Author Manuscript

Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics

Alicia Lundby[*,1,2,3], **Elizabeth J. Rossin**[*,3,4,5], **Annette B. Steffensen**[2], **Moshe Rav Acha**[6,7], **Christopher Newton-Cheh**[3,6,7], **Arne Pfeufer**[8], **Stacey N. Lynch**[6,7], **The QT-IGC Consortium**, **Søren-Peter Olesen**[2], **Søren Brunak**[1,9], **Patrick T. Ellinor**[6,7], **J.Wouter Jukema**[10,11,12], **Stella Trompet**[10,13], **Ian Ford**[14], **Peter W. Macfarlane**[15], **Bouwe P. Krijthe**[16,17], **Albert Hofman**[16,17], **Andre G. Uitterlinden**[16,17,18], **Bruno H. Stricker**[16,17,18,19,20,21], **Hendrik M. Nathoe**[22], **Wilko Spiering**[21], **Mark J. Daly**[3,4], **Folkert W. Asselbergs**[22,23,24], **Pim van der Harst**[25], **David J. Milan**[7], **Paul I.W. de Bakker**[3,26,27,#], **Kasper Lage**[1,3,4,9,28,#], and **Jesper V. Olsen**[1,#]

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3b, Copenhagen, Denmark [2]The Danish National Research Foundation Centre for Cardiac Arrhythmia, Copenhagen N, Denmark [3]The Broad Institute, Cambridge, MA, USA [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA [5]Harvard Medical School, MD/PhD Program and Health Sciences and Technology Program, Boston, USA [6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, USA [7]Center for Human Genetics Research, Massachusetts General Hospital, Boston, USA [8]Institute of Human Genetics, Technische Universität München, Munich, Germany [9]Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark [10]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands [11]Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands [12]Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands [13]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands [14]Robertson Center for Biostatistics, University of Glasgow, United Kingdom [15]Institute of Cardiovascular and Medical Sciences, University of Glasgow, United Kingdom [16]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands [17]Netherlands Consortium for Healthy Aging

Correspondence: lage.kasper@mgh.harvard.edu or jesper.olsen@cpr.ku.dk.
*These authors contributed equally to this work: Alicia Lundby and Elizabeth J Rossin
#These authors also contributed equally to this work: Paul I. de Bakker and Kasper Lage and Jesper V. Olsen.

(NCHA), The Netherlands [18]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands [19]Inspectorate for Health Care, the Hague, The Netherlands [20]Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands [21]Department of Vascular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands [22]Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands [23]Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands [24]Institute of Cardiovascular Science, faculty of Population Health Sciences, University College London, London, United Kingdom [25]University Medical Center Groningen, University of Groningen, The Netherlands [26]Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands [27]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands [28]Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA, USA

Abstract

Genome-wide association studies (GWAS) have identified thousands of loci associated wtih complex traits, but it is challenging to pinpoint causal genes in these loci and to exploit subtle association signals. We used tissue-specific quantitative interaction proteomics to map a network of five genes involved in the Mendelian disorder long QT syndrome (LQTS). We integrated the LQTS network with GWAS loci from the corresponding common complex trait, QT interval variation, to identify candidate genes that were subsequently confirmed in *Xenopus laevis* oocytes and zebrafish. We used the LQTS protein network to filter weak GWAS signals by identifying single nucleotide polymorphisms (SNPs) in proximity to genes in the network supported by strong proteomic evidence. Three SNPs passing this filter reached genome-wide significance after replication genotyping. Overall, we present a general strategy to propose candidates in GWAS loci for functional studies and to systematically filter subtle association signals using tissue-specific quantitative interaction proteomics.

Introduction

General comment

Please keep the introduction to a general description of the rational and the method, no need to go into details of which programs were used. But please make it clear what the relationship between the LQTS genes used to construct the network and the associated genes are. As written – and not being from this field, thus a good example for a non-expert reader – it is not clear to me how the LQTS genes were determined and whether they overlap with any of the 35 loci found in the GWAS. While this is explained in the Results it should already be clear from the Introduction.

Do not call out Supplementary Information, but a particular file that should be part of the SI Titles you sent (13 SI figures and 13 tables).

SI file needs to be revised;

Move methods to the main text file here and delete from the SI file.

Legend for the SI tables must be included in the Excel files of each table

GWAS has been extremely successful in identifying loci associated with numerous diseases. However, for a locus identified in a given trait it remains a major challenge to systematically identify the specific gene involved in the phenotype especially if the biology of the trait in question involves completely uncharted or largely incomplete pathways. To address this issue we have developed an integrative approach combining GWAS data with quantitative interaction proteomics to facilitate the annotation of associated loci. We apply this strategy to identify candidates that represent critical regulators of the electrocardiographic QT interval (the time between the end of the T wave and the onset of the Q wave in an electrocardiogram depicting the heart's electrical cycle).

Prolongation of the electrocardiographic QT interval reflects abnormal myocardial repolarization and is a risk factor for sudden cardiac death and drug-induced arrhythmias. Long QT syndrome (LQTS) is a Mendelian disorder caused by genetic defects in one of 12 genes resulting in major prolongation of the QT interval (>40 msec)[1]. In addition, minor variation of the QT interval (≈1-4 msec per allele) is a quantitative heritable trait in the general population[2,3] and recently 35 single nucleotide polymorphisms (SNPs) significantly associated with this phenotype were identified[4]. Due to large spans of linkage disequilibrium in the genome, these SNPs represent 35 loci (termed "common variant loci" hereafter) encoding hundreds of genes. However, despite the fact that minor and major variations of the QT interval represent different ends of the spectrum of the same phenotype, no systematic approach has yet been employed to combine LQTS-informed and experimentally derived pathways with associated SNPs to get broader insight into the biology and genetic influences on cardiac repolarization in the general population.

Five of the 35 known common variant loci harbor Mendelian LQTS genes all of which are cardiac ion channels or proteins regulating the ion channel function (Fig. 1a). Because cardiac ion channels are thought to form large protein networks with hundreds of proteins regulting the channels' functions through static and transient physical interactions, we hypothesized that systematic pathway relationships between the associated loci could be deduced by analyzing the protein network of the proteins corresponding to LQTS genes (LQTS proteins hereafter). To test this hypothesis, we investigated the protein network of five LQTS proteins in heart tissue using quantitative interaction proteomics and integrated the network with GWAS data from an analysis of QT interval variation[4]. For breadth, we chose LQTS proteins that were both ion channels and regulators of ion channels, as well as proteins residing within and outside of the 35 common variant loci as the starting point of the proteomics experiments (Fig. 1a). We cross-referenced the resulting interaction network with the 35 established common variant loci associated with QT interval variation in the general population to propose specific candidates for functional validation. We also used the network data to filter sub genome-wide significant SNPs for replication genotyping (Fig. 1b). Overall, we expand our knowledge of the molecular components and genetic variants driving cardiac repolarization. Importantly, we provide a general strategy and analytical framework to annotate GWAS loci and filter weak association signals using tissue-specific quantitative interaction proteomics.

Results

Tissue-specific protein interaction network of LQTS genes

We chose five LQTS proteins as the starting point of our analysis (i.e., KCNQ1, KCNH2, CACNA1C, SNTA1, CAV3)[5–9]. The proteins were immunoprecipitated from pooled lysates of cardiac tissue from male mice, the precipitates were separated by SDS-PAGE followed by in-gel trypsin digestion and analysis of the resulting peptide mixtures by nanoflow high-performance liquid chromatography and subjected to tandem mass spectrometry (HPLC-MS/MS)[10–12] on a LTQ-Orbitrap Velos instrument using Higher-Collisional Dissociation (HCD) fragmentation (Supplementary Figures 1 to 5)[13]. The complete set of raw MS files were processed using the MaxQuant software suite (www.maxquant.org), where peptides and proteins were identified using the Andromeda search engine at a false discovery rate (FDR) below 0.01 and quantified using the label-free quantitation approach (all quantified proteins and modification specific peptides are provided in Supplementary Tables 1 and 2). We performed triplicate immunoprecipitations (IPs) of all LQTS proteins and compared them to matched IgG control IPs, separating specific from nonspecific interactors by applying a FDR cutoff of 0.05[10,14] (Fig. 2a and b). As expected, the experimental triplicates yielded highly reproducible results for protein signal intensities (Pearson r>0.8, Supplementary Figure 6), and the LQTS proteins were among the most abundant proteins in their respective protein networks (Fig 2b).

We identified 86 protein interactors of CACNA1C, 31 of KCNH2, 116 of KCNQ1, 104 of SNTA1 and 333 for CAV3 (Supplementary Tables 3-7), and we show that at most (**Online Methods**) of these proteins were nonspecific binders due to similarity of the LQTS proteins in terms of subcellular localization in the plasma membrane. Four of the five affinity purification datasets were enriched for known interaction partners[15,16] (KCNQ1, *P*= 6.0e-3; CACNA1C, *P* = 3.1e-5; CAV3, *P* = 8.9e-3; SNTA1, *P* = 5.0e-4, **Online Methods**), and the number of interacting proteins match those reported in an analysis of CAV2 channels in rat brain, where between 97 and 161 proteins interact specifically with the tested ion channels[17]. In addition, the specificity, robustness, and high quality of the data was confirmed by applying three alternative control procedures, which were not based on IgGs (**Online Methods** and Supplementary Figure 7), and by providing biological replication in five additional mouse hearts that had not been pooled (**Online Methods**, Supplementary Figs. 8 and 9). After making individual quality controls of the pull-down datasets, we pooled the interactions of all LQTS proteins to create an integrated LQTS protein network.

The LQTS protein network points to candidate genes in GWAS loci

A recent GWAS meta-analysis in >100,000 individuals of European ancestry identified 35 genome-wide significant (GWS) SNPs to be associated with QT interval variation in the general population[4], and the corresponding 35 common variant loci span 154 genes. A locus was defined by identifying neighbor SNPs in linkage disequilibrium (r^2>0.5) to the associated SNP and expanding to the nearest recombination hotspot as previously described[18]. Strikingly, excluding LQTS genes, twelve genes in the loci (*PLN, ATP1B1, UNC45B, TRAP1, TTN, CCDC141, ATP2A2, CAV1, CAV2, GOT2, ACTR1A, MYL3*) encoded proteins in the LQTS protein network derived here. The genes represent ten of the

35 genome-wide significant loci (probability of such enrichment is P = 1.3e-6 using random sampling taking into consideration locus architecture). As a control analysis, we made analogous IPs in cardiac tissue lysates of five heart proteins involved in cardiomyopathies (RYR2, ATP1A1, DSP, MYBPC3, and DMD, Supplementary Tables 9-13) and applied the same protocols used for the LQTS proteins to derive a cardiomyopathy network. When cross-referencing the genes represented in the control network with the 35 loci reported in the GWAS meta-analysis, there was not a significant enrichment (P = 0.17, using random sampling taking into consideration locus architecture), showing that the observed enrichment was specific to the LQTS protein network and was not driven by highly heart-expressed proteins that interact nonspecifically with the antibodies used in this study. Importantly, the enrichment in QT loci is specific to the LQTS protein network and not a feature of heart networks or networks involved in cardiac diseases in general. Therefore, our results provide a strong mechanistic link at the level of protein networks between genes in which rare mutations cause LQTS and 12 specific genes (in ten loci with a total of 79 genes, Supplementary Figure 10) definitively associated with modest QT interval variation in the general population.

Functional effects of candidate genes

ATP1B1 is encoded in a locus defined by rs10919070, the most associated SNP for QT interval variation (P = 1.11e-31). We showed that ATP1B1 interacts with KCNH2, CACNA1C, KCNQ1and CAV3. ATP1B1 is well-characterized as the β-subunit for the Na^+,K^+-ATPase heterodimer. However, the α-subunit (ATP1A1) was not enriched in the protein networks, suggesting an additional function of ATP1B1, which is independent of ATP1A1.We tested the effect of ATP1B1on the KCNH2 channel by electrophysiological measurements of heterologously expressed proteins in *Xenopus laevis* oocytes. Co-expression of ATP1B1 shifts the peak of the current-voltage relationship by 10 mV to more positive potentials, slows the channel inactivation kinetics, and right-shifts the voltage-dependence of recovery from inactivation (Fig. 3). The same effects are observed in the presence of an ATP1A1 inhibitor (Supplementary Figure 11). Interestingly, pull-down experiments of ATP1A1 revealed no interaction to the KCNH2 channel (Supplementary Table 10), and together these data show that ATP1B1 has a direct functional impact on the KCNH2 channel properties independent of ATP1A1. We therefore propose a biological mechanism through which common genetic variants near or in *ATP1B1* affect QT interval variation that has not previously been shown. To directly test the effect of *ATP1B1* on cardiac repolarization, we used optical voltage-mapping to probe cardiac electrophysiology of *ATP1B1* zebrafish knockdown animals, which are a well-established model of human cardiac repolarization[19]. Morpholino knockdown of the zebrafish ortholog for *ATP1B1* (*atp1b1a*) results in shorter action potential duration compared to wildtype (P = 0.002, Fig. 3). Together these results strongly support *ATP1B1* as a candidate gene in the rs10919070 locus for further follow-up, as suggested by its interaction to KCNH2 and three other LQTS proteins.

Filtering and augmenting subtle GWAS signals using the LQTS protein network

Similar to most other complex phenotypes, the SNPs associated with QT interval variation explain only a minority of the heritability of this trait in the population. To investigate

whether proteins in the LQTS network could be used to filter modestly associated SNPs and identify a subset that is likely to influence the phenotype in the population, despite not being significant in the GWAS. We excluded genes from the 35 loci definitively associated with QT interval variation and made a composite test of genetic association across the remaining genes represented in the LQTS network. We translated all identified mouse proteins to their orthologous human genes and derived a set of association Z-scores for each gene, taking SNP density and linkage disequilibrium across and surrounding each gene into consideration[18]. Using a one-tailed Mann-Whitney rank-sum test, we compared the distribution of association scores across genes represented in the protein networks to those for all genes in the genome. Even after excluding the 12 genes from the definitively associated loci, we found that the protein networks were significantly enriched for association to QT interval variation ($P = 1.5e-4$, using a one-tailed rank sum test, Supplementary Figure 12). This suggests that proteins in the networks point to genetic variants important for QT interval variation which have so far been missed.

We used a combination of genetic and proteomic evidence to select 28 SNPs represented by proteins in the networks for replication genotyping in four cohorts comprised of 17,692 independent samples in total. Specifically, SNPs were considered for replication genotyping if the association significance in the GWAS meta-analysis was $P<1e-3$ and a protein in the LQTS networks was encoded by a gene near the SNP. We also required that the protein pointing to the SNP was abundant in the relevant LQTS IP hereby suggesting it is an in important intrearction partner of a LQTS protein. The proteins that formed the basis for the SNP selection were then plotted as a network along with information on the LQTS proteins with which they interact (Fig 4a). Twenty five SNPs were successfully tested (see **Online Methods** for filtering procedure), 18 were directionally consistent (probability of such finding using the sign-test is $P = 0.02$), 7 were nominally significant in the replication cohort (probability of such finding using permutation testing is $P = 0.0003$), and 3 reached genome-wide significance when jointly analyzed with the recent GWAS meta-analysis (*VCL* – rs10824026, $P = 1.5e-9$; *SRL* – rs889807, $P = 1.2e-8$ and *TUFM/EIF3C/EIF3CL* – rs7498491, $P = 2.2e-8$, see Table 1 and SupplementaryNotes).

Interestingly, using the LQTS networks to guide replication experiments suggested new insight into the biology of cardiac repolarization. First, *SRL* encodes the sarcolemmal Ca^{2+} binding protein sarcalumenin, which regulates Ca^{2+} reuptake into the sarcoplasmic reticulum by interaction with the Ca^{2+}-ATPase 2 (ATP2A2 also known as SERCA2)[20]. The gene encoding ATP2A2 is itself in a locus significantly associated to QT interval variation (rs17483, 3×10^{-12})[4]. The importance of *SRL* in cardiac physiology is evident from knockout mice, in which ventricular depolarization is prolonged[20]. Our data show that the mouse orthologs for ATP2A2 and SRL both interact with CAV3, and that ATP2A2 also interacts with the LQTS calcium channel CACNA1C. Second, *VCL* encodes a cytoskeletal protein, vinculin, which we show interacts with CAV3 and SNTA1. Although vinculin has previously been related to dilated cardiomyopathy[21], it has never been found to be involved in QT interval variation. We furthermore confirmed the involvement of *VCL* in cardiac repolarization by knockdown experiments of the ortholog, *vcl*, in zebrafish, which had a direct effect on cardiac repolarization *in vivo* (Supplementary Figure 13). Knockdown of

zebrafish orthologs of *TUFM* or *EIF3C* did not affect the action potential duration (data not shown).

Thus, capialitizing on the LQTS protein network to filter modestly associated SNPs for replication genotyping we identified three novel loci associated with QT interval variation in the general population (Fig. 4b). For two of these loci functional *in vivo* evidence further supports the specific gene we prioritized as driving the association signal.

Discussion

The methodological approach we have developed represents a strategy to functionally annotate loci associated to a human trait through GWAS for which the causal genes have not been identified, and to augment and filter modestly associated common variants. While it has been shown previously that generic (i.e. non-tissue specific) *in silico* protein network analyses based on public data is a powerful tool in interpreting common variants associated to disease[18], this study represents an important advance by using targeted proteomics experiments in relevant tissue types[22,23] to firmly establish the molecular interactions between proteins in the relevant biological setting. In addition, to our knowledge, our proteomics dataset represents the first analysis of the composition of protein networks involved in rare Mendelian disease and its analogous common complex trait. Therefore, the methodological and statistical framework outlined here may be applicable to a number of other complex traits to propose candidate genes for validation in future genetic studies with the ultimate goal of elucidating underlying biological systems and the specific causal genetic determinants.

By testing the interaction networks of the five LQTS proteins one-by-one for genetic enrichment in the GWAS data (Supplementary Figure 12) we show that, while individual networks can yield statistically significant results, the power of our approach lies in the integrated LQTS network obtained by pooling data from all five pull-downs. We note that this might vary depending on the genetic power of the GWAS and it is not inconceivable that similarly good results could be obtained in other traits with fewer proteins as the starting point for the proteomics experiments. We also note that the approach outlined here is not limited to complex diseases with a corresponding Mendelian phenotype. In theory any protein known or hypothesized to be involed in the trait or biology of interest could be used as the starting point of the proteomic analysis.

An interesting biological observation from our analysis together with the recent GWAS meta-analysis[4] is the involvement of calcium signaling in cardiac repolarization which is suggested both from proteomics experiments, sequencing of LQTS patients, and meta-analyses of genome-wide association studies, which all converge on a cluster of physically interacting Ca^{2+} regulating proteins, thus providing new biological insight variations of the QT interval in humans.

A limitation of our approach is that we use mouse heart tissue as the molecular components of the biology of mouse and human cardiac repolarization might differ (see **Online Methods** for discussion). For this reason, we used a variety of validation experiments including large

and robust human genetic datasets and models systems widely accepted to be relevant to human heart biology, to augment, complement, support, and filter the proteomics data. These experiments firmly establish the value of the experimental and analytical framework delineated here to gain insight into underlying molecular mechanisms of a common complex human phenotype. This approach can be extended to other complex phenotypes to help elucidate underlying biology and pinpoint candidate genes.

Online Methods

Tissue preparation and immunoprecipitations

The study was carried out following the Guide for the Care and Use of Laboratory Animals published by the United States National Institutes of Health and the Directive 2010/63/EU of the European Parliament. 6-8 weeks old male mice of strain C57BL6 were sacrificed by cervical dislocation and their hearts were harvested and snap frozen in liquid nitrogen and stored at -80 °C. Heart tissue was homogenized on a Precellys 24 and solubilized in ice-cold lysis buffer containing protease and phosphatase inhibitors. Tissue lysates were centrifuged to remove insoluble debris. For each tissue preparation produced, lysates derived from 5 mice were pooled and protein concentrations were measured by Quick Start Bradford Dye Reagent (Biorad). Solubilized heart tissue lysate was pre-cleared with Dynabeads protein G (Invitrogen) before incubation with primary antibody followed by binding to Dynabeads protein G, using either anti-KCNQ1 (10 μl SC10646, Santa Cruz), anti-CACNA1C (2 μl AC003, Alomone), anti-KCNH2 (2 μl AC062, Alomone), anti-CAV3 (2 μl ab2912, Abcam), anti-SNTA1 (2 μl ab11425, Abcam) or control IgG (1.5 μl goat IgG: SC2028, 1.5 μl rabbit IgG: SC2027, 1.5 μl mouse IgG: SC2025, Santa Cruz). After washing, bound proteins were eluted with 1× sample buffer containing 100 mM dithiothreitol (70 °C, 3 min) and separated by SDS-PAGE (4-15 % Bis-Tris gels, BioRad).

In-gel digestion

Separated proteins were fixed in the gel (40 ml water, 50 ml acetonitrile, 10 ml acetic acid, 10 min) and visualized with colloidal Coomassie staining (Invitrogen). Each gel lane was excised and separated into four slices that were minced and destained (50 % 25 mM ammonium bicarbonate, 50 % acetonitrile) in a thermomixer (3 times 20 min, 800 rpm, room temperature (RT)). Gel dices were dehydrated (acetonitrile, 10 min, 800 rpm) followed by reduction of disulfide bonds (10 mM dithiothreitol in 25 mM ammonium bicarbonate, 45 min, RT, 800 rpm) and alkylation of cysteines (55 mM chloro-acetamide in 25 mM ammonium bicarbonate, 30 min, 24 °C in darkness, 800 rpm). After washing in 25 mM ammonium bicarbonate the gel plugs were dehydrated in acetonitrile and proteins were digested by trypsin (50 ul 12.5 ng/ul sequencing grade trypsin (Promega) in 25 mM ammonium bicarbonate for 1 hour, followed by addition of 100 ul 25 mM ammonium bicarbonate, left overnight at 37 °C). Trypsin activity was quenched by acidification of the mixture with trifluoroacetic acid to pH~2 and peptides were extracted from the gel plugs with 30 % acetonitrile in 3 % trifluoroacetic acid (30 min, 800 rpm) followed by 80 % acetonitrile in 0.5 % acetic acid (30 min, 800 rpm) and finally in 100 % acetonitrile[13]. Organic solvents were removed by evaporation in a vacuum centrifuge. Extracted peptides were purified on STAGE-tips with two C_{18} filters[24].

Mass-spectrometry, LC-MS/MS

Peptides were eluted from the STAGE tips into 96 well microtiterplates with 2×10 ul 40 % acetonitrile in 0.5 % acetic acid and the acetonitrile was evaporated using a vacuum centrifuge reducing the sample volume to 4 ul. The peptide mixtures were acidified with 0.1 % trifluoroacetic acid in 2 % acetonitrile to an end volume of 9 ul and analyzed by on-line nanoflow LC-MS/MS. Peptide separation was performed by reversed-phase C_{18} HPLC on an Easy nLC system (Thermo Fisher Scientific) loading 5 ul samples with a constant flow of 750 nl/min onto 15 cm long analytical columns, packed in-house with 3 um C_{18} beads, and eluting peptides using a 135 min segmented gradient of increasing (5 %-80 %) buffer B (80 % acetonitrile in 0.5 % acetic acid) at a constant flow of 250 nl/min. The effluent from the HPLC was directly electrosprayed into an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) through a nano-spray ion source. The peptide mixture was analyzed by full-scan MS spectra (*m/z* 300-2000, resolution 30,000) in the Orbitrap analyzer after accumulation of 1,000,000 ions in the Orbitrap within a maximum fill-time of 1.000 ms with the lock mass option enabled to improve mass accuracy[25]. For every full-scan the most intense peptide ions were sequentially isolated (up to ten for every full-scan) and fragmented by higher energy collisional dissociation (HCD) in the octopole collision cell and fragments were recorded by the Orbitrap mass analyzer after accumulation of 50,000 ions with a maximum fill-time of 250 ms and using a normalized collision energy of 40%.

Mass spectrometry data analysis

The acquired data was processed by MaxQuant (version 1.1.1.25) (Max-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Munich)[14], where peptides and proteins are identified by the Andromeda search algorithm via matching of all MS and MS/MS spectra against a target/decoy-version of the mouse IPI database v. 3.68 supplemented with reversed copies of all sequences as well as frequently observed contaminants. Maximal MS/MS tolerance was 20 ppm, a maximum of 2 missed cleavages was allowed and false discovery rates were set at 0.01 both for peptides and proteins. Carbamidomethylated cysteines were set as a fixed modification, whereas N-pyroglutamine, oxidation of methionine and N-terminal acetylation were searched as variable modifications. Minimum peptide length was set at 6 amino acids. Statistical evaluation and filtering of the resulting peptide datasets were performed in MaxQuant as previously described[14]. Protein intensities were normalized and proteins were quantified between control and case experiments by the MaxQuant label-free algorithm, resulting in LFQ (label-free quantitation) protein intensities. The downstream analysis was performed with Excel (Microsoft) and Perseus (Max-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Munich) software. The triplicates of each bait IP were analyzed against the five control IPs. Protein identifications were filtered for contaminants and reverse hits. A minimum of three peptide identifications with at least one being uniquely assigned to the particular protein, and protein identification in at least three immunoprecipitations were required followed by $\log_2$ transformation of the LFQ intensities. To perform statistical analysis of the label-free bait IP experiments versus control IP experiments normal distributed values were imputed for missing values using a normal distribution with width 0.3 and a downshift of the mean by 1.8 compared to distribution of

all LFQ intensities. t-test based comparison of bait IPs versus control IPs were performed to identify significant interactors with false discovery threshold set at 0.05 and a bend of the curve value, S0, of 1[26]. LFQ protein intensity ratios of bait relative to control was plotted against the negative logarithmic P-value of the t-test as was a stipulated line representing the permutation based false discovery rate separating specific from non-specific binders. Significant interactors of the bait proteins were color coded in green and the rest were color coded in blue. For the hierarchical clustering, LFQ intensities were Z-scored and average linkage clustering was performed using Euclidian distance, and protein LFQ intensities were color-coded with blue representing low intensities and yellow representing high intensities. In general, the reporting of our mass spectrometry data acquisition, processing and search results as well as sharing of all MS raw files have been done according to the *Molecular and Cellular Proteomics Guidelines*. Raw mass spectrometric files in Thermo Scientific's *.raw format are available for download through Tranche at http://proteomecommons.org using the following Hash-key:

UpjhtcVZMgE8uKwuMa6G2qQokoYYdAs2mxUAYJmrPD6HWggQ
+WLr3DoMRQaM3wyNWHjEmFyJqjIcWxioc9NVGIRub0oAAAAAAAACiA==

with password LQT1LQT2LQT8LQT9LQT12

Association analyses

QT-IGC[4]—The QT-IGC consortium consists of 48 cohorts of European ancestry with QT-interval and genome-wide genotype data (>100,000 individuals in total). Each cohort contributed GWAS results from a linear regression of original QT-interval on genotype using RR-interval, age and sex as covariates (individuals with QRS-duration > 120ms or history of MI were excluded). The summary statistics (betas, standard errors and p-values) on 2.5 million SNPs (either directly genotyped or imputed) were then combined in a meta-analysis using the software MANTEL[27]. The non-genomic-control-corrected results were used in this analysis to match what is reported in the accompanying QT-IGC study (λ_{GC}=1.069).

To test the joint set of proteins (737 proteins in total, 436 unique proteins) derived from all LQTS protein networks for containing more GWS hits than chance expectation, taking into account that multiple GWS proteins were represented in more than one network, we simulated 10,000,000 random selections of 5 networks (each of the same number of proteins represented in the individual networks) from all genes in the genome. For each random selection of 737 total proteins, we counted the number of GWS hits. We then report an empirical P-value for the probability of selecting 22 or more GWS hits (22 represent the fact that some of the 12 GWS proteins were representing in multiple networks). To derive a P-value for each individual network we performed a hypergeometric test, since we did not need to account for proteins being represented multiple times.

The joint test for enrichment in association performed on the remaining proteins in the complexes (those that did not achieve genome-wide significance) was carried out as described previously[18]. In order to control for linkage disequilibrium (LD) between genes, we broke the genome into LD blocks as defined by recombination hotspots. We then scored

each block with the best association Z score achieved over that block (association data was from the QTIGC meta-analysis)[4]. This score was then corrected for the number of SNPs tested in the block using linear regression in R. The residuals from the regression were used as the corrected scores for each block, and genes were assigned scores according to the blocks they overlap. To test a group of proteins for enrichment in association, we compared the unique set of scores derived from the group of proteins to the unique set of scores for all genes in the genome using a 1-tailed rank-sum test, with the alternative hypothesis being that the group of proteins has higher association scores than scores from all genes in the genome.

Assessing the contribution of heart expression to association results

Because regions of the genome associated to QT interval variation are likely to code for heart-expressed genes, we assessed the probability that our association results (number of GWS proteins represented in the LQTS networks as well as enrichment in sub-genome-wide scores) were due to enrichment for association in heart-expressed proteins rather than network-specific proteins. Based on organ-wide proteomic mapping of phosphoproteins in rat hearts[28], we collected a dataset of 2000 proteins expressed in heart tissue. We assessed the likelihood of identifying 22 GWS proteins in a random selection of 5 networks (each of the same number of proteins represented in the individual networks – 737 proteins in total). After 1,000,000 permutations, we found the probability of selecting >=22 proteins to be 0.00536.

Replication genotyping and analysis

We selected 28 SNPs to replicate that met the following criteria: they are in LD with a gene that codes for one of the proteins pulled down in the 5 complexes, and either their association P-value was <1e-4 (18 SNPs) or was <1e-3 and the protein of interest passed a threshold for being abundantly present in one of the complexes (4 proteins). The selected SNPs were then genotyped or looked up in four cohorts: 5,731 independent samples were genotyped in the SMART cohort, and betas, standard errors and P-values were collected for the from the LifeLines cohort (n=4,865), the POSPER/PHASE cohort (n=5,135) and the RS3 cohort (n=1,961), for which the QT interval duration had been measured (in milliseconds) but the results had not been included in the QT-IGC meta-analysis. Each analysis performed a linear regression of the original QT measurement on genotype using RR-interval, age and sex as covariates. Individuals with QRS duration > 120ms or positive history of myocardial infarction were removed.

Cohort descriptions

SMART[29]—The Secondary Manifestations of ARTerial disease study. SMART is a prospective cohort study among patients aged 18-74 years who are referred to the University Medical Center Utrecht, The Netherlands, because of atherosclerotic vascular disease or for treatment of atherosclerotic risk factors[30]. The objective of the SMART study is to determine the prevalence of asymptomatic arterial disease and risk factors in patients presenting with a manifestation of arterial disease or known risk factor, and to study future

cardiovascular events and their predictors in these at-risk patients. Wet-lab genotyping was carried out by KBiosciences, Hertfordshire, UK, using proprietary KASPar PCR technique.

LifeLines[31]—LifeLines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity and complex genetics. The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. We thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and Medical Biobank Northern Netherlands, and the participating general practitioners and pharmacists. LifeLines Scientific Protocol Preparation: Rudolf de Boer, Hans Hillege, Melanie van der Klauw, Gerjan Navis, Hans Ormel, Dirkje Postma, Judith Rosmalen, Joris Slaets, Ronald Stolk, Bruce Wolffenbuttel; LifeLines GWAS Working Group: Behrooz Alizadeh, Marike Boezen, Marcel Bruinenberg, Noortje Festen, Lude Franke, Pim van der Harst, Gerjan Navis, Dirkje Postma, Harold Snieder, Cisca Wijmenga, Bruce Wolffenbuttel.

PROSPER/PHASE[32,33]—All data come from the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER). A detailed description of the study has been published elsewhere. PROSPER was a prospective multicenter randomized placebo-controlled trial to assess whether treatment with pravastatin diminishes the risk of major vascular events in elderly. Between December 1997 and May 1999, we screened and enrolled subjects in Scotland (Glasgow), Ireland (Cork), and the Netherlands (Leiden). Men and women aged 70-82 years were recruited if they had pre-existing vascular disease or increased risk of such disease because of smoking, hypertension, or diabetes. A total number of 5,804 subjects were randomly assigned to pravastatin or placebo. A large number of prospective tests were performed including Biobank tests and cognitive function measurements. Resting 12 lead ECGs were recorded at baseline and annually thereafter and were analyzed using the University of Glasgow analysis program. A whole genome wide screening has been performed in the sequential PHASE project with the use of the Illumina 660K beadchip. Of 5,763 subjects DNA was available for genotyping. Genotyping was performed with the Illumina 660K beadchip, after QC (call rate <95%) 5,244 subjects and 557,192 SNPs were left for analysis. These SNPs were imputed to 2.5 million SNPs based on the HAPMAP built 36 with MACH imputation software. PROSPER is supported by an investigator initiated

RS3[34]—The Rotterdam Study III (RS-III) is a prospective population-based cohort study. The cohort comprises 3,932 subjects aged 45 years and older, living in the Ommoord district in Rotterdam, the Netherlands. The rationale and design of the RS have been described in detail elsewhere. The Medical Ethics Committee of Erasmus Medical Center approved the study and written consent was obtained from all participants. Electrocardiograms were recorder on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the QRS intervals were made using the Modular ECG Analysis System (MEANS). All RS-III participants with available DNA were genotyped using Illumina Human 610 Quad array at the Department of Internal Medicine, Erasmus Medical Center following manufacturer's protocols. Participants with call rate < 97.5%, excess autosomal heterozygosity, sex mismatch, or outlying identity-by-state clustering estimates were excluded. After quality control 2,082 RS-III participants were included. Of these, 1961 participants were included in this study. The Rotterdam Study (RS) is supported by the Erasmus Medical Center and Erasmus University Rotterdam; The Netherlands Organization for Scientific Research; The Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; The Netherlands Heart Foundation; the Ministry of Education, Culture and Science; the Ministry of Health Welfare and Sports; the European Commission; and the Municipality of Rotterdam. Support for genotyping was provided by The Netherlands Organization for Scientific Research (NWO) (175.010.2005.011, 911.03.012) and Research Institute for Diseases in the Elderly (RIDE).

For the SMART data (the only data which we received as raw genotypes), we ran a linear regression in Plink[35] to test for association to the duration of the QT interval in the same manner as was done in the QT-IGC meta-analysis as well as the other 3 cohorts, controlling for age, sex and RR-interval and excluding individuals with QRS duration > 120 or past history of MI.

The meta-analysis was done with the program METAL[27] using effect size estimates and standard errors. We removed 3 SNPs due to missing data in 3 of the 4 cohorts, resulting in a total of 25 SNPs analyzed. These results are reported in the main text and as part of Figure 3d and Table 1. Association results are expressed in terms of a 1-tailed p-value in the replication cohort and a 2-tailed p-value when folded in with the meta-analysis. We assessed the results as follows: first, we counted the number of SNPs that were nominally significant ($P < 0.05$) in the replication cohort. 7 were nominally significant. 1.25 SNPs by chance are expected to be nominally significant, and this therefore represents an enrichment at P=0.0003 using a binomial test. We then did a sign-test for directional consistency, and found that the effect sizes of 18/25 SNPs were directionally consistent with QTIGC ($P = 0.02$). Then, we considered the replication p-value in addition to direction of effect by counting the number of SNPs that improved the QT-IGC meta-analysis p-value when jointly considered. 11 improved the original QT-IGC p-value, whereas on average 7.6 are expected by chance based on simulation ($P = 0.03$).

Electrophysiology and data analysis

Preparation and injection of cRNA into *Xenopus* oocytes, purchased from EcoCyte Bioscience (Castrop-Rauxel, Germany) were done as described[36]. cDNAs were verified by sequencing. GeneBank accession numbers of the clones used were NM_000238 for *hKCNH2a* and NM_001677 for hATP1B1. Currents were recorded from three batches of oocytes injected with *hKCNH2a*, *hKCNH2a+hATP1B1* or *hATP1B1* cRNA with *hKCNH2a* and *hATP1B1* injected at a 1:1 molar ratio from a holding potential of −80 mV. Electrophysiological recordings were performed at room temperature (22°C–24°C) 3 days after injection in Kulori medium (90 mM NaCl, 4 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 5 mM HEPES, pH 7.4) using a two-electrode voltage clamp amplifier (CA-1B, Dagan, Minneapolis, MN, USA). Data analysis was performed using Pulse (HEKA, Lambrecht, Germany), Igor Pro 4.04 (Wavemetrics, Lake Oswego, OR, USA), and GraphPad Prism (GraphPad Software Inc, San Diego, CA, USA). All values are displayed as mean ± SEM. Current–voltage (I/V) relations were obtained from the step-protocol by plotting the outward current at the end of the second test-pulse as a function of the test-potential. Inactivation kinetics was evaluated by the time constant derived from a monoexponential fit to the decaying phase of the current. The voltage-dependence of activation, inactivation and recovery from inactivation was determined by fitting normalized currents versus test potentials to a two-state Boltzmann distribution of the form $I(V) = 1/(1+\exp[(V_{1/2} - V)/a])$, where $V_{1/2}$ is the potential for half-maximal activation and a is the slope factor. The number of independent experiments is indicated by n. Comparison of the biophysical properties in the presence and absence of *hATP1B1* were performed using an unpaired t-test with $P < 0.05$ being considered significant.

Zebrafish experiments

All zebrafish experiments were performed in accordance with approved Institutional Animal Care and Use Committee (IACUC) protocols. TuAB or Ekwill wild type zebrafish strains were reared according to standard techniques. At the single cell stage, fertilized oocytes were injected with 1-10ng of antisense morpholino oligos targeting the transcription initiation sites of *ATP1B1a*[37], *VCL*[38], *TUFM* (5′ - GAATTTTATAACTTACCGGAGAGGC – 3′) or EIF3C (5′ – GTCTTCTCCACAAACTCACTGCTGT – 3′) dissolved in Danieau's solution (58 mM NaCl, 0.7mM KCl, 0.4 mM MgSO4, 0.6 mM Ca(NO3)2, 5.0 mM HEPES pH 7.6). Controls were injected with Danieau's solution alone. Embryo hearts were microdissected, stained with di-4-ANEPPS (Invitrogen) and imaged on a CCD Camera (Cardio-SMQ, Red Shirt Imaging) at 1000 frames per second as previously described[21]. Cardiac motion was arrested with the use of 15uM blebbistatin (Sigma), field pacing was employed to control beating frequency (Grass S48 Stimulator).

For both *ATP1B1* and *VCL*, two different morpholinos were used and knockdown was demonstrated. For *ATP1B1* the phenotype was reversed by injection of the wild type mRNA. For the morpholino targets where we did not observe any phenotype (*TUFM* and *EIF3C*) we have not yet proven knockdown, nor is there any literature-based evidence of the effect. We have added this information to the text.

Alternative control procedures not based on IgGs

We identified five proteins involved in cardiomyopathy (RYR2, ATP1A1, DSP, MYBPC3, and DMD), where we performed immunoprecipitations in heart tissue using the same methodology as for the five LQTS proteins. These proteins were analyzed analogously to the five LQTS bait IPs: i) we made triplicate IPs, ii) we separated the precipitated proteins by SDS-PAGE, iii) we in-gel digested the proteins, and iv) we analyzed the peptides by LC-MS/MS analysis (See Supplementary Tables S9-S13 for the proteins identified in the pulldowns).

We analyzed the LQTS protein network dataset using the cardiomyopathy pull-down data as the control. The resulting LQTS complexes were compared to the complexes obtained by IgG control experiments (see Supplementary Figure 7). The cardiomyopathy control data was analyzed and applied in three different ways:

First, we used the median protein intensity of the five cardiomyopathy IPs to compare the LQTS bait IPs to a 'general' cardiac protein control (labeled CM1-5_median in Supplementary Figure 7). Second, we used the average protein intensity of the five cardiomyopathy IPs to compare the LQTS bait IPs to another 'general' cardiac protein control (labeled CM1-5_average in Supplementary Figure 7). Third, we tested each of the LQTS IPs against the cardiomyopathy IP it is most similar to, where similarity is evaluated by hierarchical clustering of the data (labeled CM1 or CM2 in Supplementary Figure 7). Our results show that there is a high degree of consistency between the proteins interacting with each of the LQTS proteins when using either IgG controls or different cardiomyopathy protein controls. Using the median of all 5 cardiomyopathy pull-downs as the control, we identify between 87% and 97% (average 91%) of the interaction partners identified with the IgG control procedure. Using the average of all 5 cardiomyopathy pull-downs as the control, we identify between 83% and 90% (average 87%) of the interaction partners identified using the IgGs as the control. Testing each of the LQTS pull-downs against the most similar cardiomyopathy pull-down, we identify between 68% and 91% (average 77%) of the same interaction partners identified using the IgG control procedure. These results strongly support that the interactors identified for the five LQT baits are robust to the use of several different control procedures - including procedures based on IgGs.

Biological replication in 5 Additional Mouse Hearts

To test if our proteomics dataset is affected by the use of pooled tissue samples we generated data from individual hearts and compared those to a pooled sample. We isolated hearts from five male mouse siblings, and prepared homogenates for the individual hearts. We made four sets of IPs using antibodies against KCNQ1, KCNH2, CACNA1C and IgGs from each of the individual heart lysates as well as from a pooled sample. All sample preparation was done as described earlier with the exception that the mass spectrometric analysis was performed on Q-Exactive instrumentation instead of LTQ Orbitrap Velos. In Supplementary Figure 8 we show the hierarchical clustering of all identified proteins by their label-free quantified (LFQ) protein intensities. IPs from pooled heart samples cluster with the analogous IPs from the individual hearts. These results show that the interaction partners we identify with the different baits using technical replicates (pooled hearts), are

highly comparable to the interaction partners identified using biological replicates (hearts 1-5).

Correlation plots of LFQ intensities for the four sets of IPs (KCNQ1, KCNH2, CACNA1C and IgGs) are further supporting the high reproducibility between experiments (Supplementary Figure 9). For each plot the Pearson correlation coefficient is provided in the upper left corner. The average correlation coefficient between a pooled heart sample and the individual heart samples is 0.91 (or 0.93 for CACNA1C; 0.94 for IgG; 0.86 for KCNH2; and 0.91 for KCNQ1). We note that the correlation coefficients are comparable to the ones that we reported in the manuscript for the pooled samples, showing that the pooled samples are indeed adequate for identifying reproducible interactions using quantitative interaction proteomics.

Assessing the contribution of subcellular localization to association results

To assess if the subcellular localization of the immunoprecipitated proteins contribute significantly to the association signal we made pairwise comparisons of the three ion channel pull-downs. On average only 4% of all interaction partners are repetitive between pairs of ion channel pull-downs (specifically, the percentage of repetitive interaction partners is 6% for KCNH2 and KCNQ1; 4% for KCNH2 and CACNA1C; and 2% for KCNQ1 and CACNA1C; respectively). Thus, our data shows that protein interactors residing in the same sub-cellular domains are, at the very most, comprising ∼4% of the interactions we report. Notably, the genes corresponding to proteins that are repetitive between pairs of ion channel pull-downs are only weakly enriched in genome-wide significant loci (P= 0.041). This observation clearly demonstrates that this class of proteins does not drive the statistical enrichment of genes in genome-wide significant loci we observe across the LQTS protein complexes.

Potential weaknesses of using mouse hearts for proteomics experiments

A potential limitation of our study is that we make use of mouse heart tissue as the molecular components of the biology of mouse and human cardiac repolarization might differ. For this reason, we used a variety of validation experiments, including very large and robust human genetic datasets, to augment, complement, and filter the proteomics data. Specifically, we i) applied several statistical tests of enrichment of association to QT prolongation in a cohort of 100,000 humans, all of which showed very significant enrichment of the complexes to human QT phenotypes, and ii) we used replication genotyping in 17,500 additional individuals to confirm a handful of human genetic variants proposed by the complexes to be involved in cardiac repolarization. We went further and functionally validated a number of the specific interaction partners in well-established model systems of human cardiac repolarization by performing electrophysiological experiments in Xenopus oocytes, as well as *in-vivo* knockdowns in zebrafish. Although there are limitations to our analysis, our results clearly show that this does not preclude the identification of novel pathway relationships, new specific genes, and novel genetic variants relevant to human cardiac repolarization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Morita H, Wu J, Zipes DP. The QT syndromes: long and short. Lancet. 2008; 372:750–763. [PubMed: 18761222]

2. Newton-Cheh C, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat Genet. 2009; 41:399–406. [PubMed: 19305408]

3. Pfeufer A, et al. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. Nat Genet. 2009; 41:407–414. [PubMed: 19305409]

4. Arking DE. Genetic association study of QT interval highlights calcium signaling pathways in myocardial repolarization. Accept Nat Genet. 2014

5. Curran ME, et al. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. Cell. 1995; 80:795–803. [PubMed: 7889573]

6. Splawski I, et al. Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell. 2004; 119:19–31. [PubMed: 15454078]

7. Ueda K, et al. Syntrophin mutation associated with long QT syndrome through activation of the nNOS-SCN5A macromolecular complex. Proc Natl Acad Sci U S A. 2008; 105:9355–9360. [PubMed: 18591664]

8. Vatta M, et al. Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. Circulation. 2006; 114:2104–2112. [PubMed: 17060380]

9. Wang Q, et al. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. Nat Genet. 1996; 12:17–23. [PubMed: 8528244]

10. Hubner NC, et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol. 2010; 189:739–754. [PubMed: 20479470]

11. Olsen JV, et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics MCP. 2009; 8:2759–2769. [PubMed: 19828875]

12. Olsen JV, et al. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007; 4:709–712. [PubMed: 17721543]

13. Lundby A, Olsen JV. GeLCMS for in-depth protein characterization and advanced analysis of proteomes. Methods Mol Biol Clifton NJ. 2011; 753:143–155.

14. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008; 26:1367–1372. [PubMed: 19029910]

15. Lage K, et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proc Natl Acad Sci U S A. 2008; 105:20870–20875. [PubMed: 19104045]

16. Lage K, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol. 2007; 25:309–316. [PubMed: 17344885]

17. Müller CS, et al. Quantitative proteomics of the Cav2 channel nano-environments in the mammalian brain. Proc Natl Acad Sci U S A. 2010; 107:14950–14957. [PubMed: 20668236]

18. Rossin EJ, et al. Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology. PLoS Genet. 2011; 7:e1001273. [PubMed: 21249183]

19. Milan DJ, et al. Drug-sensitized zebrafish screen identifies multiple genes, including GINS3, as regulators of myocardial repolarization. Circulation. 2009; 120:553–559. [PubMed: 19652097]

20. Yoshida M, et al. Impaired Ca2+ store functions in skeletal and cardiac muscle cells from sarcalumenin-deficient mice. J Biol Chem. 2005; 280:3500–3506. [PubMed: 15569689]

21. Vasile VC, Edwards WD, Ommen SR, Ackerman MJ. Obstructive hypertrophic cardiomyopathy is associated with reduced expression of vinculin in the intercalated disc. Biochem Biophys Res Commun. 2006; 349:709–715. [PubMed: 16949038]

22. Lundby A, et al. In vivo phosphoproteomics analysis reveals the cardiac targets of β-adrenergic receptor signaling. Sci Signal. 2013; 6:rs11. [PubMed: 23737553]

23. Den Hoed M, et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. Nat Genet. 2013; 45:621–631. [PubMed: 23583979]

24. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007; 2:1896–1906. [PubMed: 17703201]

25. Olsen JV, et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics MCP. 2005; 4:2010–2021. [PubMed: 16249172]

26. Gavin AC, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006; 440:631–636. [PubMed: 16429126]

27. De Bakker PIW, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–128. [PubMed: 18852200]

28. Lundby A, et al. Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nat Commun. 2012; 3:876. [PubMed: 22673903]

29. Achterberg S, et al. Patients with coronary, cerebrovascular or peripheral arterial obstructive disease differ in risk for new vascular events and mortality: the SMART study. Eur J Cardiovasc Prev Rehabil Off J Eur Soc Cardiol Work Groups Epidemiol Prev Card Rehabil Exerc Physiol. 2010; 17:424–430.

30. Simons PC, Algra A, van de Laak MF, Grobbee DE, van der Graaf Y. Second manifestations of ARTerial disease (SMART) study: rationale and design. Eur J Epidemiol. 1999; 15:773–781. [PubMed: 10608355]

31. Stolk RP, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol. 2008; 23:67–74. [PubMed: 18075776]

32. Shepherd J, et al. The design of a prospective study of Pravastatin in the Elderly at Risk (PROSPER). PROSPER Study Group PROspective Study of Pravastatin in the Elderly at Risk. Am J Cardiol. 1999; 84:1192–1197. [PubMed: 10569329]

33. Shepherd J, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet. 2002; 360:1623–1630. [PubMed: 12457784]

34. Hofman A, et al. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol. 2009; 24:553–572. [PubMed: 19728115]

35. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

36. Lundby A, Ravn LS, Svendsen JH, Olesen SP, Schmitt N. KCNQ1 mutation Q147R is associated with atrial fibrillation and prolonged QT interval. Heart Rhythm Off J Heart Rhythm Soc. 2007; 4:1532–1541.

37. Blasiole B, et al. Separate Na,K-ATPase genes are required for otolith formation and semicircular canal development in zebrafish. Dev Biol. 2006; 294:148–160. [PubMed: 16566913]

38. Vogel B, et al. In-vivo characterization of human dilated cardiomyopathy genes in zebrafish. Biochem Biophys Res Commun. 2009; 390:516–522. [PubMed: 19800866]

Editorial summary

The results of genome wide association studies are combined with quantitative interaction proteomics to narrow down the list of putative causal disease genes and filter modest association signals.

A

Type of LQTS	Causal Gene	Gene in complex trait GWA locus?
LQT1	KCNQ1	Yes
LQT2	KCNH2	Yes
LQT3	SCN5A	Yes
LQT4	ANK2	No
LQT5	KCNE1	Yes
LQT6	KCNE2	No
LQT7	KCNJ2	Yes
LQT8	CACNA1C	No
LQT9	CAV3	No
LQT10	SCN4B	No
LQT11	AKAP9	No
LQT12	SNTA1	No

B

Quantitative interaction proteomics of LQTS encoded proteins

Immunoprecipitate LQTS encoded proteins from cardiac lysate and analyse protein composition by LC-MS/MS. Discriminate between specific and non-specific interactors.

Augmenting GWAS data using LQTS protein networks

Locus 1   Locus 2   Locus N   Functional validation

LQTS gene 1   LQTS gene 2   LQTS gene N

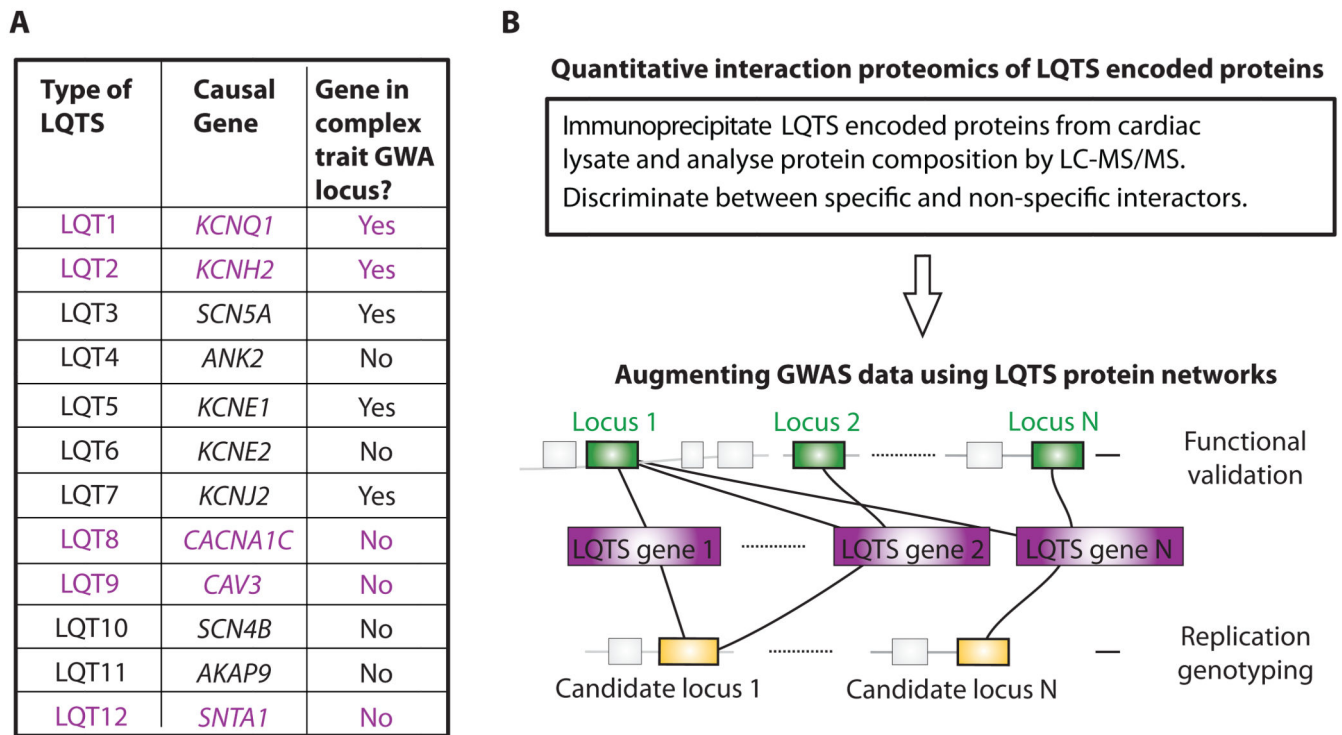Candidate locus 1   Candidate locus N   Replication genotyping

Figure 1. General design and experimental workflow of our integrated genetic and proteomic study

a) Five of the 12 LQTS genes reside in loci definitely associated with QT interval variation in the general population through GWAS. **b**) Protein interaction networks for LQTS proteins (purple boxes where physical interactions are shown as black lines) are resolved in cardiac tissue by quantitative interaction proteomics (top). Interaction partners of the LQTS proteins that reside in GWAS loci are identified and functionally validated (green boxes). Other interaction partners supported by strong proteomic evidence (yellow boxes), point to SNPs that can be prioritized for replication genotyping.
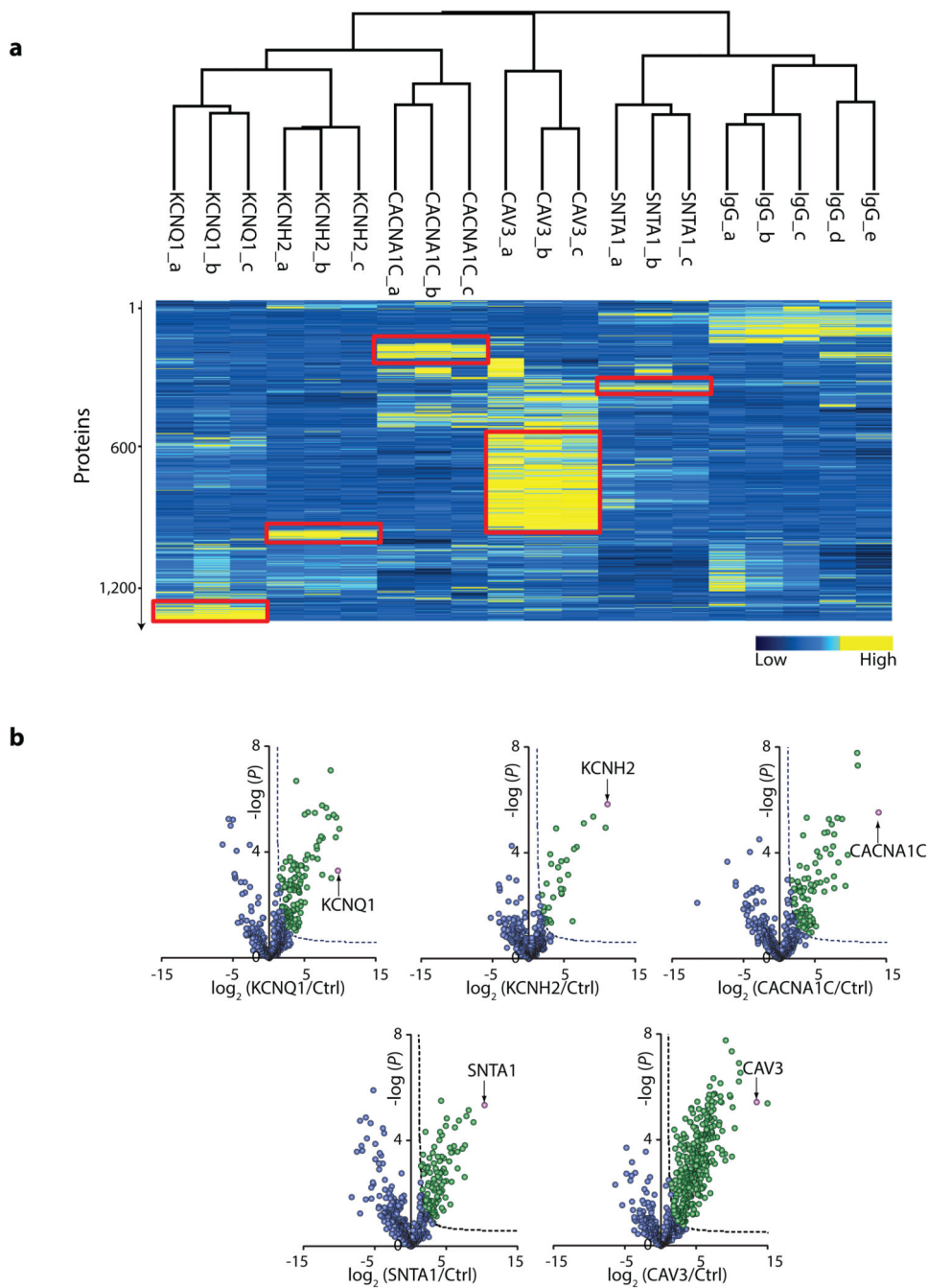
Figure 2. Quantitative interaction proteomics of five Mendelian LQTS proteins
a) Hierarchical cluster analysis of proteins identified in immunoprecipitation experiments visualizes the experimental specificity and reproducibility. Proteins are color-coded according to their mass-spectrometry signal intensity. Triplicates of the LQTS protein immunoprecipitations (**a-c**) are shown. The highlighted yellow areas indicate that each group of triplicate experiments immunoprecipitates a specific cluster of proteins. **b**) Volcano plots, representing the LQTS protein IPs versus IgG control IPs, show negative logarithmized t-test derived *P*-values (-$\log_{10}(P)$) as function of logarithmized ratios of

average protein intensities (log$_2$) for the LQTS protein relative to control. A hyperbolic curve indicates a false discovery rate cut-off of 0.05 and separates specific from nonspecific interactors. All points represent a protein. Purple indicates a LQTS protein, green represent proteins specifically interacting with the LQTS proteins, and blue represents nonspecific interactors.
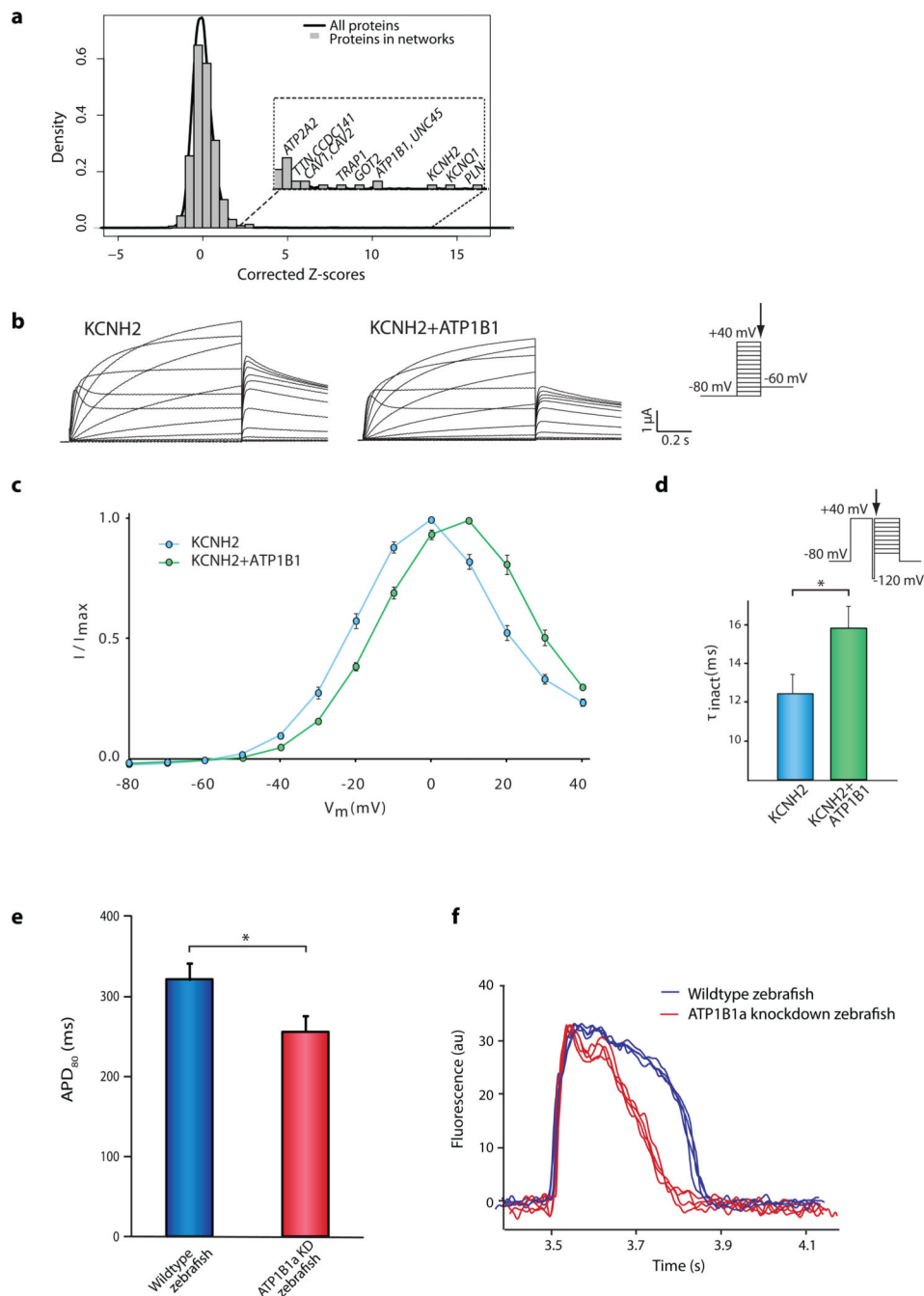
**Figure 3. Proteomic annotation of GWAS loci coupled to experimental follow up identifies
ATP1B1 as a QT variation candidate gene**

a) Distribution of association Z-scores for genes represented in the interactomes (grey bars)
to a background distribution of all genes in the genome (black line). The x-axis represents Z-
scores assigned to genes corrected for SNP density and linkage disequilibrium structure. The
insert shows a zoom-in of the tail of the distribution, illustrating that the distribution is
significantly enriched for genes at GWS loci ($P = 1.3e\text{-}6$, using random sampling, see
Online Methods). **b)** Representative current traces recorded from KCNH2 (left) and

KCNH2 +ATP1B1 (right) proteins heterologously expressed in Xenopus laevis oocytes by two-electrode voltage clamp. Step currents were elicited using the depicted voltage clamp protocol with 1s pulses to test potentials ranging from −80 to +40 mV followed by deactivation (tail) current measurements at −60 mV. **c)** Current-voltage relationships were constructed by normalizing the steady-state currents measured at the end of each voltage step to the maximum outward current and plotting it as function of the test potential (n = 11 for KCNH2, n = 9 for KCNH2+ATP1B1). **d)** Channel inactivation kinetics were evaluated from currents elicited from the indicated pulse protocol. Inactivation time constants measured at +60 mV are shown for KCNH2 in absence (n = 10) or presence (n = 14) of ATP1B1. Data points are mean ± SEM. **e)** Cardiac action potential after Morpholino knockdown of zebrafish *atp1b1a* (APD80 = 256±20 msec) compared to carrier injected controls (APD80 = 321±21 msec), n = 13 independent samples per condition. * represents *P*<0.05. **f)** Superimposed normalized traces are shown for one representative sample for *atp1b1a* knockdown (red) and control conditions (blue).
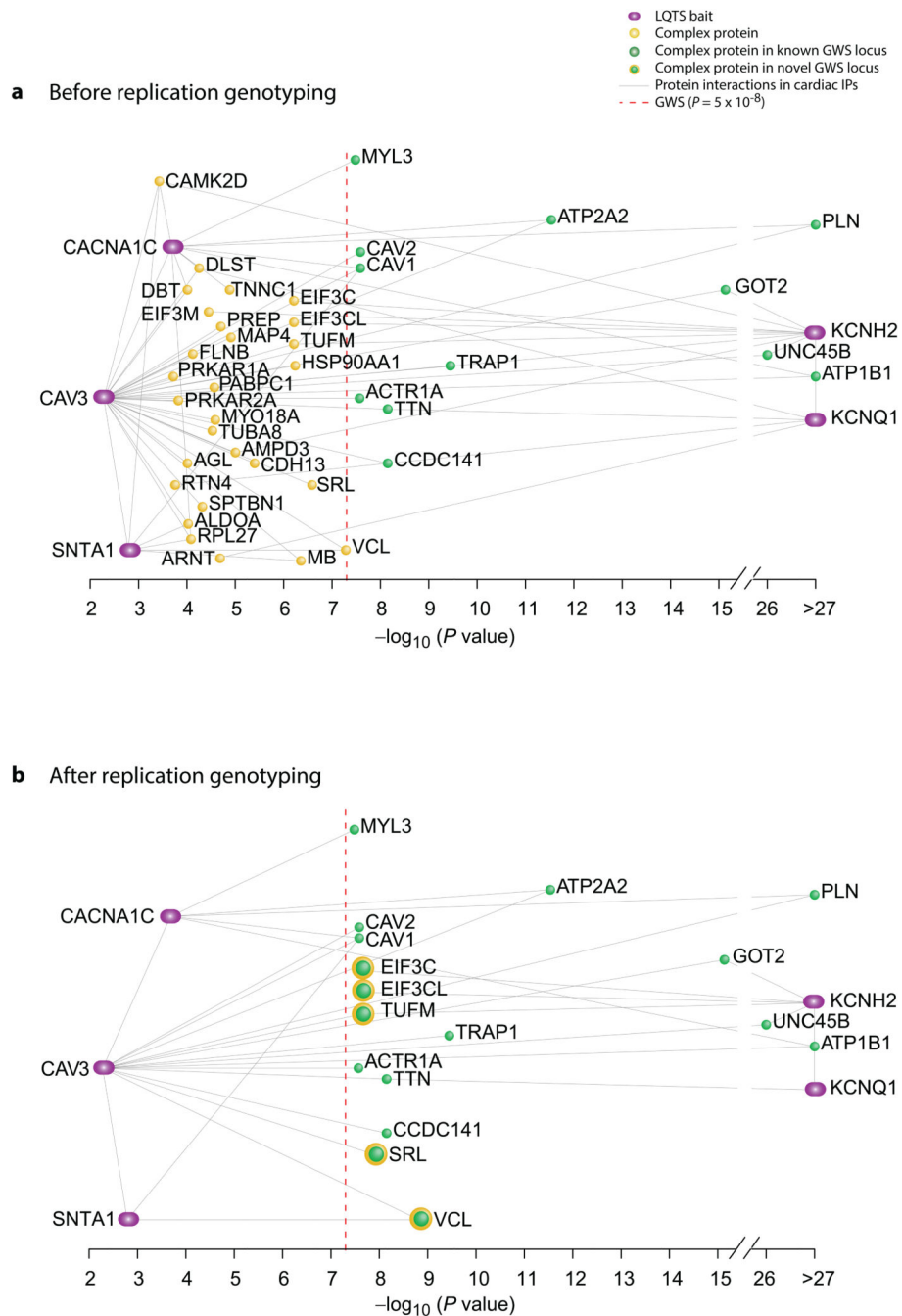
Figure 4. Integrative analysis of the LQTS protein network and GWAS data

a) Depiction of the interactions identified in the proteomics experiments between the LQTS proteins (purple) and proteins encoded by genes in genome-wide significant common variant loci (greene) as well as proteins encoded by genes that lie near the 28 SNPs filtered for replication genotyping (yellow). The proteins are plotted according to the best genetic association *P*-value of their corresponding genes in the horizontal direction after taking the negative 10 based logarithm of the *P*-value and in this depiction (for visualization purposes) we do not correct the *P*-value for multiple hypothesis testing and LD in order to preserve the

true association score as determined in the GWAS. Interactions are represented by grey lines,. The dashed red line indicates the threshold for GWS (corresponding to a *P*-value of 5.0e-8). **b)** An overview of proteins in the LQTS protein network encoded by genes in all 38 loci (green) significantly assocaied to QT variation in this study and in Arking et al.[4]. The five proteins with yellow halos represent the three SNPs that became genome-wide significant after replication genotyping in this study (locus 1, rs7498491: EIF3C, EIF3CL, TUFM; locus 2, rs889807: SRL; locus 3, rs10824026: VCL).

Table 1

Genetic replication results. The first three columns represent locus information of the 25 SNPs that were successfully tested for replication. Columns 4-12 represent the effect size in ms, standard error in ms and *P*-value of those SNPs in each of the QT-IGC GWAS meta-analysis, in the replication cohort (17,692 samples), and in the joint QT-IGC-replication meta-analysis.

Locus information			Meta-analysis			Replication			Joint		
Gene	SNP	Coded allele	Effect size	SE	P-value	Effect size	SE	P-value	Effect size	SE	P-value
Genome-wide significant loci in the joint analysis (joint P<5e-8)											
VCL	rs10824026	A	-0.71	0.13	5.20E-08	-0.72	0.27	4.23E-03	-0.71	0.12	1.49E-09
SRL	rs889807	T	-0.51	0.10	2.59E-07	-0.53	0.22	7.16E-03	-0.51	0.09	1.18E-08
TUFM/EIF3C/EIF3CL	rs7498491	A	-0.51	0.10	6.15E-07	-0.54	0.21	5.50E-03	-0.51	0.09	2.18E-08
Nominal significant loci in replication (replication P<0.05)											
CAMK2D	rs17531033	C	0.39	0.11	3.75E-04	0.66	0.24	2.79E-03	0.44	0.10	1.11E-05
TNNC1	rs352139	T	0.44	0.10	1.31E-05	0.42	0.21	2.06E-02	0.44	0.09	1.49E-06
PREP	rs7760812	A	-0.59	0.14	1.99E-05	-0.51	0.29	4.10E-02	-0.57	0.12	4.23E-06
CDH13	rs8046873	T	0.80	0.17	4.61E-06	0.75	0.45	4.92E-02	0.79	0.16	1.12E-06
Loci at P>0.05 in replication											
MB	rs17722827	A	1.00	0.20	4.42E-07	0.24	0.53	3.28E-01	0.91	0.19	1.03E-06
HSP90AA1	rs10143509	A	-0.76	0.15	5.78E-07	0.48	0.60	7.87E-01	-0.69	0.15	3.43E-06
MYO18A	rs8614	A	-0.55	0.13	2.60E-05	-0.37	0.33	1.27E-01	-0.53	0.12	1.50E-05
RPL27	rs8079855	A	0.41	0.10	8.26E-05	0.32	0.21	6.17E-02	0.39	0.09	2.55E-05
MAP4	rs777016	T	-0.46	0.11	1.23E-05	-0.13	0.22	2.79E-01	-0.40	0.09	2.85E-05
AMPD3	rs12279871	A	0.65	0.15	1.43E-05	0.17	0.31	2.92E-01	0.56	0.13	3.32E-05
DLST	rs2111705	A	0.38	0.10	5.62E-05	0.14	0.25	2.92E-01	0.35	0.09	7.40E-05
SPTBN1	rs12999048	T	-0.68	0.17	4.82E-05	-0.05	0.46	4.61E-01	-0.61	0.16	1.15E-04
PRKAR2A	rs990211	A	-0.45	0.12	1.51E-04	-0.25	0.25	1.55E-01	-0.41	0.11	1.17E-04
PABPC1	rs12114870	T	-2.00	0.48	2.72E-05	2.16	2.03	8.57E-01	-1.78	0.46	1.23E-04
ARNT	rs267734	A	-0.50	0.12	2.02E-05	0.04	0.24	5.70E-01	-0.40	0.10	1.65E-04
ALDOA	rs9924308	A	-0.38	0.10	9.44E-05	-0.11	0.20	2.84E-01	-0.33	0.09	1.67E-04
EIF3M	rs12801493	A	1.93	0.47	3.81E-05	-0.26	1.08	5.97E-01	1.58	0.43	2.32E-04
DBT/AGL	rs6682639	T	-0.79	0.23	6.76E-04	-0.76	0.54	7.93E-02	-0.78	0.21	2.34E-04
FLNB	rs6770059	A	0.68	0.17	7.60E-05	-0.03	0.44	5.27E-01	0.59	0.16	2.48E-04

Locus information			Meta-analysis			Replication			Joint		
Gene	SNP	Coded allele	Effect size	SE	P-value	Effect size	SE	P-value	Effect size	SE	P-value
PRKAR1A	rs2287301	A	0.38	0.10	1.91E-04	0.14	0.21	2.54E-01	0.33	0.09	2.65E-04
TUBA8	rs2234338	T	2.90	0.69	2.98E-05	-0.86	1.15	7.72E-01	1.89	0.59	1.43E-03
RTN4	rs6756933	T	-0.38	0.10	1.75E-04	0.71	0.28	9.95E-01	-0.25	0.10	8.95E-03