by clinical presentation, ECG findings, family history, and the background rate of the identified variant in the general population. With these factors established, the clinician can then better determine the relative risk that a particular variant is associated with disease.

## Limitations

This study is limited by the lack of phenotyping of the gnomAD cohort. It is possible that there may be some over-representation of LQTS within gnomAD. It is also possible that some pathogenic variants are present in gnomAD due to variable penetrance of the disease, subclinical disease, or normalizing effects of disease modifying variants. It is unlikely, however, that these possibilities would account for more than just a few of the more than 9600 alleles identified. Due to the limited sample size and ethnic diversity of our aggregated clinical cohort (Supplementary Table 2), we used the combined gnomAD samples and did not explore the AF distribution by ethnic groups. The estimation and classification of pathogenic variants could present differently in individuals with different genetic background. Another limitation is that the clinical cohort may include non-diseased patients. While we focused on studies that enrolled clinically diagnosed patients, we did include one study that reported clinical genetic testing results.[24] This study was included because it included phenotype data and that data suggested that patients with positive genetic tests were more likely to have phenotypic features of the LQTS.

## Conclusion

A significant number of LQTS-associated variants labelled as disease causing have an AF inconsistent with a rare disease, have amino acid changes that are predicted to be tolerated, and have insufficient evidence for pathogenicity. Caution should be used when ascribing disease causation to these variants.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

1. Schwartz PJ, Stramba-Badiale M, Crotti L, et al. Prevalence of the Congenital Long-QT Syndrome. Circulation. 2009; 120(18):1761–1767. [PubMed: 19841298]
2. Beckmann BM, Wilde AA, Kaab S. Clinical utility gene card for: long-QT syndrome (types 1–13). Eur J Hum Genet. Oct.2013 21(10)
3. Gray B, Behr ER. New Insights Into the Genetic Basis of Inherited Arrhythmia Syndromes. Circ Cardiovasc Genet. Dec; 2016 9(6):569–577. [PubMed: 27998945]
4. Tester DJ, Will ML, Haglund CM, Ackerman MJ. Effect of clinical phenotype on yield of long QT syndrome genetic testing. J Am Coll Cardiol. Feb 21; 2006 47(4):764–768. [PubMed: 16487842]
5. Crotti L, Johnson CN, Graf E, et al. Calmodulin mutations associated with recurrent cardiac arrest in infants. Circulation. Mar 05; 2013 127(9):1009–1017. [PubMed: 23388215]
6. Vatta M, Ackerman MJ, Ye B, et al. Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. Circulation. Nov 14; 2006 114(20):2104–2112. [PubMed: 17060380]
7. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet. Jan; 2014 133(1):1–9. [PubMed: 24077912]
8. [Accessed June 30, 2017] https://www.invitae.com/en/physician/category/CAT000001/#category-content-CAT000007
9. [Accessed June 30, 2017] http://www.ambrygen.com/tests/rhythmfirst-and-rhythmnext
10. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. Aug 18; 2016 536(7616):285–291. [PubMed: 27535533]
11. The Genome Aggregation Database. http://gnomad.broadinstitute.org/
12. Richards S, Aziz N, Bale S, 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. May; 2015 17(5):405–424. [PubMed: 25741868]
13. Walsh R, Thomson KL, Ware JS, et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genet Med. 2017; 19(2):192–203. [PubMed: 27532257]
14. Paludan-Muller C, Ahlberg G, Ghouse J, et al. Integration of 60,000 exomes and ACMG guidelines question the role of Catecholaminergic Polymorphic Ventricular Tachycardia-associated variants. Clin Genet. Jan; 2017 91(1):63–72. [PubMed: 27538377]
15. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat. Mar; 2016 37(3):235–241. [PubMed: 26555599]
16. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. Apr 15; 2015 24(8): 2125–2137. [PubMed: 25552646]
17. Kapa S, Tester DJ, Salisbury BA, et al. Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. Circulation. Nov 03; 2009 120(18):1752–1760. [PubMed: 19841300]
18. Gao Y, Liu W, Li C, et al. Common Genotypes of Long QT Syndrome in China and the Role of ECG Prediction. Cardiology. 2016; 133(2):73–78. [PubMed: 26496715]
19. Itoh H, Berthet M, Fressart V, et al. Asymmetry of parental origin in long QT syndrome: preferential maternal transmission of KCNQ1 variants linked to channel dysfunction. Eur J Hum Genet. Aug; 2016 24(8):1160–1166. [PubMed: 26669661]

20. Christiansen M, Hedley PL, Theilade J, et al. Mutations in Danish patients with long QT syndrome and the identification of a large founder family with p.F29L in KCNH2. BMC Med Genet. Mar 07.2014 15:31. [PubMed: 24606995]
21. Chung SK, MacCormick JM, McCulley CH, et al. Long QT and Brugada syndrome gene mutations in New Zealand. Heart Rhythm. Oct; 2007 4(10):1306–1314. [PubMed: 17905336]
22. Riuro H, Campuzano O, Berne P, et al. Genetic analysis, in silico prediction, and family segregation in long QT syndrome. Eur J Hum Genet. Jan; 2015 23(1):79–85. [PubMed: 24667783]
23. Splawski I, Shen J, Timothy KW, et al. Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. Circulation. Sep 05; 2000 102(10):1178–1185. [PubMed: 10973849]
24. Lieve KV, Williams L, Daly A, et al. Results of genetic testing in 855 consecutive unrelated patients referred for long QT syndrome in a clinical laboratory. Genet Test Mol Biomarkers. Jul; 2013 17(7):553–561. [PubMed: 23631430]
25. Kapplinger JD, Tseng AS, Salisbury BA, et al. Enhancing the Predictive Power of Mutations in the C-Terminus of the KCNQ1-Encoded Kv7.1 Voltage-Gated Potassium Channel. J Cardiovasc Transl Res. Apr; 2015 8(3):187–197. [PubMed: 25854863]
26. Ackerman MJ, Splawski I, Makielski JC, et al. Spectrum and prevalence of cardiac sodium channel variants among black, white, Asian, and Hispanic individuals: implications for arrhythmogenic susceptibility and Brugada/long QT syndrome genetic testing. Heart Rhythm. Nov; 2004 1(5): 600–607. [PubMed: 15851227]
27. 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. Eur J Hum Genet. Aug; 2012 20(8):905–908. [PubMed: 22378279]
28. Ghouse J, Have CT, Weeke P, et al. Rare genetic variants previously associated with congenital forms of long QT syndrome have little or no effect on the QT interval. Eur Heart J. Oct 01; 2015 36(37):2523–2529. [PubMed: 26159999]
29. Itan Y, Casanova JL. Can the impact of human genetic variations be predicted? Proc Natl Acad Sci. 2015; 112(37):11426–7. [PubMed: 26351682]
30. Ackerman MJ. Genetic purgatory and the cardiac channelopathies: Exposing the variants of uncertain/unknown significance issue. Heart Rhythm. Nov; 2015 12(11):2325–2331. [PubMed: 26144349]
31. Manrai AK, Funke BH, Rehm HL, et al. Genetic Misdiagnoses and the Potential for Health Disparities. N Engl J Med. Aug 18; 2016 375(7):655–665. [PubMed: 27532831]


Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

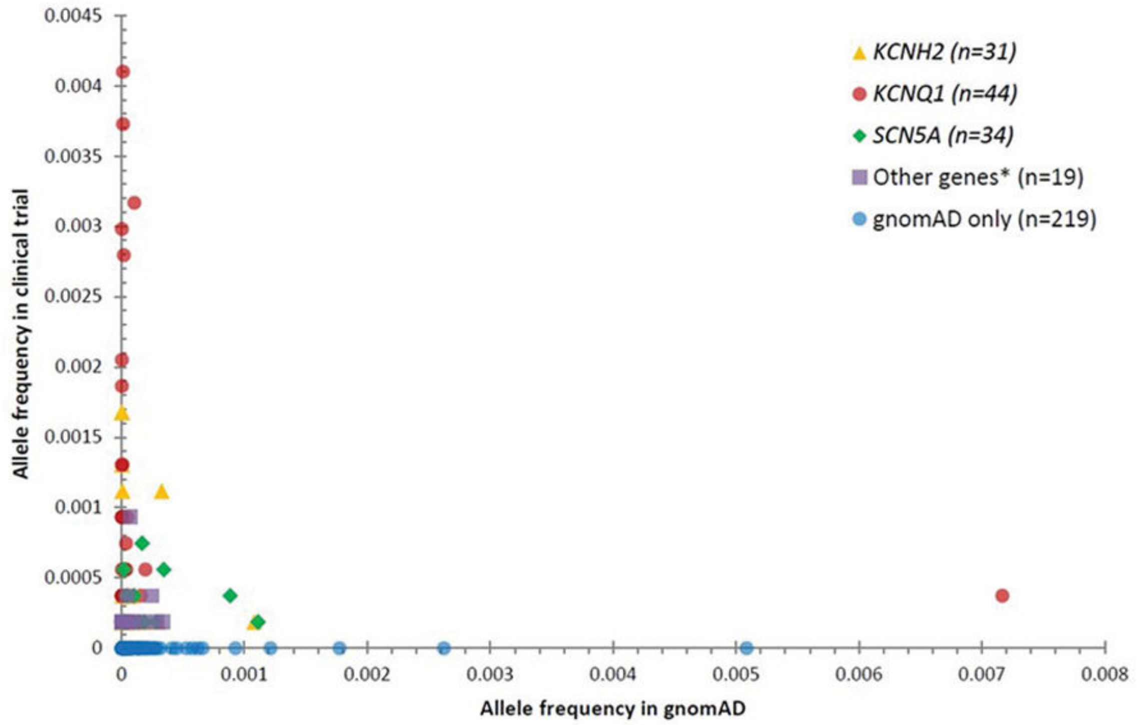


**Figure 1. Allele frequency of LQTS variants in gnomAD and the aggregated clinical cohort**

Figure 1. Scatter plot of LQTS variants with allele frequency in gnomAD cohort on X-axis and allele frequency in clinical cohort on Y-axis. Variants identified in both gnomAD and the clinical cohort are grouped by gene. Variants only found in gnomAD also displayed. Variants in other genes* includes *ANK2* (n=3), *CACNA1* (n=1), *CAV3* (n=1), *KCNE1* (n=9), *KCNE2* (n=2), *KCNJ2* (n=2), and *SCN4B* (n=1).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

**Table 1**

Number of LQTS-associated variants and genes used in this study

| Gene | In HGMD | In gnomAD | In Clinical Cohort |
|---|---|---|---|
| *AKAP9* | 5 | 4 | -- |
| *ANK2* | 17 | 13 | 3 |
| *CACNA1* | 16 | 4 | 3 |
| *CALM1* | 2 | -- | -- |
| *CALM2* | 5 | -- | -- |
| *CALM3* | 1 | -- | -- |
| *CAV3* | 6 | 2 | 1 |
| *CIT* | 2 | 1 | -- |
| *DSP* | 1 | 1 | -- |
| *KCNE1* | 34 | 28 | 12 |
| *KCNE2* | 12 | 8 | 2 |
| *KCNE3* | 1 | 1 | -- |
| *KCNH2* | 585 | 74 | 294 |
| *KCNJ2* | 57 | 5 | 4 |
| *KCNJ5* | 1 | 1 | -- |
| *KCNQ1* | 417 | 88 | 199 |
| *KIF21B* | 2 | 2 | -- |
| *NOS1AP* | 2 | -- | -- |
| *PI4KA* | 2 | -- | -- |
| *PIK3CG* | 1 | -- | -- |
| *RIMS1* | 1 | -- | -- |
| *RYR2* | 12 | 4 | -- |
| *SCN4B* | 2 | 1 | 1 |
| *SCN5A* | 214 | 100 | 61 |
| *SIRT6* | 1 | -- | -- |
| *SLC2A5* | 2 | 2 | -- |
| *SNTA1* | 4 | 4 | -- |
| *TRDN* | 1 | -- | -- |
| *UBR4* | 4 | 2 | -- |
| *UBR5* | 3 | 1 | -- |
| *WDR26* | 2 | 1 | -- |
| Total | 1415 | 347 | 580 |

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

## Table 2

Distribution and functional prediction of LQTS variants in HGMD and gnomAD

| | | In HGMD (n=1415) | | | In HGMD & gnomAD (n=347) | | |
|---|---|---|---|---|---|---|---|
| | | **Not in gnomAD** | **In gnomAD** | **p-val†** | **AF <0.0001** | **AF 0.0001** | **p-val†** |
| | **N (%)** | **1068 (75.5)** | **347 (24.5)** | | **282 (81.3)** | **65 (18.7)** | |
| **HGMD** | **Disease-causing** | 962 (90.1) | 303 (87.3) | 0.1475 | 250 (88.7) | 53 (81.5) | 0.1202 |
| | **Likely disease-causing** | 106 (9.9) | 44 (12.7) | | 32 (11.4) | 12 (18.5) | |
| **MetaSVM‡** | **Damaging** | 878 (96.2) | 248 (76.5) | 8.85E-23 | 211 (80.2) | 37 (60.7) | 0.0012 |
| | **Tolerated** | 35 (3.8) | 76 (23.5) | | 52 (19.8) | 24 (39.3) | |
| | **missing** | 155 | 23 | | 19 | 4 | |

† p-value of Chi-square tests.

‡ MetaSVM: Support vector machine (SVM) based ensemble prediction score, which incorporated 10 scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP) and the maximum frequency observed in the 1000 Genomes populations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

## Table 3

Protein topology comparison between gnomAD and aggregated clinical cohort

| Gene | Domain[†] | No. of variants used | No. of alleles[‡] (%) | gnomAD | Clinical cohort | p-val[*] |
|---|---|---|---|---|---|---|
| *KCNQ1* | NT | 2 | VT | 26 (0.0070) | 7 (0.0652) | 3.12E-05 |
| | | | WT | 370752 | 10725 | |
| | **TMLP** | 23 | VT | 71 (0.0013) | 127 (0.1029) | 1.87E-158 |
| | | | WT | 5632209 | 123291 | |
| | **CT-c** | 13 | VT | 93 (0.0030) | 53 (0.0760) | 2.66E-49 |
| | | | WT | 3131295 | 69705 | |
| | CT-nc | 6 | VT | 1818(0.1295) | 12 (0.0373) | 1.39E-07 |
| | | | WT | 1402318 | 32184 | |
| *KCNH2* | **NT-PAS** | 3 | VT | 119 (0.0166) | 3 (0.0186) | 0.7523 |
| | | | WT | 716141 | 16095 | |
| | NT-nPAS | 7 | VT | 353 (0.0233) | 13 (0.0346) | 0.169 |
| | | | WT | 1515257 | 37549 | |
| | **TMLP** | 6 | VT | 33 (0.0022) | 14 (0.0435) | 7.09E-13 |
| | | | WT | 1477151 | 32182 | |
| | **CT-cNBD** | 4 | VT | 7 (0.0007) | 18 (0.0839) | 3.47E-25 |
| | | | WT | 984579 | 21446 | |
| | CT-ncNBD | 11 | VT | 55 (0.0025) | 21 (0.0356) | 3.61E-16 |
| | | | WT | 2216311 | 59005 | |
| *SCN5A* | NT | 2 | VT | 21 (0.0043) | 2 (0.0186) | 0.0859 |
| | | | WT | 491793 | 10730 | |
| | IDL | 16 | VT | 214 (0.0055) | 21 (0.0245) | 6.14E-08 |
| | | | WT | 3884882 | 85835 | |
| | **TML** | 9 | VT | 174 (0.0079) | 13 (0.0269) | 2.26E-04 |

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

| Gene | Domain[†] | No. of variants used | No. of alleles[‡] (%) | gnomAD | Clinical cohort | p-val[*] |
|---|---|---|---|---|---|---|
| | | | WT | 2205198 | 48281 | |
| | **CT** | 6 | VT | 455 (0.0320) | 7 (0.0217) | 0.4260 |
| | | | WT | 1420515 | 32189 | |

[†]Underline signifies pathogenic domain.

[‡]VT: variant alleles; WT: wild type alleles.

[*]p-value from Fisher's Exact tests. Bonferroni-corrected significance level=0.0038.