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

Alicia Lundby1–3,28, Elizabeth J Rossin3–5,28, Annette B Steffensen2, Moshe Rav Acha6,7, Christopher Newton–Chen3,6,7, Arne Pfeufer8, Stacey N Lynch6,7, The QT Interval International GWAS Consortium (QT-IGC)27, Soren–Peter Olesen2, Soren Brunak1,9, Patrick T Ellinor6,7, J Wouter Jukema10,11, Stella Trompet10,12, Ian Ford13, Peter W Macfarlane14, Bouwe P Krijthe15,16, Albert Hofman15,16, André G Uitterlinden15–17, Bruno H Stricker15–20, Hendrik M Nathoe21, Wilko Spiering20, Mark J Daly3,4, Folkert W Asselbergs11,21,22, Pim van der Harst23, David J Milan6,7, Paul I W de Bakker3,24,25,29, Kasper Lage1,3,4,9,26,29 & Jesper V Olsen1,29

Genome-wide association studies (GWAS) have identified thousands of loci associated with 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.

GWAS have 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 that combines 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 (that is, the time between the onset of the Q wave and the end of the T 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. LQTS is a Mendelian disorder caused by genetic defects in 1 of 12 genes that result in major prolongation of the QT interval (ripe >40 ms). In addition, minor variation of the QT interval (~1–4 ms per allele) is a
quantitative heritable trait in the general population \(^2,3\); recently, 35 SNPs significantly associated with this phenotype were identified \(^4\). Owing 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 QT-interval variation in the general population according to GWAS. Baits used in the proteomics experiments in this work are indicated in purple. (b) Protein interaction networks for LQTS proteins (purple boxes with physical interactions shown as black lines) are resolved in cardiac tissue by quantitative interaction proteomics (top). Interaction partners of the LQTS proteins that reside in genome-wide significant loci are identified and functionally validated (green). Other interaction partners supported by strong proteomic evidence (yellow) point to SNPs that can be prioritized for replication genotyping.

**RESULTS**

**Tissue-specific protein interaction network of LQTS genes**

We chose five LQTS proteins as the starting point of our analysis (KCNQ1, KCNHN2, CACNA1C, SNTA1 and CAV3)\(^3,9\). The proteins were immunoprecipitated from pooled lysates of cardiac tissue from male mice; the precipitates were then separated by SDS-PAGE, which was followed by in-gel trypsin digestion and analysis of the resulting peptide mixtures by nanoflow high-performance liquid chromatography–tandem mass spectrometry (HPLC-MS/MS)\(^10–12\) on an LTQ-Orbitrap Velos instrument (Thermo Scientific) using higher-energy collisional dissociation (HCD) fragmentation\(^13\) (Supplementary Figs. 1–5). The complete set of raw mass spectrometry files were processed using the MaxQuant software suite (http://www.maxquant.org/), whereby peptides and proteins were identified using the Andromeda search engine at a false discovery rate 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 non-specific interactors by applying a false discovery rate cutoff of 0.05 (Fig. 2a, b)\(^10,14\). As expected, the experimental triplicates yielded highly reproducible results for protein signal intensities (Pearson \(r > 0.8\); Supplementary Fig. 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 KCNHN2, 116 of KCNQ1, 104 of SNTA1 and 333 for CAV3 (Supplementary Tables 3–7), and at most 4% (Online Methods) of these proteins were nonspecific binders owing to similarity of the LQTS proteins in terms of subcellular localization in the plasma membrane. Four of the five affinity purification data sets were enriched for known interaction partners in the protein–protein interaction database InWeb\(^15,16\) (KCNQ1, \(P = 6.0 \times 10^{-3}\); CACNA1C, \(P = 3.1 \times 10^{-5}\); CAV3, \(P = 8.9 \times 10^{-3}\); SNTA1, \(P = 5.0 \times 10^{-4}\); Online Methods and Supplementary Table 8), and the number of interacting proteins matches those reported in an analysis of Cav2 channels in rat brain, in which between 97 and 161 proteins interact specifically with the tested ion channels\(^17\). In addition, the specificity, robustness and high quality of the data were confirmed by applying three alternative control procedures, which were not based on IgGs (Online Methods and Supplementary Fig. 7), and by providing biological replication in five additional mouse hearts that had not been pooled (Online Methods and Supplementary Figs. 8 and 9). After making individual quality controls of the pull-down data
Figure 2 | Quantitative interaction proteomics of five Mendelian LQTS proteins. (a) Hierarchical cluster analysis of proteins identified in IP experiments visualizes the experimental specificity and reproducibility. Proteins are color coded according to their mass spectrometry signal intensity. Triplicates of the LQTS protein IPs (a, b, and 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 indicated LQTS protein IPs versus IgG control IPs, show negative logarithmized t-test–derived P values (−log10(P)) as function of logarithmized ratios of average protein intensities (log2) for the LQTS protein relative to control (ctrl). A hyperbolic curve indicates a false discovery rate cutoff of 0.05 and separates specific from nonspecific interactors. All points represent a protein. Purple indicates an LQTS protein, green represent proteins specifically interacting with the LQTS proteins, and blue represents nonspecific interactors.

sets, we pooled the interactions of all LQTS proteins to create an integrated LQTS protein network.

The LQTS protein network points to candidates 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 population4, and the corresponding 35 common variant loci span 154 genes. A locus was defined by identifying neighbor SNPs in linkage disequilibrium (r2 > 0.5) to the associated SNP and expanding to the nearest recombination hotspot as previously described15. Strikingly, excluding LQTS genes, 12 genes in the loci (PLN, ATP1B1, UNCA5B, TRAP1, TTN, CCD141, ATP2A2, CAV1, CAV2, GOT2, ACTR1A and MYL3) encoded proteins in the LQTS protein network derived here. The genes represent 10 of the 35 GWS loci (the probability of such enrichment is P = 1.3 × 10−6 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 (RTR2, 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 we cross-referenced 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. Of note, 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 Fig. 10) that are 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 strongly associated SNP for QT-interval variation (P = 1.11 × 10−31). We showed that ATP1B1 interacts with KCNH2, CACNA1C, KCNQ1 and 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, a result that suggests an additional function of ATP1B1, which is independent of ATP1A1. We tested the effect of ATP1B1 on the KCNH2 channel by electrophysiological measurements of heterologously expressed proteins in X. laevis oocytes. Coexpression of ATP1B1 shifted the peak of the current-voltage relationship by 10 mV to more positive potentials, slowed the channel inactivation kinetics, and right-shifted the voltage dependence of recovery from inactivation (Fig. 3). The same effects were observed in the presence of an ATP1A1 inhibitor (Supplementary Fig. 11). Interestingly, pulldown experiments of ATP1A1 revealed no interaction with 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 that has not previously been shown, through which common genetic variants near or in ATP1B1 affect QT-interval variation. 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 repolarization19. Morpholino knockdown of the zebrafish ortholog for ATP1B1 (atp1b1a) resulted in shorter action potential duration compared to those of wild type (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 with KCNH2 and three other LQTS proteins.
Fig. 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 LQTS protein networks (gray 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. Inset, enlarged view of the tail of the distribution, illustrating that the distribution is significantly enriched for genes at GWAS loci \( (P = 1.3 \times 10^{-6}) \), Student’s t-test, using random sampling; Online Methods). (b) Representative current traces recorded from KCNH2 (left) and KCNH2 + ATP1B1 (right) proteins heterologously expressed in X. laevis oocytes by two-electrode voltage clamp. Step currents were elicited using the depicted voltage clamp protocol with 1-s 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 \( (I) \) measured at the end of each voltage step to the maximum outward current \( (I_{\text{max}}) \) and plotting it as function of the test potential \( (V_{\text{m}}) \). (d) Channel inactivation kinetics were evaluated from currents elicited from the indicated pulse protocol. Inactivation time constants \( (\tau_{\text{inact}}) \) measured at +60 mV are shown for KCNH2 in the absence \( (n = 10) \) or presence \( (n = 14) \) of ATP1B1. (e) Duration of cardiac action potential after Morpholino knockdown (KD) of zebrafish atp1b1a (action potential duration at 80% repolarization \( (\text{APD}_{80}) = 256 \pm 20 \text{ ms} \)) compared to carrier injected controls \( (\text{APD}_{80} = 321 \pm 21 \text{ ms}) \). \( n = 13 \) independent samples per condition. \( *P < 0.05 \), Student’s t-test. (f) Superimposed normalized traces shown for one representative sample of the 13 independent samples per condition for atp1b1a knockdown (red) and control conditions (blue). Data are mean ± s.e.m. in c–e. a.u., arbitrary units.

The LQTS protein network filters subtle GWAS signals

Similarly to most other complex phenotypes, the SNPs associated with QT-interval variation do not explain all 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 to 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

Fig. 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 (GWS) common variant loci (green) 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 base-10 logarithm of the \( P \) value, and in this depiction (for visualization purposes) we do not correct the \( P \) value for multiple-hypothesis testing and linkage disequilibrium in order to preserve the true association score as determined in the GWAS. Interactions are represented by gray lines. The dashed red line indicates the threshold for genome-wide significance (corresponding to \( P = 5.0 \times 10^{-8} \)). (b) Overview of proteins in the LQTS protein network encoded by genes in all 38 loci (green) significantly associated with QT variation in this study and in Arking et al.\(^a\). The five proteins with yellow halos represent the three SNPs that became GWS after replication genotyping in this study (locus 1, rs7498491: EIF3C, EIF3CL, TUFM; locus 2, rs889807: SRL; locus 3, rs10824026: VCL).
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. Using a one-tailed Mann-Whitney rank-sum test, we compared the distribution of association scores across genes represented in the protein network to those for all genes in the genome. Even after excluding the 12 genes from the definitively associated loci, we found that the protein network was significantly enriched for association to QT-interval variation ($P = 1.5 \times 10^{-4}$; Supplementary Fig. 12). This suggests that proteins in the network point to genetic variants important for QT-interval variation that 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 composed of 17,692 independent samples in total. Specifically, SNPs were considered for replication genotyping if the association significance in the GWAS meta-analysis was $<1 \times 10^{-3}$ and if 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 and thereby suggesting it is an important interaction partner of an LQTS protein. For illustration purposes, we plotted the proteins that formed the basis for the SNP selection 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 (the probability of such a finding using the sign test is 0.02) and 7 were nominally significant in the replication cohort (the probability of such a finding using permutation testing is 0.0003). Notably, three of the tested SNPs reached genome-wide significance (VCL: rs10824026, $P = 1.5 \times 10^{-9}$; SRL: rs889807, $P = 1.2 \times 10^{-8}$; and TUFM, EIF3C, and EIF3CL; rs7498491, $P = 2.2 \times 10^{-8}$; Table 1 and Online Methods) when jointly analyzed with the recent GWAS meta-analysis (Fig. 4b).

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). The gene encoding ATP2A2 is itself in a locus significantly associated with QT-interval variation ($P = 3 \times 10^{-12}$). The importance of SRL in cardiac physiology is evident from knockout mice, in which ventricular depolarization is prolonged. 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, it has never been found to be involved in QT-interval variation. We furthermore confirmed the involvement of VCL in cardiac repolarization.

### Table 1 | Genetic replication results

| Gene          | SNP          | Coded allele | Effect size | s.e. | P value      | Effect size | s.e. | P value      | Effect size | s.e. | P value      |
|---------------|--------------|--------------|-------------|------|-------------|-------------|------|-------------|-------------|------|-------------|
| Genome-wide significant loci in the joint analysis ($P < 1 \times 10^{-8}$) |
| VCL           | rs10824026   | A            | −0.71       | 0.13 | 5.20 × 10^{-8} | −0.72       | 0.27 | 4.23 × 10^{-3} | −0.71       | 0.12 | 1.49 × 10^{-9} |
| SRL           | rs889807     | T            | −0.51       | 0.10 | 2.59 × 10^{-7} | −0.53       | 0.22 | 7.16 × 10^{-3} | −0.51       | 0.09 | 1.18 × 10^{-8} |
| TUFM/EIF3C    | rs7498491    | A            | −0.51       | 0.10 | 6.15 × 10^{-7} | −0.54       | 0.21 | 5.50 × 10^{-3} | −0.51       | 0.09 | 2.18 × 10^{-8} |
| Nominally significant loci in replication ($P < 0.05$) |
| CAMK2D        | rs17531033   | C            | 0.39        | 0.11 | 3.75 × 10^{-4} | 0.66        | 0.24 | 2.79 × 10^{-3} | 0.44        | 0.10 | 1.11 × 10^{-5} |
| TNRC1         | rs352139     | T            | 0.44        | 0.10 | 1.31 × 10^{-5} | 0.42        | 0.21 | 2.06 × 10^{-2} | 0.44        | 0.09 | 1.49 × 10^{-6} |
| PREP          | rs7760812    | A            | −0.59       | 0.14 | 1.99 × 10^{-5} | −0.51       | 0.29 | 4.10 × 10^{-3} | −0.57       | 0.12 | 4.23 × 10^{-6} |
| CDH13         | rs8046873    | T            | 0.80        | 0.17 | 4.61 × 10^{-6} | 0.75        | 0.45 | 4.92 × 10^{-3} | 0.79        | 0.16 | 1.12 × 10^{-6} |

### Loci at $P > 0.05$ in replication

| Gene          | SNP          | Coded allele | Effect size | s.e. | P value |
|---------------|--------------|--------------|-------------|------|---------|
| MB            | rs17722827   | A            | 1.00        | 0.20 | 4.42 × 10^{-7} |
| HSP90AA1      | rs10143509   | A            | −0.76       | 0.15 | 5.78 × 10^{-7} |
| MYO1A         | rs8614       | A            | −0.55       | 0.13 | 2.60 × 10^{-6} |
| RPL27         | rs807985     | A            | 0.41        | 0.10 | 8.26 × 10^{-5} |
| MAP4          | rs777016     | T            | −0.46       | 0.11 | 1.23 × 10^{-5} |
| AMPD3         | rs12279871   | A            | 0.65        | 0.15 | 1.43 × 10^{-3} |
| DLST          | rs2111705    | A            | 0.38        | 0.10 | 5.62 × 10^{-5} |
| SFMB1N1       | rs12999048   | T            | −0.68       | 0.17 | 4.82 × 10^{-5} |
| PRKAR2A       | rs990211     | A            | −0.45       | 0.12 | 1.51 × 10^{-4} |
| PABPC1        | rs12114870   | T            | −2.00       | 0.48 | 2.72 × 10^{-5} |
| ARNT          | rs267734     | A            | −0.50       | 0.12 | 2.02 × 10^{-5} |
| ALDOA         | rs9924308    | A            | −0.38       | 0.10 | 9.44 × 10^{-5} |
| EIF3M         | rs12801493   | A            | 1.93        | 0.47 | 3.81 × 10^{-5} |
| DTB/AGL       | rs6682639    | T            | −0.79       | 0.23 | 6.76 × 10^{-4} |
| FLNB          | rs6770059    | A            | 0.68        | 0.17 | 7.60 × 10^{-5} |
| PRKAR1A       | rs2287301    | A            | 0.38        | 0.10 | 1.91 × 10^{-5} |
| TUBA8         | rs2234338    | T            | 2.90        | 0.69 | 2.98 × 10^{-5} |
| RTN4          | rs6756933    | T            | −0.38       | 0.10 | 1.75 × 10^{-4} |

Columns 1–3 represent locus information of the 25 SNPs that were successfully tested for replication. Columns 4–12 represent the effect size in milliseconds, standard error (s.e.) in milliseconds 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.
repolarization by knockdown experiments of the ortholog in zebrafish, vcl, which had a direct effect on cardiac repolarization in vivo (Supplementary Fig. 13). Knockdown of zebrafish orthologs of TUFM or EIF3C did not affect the action potential duration (data not shown).

Thus, capitalizing 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. 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 with a human trait through GWAS for which the causal genes have not been identified, and to augment and filter modestly associated common variants. Although it has been shown previously that generic (i.e., non–tissue specific) in silico protein network analyses based on public data are a powerful tool in interpreting common variants associated with disease, this study advances such methods by using targeted proteomics experiments in relevant tissue types to firmly establish the molecular interactions between proteins in the relevant biological setting. In addition, to our knowledge, our proteomics data set represents the first analysis of the composition of protein networks involved in rare Mendelian disease and its analogous common complex trait.

By testing the interaction networks of the five LQTS proteins one by one for genetic enrichment in the GWAS data (Supplementary Fig. 12), we show that, although 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 pulldowns. 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 involved 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 is the involvement of calcium signaling in cardiac repolarization, a finding suggested from proteomics experiments, sequencing of LQTS patients and meta-analyses of GWAS. These results converge on a cluster of physically interacting Ca$^{2+}$ regulating proteins, thus providing new biological insight into variations of the QT interval in humans.

A potential limitation of our approach is that we used mouse heart tissue for the proteomics experiments, as the molecular biology of cardiac repolarization might differ between mice and humans (Online Methods). For this reason, we used a variety of validation experiments—including large and robust human genetic data sets and model 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. 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.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank M.B. Thomsen, N. Schmitt, H. Poulsen and P. Nissen for experimental input; S. Pult, S. Ripke and J. Cox for help with data analysis; S. Raychaudhuri, P.K. Donahoe and members of the NNF Center for Protein Research for their input on the manuscript. Research reported in this publication was supported in part by the research career programme Sapere Aude from The National Danish Research Council (A.L. and J.V.O.), the Eleanor and Miles Shore Fellowship Program from Harvard Medical School and the Harvard Medical School Junior Faculty Development Award (K.L.), US National Institute of General Medical Sciences award T32GM007753 (E.J.R.), ZonMw grant 90700342 from the Netherlands Organization for Health Research and Development (Fanar) and the EU 7th framework programme grant PRIME-XS (contract no. 262067). The research was also partially supported by the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project no. 050-060-810 and the generous donation by the Novo Nordisk Foundation to Center for Protein Research. F.W.A. is supported by UCL Hospitals NHIR Biomedical Research Centre. SMART: SMART was financially supported by BBMRI_NL from the Dutch government (W04.021.007). LifeLines: The LifeLines Cohort study and generation and management of GWAS genotype data for the LifeLines Cohort Study, is supported by the NWO (grant 175.010.2007.006), Economic Structure Enhancing Fund (FES) of the Dutch government, Ministry of Economic Affairs, Ministry of Education, Culture and Science, Ministry for Health, Welfare and Sports, Northern Netherlands Collaboration of Provinces (SNW), Province of Groningen, University Medical Center Groningen, University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. We thank B. Alizadeh, A. Boejes, M. Bruinenburg, N. Festen, I. Nolte, L. Franke and M. Valerieamahmadi for their help in creating the GWAS database; R. Bieringa, J. Keers, R. Oosterroge, R. Visser and J. Vonk for their work related to data collection and validation; 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: R. van der Horst, H. Hillege, M. van der Klauw, G. Navis, H. Oorend, D. Postma, J. Rosmalen, J. Slaets, R. Stolk and B. Wolfenbuttel; LifeLines GWAS Working Group: B. Alizadeh, M. Boezen, M. Bruinenburg, N. Festen, L. Franke, P. van der Harst, G. Navis, D. Postma, H. Smidere, C. Wijmenga and B. Wolfenbuttel. PROSPER: PROSPER is supported by an investigator-initiated grant from Bristol-Myers Squibb, the Netherlands Heart Foundation (grant 2001 D 032, J.W.J.), the EU 7th framework programme (grant 223004) and the Netherlands Genomics Initiative (Netherlands Consortium for Healthy Aging grant 050-060-810). RS3: The Rotterdam Study (RS) is supported by the Erasmus Medical Center and Erasmus University Rotterdam, Netherlands Organization for Scientific Research, Netherlands Organization for Health Research and Development (ZonMw), Research Institute for Diseases in the Elderly, Netherlands Heart Foundation, Ministry of Education, Culture and Science, Ministry of Health Welfare and Sports, European Commission and Municipality of Rotterdam. Support for genotyping was provided by the NWO (175.010.2005.011, 911.03.012) and Research Institute for Diseases in the Elderly (RIDE).

**AUTHOR CONTRIBUTIONS**

Overall idea, concept, and project coordination: A.L., E.J.R., K.L. and J.V.O. Conceived of and designed the immunoprecipitations and proteomics experiments: A.L. and J.V.O. Performed the immunoprecipitations and proteomics experiments: A.L. Analyzed the proteomics data: A.L. and J.V.O. Contributed meta-analysis GWAS data: Q.T-IGC Consortium, C.N.-C. and A.P. Conceived of and designed statistical enrichment analyses and the integration of genetic and proteomic data: E.J.R. and K.L. Performed enrichment analyses: E.J.R. Identified SNPs for replication: A.L., E.J.R., P.I.W.d.B., K.L. and J.V.O. Conceived of and designed genetic replication experiments: E.J.R., M.J.D., P.I.W.d.B. and K.L. Performed genetic meta-analysis: E.J.R. Contributed Input for the manuscript: S.B., S.-F.O.
C.-N.C., P.v.d.H. and P.I.W.d.B. Conceived of and designed electrophysiological experiments: A.L. Performed and analyzed the electrophysiological experiments: A.B.S. Conceived of and designed zebrafish experiments: A.L., P.T.E. and D.J.M. Performed and analyzed the zebrafish experiments: M.R.A. and S.N.L. Contributed data for genetic replication: SMART, F.W.A., W.S., H.M.N. and P.I.W.d.B.; LifeLines, P.v.d.H.; PROSPER-PHASE, J.W.J., S.T., I.F. and P.W.M.; RS3, B.P.K., A.G.U., B.H.S. and A.H. Wrote the paper: A.L., E.J.R., K.L. and J.V.O.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Morita, H., Wu, J. & Zipes, D.P. The QT syndromes: long and short. Lancet 372, 750–763 (2008).
2. Newton-Cheh, C. et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat. Genet. 41, 399–406 (2009).
3. Pfeuffer, A. et al. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. Nat. Genet. 41, 407–414 (2009).
4. Arking, D.E. et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. Nat. Genet. doi:10.1038/ng.3014 (22 June 2014).
5. Curran, M.E. et al. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. Cell 80, 795–803 (1995).
6. Sadowski, I. et al. CaV1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell 119, 19–31 (2004).
7. Ueda, K. et al. Syntrophin mutation associated with long QT syndrome through activation of the nNOS-SCN5A macromolecular complex. Proc. Natl. Acad. Sci. USA 105, 9355–9360 (2008).
8. Vatta, M. et al. Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. Circulation 114, 2104–2112 (2006).
9. Wang, Q. et al. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. Nat. Genet. 12, 17–23 (1996).
10. Hubner, N.C. et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J. Cell Biol. 189, 739–754 (2010).
11. Olsen, J.V. et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol. Cell. Proteomics 8, 2759–2769 (2009).
12. Olsen, J.V. et al. Higher-energy C-trap dissociation for peptide modification analysis. Nat. Methods 4, 709–712 (2007).
13. Lundby, A. & Olsen, J.V. GeLCMS for in-depth protein characterization and advanced analysis of proteomes. Methods Mol. Biol. 753, 143–155 (2011).
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. 26, 1367–1372 (2008).
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. USA 105, 20870–20875 (2008).
16. Lage, K. et al. A human phenotype-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316 (2007).
17. Müller, C.S. et al. Quantitative proteomics of the CaV2 channel nano-environments in the mammalian brain. Proc. Natl. Acad. Sci. USA 107, 14950–14957 (2010).
18. Rossin, E.J. et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 7, e1001273 (2011).
19. Milan, D.J. et al. Drug-sensitized zebrafish screen identifies multiple genes, including GINS3, as regulators of myocardial repolarization. Circulation 120, 553–559 (2009).
20. Yoshida, M. et al. Impaired Ca2+ store functions in skeletal and cardiac muscle cells from sarcalumenin-deficient mice. J. Biol. Chem. 280, 3500–3506 (2005).
21. Aspden, P. et al. Obstructive hypertrophic cardiomyopathy is associated with reduced expression of vinculin in the intercalated disc. Biochem. Biophys. Res. Commun. 349, 709–715 (2006).
22. Lundby, A. et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. Nat. Genet. 45, 621–631 (2013).
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- to 8-week-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 (50 mM Tris, pH 8.5, 150 mM NaCl, 10 mM KCl, 1% Triton X-100, 5 mM EDTA, 5 mM sodium fluoride, 5 mM β-glycerophosphate, 1 mM sodium orthovanadate, 1 complete inhibitor cocktail Tablet (Roche)) containing protease and phosphatase inhibitors. Tissue lysates were centrifuged to remove insoluble debris. For each tissue preparation produced, lysates derived from five mice were pooled, and protein concentrations were measured by Quick Start Bradford Dye Reagent (Bio-Rad). Solubilized heart tissue lysate was precleared 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-CACNB2B (2 µl SC2025, Santa Cruz), anti-CACNA1A (2 µl SC10646, Santa Cruz), anti-CACNA1D (2 µl SC10646, Santa Cruz), anti-CACNA1C (2 µl lab2912, 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, Bio-Rad).

In-gel digestion. Separated proteins were fixed in the gel (40 mM water, 50 mM acetonitrile, 10 mM 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 (three times 20 min, 800 r.p.m., room temperature (RT)). Gel dices were dehydrated (acetonitrile, 10 min, 800 r.p.m.) followed by reduction of disulfide bonds (10 mM dithiothreitol in 25 mM ammonium bicarbonate, 45 min, RT, 800 r.p.m.) and alkylation of cysteines (55 mM chloro-acetamide in 25 mM ammonium bicarbonate, 30 min, 24 °C in darkness, 800 r.p.m.). After washing in 25 mM ammonium bicarbonate, the gel plugs were dehydrated in acetonitrile and proteins were digested by trypsin (50 µl of 12.5 ng/µl sequencing-grade trypsin (Promega) in 25 mM ammonium bicarbonate for 1 h, followed by addition of 100 µl 25 mM ammonium bicarbonate, left overnight at 37 °C). Trypsin activity was quenched by acidification of the mixture with trifluoroacetic acid to pH of ~2, and peptides were extracted from the gel plugs with 30% acetonitrile in 3% trifluoroacetic acid (30 min, 800 r.p.m.) followed by 80% acetonitrile in 0.5% acetic acid (30 min, 800 r.p.m.) and finally in 100% acetonitrile. Organic solvents were removed by evaporation in a vacuum centrifuge. Extracted peptides were purified on STAGE-tips with two C18 filters.

Mass spectrometry, LC-MS/MS. Peptides were eluted from the Stage tips into 96-well microtiterplates twice with 10 µl 40% acetonitrile in 0.5% acetic acid, and the acetonitrile was evaporated using a vacuum centrifuge, reducing the sample volume to 4 µl. The peptide mixtures were acidified with 0.1% trifluoroacetic acid in 2% acetonitrile to an end volume of 9 µl and analyzed by online nanoflow LC-MS/MS. Peptide separation was performed by reversed-phase C18 HPLC on an Easy nLC system (Thermo Fisher Scientific) loading 5-µl samples with a constant flow of 750 nl/min onto 15-cm-long analytical columns, packed in-house with 3-µm C18 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 electropropayed into an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) through a nanospray ion source. The peptide mixture was analyzed by full-scan MS spectra (m/z 300–2,000, 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. 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 were processed by MaxQuant (version 1.1.1.25) (Max-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Munich), 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 p.p.m., a maximum of two missed cleavages was allowed, and false discovery rates were set at 0.01 for both 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 six amino acids. Statistical evaluation and filtering of the resulting peptide data sets were performed in MaxQuant as previously described. 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 log2 transformation of the LFQ intensities. To perform statistical analysis of the label-free bait IP experiments versus control IP experiments, we imputed normal distributed values for missing values using a normal distribution with width of 0.3 and a downshift of the mean by 1.8 compared to distribution of all LFQ intensities. t-test–based comparisons 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, SO, of 1 (ref. 10). LFQ protein intensity ratios of bait relative to control were 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 nonspecific 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 http://crpl.sund.ku.dk/datasets/proteomics/Channel_interactomes.rar.

Association analyses. QT-IGC4: The QT-IGC consortium consists of 31 cohorts plus 17 replication 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 QT interval on genotype using RR interval, age and sex as covariates (individuals with QRS duration of >120 ms or history of myocardial infarction (MI) were excluded). The summary statistics (beta values, 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 MANTAL26. The non-genomic-control–corrected results were used in this analysis to match what is reported in the accompanying QT-IGC4: The QT-IGC consortium conducted a joint test for enrichment in association performed on the genome-wide significance) was carried out as described previously4 (lambda genomic control λGC = 1.076).

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 five 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 reported 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 because 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 previously18. 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 were from the QT-IGC 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 one-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 with 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. On the basis of organ-wide proteomic mapping of phosphoproteins in rat hearts27, we collected a data set of 2,000 proteins expressed in heart tissue. We assessed the likelihood of identifying 22 GWS proteins in a random selection of five 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 five complexes, and either their association P value was <10−4 (18 SNPs), or it was <10−3 and the protein of interest passed a threshold for being abundantly present in one of the complexes (four proteins). The selected SNPs were then genotyped or looked up in four cohorts: 5,731 independent samples were genotyped in the SMART cohort; and beta values, standard errors and P values were collected from the SMART/PHASE cohort (n = 4,865), the SMART/PAS 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 of >120 ms or positive history of MI were removed.

Cohort descriptions. SMART: the Secondary Manifestations of Arterial disease study28. 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 factors29. 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 KBioscience, using proprietary KASPar PCR technique.

LifeLines30. LifeLines is a multidisciplinary prospective population-based cohort study examining in a unique three-generation design the health and health–related behaviors of 165,000 persons living in the northeast region of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, sociodemographic, behavioral, physical and psychological factors that contribute to the health and disease of the general population, with a special focus on multimorbidity and complex genetics.

PROSPER/PHASE31,32. 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 the 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 bead chip. For 5,763 subjects DNA was available for genotyping. Genotyping was performed with the Illumina 660K bead chip, 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 on the basis of the HapMap build 36 with MACH imputation software.

RS3 (ref. 33). 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 recorded on ACTA electrocardiographs (ESATO), 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 protocols. Participants with a call rate of <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, 1,961 participants were included in this study.

For the SMART data (the only data that we received as raw genotypes), we ran a linear regression in Plink34 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 three cohorts, controlling for age, sex and RR interval, and excluding individuals with QRS duration > 120 ms or past history of MI.

The meta-analysis was done with the program METAL26 using effect size estimates and standard errors. We removed three SNPs owing to missing data in at least three of the four 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 one-tailed P value in the replication cohort and a two-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 QT-IGC (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 on the basis of simulation (P = 0.03).

Electrophysiology and data analysis. Preparation and injection of cRNA into Xenopus oocytes, purchased from EcoCyte Bioscience, were done as described35. cDNAs were verified by sequencing. CB1 Reference Sequence 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 RT (22–24 °C) 3 d after injection in Kulori medium (90 mM NaCl, 4 mM KCl, 1 mM MgCl₂, 1 mM CaCl₂, 5 mM HEPES, pH 7.4) using a two-electrode voltage clamp amplifier (CA-1B). Data analysis was performed using Pulse (HEKA), Igor Pro 4.04 (WaveMetrics), and GraphPad Prism. All values are displayed as mean ± s.e.m. 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 were 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/V₅₀) − a)), where V₅₀ 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 1–10 ng of antisense Morpholino oligonucleotides targeting the transcription initiation sites of ATP1B126, VCL37, TUFM (5′-GAAATTCTATAACTTACGGGAGG-3′) or EIF3C (5′-GTC TTCCTCAAAACTCAGTTGT-3′) dissolved in Danieau’s solution (58 mM NaCl, 0.7 mM 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 charge-coupled device (CCD) camera (Cardio-SMQ, Red Shirt Imaging) at 1,000 frames/s as previously described31. Cardiac motion was arrested with the use of 15 µM blebbistatin (Sigma), and field pacing was employed to control beating frequency (Grass S48 Stimulator).

For both ATP1B1 and VCL, literature-validated Morpholinos were used. For the Morpholino targets for which we did not observe any phenotype (TUFM and EIF3C), we demonstrated knockdown by RT-PCR in triplicate.

Alternative control procedures not based on IgGs. We identified five proteins involved in cardiomyopathy (RYR2, ATP1A1, DSP, MYBP3 and DMD), for which 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 9–13 for the proteins identified in the pulldowns).

We analyzed the LQTS protein network data set using the cardiomyopathy pulldown data as the control. The resulting LQTS complexes were compared to the complexes obtained by IgG control experiments (see Supplementary Fig. 7). The cardiomyopathy control data were 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 Fig. 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 Fig. 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 Fig. 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 five cardiomyopathy pulldowns as the control, we identified between 87% and 97% (average 91%) of the interaction partners identified with the IgG control procedure. Using the average of all five cardiomyopathy pulldowns as the control, we identified between 83% and 90% (average 87%) of the interaction partners identified using the IgGs as the control. Testing each of the LQTS pulldowns against the most similar cardiomyopathy pulldown, we identified 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 five additional mouse hearts. To test whether our proteomics data set was 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 identified with the different baits using technical replicates (pooled hearts) were 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) further support the high reproducibility between experiments (Supplementary Fig. 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 report 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 whether the subcellular localization of the immunoprecipitated proteins contributed significantly to the association signal, we made pairwise comparisons of the three ion channel pulldowns. On average only 4% of all interaction partners were repetitive between pairs of ion channel pulldowns (specifically, the percentage of repetitive interaction partners was 6% for KCNH2 and KCNQ1, 4% for KCNH2 and CACNA1C, and 2% for KCNQ1 and CACNA1C, respectively). Thus, our data show that protein interactors residing in the same subcellular domains were, at the very most, comprising ~4% of the interactions we reported. Notably, the genes corresponding to proteins that were repetitive between pairs of ion channel pulldowns were only weakly enriched in GWS loci (P = 0.041). This observation clearly demonstrates that this class of proteins did not drive the statistical enrichment of genes in GWS loci we observed across the LQTS protein complexes.

Potential weaknesses of using mouse hearts for proteomics experiments. It is a potential limitation of our study that we used mouse heart tissue for the proteomics experiments, as the molecular components driving cardiac repolarization in mice and humans might differ. For this reason, we used a variety of validation experiments, including very large and robust human genetic data sets, to augment, complement and filter the proteomics data. Specifically, (i) we 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 new genetic variants relevant to human cardiac repolarization.

24. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906 (2007).
25. Olsen, J.V. et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol. Cell. Proteomics 4, 2010–2021 (2005).
26. de Bakker, P.I.W. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008).
27. Lundby, A. et al. Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nat. Commun. 3, 876 (2012).
28. Acherter, 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. 17, 424–430 (2010).
29. Simons, P.C., Algra, A., van de Laak, M.F., Grobbee, D.E. & van der Graaf, Y. Second manifestations of ARterial disease (SMART) study: rationale and design. *Eur. J. Epidemiol.* **15**, 773–781 (1999).

30. Stolk, R.P. et al. Universal risk factors for multifactorial diseases: Lifelines: a three-generation population-based study. *Eur. J. Epidemiol.* **23**, 67–74 (2008).

31. 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.* **84**, 1192–1197 (1999).

32. Shepherd, J. et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet* **360**, 1623–1630 (2002).

33. Hofman, A. et al. The Rotterdam Study: 2010 objectives and design update. *Eur. J. Epidemiol.* **24**, 553–572 (2009).

34. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

35. Lundby, A., Ravn, L.S., Svendsen, J.H., Olesen, S.-P. & Schmitt, N. KCNQ1 mutation Q147R is associated with atrial fibrillation and prolonged QT interval. *Heart Rhythm* **4**, 1532–1541 (2007).

36. Blasiolo, B. et al. Separate Na,K-ATPase genes are required for otolith formation and semicircular canal development in zebrafish. *Dev. Biol.* **294**, 148–160 (2006).

37. Vogel, B. et al. In-vivo characterization of human dilated cardiomyopathy genes in zebrafish. *Biochem. Biophys. Res. Commun.* **390**, 516–522 (2009).