

## *In silico* investigations on functional and haplotype tag SNPs associated with congenital long QT syndromes (LQTSs)

C. Sudandiradoss · Rao Sethumadhavan

Received: 10 June 2008 / Revised: 17 November 2008 / Accepted: 15 January 2009 / Published online: 12 February 2009  
© Springer Science+Business Media B.V. 2009

**Abstract** Single-nucleotide polymorphisms (SNPs) play a major role in the understanding of the genetic basis of many complex human diseases. It is still a major challenge to identify the functional SNPs in disease-related genes. In this review, the genetic variation that can alter the expression and the function of the genes, namely *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1* and *KCNE2*, with the potential role for the development of congenital long QT syndrome (LQTS) was analyzed. Of the total of 3,309 SNPs in all five genes, 27 non-synonymous SNPs (nsSNPs) in the coding region and 44 SNPs in the 5' and 3' un-translated regions (UTR) were identified as functionally significant. SIFT and PolyPhen programs were used to analyze the nsSNPs and FastSNP; UTR scan programs were used to compute SNPs in the 5' and 3' untranslated regions. Of the five selected genes, *KCNQ1* has the highest number of 26 haplotype blocks and 6 tag SNPs with a complete linkage disequilibrium value. The gene *SCN5A* has ten haplotype blocks and four tag SNPs. Both *KCNE1* and *KCNE2* genes have only one haplotype block and four tag SNPs. Four haplotype blocks and two tag SNPs were obtained for *KCNH2* gene. Also, this review reports the copy number variations (CNVs), expressed sequence tags (ESTs) and genome survey sequences (GSS) of the selected genes. These computational methods are in good agreement with experimental works reported earlier concerning LQTS.

**Keywords** Long-QT syndrome · Non-synonymous SNP · SIFT · Polyphen · 5' UTR · 3' UTR · Haplotype tag SNPs

### Introduction

Inherited mutations of ion channel proteins are prevalent, and the disorders caused by them, including epilepsy, febrile seizures, Dent's disease and cardiac arrhythmias, are now referred to as channelopathies (Marban 2002; Jentsch 2000; Schwake et al. 2001; Lossin et al. 2002; Jurkat-Rott and Lehmann-Horn 2001; Kullmann 2002). Congenital LQTS is a genetically heterogeneous disorder associated with mutations in various cardiac ion channel genes that prolongs repolarization of the ventricular myocyte. The cardiac repolarization process is known to be strongly dependent on various parameters, such as heart rate (Bazett 1920), age (Reardon and Malik 1996), sex (Yang et al. 1994; Legato 2000), plasma levels of electrolytes (Nagasaka et al. 1972), medications (Kaab et al. 2003) and inherited and acquired pathological conditions (Tomaselli and Marban 1999). The molecular basis for LQTS is delayed repolarization of the myocardium, which prolongs the cardiac action potential, increasing the QT interval measured on the surface electrocardiogram. Mutations in five ion channel genes, namely *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1* and *KCNE2*, cause the majority of cases of inherited LQTS. Recently, genetic approaches to understand diversity in cardiac function and susceptibility to cardiac arrhythmias have focused in particular on ion channels and gap junction proteins as key components in normal and abnormal cardiac electrophysiology.

One of the interests in association studies is the association between SNPs and disease development. There are millions of SNPs in the entire human genome, which creates major difficulty for planning costly population-based genotyping to target SNPs that are most likely to affect phenotypic functions and ultimately contribute to disease development. Single nucleotide polymorphism (SNP)

C. Sudandiradoss · R. Sethumadhavan (✉)  
Bioinformatics Division, School of Biotechnology, Chemical  
and Biomedical Engineering, Vellore Institute of Technology,  
Vellore, TN 632014, India  
e-mail: rsethumadhavan@vit.ac.in

markers are preferred for disease association studies because of their high abundance along with the human genome. But the current throughput of technology is inadequate for genotyping all the existing SNPs for a large number of samples. Thus, the selection of a maximally informative set of SNPs (tag SNPs) for genome-wide association studies has attracted much attention. Linkage disequilibrium (LD) patterns vary across the human genome, with some regions of high LD interspersed with regions of low LD. Such LD patterns make it possible to select a set of single nucleotide polymorphisms (SNPs; tag SNPs) for genome-wide association studies. Genome-wide association methods based on linkage disequilibrium (LD) offer a promising approach to detect genetic variation responsible for common human diseases. Several large-scale studies for dissecting LD patterns across the human genome based on SNPs have revealed that the LD patterns vary greatly across the human genome, with some regions of high LD interspersed with regions of low LD (Gabriel et al. 2002; Patil et al. 2001). In those high LD regions, which are referred to as blocks in the literature, only a small number of SNPs are sufficient to capture most of haplotype structure (Johnson et al. 2001; Patil et al. 2001).

Understanding the functions of single nucleotide polymorphisms (SNPs) can greatly help to understand the genetics of the human phenotype variation and especially the genetic basis of complex human diseases like long QT syndrome (Schork et al. 2000). Therefore, it is urgent to develop and apply methods to prioritize target SNPs. To date, at least five gene loci have been identified for the LQTS disorder, of which four encode for potassium channels, *KCNQ1*, *KCNH2*, *KCNE1* and *KCNE2*, and one encodes for the sodium channel, *SCN5A*. *KCNQ1*, the gene responsible for causing LQT1, was mapped to chromosome 11p15.5 (Keating et al. 1991), *KCNH2* to LQT2 locus on chromosome 7q35–36 (Curran et al. 1995), *SCN5A* to LQT3 locus on chromosome 3p21–24 (Jiang et al. 1994), *KCNE1* to LQT5 locus on chromosome 21q22 (Barhanin

et al. 1996) and *KCNE2* to LQT6 locus on chromosome 21q22 (Abbott et al. 1999). LQT4 has been attributed to ankyrin B mutation (Mohler et al. 2003), and its locus was mapped to chromosome 4q25–27 (Schott et al. 1995). The *ANKB* gene, which encodes for Ankyrin-B protein to cause type 4 LQTS, is not included in this work since no polymorphisms in Ankyrin-B associated with LQTS have been reported so far. In this review, computationally predicted most deleterious non-synonymous SNPs (nsSNPs) in the coding regions and SNPs in the 5' and 3' un-translated regions (UTR) of all five genes are reported. Also, the results obtained through computation were compared with an experimental work reported earlier, and this led to the conclusion that there are many other deleterious SNPs that have to be worked out experimentally in the near future. Apart from these predictions, haplotype blocks, tag SNPs, copy number variations (CNVs), expressed sequence tags (ESTs) and genome survey sequence (GSS) information about the genes with the potential role for the development of congenital long QT syndrome (LQTS) are reported.

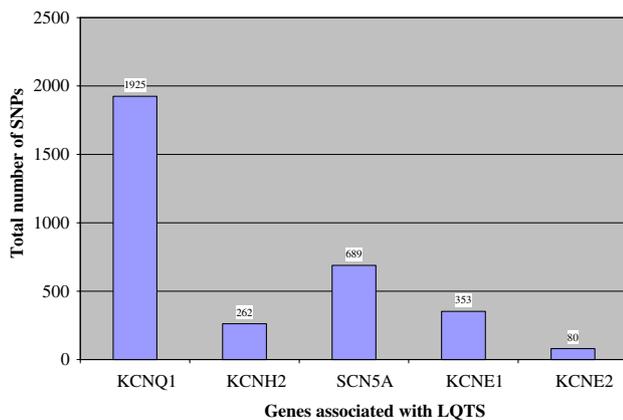
#### Distribution of total SNPs in the five selected genes

Detailed descriptions of polymorphisms and the respective mRNA sequences for the *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1* and *KCNE2* genes were obtained from the NCBI human genome protein sequence (Wheeler et al. 2006) and Swiss-Prot database (Yip et al. 2004). The information is given in Table 1. In total, 3,309 SNPs were found in all five genes. Among the 3,309 SNPs, 97 (3%) were coding non-synonymous SNPs (nsSNP), 113 were present in the 5'UTR 3' UTR regions, 71 were coding synonymous SNPs (sSNP), and the rest of the 3,028 were in the intron and non-coding exon regions. Only coding non-synonymous SNPs (nsSNPs), 5' and 3' UTR SNPs were selected for the investigations. The distribution of total SNPs in the individual selected genes is shown in Fig. 1. It can be seen

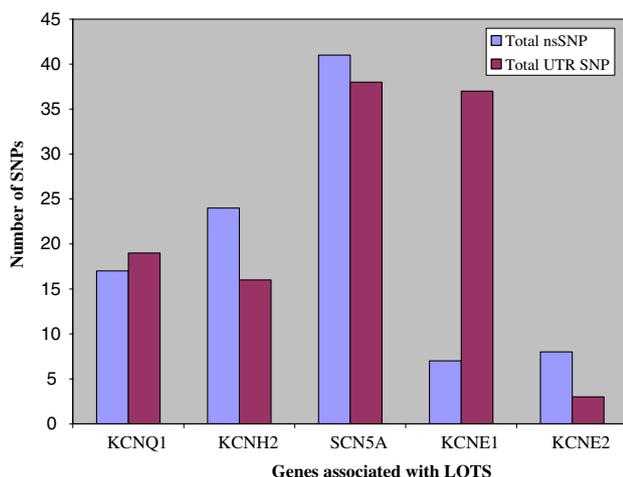
**Table 1** Detailed information about the genes associated with congenital long QT syndrome

Disorders name	Gene symbol	Protein name	Chromosome number	Nucleotide position	Swiss-Prot ID	mRNA accession number
Long QT syndrome 1	KCNQ1	Potassium voltage-gated channel subfamily KQT member 1	11	2439260–2826915	P51787	NM_181798
Long QT syndrome 2	KCNH2 or HERG	Potassium voltage-gated channel subfamily H member 2	7	150079697–150090563	Q12809	NM_000238
Long QT syndrome 3	SCN5A	Sodium channel protein type 5 subunit alpha	3	38564558–38666167	Q14524	NM_000335
Long QT syndrome 5	KCNE1	Potassium voltage-gated channel subfamily E member 1	21	34740859–34806443	P15382	NM_000219
Long QT syndrome 6	KCNE2	Potassium voltage-gated channel subfamily E member 2	21	34658193–34665310	Q9Y6J6	NM_005136

from Fig. 1 that the highest number of SNPs was present in the *KCNQ1* gene, and the lowest number of SNPs was present in the *KCNE2* gene. Figure 2 shows the distribution of nsSNPs and UTR SNPs (3' and 5' together) as a function of the five genes studied. It is interesting to note from this figure that, even though the total number of SNPs was much less in *SCN5A* gene compared to *KCNQ1* (Fig. 1), the number of nsSNPs was much higher in *SCN5A* compared to *KCNQ1*. Also, the total UTR SNPs were found to be significantly higher for *SCN5A* and *KCNE1* compared to the other three genes. No regular correlation could be observed among the total SNPs, nsSNPs and UTR SNPs present in all five genes. This prompted us to investigate the deleterious nature of the individual nsSNPs and also to determine the role of SNPs in the 5' and 3' UTR regions of the selected five ion channel genes using computational methods.



**Fig. 1** Distribution of total number of SNPs in the genes associated with LQTS



**Fig. 2** Total number of nsSNPs and UTR SNPs for genes associated with LQTS

### Deleterious coding nonsynonymous SNPs found by the SIFT program

The SIFT program (Ng and Henikoff 2002) was used to detect the deleterious coding nonsynonymous SNPs. SIFT is a sequence homology-based tool that presumes that important amino acids will be conserved in the protein family. Hence, changes at well-conserved positions tend to be predicted as deleterious (Ng and Henikoff 2002). The query has to be submitted in the form of SNP IDs or as protein sequences. The underlying principle of this program is that SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multi-step procedure that, given a protein sequence (a), searches for similar sequences (b), chooses closely related sequences that may share similar functions, (c) obtains the multiple alignment of the chosen sequences and (d) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and those greater than or equal to the cutoff are predicted to be tolerated (Ng and Henikoff 2001). The cutoff value in the SIFT program is a tolerance index of  $\geq 0.05$ . The higher the tolerance index, the less functional impact a particular amino acid substitution is likely to have. Among the 97 coding non-synonymous SNPs in the selected five genes, 31 nsSNPs were deleterious, having the tolerance index score of  $\leq 0.05$ . The results are shown in Table 2. According to the SIFT algorithm, five nsSNPs showed functionally significant scores, and the SNP with an id (rs45478697) showed a highly deleterious tolerance index score of 0.00 in the *KCNQ1* gene. Likewise, in the *KCNH2* gene, seven nsSNPs showed functionally significant scores, and the SNPs with ids (rs45607339, rs11538710 and rs731506) showed a highly deleterious tolerance index score of 0.00. In *SCN5A* gene, 11 nsSNPs have functionally significant scores, and the SNPs with ids (rs45600438, rs45589741, rs45546039 and rs6791924) showed a highly deleterious tolerance index score. In *KCNE1*, two nsSNPs with ids (rs45457092 and rs28933384) and in *KCNE2*, out of six nsSNP ids, four (namely rs45600841, rs35759083, rs16991654 and rs2234916) showed a highly deleterious tolerance index score of 0.00. The predictive power and accuracy of the SIFT program are 88.3–90.6% and 67.4–70.3% specificity and sensitivity, respectively, when tested with different datasets of human variants (Mathe et al. 2006).

### Damaged nsSNP found by the PolyPhen algorithm

Analyzing the damaged coding nonsynonymous SNPs at the structural level is considered to be very important to

**Table 2** Summary of nsSNPs that were predicted to have functional significance by SIFT algorithm

Gene name	SNP ID	Nucleotide change	Amino acid change	Tolerance index	References
KCNQ1	<b>rs45606931</b>	A → G	<b>362 K → R</b>	<b>0.01</b>	Predicted in this work
	<b>rs45594640</b>	C → T	<b>583 R → C</b>	<b>0.02</b>	Tranebjaerg et al. (1999)
	<b>rs45478697</b>	C → T	<b>341 A → V</b>	<b>0.00</b>	Tranebjaerg et al. (1999)
	<b>rs2522018</b>	A → C	<b>64 P → H</b>	<b>0.02</b>	Predicted in this work
	<b>rs179489</b>	C → G	<b>168 G → R</b>	<b>0.04</b>	Splawski et al. (1998)
KCNH2	<b>rs45607339</b>	C → T	<b>784 R → W</b>	<b>0.00</b>	Splawski et al. (2000) and Yang et al. (2002)
	<b>rs41314366</b>	A → G	<b>859 T → M</b>	<b>0.04</b>	Predicted in this work
	<b>rs41313074</b>	A → G	<b>1016 P → L</b>	<b>0.04</b>	Splawski et al. (2000)
	<b>rs36210422</b>	C → T	<b>176 R → W</b>	<b>0.01</b>	Splawski et al. (2000) and Laitinen et al. (2000)
	<b>rs28928904</b>	G → T	<b>593 I → R</b>	<b>0.01</b>	Benson et al. (1996) and Splawski et al. (2000)
SCN5A	<b>rs11538710</b>	A → T	<b>506 I → N</b>	<b>0.00</b>	Predicted in this work
	<b>rs731506</b>	G → T	<b>41 V → G</b>	<b>0.00</b>	Predicted in this work
	<b>rs45600438</b>	C → T	<b>568 R → C</b>	<b>0.00</b>	Predicted in this work
	<b>rs45589741</b>	C → T	<b>1249 F → L</b>	<b>0.00</b>	Predicted in this work
	<b>rs45563942</b>	C → T	<b>1835 I → T</b>	<b>0.05</b>	Predicted in this work
	<b>rs45546039</b>	A → G	<b>222 R → Q</b>	<b>0.00</b>	Predicted in this work
	<b>rs45514691</b>	C → T	<b>1582 R → C</b>	<b>0.02</b>	Predicted in this work
	rs41313691	G → T	524 S → Y	<b>0.04</b>	Predicted in this work
	<b>rs41313031</b>	A → G	<b>1307 L → F</b>	<b>0.02</b>	Predicted in this work
	<b>rs41276525</b>	A → G	<b>216 S → L</b>	<b>0.05</b>	Predicted in this work
KCNE1	<b>rs28937318</b>	A → G	<b>367 R → H</b>	<b>0.01</b>	Smits et al. (2002)
	<b>rs28937316</b>	A → G	<b>1643 R → H</b>	<b>0.00</b>	Splawski et al. (2000)
	<b>rs6791924</b>	A → G	<b>34 R → C</b>	<b>0.00</b>	Yang et al. (2002)
	<b>rs45457092</b>	C → G	<b>52 G → A</b>	<b>0.00</b>	Predicted in this work
	<b>rs28933384</b>	C → T	<b>7 T → I</b>	<b>0.00</b>	Schulze-Bahr et al. (1997)
KCNE2	<b>rs45600841</b>	C → T	<b>60 F → L</b>	<b>0.00</b>	Millat et al. (2006)
	rs45476493	C → T	66 A → V	0.02	Predicted in this work
	<b>rs35759083</b>	A → G	<b>8 T → I</b>	<b>0.00</b>	Abbott et al. (1999) and Millat et al. (2006)
	rs2234916	A → G	8 T → A	0.00	Abbott et al. (1999) and Millat et al. (2006)

nsSNP scores found to have functional significance by both SIFT and PolyPhen are shown in bold

understand the functional activity of the protein of concern. The PolyPhen algorithm (Ramensky et al. 2002) was used for this purpose. Input options for the PolyPhen server are protein sequence or SWALL database ID or accession number together with sequence position with two amino acid variants. The query has to be submitted in the form of protein sequence with mutational position and two amino acid variants. Sequence-based characterization of the substitution site, profile analysis of homologous sequences and mapping of the substitution site to a known protein three-dimensional structure are the parameters taken into account by the PolyPhen program to calculate the score. It calculates PSIC scores for each of the two variants and then computes the PSIC score difference between them. The higher the PSIC score difference, the higher the functional impact on a particular amino acid substitution is likely to have. SNPs can be characterized by the type of nucleotide change as well as the putative functional effect. Ninety-seven protein

sequences of nsSNPs investigated in this work were submitted as an input to the PolyPhen program, and the results are shown in Table 3. A position-specific independent count (PSIC) score difference was computed for each one, and a PSIC score difference of 1.5 and above is considered to be damaging. From the PSIC score, it was found that 46 nsSNPs (with PSIC score difference above 1.500) might significantly affect the protein structure (Table 3). Interestingly, there was a significant correlation between the SIFT and PolyPhen approach. When these two approaches were used together, 5 nsSNPs in *KCNQ1*, 7 nsSNPs in *KCNH2*, 10 nsSNPs in *SCN5A*, 3 nsSNPs in *KCNE1* and 6 nsSNPs in *KCNE2* were predicted to be most deleterious. The nsSNPs scores found to have functional significance by both SIFT and PolyPhen are shown bold in Tables 2 and 3. The results obtained by these programs partially concur with the experimental works reported earlier. Notably, the SNP with an id rs179489 in *KCNQ1* (Splawski et al. 1998),

**Table 3** Summary of nsSNPs that were predicted to have functional significance by PolyPhen algorithm

Gene name	SNP ID	Nucleotide change	Amino acid change	PSIC	References
KCNQ1	<b>rs45606931</b>	A → G	<b>362 K → R</b>	<b>1.516</b>	Predicted in this work
	<b>rs45594640</b>	C → T	<b>583 R → C</b>	<b>1.858</b>	Tranebjaerg et al. (1999)
	<b>rs45478697</b>	C → T	<b>341 A → V</b>	<b>1.656</b>	Tranebjaerg et al. (1999)
	<b>rs2522018</b>	A → C	<b>64 P → H</b>	<b>1.785</b>	Predicted in this work
	<b>rs179489</b>	C → G	<b>168 G → R</b>	<b>1.856</b>	Splawski et al. (1998)
KCNH2	<b>rs45607339</b>	C → T	<b>784 R → W</b>	<b>2.654</b>	Splawski et al. (2000) and Yang et al. (2002)
	<b>rs41314366</b>	A → G	<b>859 T → M</b>	<b>2.050</b>	Predicted in this work
	rs41313764	A → G	1018 P → S	1.800	Predicted in this work
	rs41313761	A → G	1020 P → L	1.879	Splawski et al. (2000)
	<b>rs41313074</b>	A → G	<b>1016 P → L</b>	<b>2.250</b>	Splawski et al. (2000)
	rs41307319	A → C	137 M → I	1.539	Predicted in this work
	rs41307280	A → G	1016 P → S	1.800	Splawski et al. (2000)
	rs41307271	A → G	1026 P → L	2.250	Splawski et al. (2000)
	<b>rs36210422</b>	C → T	<b>176 R → W</b>	<b>1.914</b>	Splawski et al. (2000) and Laitinen et al. (2000)
	rs36210421	G → T	1047 R → L	1.575	Predicted in this work
	rs28928905	A → G	490 A → T	1.542	Splawski et al. (2000) and Yoshida et al. (2001)
	<b>rs28928904</b>	G → T	<b>593 I → R</b>	<b>1.928</b>	Benson et al. (1996) and Splawski et al. (2000)
	<b>rs11538710</b>	A → T	<b>506 I → N</b>	<b>2.435</b>	Predicted in this work
	rs9333649	A → G	572 G → S	1.824	Splawski et al. (2000) and Larsen et al. (2000)
	rs1805123	A → C	897 K → T	1.722	Splawski et al. (2000) and Yang et al. (2002) and Laitinen et al. (2000)
SCN5A	<b>rs731506</b>	G → T	<b>41 V → G</b>	<b>1.939</b>	Predicted in this work
	rs45620037	C → T	220 T → I	1.912	Benson et al. (2003)
	<b>rs45600438</b>	C → T	<b>568 R → C</b>	<b>2.359</b>	Predicted in this work
	<b>rs45589741</b>	C → T	<b>1249 F → L</b>	<b>2.001</b>	Predicted in this work
	<b>rs45563942</b>	C → T	<b>1835 I → T</b>	<b>1.793</b>	Predicted in this work
	<b>rs45546039</b>	A → G	<b>222 R → Q</b>	<b>1.595</b>	Predicted in this work
	<b>rs45514691</b>	C → T	<b>1582 R → C</b>	<b>2.495</b>	Predicted in this work
	rs45489199	C → G	2005 P → A	1.645	Predicted in this work
	rs45475899	A → G	833 G → R	2.172	Predicted in this work
	rs45465995	C → T	1896 R → W	2.399	Predicted in this work
	rs41313681	A → G	656 P → L	2.201	Predicted in this work
	<b>rs41313031</b>	A → G	<b>1307 L → F</b>	<b>1.500</b>	Predicted in this work
	rs41311135	A → G	975 R → W	1.768	Predicted in this work
	rs41311127	A → G	1292 F → S	1.763	Predicted in this work
	rs41311061	A → G	438 M → T	2.224	Predicted in this work
	<b>rs41276525</b>	A → G	<b>216 → L</b>	<b>1.998</b>	Predicted in this work
	<b>rs28937318</b>	A → G	<b>367 R → H</b>	<b>1.820</b>	Smits et al. (2002)
	<b>rs28937316</b>	A → G	<b>1643 R → H</b>	<b>1.820</b>	Splawski et al. (2000)
	rs7626962	G → T	1102 S → Y	1.604	Predicted in this work
	<b>rs6791924</b>	A → G	<b>34 R → C</b>	<b>2.008</b>	Yang et al. (2002)
KCNE1	<b>rs45457092</b>	C → G	<b>52 G → A</b>	<b>1.824</b>	Predicted in this work
	<b>rs28933384</b>	C → T	<b>7 T → I</b>	<b>2.050</b>	Schulze-Bahr et al. (1997)
	rs17857111	C → T	32 R → H	1.704	The MGC Project Team (2004), Splawski et al. (2000)
KCNE2	<b>rs45600841</b>	C → T	<b>60 F → L</b>	<b>1.775</b>	Millat et al. (2006)
	<b>rs35759083</b>	A → G	<b>8 T → I</b>	<b>1.711</b>	Abbott et al. (1999) and Millat et al. (2006)

nsSNPs scores found to have functional significance by both SIFT and PolyPhen are shown in bold

rs36210422 (Laitinen et al. 2000) and rs41313074 (Splawski et al. 2000) in *KCNH2*, rs28937316 (Splawski et al. 2000), rs6791924 (Yang et al. 2002) and rs28937318 (Smits et al. 2002) in *SCN5A*, rs28933384 (Schulze-Bahr et al. 1997) in *KCNE1*, and rs2234916 (Millat et al. 2006), rs16991654 (Millat et al. 2006) and rs35759083 (Millat et al. 2006) in *KCNE2* gene reported through experimental work were also predicted to be the deleterious mutations by the SIFT and PolyPhen programs. There are a few other mutations, which are depicted as ‘Predicted in this work’ in Tables 2 and 3, that have not been reported experimentally so far, but must be worked out in the future.

### Functional SNPs in un-translated regions (UTR) found by the FastSNP program

Recent studies show that SNPs have functional effects on protein structure by a single change in the amino acid (Cargill et al. 1999; Sunyaev et al. 2000) and on transcriptional regulation (Prokunina and Alarcn-Riquelme 2004; Prokunina et al. 2002). The Web-based algorithm FastSNP (Yuan et al. 2006) was used for predicting the functional significance of the 5′ and 3′ UTRs of the selected genes. The FastSNP program follows the decision tree principle with external Web service access to TFSearch, which predicts whether a noncoding SNP alters the transcription factor-binding site of a gene. The score will be given by this program on the basis of levels of risk with a ranking of 0, 1, 2, 3, 4 or 5. This signifies the levels of no, very low, low, medium, high and very high effect, respectively. Table 4 shows the list of SNPs in the 5′ untranslated region that are predicted to be functionally significant in the five selected genes. According to FastSNP, only five SNPs with ids (rs41315349, rs41314819, rs4131547, rs41315473 in *KCNE1* and rs41260744 in *KCNE2*) have possible functional effects in the 5′ UTR regions of *KCNE1* and *KCNE2* genes. Among these five SNPs, the SNP with an id rs41315349 shows moderate to high levels of risk, and the remaining SNPs shows very low to medium levels of risk. No SNPs were predicted to have a functional effect by the FastSNP program in the remaining three genes. However, this algorithm did not find any

functional significance for the 3′ UTR, and hence the UTR scan algorithm was used to check the functional significance in the 3′ and 5′ un-translated regions.

### Functional SNPs in UTR found by the UTRscan

The 5′ and 3′ UTRs are involved in various biological processes, such as post-transcriptional regulatory pathways, stability and translational efficiency (Sonenberg 1994; Nowak 1994). The UTRscan program (Pesole and Liuni 1999) allows one to search the user-submitted sequences for any of the patterns collected in the UTR site. UTRsite is a collection of functional sequence patterns located in 5′ or 3′ UTR sequences. Briefly, two or three sequences of each UTR SNP that have a different nucleotide at an SNP position are analyzed by UTRscan, which looks for UTR functional elements by searching through user-submitted sequence data for the patterns defined in the UTRsite and UTR databases. If different sequences for each UTR SNP are found to have different functional patterns, that particular UTR SNP is predicted to have functional significance. The Internet resources for UTR analysis are UTRdb and UTRsite. UTRdb contains experimentally proven biological activity of functional patterns of UTR sequences from eukaryotic mRNAs (Pesole et al. 2002). The UTRsite has the data collected from UTRdb and also is continuously enriched with new functional patterns. The different patterns include 15-lipoxygenase differentiation control element (15-LOX-DICE) (Ostareck-Lederer et al. 1994, 1998; Ostareck et al. 1997), the internal ribosome entry site (IRES) (Le and Maizel 1997), the GY box (Lai et al. 2000), alcohol dehydrogenase 3′UTR downregulation control element (ADH\_DRE) (Parsch et al. 1999, 2000), cytoplasmic polyadenylation element (CPE) (Vassalli and Stutz 1996; Verrotti et al. 1996) and terminal oligopyrimidine tract (TOP) (Kato et al. 1994; Levy et al. 1991; Kaspar et al. 1992; Meyuhas et al. 1996). Polymorphisms in the 3′ UTR affect gene expression by affecting the ribosomal translation of mRNA or by influencing the RNA half-life (Van Deventer 2000). The UTRscan program results are depicted in Table 5. There were 113 SNPs in the UTR regions of the five selected genes, out of which 71 were in the 3′ UTR and

**Table 4** List of SNPs (UTR mRNA) predicted to be functionally significant by FastSNP

Gene	SNP ID	Nucleotide change	mRNA position	UTR position	Level of risk	Possible functional effects
KCNE1	rs41315349	C → T	251	5′UTR	Moderate–high (3–4)	Splicing site
	rs41314819	A → G	526	5′UTR	Very low–medium (1–3)	Promoter/regulatory region
	rs41315471	C → T	373	5′UTR	Very low–medium (1–3)	Promoter/regulatory region
	rs41315473	A → G	95	5′UTR	Very low–medium (1–3)	Promoter/regulatory region
KCNE2	rs41260744	A → G	62	5′UTR	Very low–medium (1–3)	Promoter/regulatory region

**Table 5** Summary of nsSNPs that were predicted to have functional significance by UTR scan algorithm

Gene name	SNP ID	Nucleotide change	UTR position	Functional element change
KCNQ1	rs45579540	C → T	3' UTR	15-LOX-DICE
	rs45570140	A → C	3' UTR	15-LOX-DICE
	rs45510192	A → G	3' UTR	15-LOX-DICE
	rs45477500	A → G	3' UTR	15-LOX-DICE
	rs45460605	C → T	3' UTR	15-LOX-DICE
	rs3889377	A → G	5' UTR	15-LOX-DICE
	rs2519184	A → G	3' UTR	15-LOX-DICE
	rs422976	C → G	3' UTR	15-LOX-DICE
	rs8234	A → G	3' UTR	15-LOX-DICE
	rs45504094	A → G	5' UTR	IRES
	rs8178998	A → C	5' UTR	IRES
	rs3889377	A → G	5' UTR	ADH_DRE
	rs10798	A → G	3' UTR	GY-Box
	rs8234	A → G	3' UTR	GY-Box
	KCNH2	rs41314396	C → T	5' UTR
rs41314393		A → G	5' UTR	15-LOX-DICE
rs41314363		A → G	3' UTR	15-LOX-DICE
rs41311012		C → T	5' UTR	15-LOX-DICE
rs41307289		A → G	3' UTR	15-LOX-DICE
rs41314396		C → T	5' UTR	IRES
rs41314393		A → G	5' UTR	IRES
SCN5A	rs45593136	A → G	3' UTR	CPE
	rs45601739	A → G	5' UTR	15-LOX-DICE
	rs45589543	C → T	3' UTR	15-LOX-DICE
	rs45548632	C → T	3' UTR	15-LOX-DICE
	rs45512996	A → G	3' UTR	15-LOX-DICE
	rs45503498	C → G	3' UTR	15-LOX-DICE
	rs45502198	A → G	3' UTR	15-LOX-DICE
	rs45474195	A → G	3' UTR	15-LOX-DICE
	rs45441103	A → G	3' UTR	15-LOX-DICE
	rs41315491	C → T	3' UTR	15-LOX-DICE
	rs45601739	A → G	5' UTR	IRES
KCNE1	rs41315471	C → T	5' UTR	15-LOX-DICE
	rs41315349	C → T	5' UTR	15-LOX-DICE
	rs41314803	G → T	3' UTR	15-LOX-DICE
	rs41314793	C → T	3' UTR	15-LOX-DICE
	rs41312995	C → T	3' UTR	15-LOX-DICE
	rs41312993	A → G	3' UTR	15-LOX-DICE
	rs41312367	A → G	3' UTR	15-LOX-DICE
	rs41315473	A → G	5' UTR	TOP
	rs41315471	C → T	5' UTR	IRES
	rs41315349	C → T	5' UTR	K-Box
rs41314819	A → G	5' UTR	Brd-Box	
KCNE2	rs41260744	A → G	5' UTR	15-LOX-DICE

42 were in the 5' UTR regions. UTRscan was applied to prioritize 113 UTR region SNPs. It found 44 of them to have different patterns for each sequence, which were predicted to have functional significance. Among the 44 UTR region SNPs, 18 were present in the 5' UTR region,

and 26 were present in the 3' UTR region. Also, 31 of them (Table 5) were related to the functional pattern change of 15-LOX-DICE; 6 functional SNPs out of 44 were related to the functional pattern change of IRES, 2 of them were related to the functional pattern change of GY-Box and 5 of

them related to the functional pattern change of ADH\_DRE, CPE, TOP, K-Box and Brd-Box, respectively.

### Selection of haplotype tag SNPs

Haplotypes are common single nucleotide polymorphisms (SNPs) that have important implications for mapping of disease genes and human traits. Often only a small subset of the SNPs is sufficient to capture the full haplotype information. Such subsets of markers are called haplotype tagging SNPs (htSNPs). The HapMap website at <http://www.hapmap.org> is the primary portal to genotype data produced as part of the International HapMap Project (Gibbs et al. 2003). The Haploview program (Barrett et al. 2005) was used for analyzing the number of haplotype blocks and selecting the haplotype tag SNPs. Haploview is a tool for the selection and evaluation of tag SNPs from genotype data, such as those from the International HapMap Project. It combines the simplicity of pairwise tagging methods with the efficiency benefits of multimarker haplotype approaches. The genotype data of all the five genes, namely *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1* and *KCNE2*, have to be uploaded in raw HapMap format to the Haploview program, and the linkage disequilibrium patterns and number of haplotype blocks in each gene can be calculated. The two most common measures are the absolute value of  $D'$  and  $r^2$ . The absolute value of  $D'$  is determined by dividing  $D$  by its maximum possible value for the given allele frequencies at two loci. The case of  $D' = 1$  is known as complete LD, and the values of  $D' < 1$  indicate that the

complete ancestral LD has been disrupted. The magnitude of values of  $D' < 1$  has no clear interpretation (Lewontin 1964; Hill and Robertson 1968; Nachman 2002). Estimates of  $D'$  are strongly inflated in small samples. Therefore, statistically significant values of  $D'$  near one provide a useful indication of minimal historical recombination, but intermediate values should not be used to compare the strength of LD between studies or to measure the extent of LD. The measurement of  $r^2$  is in some ways complementary to  $D'$ .  $r^2$  is equal to  $D^2$  divided by the product of the allele frequencies at the two loci.  $r^2 > 0.9$  is complete linkage disequilibrium, and  $< 0.9$  has no significance. Hill and Robertson (1968) deduced that  $E[r^2] = 1/1 + 4Nc$ , where  $c$  is the recombination rate between the two markers, and  $N$  is the effective population size. This equation illustrates two important properties of LD. First, expected levels of LD are a function of recombination. The more recombination between two sites, the more they are shuffled with respect to one another, decreasing LD. Second, LD is a function of  $N$ , emphasizing that LD is a property of populations. To arrive at this equation, Hill and Robertson assumed that the population was an “ideal,” large, random-mating population without natural selection and mutation. Another approach for quantifying LD is through the population recombination parameter  $4N_e c(\rho)$ . This approach avoids reliance on pairwise measures of LD, which differ from marker to marker, and facilitates comparisons between regions. It shows considerable promise for quantifying the strength of LD in a region. The number of haplotype blocks, haplotype tag SNPs and their  $D'$ ,  $r^2$  values in the genes associated with congenital long QT syndrome are depicted in Table 6. Among the five

**Table 6** Number of haplotype blocks and haplotype tag SNPs in the genes associated with congenital long QT syndrome (LQTS)

Gene ID	Number of haplotype blocks	rs IDs	Nucleotide change	Functional role	Linkage disequilibrium	
					$D'$	$r^2$
KCNQ1	26	rs10741726	G → T	Intron	1.0	0.96
		rs11024304	C → T	Intron	1.0	0.9
		rs2237898	A → C	Intron	1.0	0.9
		rs2237899	C → T	Intron	1.0	0.93
		rs10798	A → G	3' UTR	1.0	0.96
		rs8234	A → G	3' UTR	1.0	0.92
KCNH2	4	rs3807375	A → G	Intron	1.0	0.9
SCN5A	10	rs2072413	A → G	Intron	0.547	0.71
		rs6795580	C → G	Intron	0.104	0.64
		rs6599229	A → G	Intron	0.599	0.72
		rs7427106	C → G	Intron	1.0	0.97
		rs6768664	A → C	Intron	1.0	0.97
KCNE1 and KCNE2	1	rs2834485	A → G	3' UTR	1.0	0.96
		rs11702354	A → G	3' UTR	1.0	0.96
		rs9305548	C → T	Intron	0.871	0.71
		rs9984281	A → G	Intron	0.871	0.59

selected genes, *KCNQ1* has the highest number of haplotype blocks with 26, as well as the maximum number of 6 tag SNPs with complete  $D'$  and  $r^2$  values. Among the six tag SNPs, rs10798 and rs8234 SNPs were present in the 3' UTR and predicted to be deleterious SNPs by SIFT and Polyphen programs also (Tables 2, 3). Remaining tag SNPs in *KCNQ1* are present in the intron regions. The gene *SCN5A* has ten haplotype blocks and four tag SNPs, out of which two tag SNPs, namely rs6795580 and rs6599229, have the ancestral LD disrupted values, and the other SNPs, rs7427106 and rs6768664, have complete  $D'$  and  $r^2$  values. All four tag SNPs are present in the intron regions. Since both *KCNE1* and *KCNE2* genes are present in the same 21st chromosome, similar haplotype block and tag SNPs were obtained. Out of four tag SNPs, two are present in the intron region, and the other two are present in the 3' UTR. The SNPs rs2834485 and rs11702354 show complete LD values, and rs9305548 and rs9984281 have the ancestral LD disrupted values. Four haplotype blocks and two tag SNPs, namely rs3807375 and rs2072413, are present in the *KCNH2* gene.

### Copy number variation assessment

Copy number variation (CNV) assessment should now become standard in the design of all studies of the genetic basis of phenotypic variation, including disease susceptibility. CNV in the human genome takes many forms, ranging from large, microscopically visible chromosome anomalies to single nucleotide changes. Recently, multiple studies have discovered an abundance of submicroscopic copy number variations of DNA segments ranging from kilobases (kb) to megabases (Mb) in size (Iafate et al. 2004; Sebat et al. 2004; Sharp et al. 2005; Tuzun et al. 2005). Deletions, insertions, duplications and complex multi-site variants (Fredman et al. 2004), collectively termed copy number variations (CNVs) or copy number polymorphisms (CNPs), are found in all humans (Feuk et al. 2006) and other mammals (Freeman et al. 2006) as well. Copy number variation (CNV) of DNA sequences is

functionally significant, but has yet to be fully ascertained. A CNV can be simple in structure, such as tandem duplication, or may involve complex gains or losses of homologous sequences at multiple sites in the genome. Most CNVs are benign variants that will not directly cause disease. However, there are several instances where CNVs that affect critical developmental genes do cause disease. Since the discovery of CNVs is so new, bioethics studies are just now underway. Compared to other genetic variants, CNVs are larger in size and can often involve complex repetitive DNA sequences. They can also encompass entire genes, many of which have a specific function ascribed to them. For these reasons CNV data could potentially be more amenable to misinterpretation. Some CNVs could be employed to add discrimination power in forensics, but typing them is usually less efficient than other types of genetic markers. As with all types of genetic variation, CNVs can vary in frequency and occurrence between populations. As a result of recent common origin, the vast majority of copy-number variations—around 89%—is shared among the diverse human populations studied. Copy number variations (CNVs) can be retrieved from the Database of Genomic Variants. There were three copy number variants, namely Variation\_3710, Variation\_22718 and Variation\_38031, which are present at the cytogenetic band of 7q36.1 in *KCNH2* gene. The genes *KCNE1* and *KCNE2* are present in the same 21st chromosome and have two variants, namely Variation\_34534 and Variation\_26957, at the cytogenetic bands between 21q22.11 and 21q22.12. The variants Variation\_29897 and Variation\_36146 are present in *KCNQ1* and *SCN5A* genes at the cytogenetic bands of 11p15.5 and 3p22.2, respectively. The details of copy number variations are depicted in Table 7.

### Expressed sequence tag and genome survey sequence database screening

The human expressed sequence tag (EST) database provides a wealth of resources, which can be used to rapidly

**Table 7** Copy number variations (CNVs) in the genes associated with congenital long QT syndrome (LQTS)

Gene ID	NCBI copy number variation ID	Genomic position	Cytogenetic band	References
KCNQ1	Variation_29897	chr11:2,486,738..2,516,911	11p15.5	Jakobsson et al. (2008)
KCNH2	Variation_3710	chr7:149,882,882..150,094,737	7q36.1	Redon et al. (2006)
	Variation_22718	chr7:149,931,990..149,938,943	7q36.1	Korbel et al. (2007)
	Variation_38031	chr7:149,916,734..149,932,502	7q36.1	McCarroll et al. (2008)
SCN5A	Variation_36146	chr3:38,587,021..38,616,870	3p22.2	Kidd et al. (2008)
KCNE1 and KCNE2	Variation_34534	chr21:34,648,096..34,829,283	21q22.11–21q22.12	Zogopoulos et al. (2007), Levy et al. (2007)
	Variation_26957	chr21:34,700,265..34,700,265		

screen for potential polymorphisms in proteins of physiological interest. The human expressed sequence tags (ESTs) database consists of >3,700,000 entries of partial cDNA sequences. These sequences have been generated from many different tissues and are derived from a range of individuals. ESTs can reflect a part or all of the transcribed sequence of a gene, which includes the coding sequences as well as the 5' and 3' un-translated regions (UTRs). Currently, the ESTs database is accessible online from the website of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/dbEST/>). Database screening can be performed using gapped BLAST programs (Ulrich et al. 2000), which are obtainable from the homepage of the NCBI (Altschul et al. 1997). The genome survey sequence (GSS) division of GenBank is similar to the EST division, with the exception that most of the sequences are genomic in origin, rather than cDNA (mRNA). It should be noted that two classes (exon trapped products and gene trapped products) may be derived via cDNA intermediate. Although dbGSS sequences are incorporated into the GSS Division of GenBank, annotation in dbGSS is more comprehensive and includes detailed information about the contributors, experimental conditions and genetic map locations. The EST and GSS data were retrieved from dbEST database and dbGSS database, respectively, for the selected five genes, and the information is given in Table 8.

## Summary and conclusions

The congenital long QT syndrome is a potentially life-threatening condition caused by mutations in genes encoding cardiac ion channels. The genes, namely *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1* and *KCNE2*, encoding cardiac ion channels with a potential role for the cause of LQTS were investigated by evaluating the influence of functional SNPs through computation methods. Although the literature survey showed that there is a wide range of material on these genes related to LQTS, there have been no computational studies undertaken for an investigation of the nsSNP mutations. Of the total 3,309 SNPs in all five genes, 27 nsSNPs by SIFT and PolyPhen programs, and 44 SNPs by FastSNP and UTR Scan programs were found to have functional significance. Of 27 functionally significant nsSNPs in the coding region, 5 nsSNPs belong to the *KCNQ1* gene, 7 belong to the *KCNH2* gene, 10 belong to the *SCN5A* gene, 2 belong to the *KCNE1* gene, and 3 belong to the *KCNE2* gene. In 44 functionally significant SNPs in the 5' and 3' UTR regions, 14 belong to the *KCNQ1* gene, 7 belong to the *KCNH2*, 11 belong to the *SCN5A*, 11 belong to the *KCNE1*, and 1 belongs to the *KCNE2* gene. Among the five selected genes, *KCNQ1* has the highest number of haplotype blocks of 26 and 6 tag SNPs with a complete LD value of 1.0. Among the six tag SNPs, rs10798 and rs8234 SNPs were also predicted to be deleterious SNPs by the SIFT and PolyPhen programs. The

**Table 8** Database information on expressed sequence tag (EST) and genome survey sequences (GSS) of genes associated with congenital long QT syndrome (LQTS)

Gene ID	Expressed sequence tag (EST)				Genome survey sequences (GSS)			
	dbEST Id	EST name	GeneBank Acc	GenBank gi	dbGSS Id	GSS name	GeneBank Acc	GenBank gi
KCNQ1	26466745	PDUts1166B09	CV877550	84131510	14246920	AY404834	AY404834	39760811
	26460300	PDUts1087E06	CV871105	84125065	14246919	AY404833	AY404833	39760810
	26456223	PDUts1037G10	CV867028	84120988	14246918	AY404832	AY404832	39760809
KCNH2	34047624	ILLUMIGEN_MCQ_68905	DV769844	82696501	14245708	AY407435	AY407435	39763406
	28907670	TC124103	DR003775	66263648	14245707	AY407434	AY407434	39763405
	28907655	TC123979	DR003760	66263633	14245706	AY407433	AY407433	39763404
	28902580	TC109366	DN998684	66258511				
	24623702	ILLUMIGEN_MCQ_48787	CO580350	50411474				
SCN5A	28907665	TC124054	DR003770	66263643	14279954	DQ052008	DQ052008	66905428
KCNE1	EST not available				14279953	DQ052007	DQ052007	66905427
					14253844	AY419426	AY419426	39775383
					14253843	AY419425	AY419425	39775382
					14253842	AY419424	AY419424	39775381
KCNE2	EST not available				14253841	AY419423	AY419423	39775380
					14253840	AY419422	AY419422	39775379
					14253839	AY419421	AY419421	39775378

gene *SCN5A* has ten haplotype blocks and four tag SNPs. Both *KCNE1* and *KCNE2* genes have similar haplotype block and tag SNPs. Two tag SNPs and four haplotype blocks were obtained for the *KCNH2* gene. Results on these five ion channel genes provide excellent insight into the disease causing functional and haplotype tag SNPs related to LQTS. The reported data indicate that bioinformatic tools are indeed useful in predicting the functional impact of SNPs. Although these algorithms have been developed based on empirical data, correlations between predictive scores and findings from human studies have not been explored, with the exception of SIFT. The results of this study have therefore provided novel evidence of the correlation using human data, in turn facilitating genotyping efforts in future molecular epidemiological studies and providing targets for phenotypic analysis of genetic variants. These results can also be used to refine the bioinformatic algorithms. Also, these findings warrant a more comprehensive approach and more available bioinformatic tools in future analyses.

Genetic polymorphisms in the human population have been studied in order to gain insight into their influence on the activity of specific genes involved in disease susceptibility. Finding previously unknown polymorphisms has often relied on the detection of a related phenotype. This is a time-consuming task, usually requiring months or years at the bench to identify a novel polymorphism. Moreover, many polymorphisms may exist in the human genome that have not been identified and characterized because of problems of methodology. The use of computational algorithms and human genomic variation databases to find novel genetic polymorphism provides an alternative opportunity to investigate the consequences of polymorphism on the gene and the protein activity. The genetic screening of symptomatic patients or asymptomatic family members may identify patients at risk for life-threatening congenital long QT syndromes. More specifically, the mutation carriers without symptoms or ECG characteristics of the congenital long QT syndrome are at great risk. The family members considered normal on clinical and ECG grounds could be silent gene carriers displaying a very mild phenotype. They would be unexpectedly at risk for generating affected offspring and also for developing arrhythmias if exposed to either cardiac or noncardiac drugs that block potassium channels. Molecular screening is therefore recommended in all family members of positively genotyped patients. This report will surely help for such molecular screening work. Also, recent studies suggest that genotype-specific treatment of the congenital long QT syndrome will be feasible in the near future. These results, based on the application of computational tools, such as SIFT, PolyPhen, FASTSNP, UTR Scan and Haploview analysis, might provide an excellent approach to selecting target SNPs in genotype-

specific treatment of congenital long QT syndrome. Also, the applications of these computational algorithms in association studies will greatly strengthen the understanding of inheritance of complex human phenotypes. Therefore, this kind of analysis will provide useful information in selecting SNPs that are likely to have potential functional impact and ultimately contribute to an individual's susceptibility to LQTS by the *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1* and *KCNE2* genes.

**Acknowledgments** The authors thank the management of Vellore Institute of Technology, Vellore, for providing the facilities to carry out this work. The authors also heartily thank the Editor-in-chief and the reviewers for their valuable suggestions in the improvement of this manuscript.

## References

- Abbott GW, Sesti F, Splawski I et al (1999) MiRP1 forms Ikr potassium channels with HERG and is associated with cardiac arrhythmia. *Cell* 97:175–187. doi:10.1016/S0092-8674(00)80728-X
- Altschul SF, Madden TL, Schaffer A et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. doi:10.1093/nar/25.17.3389
- Barhanin J, Lesage F, Guillemare E et al (1996) KvLQT1 and IsK (mink) proteins associate to form the Iks cardiac potassium current. *Nature* 384:78–80. doi:10.1038/384078a0
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265. doi:10.1093/bioinformatics/bth457
- Bazett HC (1920) An analysis of the time relationship of electrocardiograms. *Heart* 7:353–370
- Benson DW, MacRae CA, Vesely MR et al (1996) Missense mutation in the pore region of HERG causes familial long QT syndrome. *Circulation* 93:1791–1795
- Benson DW, Wang DW, Dyment M et al (2003) Congenital sick sinus syndrome caused by recessive mutations in the cardiac sodium channel gene (*SCN5A*). *J Clin Invest* 112:1019–1028
- Cargill M, Altshuler D, Ireland J et al (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238. doi:10.1038/10290
- Curran ME, Splawski I, Timothy KW et al (1995) A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* 80:795–803. doi:10.1016/0092-8674(95)90358-5
- Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7:85–97. doi:10.1038/nrg1767
- Fredman D, White SJ, Potter S et al (2004) Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet* 36:861–866. doi:10.1038/ng1401
- Freeman JL, George HP, Feuk Lars et al (2006) Copy number variation: new insights in genome diversity. *Genome Res* 16:949–961. doi:10.1101/gr.3677206
- Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
- Gibbs RAJW, Belmont P, Hardenbol TD et al (2003) The international HapMap project. *Nature* 426:789–796. doi:10.1038/nature02168
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231. doi:10.1007/BF01245622

- Iafraite AJ, Feuk L, Rivera MN et al (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951. doi:[10.1038/ng1416](https://doi.org/10.1038/ng1416)
- Jakobsson M, Scholz SW, Scheet P (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003. doi:[10.1038/nature06742](https://doi.org/10.1038/nature06742)
- Jentsch TJ (2000) Neuronal KCNQ potassium channels: physiology and role in disease. *Nat Rev Neurosci* 1:21–30. doi:[10.1038/35036198](https://doi.org/10.1038/35036198)
- Jiang C, Atkinson D, Towbin JA et al (1994) Two long QT syndrome loci map to chromosomes 3 and 7 with evidence for further heterogeneity. *Nat Genet* 8:141–147. doi:[10.1038/ng1094-141](https://doi.org/10.1038/ng1094-141)
- Johnson GCL, Esposito L, Barratt BJ et al (2001) High-resolution haplotype structure in the human genome. *Nat Genet* 29:233–237
- Jurkat-Rott K, Lehmann-Horn F (2001) Human muscle voltage-gated ion channels and hereditary disease. *Curr Opin Pharmacol* 1:280–287. doi:[10.1016/S1471-4892\(01\)00050-9](https://doi.org/10.1016/S1471-4892(01)00050-9)
- Kaab S, Hinterseer M, Nabauer M et al (2003) Sotalol testing unmasks altered repolarization in patients with suspected acquired long-QT syndrome—a case-control pilot study using i.v. sotalol. *Eur Heart J* 24:649–657. doi:[10.1016/S0195-668X\(02\)00806-0](https://doi.org/10.1016/S0195-668X(02)00806-0)
- Kaspar RL, Kakegawa T, Cranston H et al (1992) A regulatory cis element and a specific binding factor involved in the mitogenic control of murine ribosomal protein L32 translation. *J Biol Chem* 267:508–514
- Kato S, Sekine S, Oh SW et al (1994) Construction of a human full-length cDNA bank. *Gene* 150:243–250. doi:[10.1016/0378-1119\(94\)90433-2](https://doi.org/10.1016/0378-1119(94)90433-2)
- Keating M, Atkinson D, Dunn C et al (1991) Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey ras-1 gene. *Science* 252:704–706. doi:[10.1126/science.1673802](https://doi.org/10.1126/science.1673802)
- Kidd JM, Cooper GM, Donahue WF (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64. doi:[10.1038/nature06862](https://doi.org/10.1038/nature06862)
- Korbel JO, Urban AE, Affourtit JP (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318:420–426. doi:[10.1126/science.1149504](https://doi.org/10.1126/science.1149504)
- Kullmann DM (2002) The neuronal channelopathies. *Brain* 125:1177–1195. doi:[10.1093/brain/awf130](https://doi.org/10.1093/brain/awf130)
- Lai EC, Bodner R, Kavalier J, Freschi G et al (2000) Antagonism of Notch signaling activity by members of a novel protein family encoded by the bearded and enhancer of split gene complexes. *Development* 127:291–306
- Laitinen P, Fodstad H, Piippo K et al (2000) Survey of the coding region of the HERG gene in long QT syndrome reveals six novel mutations and an amino acid polymorphism with possible phenotypic effects. *Hum Mutat* 15:580–581. doi:[10.1002/1098-1004\(200006\)15:6<580::AID-HUMU16>3.0.CO;2-0](https://doi.org/10.1002/1098-1004(200006)15:6<580::AID-HUMU16>3.0.CO;2-0)
- Larsen LA, Svendsen IH, Jensen AM et al (2000) Long QT syndrome with a high mortality rate caused by a novel G572R missense mutation in KCNH2. *Clin Genet* 57:125–130. doi:[10.1034/j.1399-0004.2000.570206.x](https://doi.org/10.1034/j.1399-0004.2000.570206.x)
- Le SY, Maizel JV (1997) A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucleic Acids Res* 25:362–369. doi:[10.1093/nar/25.2.362](https://doi.org/10.1093/nar/25.2.362)
- Legato MJ (2000) Gender and the heart: sex-specific differences in normal anatomy and physiology. *J Gend Specif Med* 3:15–18
- Levy S, Avni D, Hariharan N et al (1991) Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc Natl Acad Sci USA* 88:3319–3323. doi:[10.1073/pnas.88.8.3319](https://doi.org/10.1073/pnas.88.8.3319)
- Levy S, Sutton G, Ng PC (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:254. doi:[10.1371/journal.pbio.0050254](https://doi.org/10.1371/journal.pbio.0050254)
- Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49:49–67
- Lossin CDW, Wang TH, Rhodes CG et al (2002) Molecular basis of an inherited epilepsy. *Neuron* 34:877–884. doi:[10.1016/S0896-6273\(02\)00714-6](https://doi.org/10.1016/S0896-6273(02)00714-6)
- Marban E (2002) Cardiac channelopathies. *Nature* 415:213–218. doi:[10.1038/415213a](https://doi.org/10.1038/415213a)
- Mathe E, Olivier M, Kato S et al (2006) Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res* 34:1317–1325. doi:[10.1093/nar/gkj518](https://doi.org/10.1093/nar/gkj518)
- McCarroll SA, Kuruvilla FG, Korn JM (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 40:1166–1174. doi:[10.1038/ng.238](https://doi.org/10.1038/ng.238)
- Meyuhos O, Avni D, Shama S (1996) Translational control of ribosomal protein mRNAs in eukaryotes. Translational control. Cold Spring Harbor, Laboratory press, USA, pp 363–368
- Millat G, Chevalier P, Restier-Miron L et al (2006) Spectrum of pathogenic mutations and associated polymorphisms in a cohort of 44 unrelated patients with long QT syndrome. *Clin Genet* 70:214–227. doi:[10.1111/j.1399-0004.2006.00671.x](https://doi.org/10.1111/j.1399-0004.2006.00671.x)
- Mohler PJ, Schott JJ, Gramolini AO et al (2003) Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* 421:634–639. doi:[10.1038/nature01335](https://doi.org/10.1038/nature01335)
- Nachman MW (2002) Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev* 12(6):657–663. doi:[10.1016/S0959-437X\(02\)00358-1](https://doi.org/10.1016/S0959-437X(02)00358-1)
- Nagasaka M, Yokosuka H, Yamanaka T (1972) QT duration and plasma electrolytes (Ca, Na, and K) in uremic patients. *Jpn Heart J* 13:187–194
- Ng CP, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11:863–874. doi:[10.1101/gr.176601](https://doi.org/10.1101/gr.176601)
- Ng CP, Henikoff S (2002) Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 12:436–446. doi:[10.1101/gr.212802](https://doi.org/10.1101/gr.212802)
- Nowak R (1994) Mining treasures from 'junk DNA'. *Science* 263:608–610. doi:[10.1126/science.7508142](https://doi.org/10.1126/science.7508142)
- Ostareck DH, Ostareck-Lederer AM, Wilm BJ et al (1997) Silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-Lipoxygenase translation from the 3' end. *Cell* 89:597–606
- Ostareck-Lederer A, Ostareck DH, Standart N et al (1994) Translation of 15-lipoxygenase mRNA is inhibited by a protein that binds to a repeated sequence in the 3' untranslated region. *EMBO J* 13:1476–1481
- Ostareck-Lederer A, Ostareck DH, Hentze MW (1998) Cytoplasmic regulatory functions of the KH-domain protein hnRNPs K and E1/E2. *Trends Biochem Sci* 23:409–411. doi:[10.1016/S0968-0004\(98\)01301-2](https://doi.org/10.1016/S0968-0004(98)01301-2)
- Parsch J, Stephan W, Tanda S (1999) A highly conserved sequence in the 3'-untranslated region of the drosophila Adh gene plays a functional role in Adh expression. *Genetics* 151(2):667–674
- Parsch J, Russell JA, Beerman I et al (2000) Deletion of a conserved regulatory element in the Drosophila Adh gene leads to increased alcohol dehydrogenase activity but also delays development. *Genetics* 156:219–227
- Patil N, Bero AJ, Hinds DA et al (2001) Blocks of limited haplotype diversity revealed by high resolution scanning of human chromosome 21. *Science* 294:1719–1722
- Pesole G, Liuni S (1999) Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNA. *Trends Genet* 15:378. doi:[10.1016/S0168-9525\(99\)01795-3](https://doi.org/10.1016/S0168-9525(99)01795-3)
- Pesole G, Liuni S, Grillo G et al (2002) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 30:335–340. doi:[10.1093/nar/30.1.335](https://doi.org/10.1093/nar/30.1.335)

- Prokunina L, Alarcón-Riquelme ME (2004) Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Rev Mol Med* 6:1–15
- Prokunina L, Castillejo-Lopez C, Oberg F et al (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* 32:666–669. doi:10.1038/ng1020
- Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30:3894–3900. doi:10.1093/nar/gkf493
- Reardon M, Malik M (1996) QT interval change with age in an overtly healthy older population. *Clin Cardiol* 19:949–950
- Redon R, Ishikawa S, Fitch KR (2006) Global variation in copy number in the human genome. *Nature* 444:444–454. doi:10.1038/nature05329
- Schork NJ, Fallin D, Lanchbury JS (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet* 58:250–264. doi:10.1034/j.1399-0004.2000.580402.x
- Schott JJ, Charpentier F, Peltier S et al (1995) Mapping of a gene for long QT syndrome to chromosome 4q25–27. *Am J Hum Genet* 57:1114–1122
- Schulze-Bahr E, Wang Q, Wedekind H et al (1997) KCNE1 mutations cause Jervell and Lange-Nielsen syndrome. *Nat Genet* 17:267–268. doi:10.1038/ng1197-267
- Schwake M, Friedrich MT, Jentsch TJ (2001) An internalization signal in CIC-5, an endosomal Cl-channel mutated in dent's disease. *J Biol Chem* 276:12049–12054
- Sebat J, Lakshmi B, Troge J et al (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528. doi:10.1126/science.1098918
- Sharp AJ, Locke DP, McGrath SD et al (2005) Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77:78–88. doi:10.1086/431652
- Smits JPP, Eckardt L, Probst V et al (2002) Genotype–phenotype relationship in Brugada syndrome: electrocardiographic features differentiate SCN5A-related patients from non-SCN5A-related patients. *J Am Coll Cardiol* 40:350–356. doi:10.1016/S0735-1097(02)01962-9
- Sonenberg N (1994) mRNA translation: influence of the 5' and 3' untranslated regions. *Curr Opin Genet Dev* 4:310–315. doi:10.1016/S0959-437X(05)80059-0
- Splawski I, Shen J, Timothy KW et al (1998) Genomic structure of three long QT syndrome genes: KVLQT1, HERG, and KCNE1. *Genomics* 51:86–97. doi:10.1006/geno.1998.5361
- Splawski I, Shen J, Timothy KW et al (2000) Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation* 102:1178–1185
- Sunyaev S, Ramensky V, Bork P (2000) Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* 16:198–200. doi:10.1016/S0168-9525(00)01988-0
- The MGC Project Team (2004) The status, quality, and expansion of the NIH full-length cDNA project: the mammalian gene collection (MGC). *Genome Res* 14:2121–2127. doi:10.1101/gr.2596504
- Tomaselli GF, Marban E (1999) Electrophysiological remodeling in hypertrophy and heart failure. *Cardiovasc Res* 42:270–283. doi:10.1016/S0008-6363(99)00017-6
- Tranebjaerg L, Bathen J, Tyson J et al (1999) Jervell and Lange-Nielsen syndrome: a Norwegian perspective. *Am J Med Genet* 89:137–146. doi:10.1002/(SICI)1096-8628(19990924)89:3<137::AID-AJMG4>3.0.CO;2-C
- Tuzun E, Sharp AJ, Bailey JA et al (2005) Fine-scale structural variation of the human genome. *Nat Genet* 37:727–732. doi:10.1038/ng1562
- Ulrich CM, Bigler J, Velicer CM et al (2000) Searching expressed sequence tag databases: discovery and confirmation of a common polymorphism in the Thymidylate synthase gene. *Cancer Epidemiol Biomarkers Prev* 9:1381–1385
- Van Deventer S (2000) Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Med* 26:S98–S102. doi:10.1007/s001340051125
- Vassalli J-D, Stutz A (1996) Translational control: awakening dormant mRNAs. *Curr Biol* 5:476–479. doi:10.1016/S0960-9822(95)00095-9
- Verrotti AC, Thompson SR, Wreden C et al (1996) Evolutionary conservation of sequence elements controlling cytoplasmic polyadenylation. *Proc Natl Acad Sci USA* 93:9027–9032. doi:10.1073/pnas.93.17.9027
- Wheeler DL, Barrett T, Benson DA et al (2006) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 34:D173–D180. doi:10.1093/nar/gkj158
- Yang H, Elko P, LeCarpentier GL et al (1994) Sex differences in the rate of cardiac repolarization. *J Electrocardiol* 27:72–73. doi:10.1016/S0022-0736(94)80052-9
- Yang P, Kanki H, Drolet B et al (2002) Allelic variants in long-QT disease genes in patients with drug-associated torsades de pointes. *Circulation* 105:1943–1948. doi:10.1161/01.CIR.000.014448.19052.4C
- Yip YL, Scheib H, Diemand AV et al (2004) The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. *Hum Mutat* 23:464–470. doi:10.1002/humu.20021
- Yoshida H, Horie M, Otani H et al (2001) Bradycardia-induced long QT syndrome caused by a de novo missense mutation in the S2–S3 inner loop of HERG. *Am J Med Genet* 98:348–352. doi:10.1002/1096-8628(20010201)98:4<348::AID-AJMG1109>3.0.CO;2-A
- Yuan HY, Chiou JJ, Tseng WH et al (2006) FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* 34:W635–W641. doi:10.1093/nar/gkl236
- Zogopoulos G, Ha KC, Naqib F (2007) Germ-line DNA copy number variation frequencies in a large North American population. *Hum Genet* 122:345–353. doi:10.1007/s00439-007-0404-5