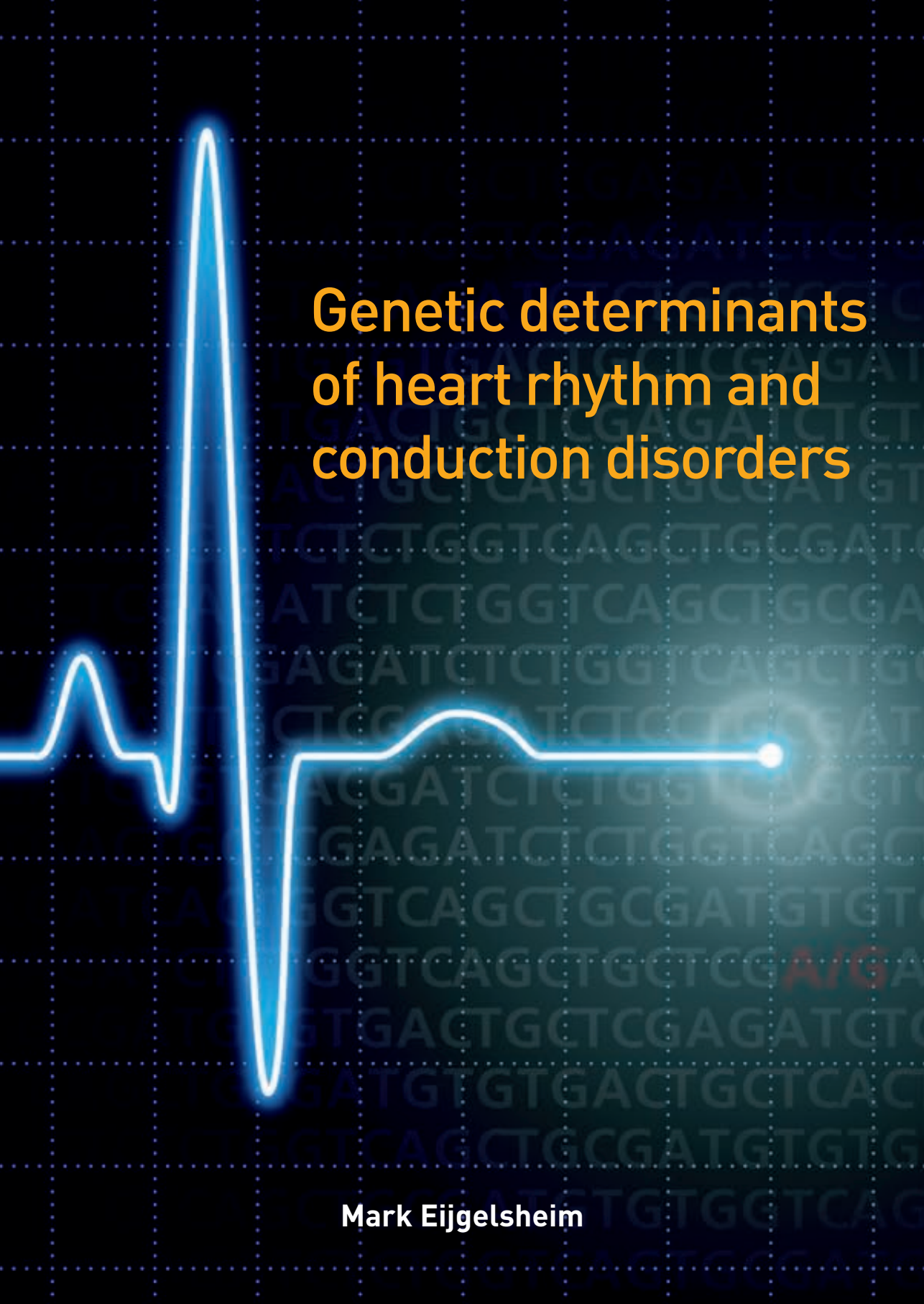

Genetic determinants of heart rhythm and conduction disorders
Mark Eijgelsheim

Genetic determinants of Heart Rhythm and Conduction Disorders

The work described in this thesis was conducted at the Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.

The Rotterdam study is supported by the Erasmus Medical Center and Erasmus University Rotterdam, the Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII), the Netherlands Genomics Initiative (NGI), and the municipality of Rotterdam.

The contributions of the inhabitants, general practitioners and pharmacists of the Ommoord district to the Rotterdam Study are gratefully acknowledged. The role of the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium in development of and support for parts of the work described in this thesis is gratefully acknowledged.

The Department of Epidemiology, Erasmus MC and the Erasmus University financially supported publication of this thesis.

The research described in this thesis was supported by grants of the Netherlands Heart Foundation (NHF-2007B221 and NHF-2009R014).

Cover design and lay-out: Luci van Engelen (www.cviiiontwerpers.nl)
Printed by: Haveka BV, Alblasserdam

ISBN/EAN: 978-94-6050-003-9

Genetic determinants of Heart Rhythm and Conduction Disorders

Genetische determinanten van hartritme en conductie stoornissen

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 9 maart 2011 om 15.30 uur

door

Mark Eijgelsheim

geboren te Rotterdam.

PROMOTIECOMMISSIE

Promotoren:	Prof.dr. B.H.Ch. Stricker
	Prof.dr. A.G. Uitterlinden
Overige leden:	Prof.dr. C.M. van Duijn
	Prof.dr. J.W. Jukema
	Prof.dr. J.L.C.M. van Saase

Financial support by the Netherlands Heart Foundation for the publication of this thesis is gratefully acknowledged.

Contents

PART 1
Introduction

1 Introduction, outline and scope of the thesis

Sudden cardiac arrest (SCA) and sudden cardiac death (SCD) are terms used to describe a physical state in which there is a sudden loss of cardiac activity with hemodynamic collapse. The event is referred to as SCD if the patient dies as a consequence and SCA when death is prevented, either spontaneously or by an intervention[1]. According to the European Society of Cardiology, SCD is defined as a natural death due to cardiac causes, heralded by abrupt loss of consciousness, within one hour after the onset of acute symptoms or an unwitnessed, unexpected death of someone seen in a stable medical condition less than 24 hours previously with no evidence of a non-cardiac cause[2]. However, by convention the term SCD is used by many to describe both fatal and non-fatal SCA events.

ETIOLOGY OF SCD

The majority (65-70%) of events of SCD occur in the presence of underlying coronary heart disease, either acute or chronic. Other causes of SCD include non-ischemic structural heart disease, electrophysiological heart disease in the absence of structural heart disease and cardiac arrest of non-cardiac origin[3-5]. Most commonly, out-of-hospital events are the result of ventricular tachyarrhythmias in the presence of ischemia[6]. Most events occur unexpected in individuals unrecognized to be at specific risk for SCD with two thirds of all SCDs occuring as the first clinical manifestation or in the presence of known disease but without strong risk predictors[7]. Familial aggregation of SCD, independent of other risk factors that cluster within families (for example hyperlipidemia, hypertension, obesity and smoking), suggest that there is a genetic component in SCD risk[8-12].

HUMAN GENETIC VARIATION

Short and Long QT Syndrome (SQTS and LQTS, respectively) and SCD associated with extreme QT interval duration are often caused by rare mutations with large effects, commonly in genes encoding ion channels involved in myocardial repolarization[13] and explain only little of the population variation and familial aggregation in SCD risk. Similar to other common disorders, SCD in the general population can be considered as a complex genetic trait with common ge-

netic variants within many genes each having a small effect on disease risk[14,15]. The hypothesis underlying this consideration is called the 'common variant, common disease' hypothesis[15]. Essentially, this assumes that common complex disorders have a normally distributed genetic liability conferred by Mendelian inherited genes and thus the trait would be normally distributed as a quantitative trait, something already formalized by R.A. Fisher in 1918[16]. This is an important reason to study the quantitative traits that underlie the common disorder of interest[14].

OPERATIONAL DEFINITIONS OF AND QUANTITATIVE TRAITS UNDERLYING SCD

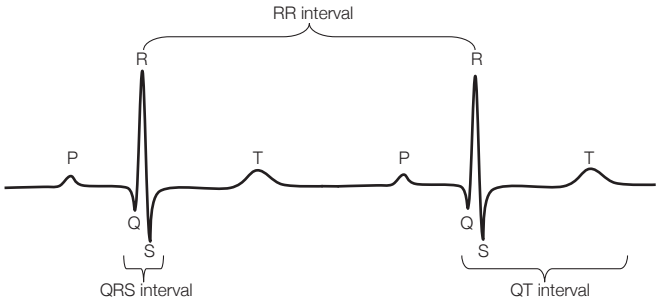
As outlined above, SCD consists of a heterogeneous collection of causes of sudden unexpected loss of cardiac activity with fatal hemodynamic collapse. Operational definitions were formulated to facilitate scientific studies on SCD. The operational definition from the European Society of Cardiology does not depend on the specific cause of the loss of cardiac activity and the rhythm at the time of the event since this is often unknown. The definition is based on the duration of symptoms and the absence of evidence for a non-cardiac cause and consists of two separate descriptions for witnessed and unwitnessed events that read as follows: 1) a witnessed death by a natural cardiac cause preceded by an abrupt loss of consciousness within one hour of symptom onset, or 2) an unwitnessed and unexpected death of a person seen in a stable medical condition within 24 hours prior with no evidence of a non-cardiac cause[2]. Inherent to these definitions they do not distinguish between mechanisms of SCD of cardiac origin that are outlined above introducing heterogeneity in underlying etiology of the studied cases. The degree of heterogeneity might differ between studies due to differences in study design and source populations. Due to this heterogeneous case mix within and between studies, in addition to the relatively small size of existing collections, the statistical power to study genetic risk factors for SCD is limited, especially if the genetic risks differ between the etiological pathways towards SCD. However, quantitative traits can serve as endophenotypical risk factors for SCD.

For SCD, several of such endophenotypical risk factors measured on electrocardiographic (ECG) recordings are available, namely: 1) the QT interval, 2) the QRS duration and 3) the RR interval (Figure 1).

The QT interval is one of the most frequently studied quantitative predictors of arrhythmogenesis and SCD. The QT interval is a non-invasive measure of ventricular myocardial repolarization duration with high clinical relevance. Approximately 35% of the variation in the QT interval in the general population is heritable[10,17]. It is known that a heterogeneous group of mutations in ion channels are responsible for Mendelian disorders in repolarization with increased risk of SCD as the result of ventricular arrhythmias[34-37], such as in the SQTS and LQTS disorders[13]. In addition, non-syndromal prolonged QT intervals are associated with cardiovascular morbidity and mortality, including SCD, in the general population[18-22].

The QRS duration reflects ventricular depolarization and is a function of electrophysiological properties within the specialized conduction system (His-Purkinje system) and the ventricular myocardium. A longer QRS duration – also in the absence of a bundle-branch block – is

Figure 1 *ECG and several of the measures that can be derived from the recording.*



associated with increased mortality and SCD risk[23-25]. Twin and family studies suggest a genetic contribution to QRS duration, with heritability estimates of up to 40%[26,27].

The duration of the interval between two R peaks reflects the time elapsed between two cardiac contractions and therefore the RR interval is the reflection of heart rate on the ECG. Higher resting heart rate (shorter RR interval) is associated with increased risk of cardiovascular mortality[28,29], including sudden death[30-33], independent of traditional risk factors. Heredity plays a substantial role in the inter-individual variation of resting heart rate, accounting for 26–32% of heart rate variation in prior studies[34-37] with twin studies reporting even higher heritability estimates up to 55–63%[38,39].

HOW TO IDENTIFY COMMON GENETIC VARIATION UNDERLYING COMPLEX TRAITS

Until recently, the search for genetic factors with smaller effects on the risk of SCD in the general population was limited to candidate genes known to have a role in arrhythmogenesis on the basis of their involvement in LQTS[40-43]. Only a few years ago, it has become possible to study genetic variation across the whole genome and their association with human traits of interest without a prior hypothesis on the involvement of a certain variant in a biological candidate gene. This was the result of: 1) the completion of the human International HapMap Project (a catalogue of human genetic variants and their underlying correlations)[44-46], 2) technical methods to genotype these variants in large groups of subjects, and 3) the appropriate methods to analyze the data[47]. The QT interval was among the earliest studied phenotypes and the *NOS1AP* gene locus was identified as relevant for cardiac repolarization duration[48]. Subsequently, this finding was consistently validated by several other studies[49-53] and the effects of *NOS1AP* on ion channels were partly elucidated in follow up studies[54,55]. Prior to this initial study, no one ever hypothesized that this gene and its protein product was involved in cardiac electrophysiology.

It has become clear over the course of time that the sample used to perform these genome-wide association analyses should be of sufficient size to identify common variants affecting common diseases. Both false positive as well as false negative results are relevant phenomena that have to be dealt with. Stringent thresholds for statistical significance need to be applied to

avoid false positives due to the many variants tested for an association with the outcome of interest. By chance alone, due to multiple testing, many variants will be associated at conventional statistical significance thresholds and result in false positive associations. It has been shown that a genome-wide association analysis corresponds with approximately 1 million independent hypotheses tests[56]. Adjustment of the conventional significance threshold of $P=0.05$ (i.e. 5×10^{-2}) for this number results in a significance threshold of $P=5 \times 10^{-8}$. False negative findings on the other hand are partly due to insufficient power to detect the true association at this stringent significance threshold. Several studies recognized the need for large samples at an early stage in their preparations for performing genome-wide association analysis in order to generate enough power. Therefore, five population based cohort studies formed the Cohorts for Heart and Ageing Research in Genetic Epidemiology[57] (CHARGE) consortium early 2007 based on the similar design and available data collections. The founding cohorts of this consortium are the Age, Gene/Environment Susceptibility study (AGES), Atherosclerosis Risk in Elderly Study (ARIC), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS) and the Rotterdam Study (RS). Other consortia formed based on a common interest in certain diseases (e.g. the Welcome Trust Case Control Consortium (WTCCC))[58,59] or a central funding agency (e.g. European Network for Genetic and Genomic Epidemiology (ENGAGE) funded by the European Union).

TRAIT CORRELATION AND GWAS

At this moment, the analytical framework and international consortia for genome-wide association studies are well established and many genome-wide association studies have been performed on common diseases, anthropomorphic measures and other quantitative traits. These quantitative traits are often chosen since they are associated with a common disease. Examples include high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol levels for ischemic heart disease, insulin and glucose levels for diabetes mellitus and the ECG measures mentioned in this introduction for SCD. Several of such traits are correlated and studying related quantitative traits might result in more information about the underlying common disease liability. Analysing these traits independently from each other ignores the additional information that is available in this correlation. This led to the development of multivariate genome-wide association analysis methods that can increase the power of detecting true associations[60]. In addition, such a method would have increased power for detecting loci that influence multiple traits (pleiotropy), especially if the variant has opposite effects in two traits, irrespective of trait correlation. However, in the current large consortia the multivariate association analysis methods are not easily applied. Novel methods have to be developed that make such an approach easily applicable, ideally making use of the already available results at the consortium aggregate level to limit the analytical burden on human and computational resources.

PHARMACOGENETICS

Pharmacogenetics is the term used to describe the role of genetics in drug response, which can mean efficacy of the drug or the occurrence of adverse reactions to the drug. With respect to SCD there is special interest in pharmacogenetics since many cardiac and non-cardiac medications are associated with ventricular arrhythmias and SCD. While there is a clear relation between rare mutations causing LQTS or sub-clinical LQTS and drug-induced arrhythmia[61], no clear associations exist between common variants and drug-induced arrhythmia/SCD risk in the general population. This risk of drug-induced arrhythmia is a major problem in drug development and safety[62-64]. The results from genome-wide association studies on risk factors for SCD might generate novel loci for further candidate gene study with respect to gene-drug interactions physiology. In addition, applying genome-wide association studies to identify novel drug by gene interactions might identify the genetic variation that is partly responsible for inter-individual variation in drug response[65].

AIM AND OUTLINE

The main objectives of this thesis were: 1) refining the *NOS1AP*-QT association signal and assessing the relation with SCD, 2) to identify, by means of genome-wide association studies, the genetic variants underlying the genetically determined variation within quantitative endophenotypes of SCD, 3) to develop methodology to enhance our ability to do so and, finally, 4) to study hypothesis driven gene by drug interactions. Part 2 focuses on the role of *NOS1AP* on QT interval variation in the population and the association with SCD. Chapter 2.1 will describe the refinement of the association signal in *NOS1AP* with QT interval identified by the first genome-wide association study result published for QT interval and in chapter 2.2 the relation between genetic variation within *NOS1AP* and SCD is studied in an elderly population. Part 3 consists of several studies on identifying novel genetic loci influencing QT interval (chapter 3.1), RR interval (chapter 3.2) and QRS duration (chapter 3.3), using genome-wide association studies. Chapter 4 continues with the description of a novel method that was developed to combine the results of individual genome-wide association studies to increase the power to identify genetic loci associated with correlated traits. Finally, in chapter 5, a pharmacogenetic interaction effect on heart rate is described between cardiac calcium channel blockers and genetic variation in the *CACNA1C* gene encoding the protein targeted by these drugs. Chapter 6 will provide a general discussion.

REFERENCES

1. Buxton, A.E., Calkins, H., Callans, D.J., DiMarco, J.P., Fisher, J.D., Greene, H.L., Haines, D.E., Hayes, D.L., Heidenreich, P.A., Miller, J.M. *et al...* (2006) ACC/AHA/HRS 2006 key data elements and definitions for electrophysiological studies and procedures: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (ACC/AHA/HRS Writing Committee to Develop Data Standards on Electrophysiology). *Circulation*, 114, 2534-2570.
2. Priori, S.G., Aliot, E., Blomstrom-Lundqvist, C., Bossaert, L., Breithardt, G., Brugada, P., Camm, A.J., Cappato, R., Cobbe, S.M., Di Mario, C. *et al...* (2001) Task Force on Sudden Cardiac Death of the European Society of Cardiology. *Eur Heart J*, 22, 1374-1450.
3. Chugh, S.S., Jui, J., Gunson, K., Stecker, E.C., John, B.T., Thompson, B., Ilias, N., Vickers, C., Dogra, V., Daya, M. *et al...* (2004) Current burden of sudden cardiac death: multiple source surveillance versus retrospective death certificate-based review in a large U.S. community. *J Am Coll Cardiol*, 44, 1268-1275.
4. Zheng, Z.J., Croft, J.B., Giles, W.H. and Mensah, G.A. (2001) Sudden cardiac death in the United States, 1989 to 1998. *Circulation*, 104, 2158-2163.
5. Kuisma, M. and Alaspaa, A. (1997) Out-of-hospital cardiac arrests of non-cardiac origin. Epidemiology and outcome. *Eur Heart J*, 18, 1122-1128.
6. Huikuri, H.V., Castellanos, A. and Myerburg, R.J. (2001) Sudden death due to cardiac arrhythmias. *N Engl J Med*, 345, 1473-1482.
7. Myerburg, R.J. (2001) Sudden cardiac death: exploring the limits of our knowledge. *Journal of cardiovascular electrophysiology*, 12, 369-381.
8. Friedlander, Y., Siscovick, D.S., Arbogast, P., Psaty, B.M., Weinmann, S., Lemaitre, R.N., Raghunathan, T.E. and Cobb, L.A. (2002) Sudden death and myocardial infarction in first degree relatives as predictors of primary cardiac arrest. *Atherosclerosis*, 162, 211-216.
9. Friedlander, Y., Siscovick, D.S., Weinmann, S., Austin, M.A., Psaty, B.M., Lemaitre, R.N., Arbogast, P., Raghunathan, T.E. and Cobb, L.A. (1998) Family history as a risk factor for primary cardiac arrest. *Circulation*, 97, 155-160.
10. Hong, Y., Rautaharju, P.M., Hopkins, P.N., Arnett, D.K., Djousse, L., Pankow, J.S., Sholinsky, P., Rao, D.C. and Province, M.A. (2001) Familial aggregation of QT-interval variability in a general population: results from the NHLBI Family Heart Study. *Clin Genet*, 59, 171-177.
11. Jouven, X., Desnos, M., Guerot, C. and Ducimetiere, P. (1999) Predicting sudden death in the population: the Paris Prospective Study I. *Circulation*, 99, 1978-1983.
12. Kaikkonen, K.S., Kortelainen, M.L., Linna, E. and Huikuri, H.V. (2006) Family history and the risk of sudden cardiac death as a manifestation of an acute coronary event. *Circulation*, 114, 1462-1467.
13. Splawski, I., Shen, J., Timothy, K.W., Lehmann, M.H., Priori, S., Robinson, J.L., Moss, A.J., Schwartz, P.J., Towbin, J.A., Vincent, G.M. *et al...* (2000) Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation*, 102, 1178-1185.
14. Plomin, R., Haworth, C.M. and Davis, O.S. (2009) Common disorders are quantitative traits. *Nat Rev Genet*, 10, 872-878.
15. Reich, D.E. and Lander, E.S. (2001) On the allelic spectrum of human disease. *Trends Genet*, 17, 502-510.
16. Fisher, R.A. (1918) The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinb.*, 52, 399-433.
17. Newton-Cheh, C., Larson, M.G., Corey, D.C., Benjamin, E.J., Herbert, A.G., Levy, D., D'Agostino, R.B. and O'Donnell, C.J. (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Heart Rhythm*, 2, 277-284.
18. Algra, A., Tijssen, J.G., Roelandt, J.R., Pool, J. and Lubsen, J. (1993) QT interval variables from 24 hour electrocardiography and the two year risk of sudden death. *Br Heart J*, 70, 43-48.
19. Karjalainen, J., Reunanen, A., Ristola, P. and Viitasalo, M. (1997) QT interval as a cardiac risk factor in a middle aged population. *Heart*, 77, 543-548.
20. Schouten, E.G., Dekker, J.M., Meppelink, P., Kok, F.J., Vandenbroucke, J.P. and Pool, J. (1991) QT interval prolongation predicts cardiovascular mortality in an apparently healthy population. *Circulation*, 84, 1516-1523.
21. Siscovick, D.S., Raghunathan, T.E., Rautaharju, P., Psaty, B.M., Cobb, L.A. and Wagner, E.H. (1996) Clinically silent electrocardiographic abnormalities and risk of primary cardiac arrest among hypertensive patients. *Circulation*, 94, 1329-1333.

22. Straus, S.M., Kors, J.A., De Bruin, M.L., van der Hooft, C.S., Hofman, A., Heeringa, J., Deckers, J.W., Kingma, J.H., Sturkenboom, M.C., Stricker, B.H. *et al...* (2006) Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. *J Am Coll Cardiol*, 47, 362-367.
23. Desai, A.D., Yaw, T.S., Yamazaki, T., Kaykha, A., Chun, S. and Froelicher, V.F. (2006) Prognostic Significance of Quantitative QRS Duration. *Am J Med*, 119, 600-606.
24. Elhendy, A., Hammill, S.C., Mahoney, D.W. and Pellikka, P.A. (2005) Relation of QRS duration on the surface 12-lead electrocardiogram with mortality in patients with known or suspected coronary artery disease. *Am J Cardiol,* 96, 1082-1088.
25. Oikarinen, L., Nieminen, M.S., Viitasalo, M., Toivonen, L., Jern, S., Dahlof, B., Devereux, R.B. and Okin, P.M. (2004) QRS duration and QT interval predict mortality in hypertensive patients with left ventricular hypertrophy: the Losartan Intervention for Endpoint Reduction in Hypertension Study. *Hypertension*, 43, 1029-1034.
26. Busjahn, A., Knoblauch, H., Faulhaber, H.D., Boeckel, T., Rosenthal, M., Uhlmann, R., Hoehe, M., Schuster, H. and Luft, F.C. (1999) QT interval is linked to 2 long-QT syndrome loci in normal subjects. *Circulation*, 99, 3161-3164.
27. Hanson, B., Tuna, N., Bouchard, T., Heston, L., Eckert, E., Lykken, D., Segal, N. and Rich, S. (1989) Genetic factors in the electrocardiogram and heart rate of twins reared apart and together. *Am J Cardiol*, 63, 606-609.
28. Greenland, P., Daviglus, M.L., Dyer, A.R., Liu, K., Huang, C.F., Goldberger, J.J. and Stamler, J. (1999) Resting heart rate is a risk factor for cardiovascular and noncardiovascular mortality: the Chicago Heart Association Detection Project in Industry. *Am J Epidemiol*, 149, 853-862.
29. Nauman, J., Nilsen, T.I., Wisloff, U. and Vatten, L.J. (2010) Combined effect of resting heart rate and physical activity on ischaemic heart disease: mortality follow-up in a population study (the HUNT study, Norway). *Journal of epidemiology and community health*, 64, 175-181.
30. Dyer, A.R., Persky, V., Stamler, J., Paul, O., Shekelle, R.B., Berkson, D.M., Lepper, M., Schoenberger, J.A. and Lindberg, H.A. (1980) Heart rate as a prognostic factor for coronary heart disease and mortality: findings in three Chicago epidemiologic studies. *Am J Epidemiol*, 112, 736-749.
31. Jouven, X., Zureik, M., Desnos, M., Guerot, C. and Ducimetiere, P. (2001) Resting heart rate as a predictive risk factor for sudden death in middle-aged men. *Cardiovasc Res*, 50, 373-378.
32. Kannel, W.B., Kannel, C., Paffenbarger, R.S., Jr. and Cupples, L.A. (1987) Heart rate and cardiovascular mortality: the Framingham Study. *Am Heart J*, 113, 1489-1494.
33. Shaper, A.G., Wannamethee, G., Macfarlane, P.W. and Walker, M. (1993) Heart rate, ischaemic heart disease, and sudden cardiac death in middle-aged British men. *Br Heart J*, 70, 49-55.
34. Laramie, J.M., Wilk, J.B., Hunt, S.C., Ellison, R.C., Chakravarti, A., Boerwinkle, E. and Myers, R.H. (2006) Evidence for a gene influencing heart rate on chromosome 5p13-14 in a meta-analysis of genome-wide scans from the NHLBI Family Blood Pressure Program. *BMC Med Genet*, 7, 17.
35. Martin, L.J., Comuzzie, A.G., Sonnenberg, G.E., Myklebust, J., James, R., Marks, J., Blangero, J. and Kissebah, A.H. (2004) Major quantitative trait locus for resting heart rate maps to a region on chromosome 4. *Hypertension*, 43, 1146-1151.
36. Russell, M.W., Law, I., Sholinsky, P. and Fabsitz, R.R. (1998) Heritability of ECG measurements in adult male twins. *J Electrocardiol*, 30 Suppl, 64-68.
37. Singh, J.P., Larson, M.G., O'Donnell, C.J., Tsuji, H., Evans, J.C. and Levy, D. (1999) Heritability of heart rate variability: the Framingham Heart Study. *Circulation*, 99, 2251-2254.
38. Dalageorgou, C., Ge, D., Jamshidi, Y., Nolte, I.M., Riese, H., Savelieva, I., Carter, N.D., Spector, T.D. and Snieder, H. (2008) Heritability of QT interval: how much is explained by genes for resting heart rate? *Journal of cardiovascular electrophysiology*, 19, 386-391.
39. De Geus, E.J., Kupper, N., Boomsma, D.I. and Snieder, H. (2007) Bivariate genetic modeling of cardiovascular stress reactivity: does stress uncover genetic variance? *Psychosomatic medicine*, 69, 356-364.
40. Bezzina, C.R., Verkerk, A.O., Busjahn, A., Jeron, A., Erdmann, J., Koopmann, T.T., Bhuiyan, Z.A., Wilders, R., Mannens, M.M., Tan, H.L. *et al...* (2003) A common polymorphism in KCNH2 (HERG) hastens cardiac repolarization. *Cardiovasc Res*, 59, 27-36.
41. Gouas, L., Nicaud, V., Berthet, M., Forhan, A., Tiret, L., Balkau, B., Guicheney, P. and Group, D.E.S.I.R.S. (2005) Association of KCNQ1, KCNE1, KCNH2 and SCN5A polymorphisms with QTc interval length in a healthy population. *Eur J Hum Genet*, 13, 1213-1222.
42. Newton-Cheh, C., Guo, C.Y., Larson, M.G., Musone, S.L., Surti, A., Camargo, A.L., Drake, J.A., Benjamin, E.J., Levy, D., D'Agostino, R.B., Sr. *et al.*. (2007) Common genetic variation in KCNH2 is associated with QT interval duration: the Framingham Heart Study. *Circulation*, 116, 1128-1136.

43. Pfeufer, A., Jalilzadeh, S., Perz, S., Mueller, J.C., Hinterseer, M., Illig, T., Akyol, M., Huth, C., Schopfer-Wendels, A., Kuch, B. *et al.*. (2005) Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. *Circ Res*, 96, 693-701.
44. The International HapMap, C. (2003) The International HapMap Project. *Nature*, 426, 789-796.
45. International HapMap, C. (2005) A haplotype map of the human genome. *Nature,* 437, 1299-1320.
46. Frazer, K.A. and Ballinger, D.G. and Cox, D.R. and Hinds, D.A. and Stuve, L.L. and Gibbs, R.A. and Belmont, J.W. and Boudreau, A. and Hardenbol, P. and Leal, S.M. *et al.*. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449, 851-861.
47. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*, 17, R122-128.
48. Arking, D.E., Pfeufer, A., Post, W., Kao, W.H., Newton-Cheh, C., Ikeda, M., West, K., Kashuk, C., Akyol, M., Perz, S. *et al.*. (2006) A common genetic variant in the NOS1 regulator *NOS1AP* modulates cardiac repolarization. *Nat Genet*, 38, 644-651.
49. Aarnoudse, A.J., Newton-Cheh, C., de Bakker, P.I., Straus, S.M., Kors, J.A., Hofman, A., Uitterlinden, A.G., Witteman, J.C. and Stricker, B.H. (2007) Common *NOS1AP* variants are associated with a prolonged QTC interval in the Rotterdam Study. *Circulation*, 116, 10-16.
50. Lehtinen, A.B., Newton-Cheh, C., Ziegler, J.T., Langefeld, C.D., Freedman, B.I., Daniel, K.R., Herrington, D.M. and Bowden, D.W. (2008) Association of *NOS1AP* genetic variants with QT interval duration in families from the Diabetes Heart Study. *Diabetes*, 57, 1108-1114.
51. Marroni, F., Pfeufer, A., Aulchenko, Y.S., Franklin, C.S., Isaacs, A., Pichler, I., Wild, S.H., Oostra, B.A., Wright, A.F., Campbell, H. *et al.*. (2009) A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: the EUROSPAN project. *Circ Cardiovasc Genet*, 2, 322-328.
52. Raitakari, O.T., Blom-Nyholm, J., Koskinen, T.A., Kahonen, M., Viikari, J.S. and Lehtimaki, T. (2008) Common variation in *NOS1AP* and KCNH2 genes and QT interval duration in young adults. The Cardiovascular Risk in Young Finns Study. *Ann Med*, 1-8.
53. Tobin, M.D., Kahonen, M., Braund, P., Nieminen, T., Hajat, C., Tomaszewski, M., Viik, J., Lehtinen, R., Ng, G.A., Macfarlane, P.W. *et al.*. (2008) Gender and effects of a common genetic variant in the *NOS1* regulator *NOS1AP* on cardiac repolarization in 3761 individuals from two independent populations. *Int J Epidemiol*, 37, 1132-1141.
54. Barouch, L.A., Harrison, R.W., Skaf, M.W., Rosas, G.O., Cappola, T.P., Kobeissi, Z.A., Hobai, I.A., Lemmon, C.A., Burnett, A.L., O'Rourke, B. *et al.*. (2002) Nitric oxide regulates the heart by spatial confinement of nitric oxide synthase isoforms. *Nature*, 416, 337-339.
55. Chang, K.C., Barth, A.S., Sasano, T., Kizana, E., Kashiwakura, Y., Zhang, Y., Foster, D.B. and Marban, E. (2008) CAPON modulates cardiac repolarization via neuronal nitric oxide synthase signaling in the heart. *Proc Natl Acad Sci USA*, 105, 4477-4482.
56. Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32, 381-385.
57. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., Boerwinkle, E. *et al.*. (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts, *Circ. Cardiovasc. Genet.*, 2, 73-80.
58. Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M. *et al.*. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science,* 316, 1336-1341.
59. Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. *et al.*. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316, 1331-1336.
60. Ferreira, M.A. and Purcell, S.M. (2009) A multivariate test of association. *Bioinformatics*, 25, 132-133.
61. Paulussen, A.D., Gilissen, R.A., Armstrong, M., Doevendans, P.A., Verhasselt, P., Smeets, H.J., Schulze-Bahr, E., Haverkamp, W., Breithardt, G., Cohen, N. *et al.*. (2004) Genetic variations of KCNQ1, KCNH2, SCN5A, KCNE1, and KCNE2 in drug-induced long QT syndrome patients. *J Mol Med*, 82, 182-188.
62. De Ponti, F., Poluzzi, E., Cavalli, A., Recanatini, M. and Montanaro, N. (2002) Safety of non-antiarrhythmic drugs that prolong the QT interval or induce torsade de pointes: an overview. *Drug Saf,* 25, 263-286.
63. Roden, D.M. (2004) Drug-induced prolongation of the QT interval. *N Engl J Med*, 350, 1013-1022.
64. Roden, D.M. (2008) Cellular basis of drug-induced torsades de pointes. *Br J Pharmacol*, 154, 1502-1507.
65. Daly, A.K. Genome-wide association studies in pharmacogenomics. *Nat Rev Genet*, 11, 241-246.

PART 2

NOS1AP, QT interval duration and Sudden Cardiac Death

2.1 Identification of a common variant at the *NOS1AP* locus strongly associated to QT interval

Eijgelsheim M, Aarnoudse AL, Rivadeneira F, Kors JA, Witteman JC, Hofman A, van Duijn CM, Uitterlinden AG, Stricker BHCh

Hum Mol Genet. 2009 Jan 15; 18(2):347-357.

ABSTRACT

QT interval prolongation is an electrophysiologic phenomenon associated with sudden cardiac death. The QT interval in the general population is for about 35% heritable. In genome-wide association studies a common variant (rs10494366T>G) within the nitric oxide synthase 1 adaptor protein (*NOS1AP*) gene was identified and consistently associated with QT interval duration. Yet, the causal variant remains unclear. Therefore, we performed fine mapping of the association of the *NOS1AP* locus with QT interval within the Rotterdam Study, a population-based, prospective cohort study of individuals ≥55 years of age. First, we tested the association of single-nucleotide polymorphisms (SNPs) in or within ±100kb of the *NOS1AP* gene with QT interval duration, using sex-specific unstandardized residuals after regression on age and RR interval, in 385 individuals using the combined set of SNPs present in the Affymetrix 500k and Illumina 550k chip arrays. Subsequently, we examined correspondence of the association signals in 4,606 individuals using the Illumina 550k array. A *C*-to-*T* SNP at chromosome 1 position 160300514 (rs12143842, T-allele frequency=24%) was associated with a QT interval duration increase of 4.4-ms. per additional T-allele ($P=4.4\times10^{-28}$). For comparison, the most strongly associated variant to date, rs10494366T>G, was associated with a 3.5-ms. increase ($P=1.6\times10^{-23}$) per additional G-allele. None of the inferred haplotypes showed a stronger effect than the individual rs12143842C>T SNP. In conclusion, we found rs12143842 6kb upstream distance of *NOS1AP* to be more strongly associated to QT interval duration than rs10494366T>G. Functional analysis of this marker is warranted.

INTRODUCTION

Sudden cardiac death (SCD) is one of the major causes of cardiovascular mortality in developed countries. Most cases occur in individuals unrecognized to be at risk[1]. Familial aggregation of SCD, independent of other risk factors, suggests a component of genetic variation in SCD risk[2-6]. Due to the relatively small size of SCD collections and etiologic heterogeneity the statistical power to study this dichotomous trait is limited. The focus has thus been on quantitative outcomes such as endophenotypes for SCD risk. The QT interval duration as measured on the electrocardiogram (ECG) is one of the most studied quantitative predictors of arrhythmogenesis and SCD. The QT interval is a non-invasive measure of ventricular repolarization with high clinical relevance. Variation of the QT interval in the general population is approximately 35% heritable[7,8]. It is known that a heterogeneous group of mutations in ion channels are responsible for Mendelian disorders in repolarization, such as Long-QT Syndrome (LQTS) and Short-QT Syndrome (SQTS)[9], with increased risk of SCD as the result of ventricular arrhythmias[10-13]. But these rare disorders explain only little of the population burden of SCD. In addition, non–syndromal shortened and prolonged QT intervals are associated with cardiovascular morbidity and mortality, including SCD[14-18], and multiple non-cardiac medications are associated with ventricular arrhythmias and SCD due to QT prolongation[19,20]. Therefore, QT interval is an interesting quantitative risk factor for SCD to study. Until recently, research of genetic factors influencing SCD risk through variation in electrogenesis was limited to candidate genes known to have a role in arrhythmogenesis on the basis of their involvement in LQTS[21-29]. Using genome wide analysis a common variant (rs10494366T>G, G-allele frequency 38%) in the nitric oxide synthase 1 adaptor protein (*NOS1AP*) gene was identified and consistently associated with QT interval variation across independent replication studies, including ours[30-34]. Since rs10494366T>G might not be the causative common variant we studied the region using dense genome coverage genotype information to localize the region associated with QT interval duration.

MATERIALS & METHODS

Setting and design

The Rotterdam Study is an ongoing prospective population-based cohort study of chronic diseases in Caucasian elderly, which started in 1990. All inhabitants of Ommoord, a Rotterdam suburb in the Netherlands, aged 55 years and over (n=10,278) were invited to participate. Of them, 78% (n=7,983) gave their written informed consent for participation, including retrieval of medical records, use of blood and DNA for research purposes and publication of results. Baseline examinations took place from March 1990 through July 1993. Follow up examinations are carried out periodically. Furthermore, exposure to medication is continuously monitored since January 1, 1991, through computerized pharmacy records of the pharmacies in the Ommoord district. The pharmacy data include the Anatomical Therapeutical Chemical (ATC)-code, the dispensing date, the total amount of drug units per prescription, the prescribed daily number of units, and product name of the drugs. This provides us with information on start and duration

of use of all prescribed medication. Detailed information on design, objectives and methods of the Rotterdam Study was described elsewhere[35,36]. The Medical Ethics Committee of the Erasmus University approved the study.

Phenotype: assessment of QT interval, other alectrocardiographic measurements and phenotype modeling

As described in earlier studies on QT duration in the Rotterdam Study[18] we used a 10-second resting 12-lead ECG (average 8-10 beats), which was recorded on an ACTA ECG (ESAOTE, Florence, Italy) at a sampling frequency of 500 Hz and stored digitally. All ECGs were processed by the Modular ECG Analysis System (MEANS) to obtain ECG measurements, in agreement with the FDA guidance for clinical evaluation of QT/QTc interval prolongation (http://www.fda.gov/cder/guidance/6922fnl.pdf). The MEANS program determines the QT duration from the start of the QRS complex until the end of the T wave. MEANS also determines the presence of right or left bundle-branch block and QRS duration. ECGs with right or left bundle-branch block, QRS duration >120 milliseconds, atrial fibrillation or missing data were excluded from the analysis. In addition, all ECGs taken while under QT prolonging drugs were excluded to minimize confounding by non-genetic factors. Drugs were considered as potentially QT interval prolonging if they appeared on any of the list 1 through 4 at the QT Drugs website (http://www.qtdrugs.org/). We also excluded ECGs if subjects used flupentixol, levomepromazine, mefloquine, olanzapine, or sertindole, drugs which may all prolong QT interval[37-40]. In addition, we excluded those ECGs taken under exposure to digoxin, which may shorten QT interval[41,42]. An ECG was considered taken during drug exposure if a prescription of any of above-mentioned drugs overlapped with the date of ECG measurement. The phenotype of interest was the residual QT interval duration in milliseconds estimated from the regression on age (in years) and heart rate (RR interval) in sex specific strata. This was done to allow for different effects for age and RR interval in men and women. Up to 4 ECGs per individual were available for QT phenotype modeling. Phenotype modeling was performed on all eligible ECGs using PROC MIXED in SAS 9.1.3 (SAS Institute Inc., Cary, NC) for repeated measurements to enhance precision of adjustment.

Genotyping, data cleaning and SNP extraction

Genomic DNA was extracted from whole blood samples using the salting out method[43]. Microarray genotyping was performed in the whole original Rotterdam Study cohort with proper quality DNA samples (n=6,449) using the Infinium II HumanHap550K Genotyping BeadChip® version 3 (Illumina). Any samples with a call rate below 97.5% (n=209), excess autosomal heterozygosity >0.336~FDR<0.1% (n=21), mismatch between called and phenotypic gender (n=36), or if there were outliers identified by the IBS clustering analysis (see below) clustering > 3 standard deviations away from the population mean (n=102) or IBS probabilities > 97% (n=129) were excluded from the analysis. In total, 5,974 samples met quality control inclusion criteria. The GeneChip® Human Mapping 500K Duo Array Set (Affymetrix) was used in a subset of 509 women selected for a pilot study on quantitative traits on the basis of absence of major chronic diseases and/or use of medication. Genotyping procedures were followed according to Illumina and Affymetrix

manufacturer's protocols respectively. Any samples with a call rate below 95% (n=15), mismatch between called and phenotypic gender (n=3), or if there were outliers identified by the IBS clustering analysis clustering > 3 standard deviations away from the population mean (n=10), or if either of the NspI or StyI arrays was missing (n=22) were excluded from the analysis; in total, 470 samples were successfully genotyped as described previously[44]. A subset of 423 selected healthy women from the pilot study had complete genotyping for both the Illumina and Affymetrix platforms.

All SNPs present within the *NOS1AP* gene area (Chr. 1; Base pair 160306205 to 160604864; NCBI build 36.2) +/- 100kb were extracted from both datasets and merged into one dataset. This allowed for a higher SNP density by combining SNPs genotyped on either platform, considering that Illumina and Affymetrix use different SNP selection criteria. Illumina selected haplotype tagging SNPs based on HapMap Phase II linkage disequilibrium (LD) structure determined in the Caucasian (CEU) population. In contrast, Affymetrix selected SNPs randomly across the human genome. The combination of SNPs across platforms yielded 314 unique SNPs in a 500 kb region containing *NOS1AP*. With the dual platform data we cover 72%, 91% and 95% of the common variation as present in the Caucasian HapMap reference population at r^2 thresholds of 1, ≥ 0.7 and ≥ 0.5, respectively. For the Illumina 550k platform only, there were 156 htSNPs present in this region. Markers were excluded if they deviated significantly from Hardy-Weinberg equilibrium ($P<1\times10^{-4}$, n=1 SNPs for the combined dataset and n=0 SNPs for the Illumina dataset), if they had a minor allele frequency <5% (n=59 and 19 SNPs, respectively), or if they had a SNP call rate <95% within the samples (n=14 and 1 SNPs, respectively). This resulted in 242 SNPs in the healthy women subset and 136 SNPs in the full Rotterdam Study population available for association analysis.

Association analysis

For the SNP association analyses we followed a two-stage approach. In the first stage, we estimated the effect and *P*-values for the association between the selected common variants and residual QT interval duration in the subset of women with dual platform, high density, genotype data using additive linear regression models. This allowed for evaluation of coverage by Illumina genotyped htSNPs compared to the genotype data enriched by non-tagging SNPs genotyped on the Affymetrix platform. In the second stage the same analysis methods were applied in the total study population genotyped on the Illumina platform. Additional analyses were performed, such as sex-stratified analyses and haplotype analysis. Haplotype analysis was performed using several approaches. First, a sliding window haplotype analysis using the expectation/maximization algorithm was used to estimate haplotype effects on QT interval duration[45,46]. Second, a fixed marker set based on the CEU Phase 2 HapMap r-squared was used[47]. Additional haplotype analyses were performed for combinations of top hits from single SNP analysis For the gene-wide association analysis and all other haplotype analyses we used PLINK v.1.01[48]. Since the software used does not allow for repeated measurements we selected the first eligible ECG for association testing.

Finally, for the most significant single SNP result and the earlier reported rs10494366T>G we

performed linear regression analysis for multiple measurements using the PROC MIXED procedure in SAS 9.1.3 to make full benefit of all available ECG data. This analysis included general genotypic and allelic models. Again gender-stratified analyses were performed to assess differential effects between men and women. To test statistical independence of the most significant single SNPs and rs10494366, conditional regression analysis was performed by assessing the effect of each SNP within strata of homozygous reference allele carriers of the other SNP. Results are presented as delta QT in milliseconds when compared to the reference allele with corresponding 95% confidence intervals and *P*-values.

RESULTS

Study population

For QT interval duration association analysis, the source population consisted of all subjects with at least one eligible QT interval measurement (n=5,442). The final number of subjects available for QT association analysis included 1,854 men and 2,752 women (total n=4,606) since not all eligible subjects had genotype data. Of the 423 female subjects additionally genotyped on the Affymetrix 500K, 385 had an ECG with a valid phenotype measurement. On average, these women were younger than the total study sample. Of all genotyped subjects with a valid phenotype 11,535 eligible ECGs were available, averaging 2.5 ECGs per individual. The QT interval duration, RR interval, residual QT duration and QTc presented in Table 1 are based on first eligible ECG measurements.

Table 1 *Baseline characteristics of the study population*

	Total QT Sample		**Dual Platform QT sample**	**Illumina platform QT sample**	
Characteristic	**Men**	**Women**	**Women**	**Men**	**Women**
N (%)	2,139 (39.3)	3,303 (60.7)	385 (100)	1,854 (40.3)	2,752 (59.7)
Eligible ECGs, n (%)	5,263 (39.2)	8,155 (60.8)	1,244 (100)	4,642 (40.2)	6,893 (59.8)
Age, years, mean (SD)	67.6 (7.9)	69.5 (9.0)	62.5 (4.3)	67.5 (7.8)	69.4 (8.9)
QT, msec., mean (SD)	397.7 (28.6)	399.0 (28.6)	402.0 (25.7)	397.6 (28.6)	399.0 (28.9)
RR, msec., mean (SD)	905.0 (154.6)	864.8 (138.5)	889.1 (136.3)	903.6 (154.5)	864.8 (140.3)
QT residual, mean (SD)*	-0.4 (18.4)	-0.7 (18.4)	1.0 (15.8)	-0.3 (18.1)	-0.7 (17.8)
QTc, mean (SD)#	420.3 (23.5)	430.8 (21.6)	428.0 (19.6)	420.5 (23.3)	430.9 (21.5)

Shown are first eligible ECG characteristics of all individuals with eligible ECG data, of the subset with genotype data on both genotyping platforms, and of the total sample having genotype data.
* Residual QT duration in milliseconds on first eligible ECG after regression on age in years and RR interval in milliseconds in sex-specific strata using repeated measurements on all eligible ECGs,
# Calculated using Bazett's formula (QTc= QT/√RR).

NOS1AP variants and residual QT duration; dual platform sample

In this first stage, multiple variants within *NOS1AP* were significantly associated with QT interval duration. All 22 SNPs from both platforms associated with *P*-values <0.01 are presented in Table 2. The original SNP related to QT duration as reported by Arking *et al.*[31] through GWAS, rs10494366T>G, a positive control for our data, reached a *P*-value of 8.2×10^{-3}. The top result SNP from our analyses in these 385 women, rs10919117A>G, was genotyped on both the Affymetrix and the Illumina platform. A second signal situated 5' of the gene was observed. Here the most significant SNPs, rs12036340A>G (Affymetrix) and rs12134842C>T (Illumina), are array specific but in high linkage disequilibrium (LD) (r^2=0.91, D'=0.97). Several SNPs in the top results were in strong LD, and may flag the same association (Figure 1). These results allow to proceed fine mapping in the larger sample with the Illumina array genotypes only, since all of these associations are covered by the Illumina htSNPs.

Table 2 *Gene wide results of Illumina 550K and Affymetrix 500K genotypes for association with QT interval in 385 women from the Rotterdam Study. SNPs with P<1×10⁻² are presented ranked by P-value.*

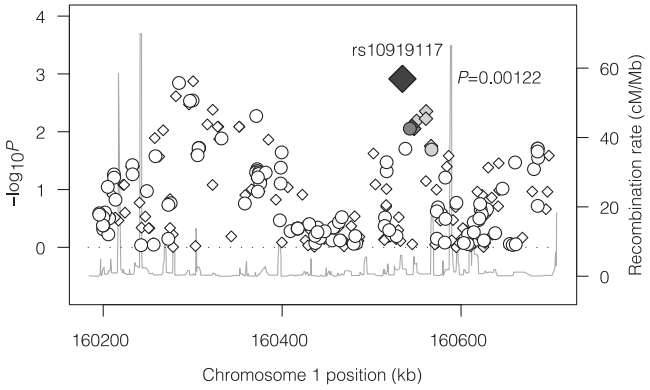
SNP (rs #)	BP pos	Minor allele	MAF (%)	Platform	Subjects	ΔQT (ms)	*P*-value
rs10919117	160534447	G	47.4	I + A	384	-3.7	1.2×10^{-3}
rs12143842	160300514	T	28.2	I	385	4.1	1.3×10^{-3}
rs12036340	160282364	G	28.9	A	383	4.1	1.5×10^{-3}
rs2880058	160281256	G	36.4	I	385	3.7	2.4×10^{-3}
rs7546009	160297525	C	34.9	A	384	3.6	3.1×10^{-3}
rs7547308	160294616	G	36.5	A	385	3.5	3.1×10^{-3}
rs7550692	160295915	T	34.6	I	380	3.6	3.4×10^{-3}
rs6670339	160322430	C	39.2	I	385	3.3	4.2×10^{-3}
rs12046924	160560906	T	33.6	I	385	3.4	4.4×10^{-3}
rs4657150	160368688	C	41.8	A	373	3.2	5.7×10^{-3}
rs12048222	160560718	G	25.6	I	385	3.7	6.0×10^{-3}
rs10919166	160550291	A	29.9	I	384	3.4	6.5×10^{-3}
rs4656349	160316448	G	36.0	I + A	385	3.1	7.5×10^{-3}
rs2819316	160545176	G	43.8	I	385	-3.1	8.0×10^{-3}
rs4657140	160327889	T	40.1	I + A	385	3.1	8.2×10^{-3}
rs1415257	160328668	G	40.1	I + A	385	3.1	8.2×10^{-3}
rs1415259	160351933	G	40.1	I	385	3.1	8.2×10^{-3}
rs10494366*	160352309	G	40.1	I	385	3.1	8.2×10^{-3}
rs2661810	160543056	T	46.1	A	383	3.1	8.7×10^{-3}
rs2819318	160547349	T	46.4	I	385	3.1	8.9×10^{-3}
rs2819322	160547652	T	46.4	I	385	3.1	8.9×10^{-3}
rs7514121	160267056	G	19.4	I + A	385	3.7	9.4×10^{-3}

Presented are: SNP identification (rs-number), base pair position on chromosome 1, minor allele, minor allele frequency (MAF, calculated within 385 subjects), presence of SNP on genotyping platform (I=Illumina, A=Affymetrix), number of subjects successfully genotyped for SNP, delta QT in milliseconds per additional coding allele and *P*–value.

* Original reported SNP associated with QT interval[31].

Figure 1 *NOS1AP Gene wide association results to QT interval of Illumina 550k and Affymetrix 500k genotype data in 385 women of the Rotterdam Study. (A) P-value plotted at a –log(p) scale as a function of genomic position (NCBI Build 36). Rs10919117 is listed (black diamond). Estimated recombination rates (HapMap) plotted to reflect the local LD structure around the associated SNPs and correlated SNPs in grey scales (dark grey: r^2>0.8; medium grey: 0.5<r^2<0.8; light grey:0.2<r^2<0.5; white: r^2<0.2). Diamonds represent SNPs present on Illumina 550K platform. Circles indicate Affymetrix 500K unique SNPs. (B) Like A, for rs12143842. (C) Linkage blocks, corresponding to A and B, of associated SNPs significantly associated to QT interval presenting r^2-values.*

1A *NOS1AP region, Illumina 550K + Affymetrix 500K*



1B *NOS1AP region, Illumina 550K + Affymetrix 500K*

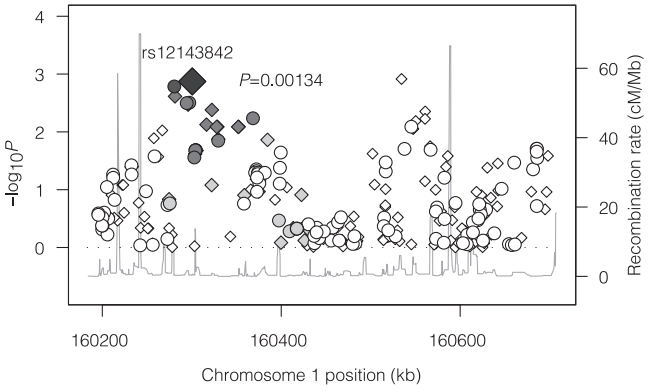
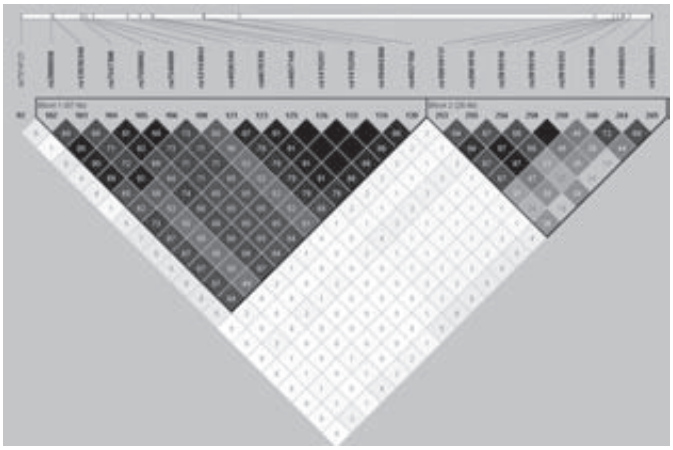


Figure 1C



NOS1AP variant and residual QT duration: Illumina 550K platform

Single SNP association analysis

Of the 136 SNPs tested, we observed multiple highly significant associations in the sex-pooled analysis. The top ten hits are reported in Table 3 and a graphical representation of all SNPs is given in Figure 2. All top hits arose from within one block, corresponding to the rs12143842 linkage block also seen in the 385 women subset. The previously described rs10494366T>G reached a *P*-value of $1.2x10^{-19}$. There was full LD (r^2=1) between rs10494366 and three other SNPs (i.e., rs1415257, rs1415259, rs4657140). Only two other SNPs showed stronger associations (i.e., rs6670339T>C and rs12143842C>T). The top hit, rs12143842C>T, was associated with a 4.5-ms. increase in QT duration per additional T-allele (*P*-value of 4.9×10^{-25}, minor allele frequency 24%). There exists some degree of LD between rs12143842C>T and rs10494366T>G, with an r^2 of 0.46 and a D' = 0.91. The rs6670339T>C variant is in high, though not full, LD with rs10494366T>G (r^2 of 0.91; D' of 0.98).

Compared to the sex-pooled results, only some rank order differences are observed in the sex-stratified analysis, but rs12143842C>T is consistently the strongest associated SNP in both men and women. The top result (rs10919117A>G), and other significant SNPs from within the second block from the first stage association analysis lost all significance in the larger sample.

Multimarker haplotype association analysis

Haplotype analysis results for rs12143842C>T | rs6670339T>C (4.62-ms.; *P*-value 4.85×10^{-25} for the TC haplotype, freq: 22.8%) and rs12143842C>T | rs10494366T>G (4.68-ms.; *P*-value 1.53×10^{-25} for the TG haplotype; freq: 22.5%) are consistent with the single rs12143842 T-allele effect. Using a sliding window the region was scanned for haplotypes associated with QT duration.

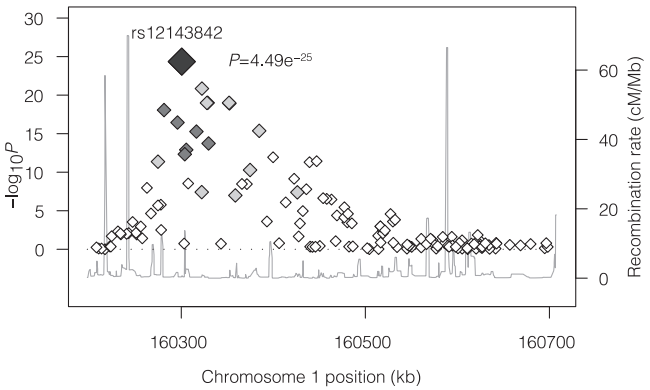
Table 3 *Gene wide results of Illumina 550K genotypes for association with QT interval in 4,606 men and women from the Rotterdam Study. Top ten SNPs are presented and ranked by P-value.*

SNP(rs†)	BP position	Minor allele	MAF (%)	Subjects (n)	ΔQT (ms)	*P*–value	r^2 to rs12143842
rs12143842	160300514	T	24.0	4,592	4.5	4.9×10^{-25}	-
rs6670339	160322430	C	35.0	4,606	3.7	1.0×10^{-21}	0.50
rs1415257a	160328668	G	36.2	4,606	3.5	1.0×10^{-19}	0.46
rs1415259a	160351933	G	36.2	4,606	3.5	1.0×10^{-19}	0.46
rs4657140*	160327889	T	36.2	4,605	3.5	1.0×10^{-19}	0.46
rs10494366†	160352309	G	36.2	4,603	3.5	1.2×10^{-19}	0.46
rs2880058	160281256	G	32.5	4,606	3.5	8.6×10^{-19}	0.63
rs7550692	160295915	T	31.2	4,490	3.4	3.6×10^{-17}	0.69
rs1932933	160384670	T	37.2	4,605	3.1	4.3×10^{-16}	0.42
rs4656349	160316448	G	31.5	4,602	3.2	5.3×10^{-16}	0.57

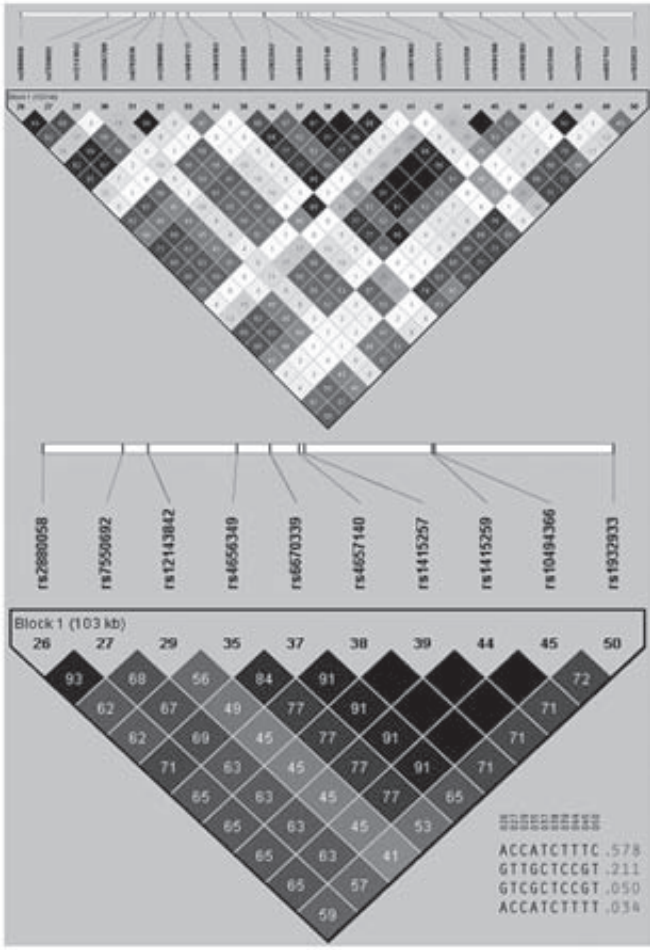
Presented are: SNP identification (rs-number), base pair position on chromosome 1, minor allele, minor allele frequency (MAF, calculated within 4,606 subjects), number of subjects successfully genotyped for SNP, delta QT in milliseconds per additional coding allele, P–value and extend of LD (expressed as r^2) to rs12143842.
* r^2=1 to rs10494366.
† Original reported SNP associated with QT interval[31].

Figure 2 *NOS1AP Gene wide association results to QT interval of Illumina 550K genotype and imputed data in 4,606 subjects of the Rotterdam Study. (A) P-values of directly genotyped SNPs plotted at a –log(p) scale as a function of genomic position (NCBI Build 36). The SNP with the most significant association in the analysis is listed (rs12143842, black diamond)). Estimated recombination rates plotted to reflect the local LD structure around the associated SNPs and correlated SNPs in grey scales (dark grey: r^2>0.8; medium grey: 0.5<r^2<0.8; light grey:0.2<r^2<0.5; white: r^2<0.2). (B) Detail of LD structure of gene region containing top 10 associated SNPs and the linkage structure of top 10 associated directly genotyped SNPs (r^2-values).*

2A *NOS1AP region, genotyped Illumina 550k SNPs.*



2B



The top two haplotypes for each window width (rs7550692T | rs12143842T and rs12143842T | rs1267209G for window width = 2; rs2880058G | rs7550692T | rs12143842T and rs7550692T | rs12143842T | rs12567029G for window width = 3) all showed associations similar to the individual rs12143842C>T effect. Finally, using the fixed marker set, the rs2880058G | rs12143842T haplotype reached the highest level of significance, but again results were consistent with the rs12143842 T-allele effect seen in single SNP association analysis. Adding rs10494366T>G to any of the haplotypes did not change these results.

Repeated measurement association analysis

We used all eligible ECGs in a repeated measurement analysis of rs12143842C>T and, for comparison, rs10494366G>T. The rs12143842C>T variant, was associated with 4.4-ms. increase in

QT duration per additional T-allele (95%CI 3.6–5.1; $P=4.4\times10^{-28}$). In contrast, rs10494366T>G was associated with a 3.5-ms. increase of QT duration per additional G-allele (95%CI 2.8–4.2; $P=1.6\times10^{-23}$). In genotypic analysis, heterozygous and homozygous carriage of the rs12143842 T-allele were associated with a 5.0-ms. (95%CI 4.0–6.0) and 7.3-ms. (95% CI 5.3–9.4) increase in QT duration, respectively. In sex specific analysis the QT prolonging effect of rs12134842C>T was 13% longer in women compared to men. (Table 4A) With conditional regression analysis independent effects were estimated for rs12143842C>T, rs6670339T>C and rs10494366T>G. Within subjects with the rs12143842-CC genotype both rs6670339T>C (+2.6-ms.; *P*-value 1.1×10^{-5}) and rs10494366T>G (+2.0-ms.; *P*-value 3.0×10^{-4}) remained associated with QT interval. Likewise, within rs10494366-TT and rs6670339-TT genotypes, rs12143842C>T was associated to QT interval

Table 4 *Repeated measurement analysis for genotypic and allelic effects of rs12143842 and rs10494366 on QT interval and conditional analysis, including 11,535 ECGs from 4,606 men and women from the Rotterdam Study.*

4A

SNP (rs–number)	**Genotypic**			***P*-value**	**Allelic model**	***P*-value**
rs12143842C>T	**CC**	**CT**	**TT**		**Per T-allele**	
All subjects, ECGs (n)	2643 (6604)	1697 (4227)	252 (666)		4592 (11497)	
Δ QT, ms (95% CI)	reference	5.0 (4.0–6.0)	7.3 (5.3–9.4)	7.9×10^{-28}	4.4 (3.6–5.1)	4.4×10^{-28}
Men, ECGs (n)	1052 (2636)	696 (1717)	103 (281)		1851 (4634)	
Δ QT, ms (95% CI)	reference	4.7 (3.2–6.3)	6.8 (3.6–10.1)	2.5×10^{-10}	4.1 (2.9–5.3)	7.0×10^{-11}
Women, ECGs (n)	1591 (3968)	1001 (2510)	149 (385)		2741 (6863)	
Δ QT, ms (95% CI)	reference	5.2 (3.9–6.4)	7.7 (5.0–10.3)	2.7×10^{-18}	4.6 (3.5–5.6)	8.3×10^{-19}
rs10494366T>G*	**TT**	**TG**	**GG**		**Per G-allele**	
All subjects, ECGs (n)	1883 (4681)	2110 (5273)	610 (1575)		4603 (11529)	
Δ QT, ms (95% CI)	reference	4.1 (3.1–5.1)	6.6 (5.1– 8.1)	5.8×10^{-23}	3.5 (2.8–4.2)	1.6×10^{-23}
Men, ECGs (n)	755 (1897)	848 (2083)	249 (658)		1852 (4638)	
Δ QT, ms (95% CI)	reference	3.7 (2.1–5.3)	5.8 (3.5–8.1)	8.6×10^{-8}	3.1 (2.0–4.2)	1.9×10^{-8}
Women, ECGs (n)	1128 (2784)	1262 (3190)	361 (917)		2751 (6891)	
Δ QT, ms (95% CI)	reference	4.4 (3.1–5.7)	7.1 (5.2–9.0)	4.2×10^{-16}	3.8 (2.9–4.7)	8.3×10^{-17}

4B

Condition	**ECGs n**	**Δ QT, ms (95% CI) Per rs10494366 G-allele**	***P*–value**	**Δ QT, ms (95%CI) Per rs6670339 C-allele**	***P* –value**	**Δ QT, ms (95% CI) Per rs12143842 T-allele**	***P*–value**
rs10494366 TT	4650			0.8 (-5.7–7.4)	8.1×10^{-1}	4.0 (0.8 –7.2)	1.3×10^{-2}
rs6670339 TT	4843	-1.0 (-4.2–2.2)	5.5×10^{-1}			4.8 (1.5 – 8.2)	5.0×10^{-3}
rs12143842 CC	6602	2.0 (0.9–3.1)	3.0×10^{-4}	2.6 (1.4–3.7)	1.1×10^{-5}		

Presented are: A) Genotypic effect and allelic effects of rs12143842 and rs10494366. Both pooled and sex-specific effect estimates are presented with corresponding 95% confidence intervals and *P*-values.

* Original reported SNP associated with QT interval[31].

B) Allelic effect within strata of homozygous non variant carriers of the alternate SNP, testing statistical independence.

duration (+4.0-ms.; *P*-value 1.3×10^{-2} and +4.8-ms; *P*-value 5.0×10^{-3}, respectively). rs10494366T>G and rs6670339T>C did not show independence, which is expected since they are highly linked (r^2=0.91). (Table 4B)

DISCUSSION

Main findings and considerations

Using high density genotype data of the *NOS1AP* gene, consisting of common variants selected on the basis of LD structure (Illumina 550K), enriched with genotype data of randomly selected common variants (Affymetrix 500K) in a limited number of subjects, we observed a novel strong association of the common variant rs12143842C>T (+4.4-ms. per additional allele, *P*-value 4.4×10^{-28}) with QT interval duration. Compared to the previously described rs10494366G>T (+3.5-ms. per additional allele; *P*-value 1.6×10^{-23}) variant, both the strength of association and the magnitude of effect are considerably larger. Given the certain extent of LD between rs12143842 and rs10494366 (r^2=0.46) and a minor allele frequency of 36.2% for rs10494366 compared to 24.0% for rs12143842, it is possible that the strong association of rs10494366 is largely explained by the much stronger association of the phenotype with rs12143842C>T, since the *P*-value is a function of effect size, sample size (power), SNP minor allele frequency and LD with the causative SNP. Therefore, LD in the human genome allows identification of functional loci without directly genotyping the causative allele, but makes discrimination between the functional variant and all correlated variants difficult. In line with this argument it is very well possible that also rs12143842 is not the causal SNP. Two other common variants in the HapMap CEU database are in considerable LD (r^2>0.8) with rs12143842, namely rs12036340 (genotyped on the Affymetrix platform which showed equal effect sizes and rs16847548T>C (not genotyped). None of the inferred haplotypes showed stronger associations than rs12143842 alone. However, in a conditional regression analysis of these SNPs, statistical independence between rs12143842C>T and rs10494366T>G / rs6670339T>C was shown, which might indicate the presence of another (causal) SNP linked to both SNPs. Both rs12143842C>T and rs10494366T>G variant alleles might have, to a certain extent, the ability to predict this variant in the absence of one another. However, the presence of two independent causal variants at the *NOS1AP* locus cannot be ruled out completely. It can be noted that the rs6670339 C-allele is a better predictor of QT interval duration than the rs10494366 G-allele in subjects with an rs12143842 CC-genotype.

The rs12143842C>T common variant

The rs12143842C>T common variant is located in the 5' region of *NOS1AP*. Remarkably, the first fine map attempt in the original paper describing the association between *NOS1AP* and QT interval duration already suggested that this region harbors a potential variant more strongly related to QT interval duration[31]. This strengthens the validity of our results. The rs12143842 T-allele frequency reported in the HapMap CEU sample (15.8%) was lower than that observed in our population (24.0%). These differences might be due to the lower number of subjects in the HapMap sample or to differences in ancestry, which are also reflected in different r^2 values for

rs12143842 | rs10494366 in the HapMap sample (r^2=0.11) compared to our sample (r^2=0.46).

The functionality of this SNP is not clear. Given its location in a potential regulatory region we looked for potential transcription factors binding sites at this position using the TESS web tool (http://www.cbil.upenn.edu/cgi-bin/tess/tess)[49]. For rs12143842 the C to T substitution resulted in two changes in predicted transcription factor binding sites. First, a loss of c-Myb as a potential transcription factor known to play a role in development of malignancies and in normal haematopoietic regulation[50-52]. Second, more interesting, the C to T substitution results in gaining the potential binding of myocyte-specific enhancer factor 2 (MEF2), alternative-myocyte-specific enhancer factor 2 (aMEF2) and Related to Serum Response factor C4 (RSRFC4) transcription factors that are the product from the MEF2 gene and bind conserved sequences found in growth factor-inducible and muscle-specific promoters with preferential expression in skeletal muscle, heart and brain[53,54]. We found no evidence that the highly linked rs12036340 had functional properties over possible human transcription factor binding sites. Another linked SNP, which was not directly genotyped in our study on either platform, rs16847548T>C (HapMap: minor allele frequency 12.7%; r^2 to rs12143842 of 0.82), is also situated in the 5' region of *NOS1AP* but does not seem to result in a change of the potential to bind a transcription factor. Any additional conclusions on functionality derived from in silico information needs to be tested by functional experiments in order to prove that *NOS1AP* is truly related to QT interval prolongation through transcription factor binding differences caused by the rs12143842C>T polymorphism. The identification of the causative variant is difficult and requires laborious and resource intensive methods. Future experiments that can be thought of include zebra fish and mice models for cardiac repolarization, which has been successful in the past for QT interval research, where the polymorphism can be knocked in to observe the effect on the phenotype[55-57].

Study strengths and study limitations

Major advantages of our study are the large community based sample size, with up to 4 ECGs per subjects increasing the number of eligible subjects and ECGs available for analysis. Furthermore, the use of digital ECG recordings processed by the MEANS system reduces information bias and inter-observer differential assessment of QT interval duration. Detailed pharmacy data on day-to-day drug exposure allowed exclusion of ECGs taken during administration of QT interval prolonging or shortening drugs. With respect to genotype data, we had the advantage of having additional data from the Affymetrix 500k array in a subset of our total study population, allowing for enrichment of genotyped SNP density. This increased the coverage of genome variability in this region, reducing the chance of missing a strong association arising from variants in low LD with the Illumina SNPs. However, even if coverage is high, difficulty in discrimination between signals remains a problem that cannot be solved by solely increasing genome coverage. This raises the limitation that our results are restricted to a European ancestry population. The rs12143842 common variant shows LD with other SNPs at the locus in the HapMap CEU sample. Since LD at this locus is less an issue in the HapMap Yoruban sample (e.g. rs12036340 | rs12143842: YRI r^2=0.45, CEU r^2=0.94, 385 women sample r^2=0.91), confounded results due to LD are less likely if common variants in the *NOS1AP* locus are tested for association with QT interval

duration in a population of African ancestry. In theory such admixture mapping would lead to narrowing the area of association further, given that this locus is of relevance for QT interval duration in non-Caucasians.

CONCLUSION

We have successfully fine mapped the prior association of rs10494366T>G at the *NOS1AP* locus with QT interval duration by newly identifying rs12143842C>T in the 5' region of *NOS1AP* to be strongly associated. This SNP is promising in silico regarding transcription factor binding abilities. However, we have also showed that rs6670339C>T remained an independent predictor of QT interval within rs12143842 CC genotype carriers Therefore, the presence of another causal SNP linked to both rs12143842 and rs6670339 or multiple causal variants at his locus remain possible and should be explored in future research. Until the next step is made, rs12143842C>T is the best indicator in a Caucasian population for the effect of common variants in the *NOS1AP* locus on QT interval duration.

REFERENCES

1. Myerburg, R.J. and Castellanos, A. (2006) Emerging paradigms of the epidemiology and demographics of sudden cardiac arrest. *Heart Rhythm*, 3, 235-239.
2. Dekker, L.R., Bezzina, C.R., Henriques, J.P., Tanck, M.W., Koch, K.T., Alings, M.W., Arnold, A.E., de Boer, M.J., Gorgels, A.P., Michels, H.R. *et al.* (2006) Familial sudden death is an important risk factor for primary ventricular fibrillation: a case-control study in acute myocardial infarction patients. *Circulation*, 114, 1140-1145.
3. Friedlander, Y., Siscovick, D.S., Weinmann, S., Austin, M.A., Psaty, B.M., Lemaitre, R.N., Arbogast, P., Raghunathan, T.E. and Cobb, L.A. (1998) Family history as a risk factor for primary cardiac arrest. *Circulation*, 97, 155-160.
4. Friedlander, Y., Siscovick, D.S., Arbogast, P., Psaty, B.M., Weinmann, S., Lemaitre, R.N., Raghunathan, T.E. and Cobb, L.A. (2002) Sudden death and myocardial infarction in first degree relatives as predictors of primary cardiac arrest. *Atherosclerosis*, 162, 211-216.
5. Jouven, X., Desnos, M., Guerot, C. and Ducimetiere, P. (1999) Predicting sudden death in the population: the Paris Prospective Study I. *Circulation*, 99, 1978-1983.
6. Kaikkonen, K.S., Kortelainen, M.L., Linna, E. and Huikuri, H.V. (2006) Family history and the risk of sudden cardiac death as a manifestation of an acute coronary event. *Circulation*, 114, 1462-1467.
7. Newton-Cheh, C., Larson, M.G., Corey, D.C., Benjamin, E.J., Herbert, A.G., Levy, D., D'Agostino, R.B. and O'Donnell, C.J. (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Heart Rhythm*, 2, 277-284.
8. Hong, Y., Rautaharju, P.M., Hopkins, P.N., Arnett, D.K., Djousse, L., Pankow, J.S., Sholinsky, P., Rao, D.C. and Province, M.A. (2001) Familial aggregation of QT-interval variability in a general population: results from the NHLBI Family Heart Study. *Clin Genet*, 59, 171-177.
9. Splawski, I., Shen, J., Timothy, K.W., Lehmann, M.H., Priori, S., Robinson, J.L., Moss, A.J., Schwartz, P.J., Towbin, J.A., Vincent, G.M. *et al.* (2000) Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation*, 102, 1178-1185.
10. Chugh, S.S., Senashova, O., Watts, A., Tran, P.T., Zhou, Z., Gong, Q., Titus, J.L. and Hayflick, S.J. (2004) Postmortem molecular screening in unexplained sudden death. *J Am Coll Cardiol*, 43, 1625-1629.
11. Doolan, A., Langlois, N., Chiu, C., Ingles, J., Lind, J.M. and Semsarian, C. (2007) Postmortem molecular analysis of KCNQ1 and SCN5A genes in sudden unexplained death in young Australians. *Int J Cardiol.*
12. Gaita, F., Giustetto, C., Bianchi, F., Wolpert, C., Schimpf, R., Riccardi, R., Grossi, S., Richiardi, E. and Borggrefe, M. (2003) Short QT Syndrome: a familial cause of sudden death. *Circulation*, 108, 965-970.
13. Tester, D.J. and Ackerman, M.J. (2007) Postmortem long QT syndrome genetic testing for sudden unexplained death in the young. *J Am Coll Cardiol*, 49, 240-246.

14. Algra, A., Tijssen, J.G., Roelandt, J.R., Pool, J. and Lubsen, J. (1993) QT interval variables from 24 hour electrocardiography and the two year risk of sudden death. *Br Heart J*, 70, 43-48.
15. Karjalainen, J., Reunanen, A., Ristola, P. and Viitasalo, M. (1997) QT interval as a cardiac risk factor in a middle aged population. *Heart,* 77, 543-548.
16. Schouten, E.G., Dekker, J.M., Meppelink, P., Kok, F.J., Vandenbroucke, J.P. and Pool, J. (1991) QT interval prolongation predicts cardiovascular mortality in an apparently healthy population. *Circulation*, 84, 1516-1523.
17. Siscovick, D.S., Raghunathan, T.E., Rautaharju, P., Psaty, B.M., Cobb, L.A. and Wagner, E.H. (1996) Clinically silent electrocardiographic abnormalities and risk of primary cardiac arrest among hypertensive patients. *Circulation*, 94, 1329-1333.
18. Straus, S.M., Kors, J.A., De Bruin, M.L., van der Hooft, C.S., Hofman, A., Heeringa, J., Deckers, J.W., Kingma, J.H., Sturkenboom, M.C., Stricker, B.H. *et al.* (2006) Prolonged QTC interval and risk of sudden cardiac death in a population of older adults. *J Am Coll Cardiol*, 47, 362-367.
19. Roden, D.M. (2004) Drug-induced prolongation of the QT interval. *N Engl J Med*, 350, 1013-1022.
20. Yap, Y.G. and Camm, A.J. (2003) Drug induced QT prolongation and torsades de pointes. *Heart,* 89, 1363-1372.
21. Akyol, M., Jalilzadeh, S., Sinner, M.F., Perz, S., Beckmann, B.M., Gieger, C., Illig, T., Wichmann, H.E., Meitinger, T., Kaab, S. *et al.* (2007) The common non-synonymous variant G38S of the KCNE1-(minK)-gene is not associated to QT interval in Central European Caucasians: results from the KORA study. *Eur Heart J*, 28, 305-309.
22. Aydin, A., Bahring, S., Dahm, S., Guenther, U.P., Uhlmann, R., Busjahn, A. and Luft, F.C. (2005) Single nucleotide polymorphism map of five long-QT genes. *J Mol Med*, 83, 159-165.
23. Bezzina, C.R., Verkerk, A.O., Busjahn, A., Jeron, A., Erdmann, J., Koopmann, T.T., Bhuiyan, Z.A., Wilders, R., Mannens, M.M., Tan, H.L. *et al.* (2003) A common polymorphism in KCNH2 (HERG) hastens cardiac repolarization. *Cardiovasc Res*, 59, 27-36.
24. Friedlander, Y., Vatta, M., Sotoodehnia, N., Sinnreich, R., Li, H., Manor, O., Towbin, J.A., Siscovick, D.S. and Kark, J.D. (2005) Possible association of the human KCNE1 (minK) gene and QT interval in healthy subjects: evidence from association and linkage analyses in Israeli families. *Ann Hum Genet,* 69, 645-656.
25. Gouas, L., Nicaud, V., Berthet, M., Forhan, A., Tiret, L., Balkau, B., Guicheney, P. and Group, D.E.S.I.R.S. (2005) Association of KCNQ1, KCNE1, KCNH2 and SCN5A polymorphisms with QTC interval length in a healthy population. *Eur J Hum Genet*, 13, 1213-1222.
26. Gouas, L., Nicaud, V., Chaouch, S., Berthet, M., Forhan, A., Tichet, J., Tiret, L., Balkau, B. and Guicheney, P. (2007) Confirmation of associations between ion channel gene SNPs and QTC interval duration in healthy subjects. *Eur J Hum Genet*, 15, 974-979.
27. Newton-Cheh, C., Guo, C.Y., Larson, M.G., Musone, S.L., Surti, A., Camargo, A.L., Drake, J.A., Benjamin, E.J., Levy, D., D'Agostino, R.B., Sr. *et al.* (2007) Common genetic variation in KCNH2 is associated with QT interval duration: the Framingham Heart Study. *Circulation*, 116, 1128-1136.
28. Pfeufer, A., Jalilzadeh, S., Perz, S., Mueller, J.C., Hinterseer, M., Illig, T., Akyol, M., Huth, C., Schopfer-Wendels, A., Kuch, B. *et al.* (2005) Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. *Circ Res*, 96, 693-701.
29. Pietila, E., Fodstad, H., Niskasaari, E., Laitinen, P.P., Swan, H., Savolainen, M., Kesaniemi, Y.A., Kontula, K. and Huikuri, H.V. (2002) Association between HERG K897T polymorphism and QT interval in middle-aged Finnish women. *J Am Coll Cardiol*, 40, 511-514.
30. Aarnoudse, A.J., Newton-Cheh, C., de Bakker, P.I., Straus, S.M., Kors, J.A., Hofman, A., Uitterlinden, A.G., Witteman, J.C. and Stricker, B.H. (2007) Common *NOS1AP* variants are associated with a prolonged QTC interval in the Rotterdam Study. *Circulation,* 116, 10-16.
31. Arking, D.E., Pfeufer, A., Post, W., Kao, W.H., Newton-Cheh, C., Ikeda, M., West, K., Kashuk, C., Akyol, M., Perz, S. *et al.* (2006) A common genetic variant in the *NOS1* regulator *NOS1AP* modulates cardiac repolarization. *Nat Genet*, 38, 644-651.
32. Lehtinen, A.B., Newton-Cheh, C., Ziegler, J.T., Langefeld, C.D., Freedman, B.I., Daniel, K.R., Herrington, D.M. and Bowden, D.W. (2008) Association of *NOS1AP* genetic variants with QT interval duration in families from the Diabetes Heart Study. *Diabetes*, 57, 1108-1114.
33. Post, W., Shen, H., Damcott, C., Arking, D.E., Kao, W.H., Sack, P.A., Ryan, K.A., Chakravarti, A., Mitchell, B.D. and Shuldiner, A.R. (2007) Associations between genetic variants in the *NOS1AP* (CAPON) gene and cardiac repolarization in the old order Amish. *Hum Hered*, 64, 214-219.

34. Tobin, M.D., Kahonen, M., Braund, P., Nieminen, T., Hajat, C., Tomaszewski, M., Viik, J., Lehtinen, R., Ng, G.A., Macfarlane, P.W. *et al.* (2008) Gender and effects of a common genetic variant in the *NOS1* regulator *NOS1AP* on cardiac repolarization in 3761 individuals from two independent populations. *Int J Epidemiol*, 37, 1132-1141.
35. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22, 819-829.
36. Hofman, A., Grobbee, D.E., de Jong, P.T. and van den Ouweland, F.A. (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*, 7, 403-422.
37. De Ponti, F., Poluzzi, E., Cavalli, A., Recanatini, M. and Montanaro, N. (2002) Safety of non-antiarrhythmic drugs that prolong the QT interval or induce torsade de pointes: an overview. *Drug Saf*, 25, 263-286.
38. Kang, J., Chen, X.L., Wang, L. and Rampe, D. (2001) Interactions of the antimalarial drug mefloquine with the human cardiac potassium channels KVLQT1/minK and HERG. *J Pharmacol Exp Ther*, 299, 290-296.
39. Kitayama, H., Kiuchi, K., Nejima, J., Katoh, T., Takano, T. and Hayakawa, H. (1999) Long-term treatment with antipsychotic drugs in conventional doses prolonged QTc dispersion, but did not increase ventricular tachyarrhythmias in patients with schizophrenia in the absence of cardiac disease. *Eur J Clin Pharmacol*, 55, 259-262.
40. Stollberger, C., Huber, J.O. and Finsterer, J. (2005) Antipsychotic drugs and QT prolongation. *Int Clin Psychopharmacol*, 20, 243-251.
41. Cheng, T.O. (2004) Digitalis administration: an underappreciated but common cause of short QT interval. *Circulation*, 109, e152; author reply e152.
42. Saner, H.E., Lange, H.W., Pierach, C.A. and Aeppli, D.M. (1988) Relation between serum digoxin concentration and the electrocardiogram. *Clin Cardiol,* 11, 752-756.
43. Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16, 1215.
44. Kayser, M., Liu, F., Janssens, A.C., Rivadeneira, F., Lao, O., van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M.M., van Ijcken, W.F. *et al.* (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet*, 82, 411-423.
45. Stephens, M. and Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*, 73, 1162-1169.
46. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, 68, 978-989.
47. Pe'er, I., de Bakker, P.I., Maller, J., Yelensky, R., Altshuler, D. and Daly, M.J. (2006) Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet*, 38, 663-667.
48. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559-575.
49. Schug, J. (2003) Using TESS to Predict Transcription factor Binding Sites in DNA Sequence. In Baxevanis, A.D. (ed.), *Current Protocols in Bioinformatics*. J. Wiley and Sons, p. .
50. Melotti, P. and Calabretta, B. (1996) The transcription factors c-myb and GATA-2 act independently in the regulation of normal hematopoiesis. *Proc Natl Acad Sci USA*, 93, 5313-5318.
51. Slamon, D.J., Boone, T.C., Murdock, D.C., Keith, D.E., Press, M.F., Larson, R.A. and Souza, L.M. (1986) Studies of the human c-myb gene and its product in human acute leukemias. *Science*, 233, 347-351.
52. Sureau, A., Soret, J., Vellard, M., Crochet, J. and Perbal, B. (1992) The PR264/c-myb connection: expression of a splicing factor modulated by a nuclear protooncogene. *Proc Natl Acad Sci USA*, 89, 11683-11687.
53. Pollock, R. and Treisman, R. (1991) Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev*, 5, 2327-2341.
54. Yu, Y.T., Breitbart, R.E., Smoot, L.B., Lee, Y., Mahdavi, V. and Nadal-Ginard, B. (1992) Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. *Genes Dev*, 6, 1783-1798.
55. Milan, D.J., Jones, I.L., Ellinor, P.T. and MacRae, C.A. (2006) In vivo recording of adult zebrafish electrocardiogram and assessment of drug-induced QT prolongation. *Am J Physiol Heart Circ Physiol*, 291, H269-273.
56. Milan, D.J., Peterson, T.A., Ruskin, J.N., Peterson, R.T. and MacRae, C.A. (2003) Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation*, 107, 1355-1358.
57. Remme, C.A., Verkerk, A.O., Nuyens, D., van Ginneken, A.C., van Brunschot, S., Belterman, C.N., Wilders, R., van Roon, M.A., Tan, H.L., Wilde, A.A. *et al.* (2006) Overlap syndrome of cardiac sodium channel disease in mice carrying the equivalent mutation of human SCN5A-1795insD. *Circulation*, 114, 2584-2594.

2.2 Genetic variation in *NOS1AP* is associated with sudden cardiac death: evidence from the Rotterdam Study

Eijgelsheim M, Newton-Cheh C, Aarnoudse AL, van Noord C, Witteman JC, Hofman A, Uitterlinden AG, Stricker BHCh

Hum Mol Genet. 2009 Nov 1;18(21):4213-4218.

ABSTRACT

Common variation within the Nitric Oxide-1 Synthase Activator Protein (*NOS1AP*) locus is strongly related to QT interval, a sudden cardiac death (SCD) risk factor. A recent report describes common variation in *NOS1AP* associated with SCD in a US population of European ancestry. The objective of the current study was to obtain additional evidence by investigating the association between *NOS1AP* variants and SCD in the prospective population based Rotterdam Study. The study population consisted of 5,974 European ancestry subjects, aged 55 years and older, genotyped on Illumina arrays. SCD was defined according to European Society of Cardiology guidelines. Smoking, body mass index, diabetes mellitus, hypertension, heart failure, and myocardial infarction were used as covariates in Cox proportional hazard models. Results were combined with reported evidence using inverse-variance weighted meta-analysis. 208 (109 witnessed) cases of SCD occurred during a mean follow up of 10.4 years. Within the Rotterdam Study alone no significant associations were observed. Upon pooling of results with existing data we observed strengthening of existing evidence for rs16847549 (US data hazard ratio (HR)=1.31, *P*=0.0024; Rotterdam Study HR=1.18, *P*=0.16; joint HR=1.26, *P*=0.0011). When the case definition in the Rotterdam Study was restricted to witnessed SCD, association of rs16847549 with SCD became stronger (joint *P*=0.00019) and additionally the association between rs12567209 and SCD gained significance (US data HR=0.57, *P*=0.0035; Rotterdam Study HR=0.69, *P*=0.23; joint HR=0.60, *P*=0.0018). In conclusion, this study provided additional evidence for association between genetic variation within *NOS1AP* and SCD. The mechanism by which this effect is exerted remains to be elucidated.

INTRODUCTION

Sudden cardiac death (SCD) is one of the major causes of cardiovascular mortality in developed countries and mostly occurs in individuals unrecognised to be at risk[1]. Familial aggregation of SCD, independent of other risk factors, suggests a genetic role in SCD risk[2-5]. By the use of genome-wide association studies (GWAS) it is possible to search for common variants associated with complex traits. Due to the small size of existing SCD collections, the statistical power to study this trait through GWA studies is limited. The focus has been on intermediate quantitative measures predictive of SCD. One of the quantitative predictors of arrhythmogenesis is the electrocardiographic (ECG) QT interval duration, a non-invasive measure of ventricular repolarisation. Variation of the QT interval in the general population is approximately 35% heritable and prolonged QT intervals are associated with increased cardiovascular morbidity and mortality, including SCD[6-8]. Using GWAS, a common variant (rs10494366) in the nitric oxide synthase 1 adaptor protein (*NOS1AP*) gene was identified and reproducibly associated with QT interval duration[9-13]. Previously, we have reported a fine mapping effort for this association in the Rotterdam Study, that refined the signal of association to rs12143842[14]. This SNP was consistently the strongest associated variant in recent GWAS reports on QT interval[15-17].

Recently, two SNPs with very low linkage disequilibrium (rs16847548 and rs12567209) at the *NOS1AP* locus were reported to be independently associated with SCD risk in a large US community based population study of European ancestry[18], but to date no replication has been reported. The rs12143842 common variant is in considerable linkage disequilibrium with rs16847548 (HapMap CEU $r^2 = 0.82$) but not rs12567209 (HapMap CEU $r^2 = 0.027$). The goal of the present study was to test for association of rs12143842 or its recently reported proxy rs16847548 and the second independent variant rs12567209 with SCD in the Rotterdam Study, to attempt to validate these associations.

MATERIALS & METHODS

Setting and design

The Rotterdam Study is a population-based cohort study of chronic diseases in the elderly, which started in 1990. All inhabitants of Ommoord, a Rotterdam suburb in the Netherlands, aged ≥ 55 years (n=10,278) were invited to participate. Of them, 78% (n=7,983) gave their written informed consent. Baseline examinations took place from March 1990 through July 1993. Follow up examinations are carried out periodically. Participants are continuously monitored for major events through automated linkage with general practitioner's files. Clinical characteristics, including smoking, body mass index, diabetes mellitus, hypertension, heart failure, and myocardial infarction were ascertained as previously described[8,19-22]. Detailed information on design, objectives and methods has been described elsewhere[23,24]. The Medical Ethics Committee of the Erasmus Medical Center approved the study and all participants gave written informed consent.

Adjudication of SCD

The ascertainment of SCD cases in the Rotterdam Study was performed according to European Society of Cardiology guidelines[25] and has been described previously[8]. SCD cases were defined as a witnessed natural death attributable to cardiac causes, heralded by abrupt loss of consciousness, within one hour of onset of acute symptoms, or as an unwitnessed, unexpected death of a person seen in a stable medical condition within 24 hours before death without evidence of a non-cardiac cause.

QT, QRS and RR interval measurement

As described in earlier studies of ECG parameters in the Rotterdam Study[8], we used a 10 second resting 12-lead ECG (average 8–10 beats), which was recorded on an ACTA ECG (ESAOTE, Florence, Italy) at a sampling frequency of 500 HZ and stored digitally. All ECGs were processed by the Modular ECG Analysis System (MEANS), which has been evaluated extensively, to obtain QT, QRS and RR duration measurements[26-28]. MEANS also determines the presence of right or left bundle-branch block. For the current analyses, we used measurements from baseline ECG recordings.

Genotyping of *NOS1AP* common variants

Genotyping, quality control and imputation procedures in the Rotterdam Study have been previously described[15]. The Rotterdam Study is a homogenous European sample and quality control procedures removed population outliers from the GWAS dataset. Briefly, of all 7,983 participants, 5,974 subjects of European descent were successfully genotyped on the Infinium II HumanHap550K Genotyping BeadChip® version 3 (Illumina) as part of a large population-based project on genetics of complex traits and diseases. In addition to the direct genotypes generated, we have access to ~2.5 million autosomal genotypes that were imputed based on linkage disequilibrium patterns observed in HapMap CEU reference samples (Utah residents of Northern and Western European descent) using Mach 1.0.15[29]. Imputed was the genotype dosage, a value between 0–2, that reflects the expected number of alleles.

For all subjects direct genotype information of rs12143842 and rs12567209 was extracted from the dataset. For rs16847548 the imputed genotype dosage was retrieved. The metric used to reflect imputation quality is the observed over expected variance of the allele frequency[30]. This ratio ranges between 0 and 1, where a value of 1 reflects perfect imputation. No other genotypes were extracted and tested for an association with SCD.

Association analysis

Genotype frequencies of genotyped SNPS were tested for Hardy-Weinberg equilibrium by calculating exact *P*-values. Hazard ratios for time to SCD from baseline were estimated with Cox proportional hazards models. Cox proportional hazards analyses were performed with SPSS for Windows, 15.0 (SPSS Inc, Chicago, Ill). The proportional hazards assumption was assessed using log-minus-log plots. Overall, additive genetic models were used but for rs12567209 a dominant model was additionally fitted due to the lower minor allele frequency, and thus the small

number of homozygous variant carriers. In addition to the genotype, known risk factors such as gender, smoking and hypertension at baseline and time-dependent age, body mass index, diabetes mellitus, heart failure, and myocardial infarction were included in the model. To assess if the SNPs exert their effect on SCD through modulation of the heart rate corrected QT interval, both QT interval and heart rate measured at baseline were included in the model.

The primary outcome definition included both witnessed and unwitnessed SCD. Since potential random misclassification of the outcome was expected in unwitnessed cases, additional analyses describing the association between the genetic variants and SCD were performed using a restricted definition, only including witnessed SCD. This was regarded as a secondary outcome. We additionally estimated the effect of QT interval and gender, established SCD risk factors, using both the primary (all) and secondary (witnessed) outcome definition. Larger effects are expected when using a stricter case definition due to a reduction in non-differential misclassification.

Since the presented sample was smaller than the sample in the original report describing the association between these SNPs and SCD[18], non-significant findings in the current analysis might be due to a lack of power. Nevertheless, this sample may add valuable evidence for association by combination with earlier results. Hereto, results were combined with the results from the literature using inverse variance weighted meta-analyses methods and the overall association evidence is presented.

RESULTS

Study subjects and genotyping

Baseline characteristics for the study population, consisting of all genotyped Rotterdam Study participants (n=5,974), and SCD cases are summarized in Table 1. During a mean of 10.4 years of follow-up 208 sudden cardiac deaths, 109 of which witnessed, were identified. Successful genotype calls were made in 99.7% (rs12143842) and 99.8% (rs12567209) of the subjects, re-

Table 1 *Baseline characteristics of the Rotterdam Study study population*

	Genotyped Sample		All SCD cases		Witnessed SCD cases	
	Men n=2,427 (40.6%)	Women n=3,547 (59.4%)	Men n=104 (50%)	Women n=104 (50%)	Men n=67 (61.5%)	Women n=42 (38.5%)
Age, years, mean (SD)	68.1 (8.2)	70.3 (9.6)	71.5 (7.5)	74.0 (7.7)	70.1 (6.9)	72.7 (8.2)
Follow-up time, years, mean (SD)	10.0 (3.8)	10.6 (3.7)	6.4 (3.7)	7.7 (3.7)	6.4 (3.5)	6.7 (3.6)
Current smoking, n (%)	716 (29.5)	623 (17.6)	28 (26.9)	14 (13.5)	18 (26.9)	5 (11.9)
Past smoking, n (%)	1474 (60.7)	951 (26.8)	69 (66.4)	35 (33.7)	44 (65.7)	19 (45.2)
Body mass index, kg/m^2, mean (SD)	25.7 (3.0)	26.7 (4.1)	25.4 (3.0)	27.2 (4.0)	25.4 (3.0)	27.1 (4.2)
Hypertension, n (%)	714 (29.4)	1283 (36.2)	49 (47.1)	57 (54.8)	31 (46.3)	25 (59.5)
Diabetes Mellitus, n (%)	248 (10.2)	383 (10.8)	13 (12.5)	23 (22.1)	11 (16.4)	13 (31.0)
Myocardial infarction, n (%)	412 (17.0)	284 (8.0)	42 (40.4)	16 (15.4)	26 (38.8)	6 (14.3)
Heart failure, n (%)	74 (3.0)	120 (3.4)	17 (16.3)	6 (5.8)	10 (14.9)	1 (2.4)

spectively. The genotype frequencies are in Hardy-Weinberg equilibrium for both rs12143842 (T-allele frequency=24.0%) and rs12567209 (A-allele frequency=14.9%) with P=0.72 and P=0.85, respectively. The imputed genotype for rs16847548 was available for all subjects. The imputation quality of this SNP was very good with an observed/expected variance ratio of 0.993.

For secondary analyses exploring the effect of a differential outcome definition on established risk factors the study population consisted of 5,661 subjects from the Rotterdam Study, all with available QT measurements and without a left or right bundle branch block, atrial fibrillation and a QRS duration <120 milliseconds, to minimize QT measurement errors. During a mean follow up of 10.8 years 154 cases of SCD were identified, 85 of which witnessed.

NOS1AP variants association with SCD in the Rotterdam Study

The proportional hazards assumption for a constant hazard ratio (HR) over time was met. After adjustment for age and sex, rs12143842, rs16847548 and rs12567209 variants showed non-significant trends for association with SCD under an additive genetic model in the Rotterdam Study alone.(Table 2) Since rs16847548 is in linkage disequilibrium with rs12143842 (HapMaP CEU r^2=0.82) the results are very similar for these SNPs (Table 3). For simplicity in subsequent analyses we show results for the genotyped rs12143842 SNP only. Upon inclusion of both rs12567209 and rs12143842 in the same model, in addition to age and sex, the hazard ratio (HR) for rs12567209 changed from 0.85 (95% confidence interval (CI) 0.57 – 1.26) to 0.86 (95% CI 0.57 – 1.28) and for rs12143842 from 1.18 (95% CI 0.95 – 1.47) to 1.17 (95% CI 0.93 – 1.45). When QT interval, heart rate (RR interval) and QRS duration were subsequently added as continuous variables in the model,

Table 2 *Association results for rs12567209 and rs12143842 with all and witnessed Sudden Cardiac Death in the Rotterdam Study.*

	Cases	rs12567209		rs12143842	
	n	GG	HR (95% CI) per additional A-allele	CC	HR (95% CI) per additional T-allele
All SCD, model 1*	208	1.00 (ref)	0.85 (0.57–1.26) P=0.42	1.00 (ref)	1.18 (0.95–1.47) P=0.13
All SCD, model 2†	208	1.00 (ref)	0.86 (0.57–1.28) P=0.18	1.00 (ref)	1.17 (0.93–1.45) P=0.18
All SCD, model 3‡	178	1.00 (ref)	0.74 (0.47–1.17) P=0.20	1.00 (ref)	1.06 (0.82–1.35) P=0.66
All SCD, model 4§	175	1.00 (ref)	0.74 (0.46–1.18) P=0.21	1.00 (ref)	1.11 (0.87–1.43) P=0.40
All SCD, model 5°	126	1.00 (ref)	0.77 (0.45–1.34) P=0.36	1.00 (ref)	1.14 (0.86–1.52) P=0.37
Witnessed SCD, model 1*	109	1.00 (ref)	0.69 (0.37–1.27) P=0.23	1.00 (ref)	1.42 (1.06–1.89) P=0.018
Witnessed SCD, model 2†	109	1.00 (ref)	0.73 (0.40–1.35) P=0.27	1.00 (ref)	1.39 (1.04–1.86) P=0.027
Witnessed SCD, model 3‡	99	1.00 (ref)	0.72 (0.38–1.37) P=0.31	1.00 (ref)	1.27 (0.93–1.74) P=0.14
Witnessed SCD, model 4§	95	1.00 (ref)	0.67 (0.34–1.33) P=0.25	1.00 (ref)	1.34 (0.97–1.85) P=0.08
Witnessed SCD, model 5°	66	1.00 (ref)	0.79 (0.36–1.72) P=0.55	1.00 (ref)	1.36 (0.93–1.99) P=0.11

* Model 1: adjusted for age and sex;
† Model 2: Model1 + alternate SNP;
‡ Model 3: Model 2 + QRS, QT and RR interval;
§ Model 4: Model3 + hypertension at baseline and time dependent, body mass index, diabetes mellitus, heart failure and myocardial infarction;
° Model 5: Model 4, after exclusion of subjects with left or right bundle branch block or QRS duration over 120 milliseconds.

the HRs for rs12567209 and rs12143842 were 0.74 (95% CI 0.47 – 1.17) and 1.06 (95% CI 0.82 – 1.35), respectively. (Table 2)

To compare our results with the results from Kao *et al.*[17], QT interval was also modelled as a categorical variable (quintiles). This showed similar results compared to analysis including the linear QT variable for both rs12567209 (HR=0.73; 95% CI 0.46 – 1.16) and rs12143842 (HR =1.07; 95% CI 0.84 – 1.36). Additional adjustment for myocardial infarction, diabetes, heart failure, hypertension, smoking and body mass index did not change the effect estimates. To account for potential measurement errors of the QT interval, additional analyses were performed with exclusion of those subjects who showed a left or right bundle branch block, QRS duration over 120 milliseconds or had prevalent atrial fibrillation. This did not change the results substantially. (Table 2)

Secondary analyses

Since some non-differential misclassification was expected in non-witnessed SCD, analyses were repeated using the restricted witnessed SCD case definition. The common variants studied here and established SCD risk factors were analyzed to assess the effect of the more stringent case definition. Stronger effects were observed for established SCD risk factors including QT interval and sex when using this restricted definition despite the reduction in case sample size. (Table 3) When the genetic variants were tested for association with witnessed SCD, significant associations were observed for rs12143842 (and the correlated SNP rs16847548). Under the assumption of an additive genetic model the HR per additional T-allele of rs12143842 was 1.42 (95% CI 1.06 – 1.89). Effect estimates for rs12567209 moved away from the null, but did not reach statistical significance, with an observed HR of 0.69 (95% CI 0.34 – 1.27).

Meta-analyses

The findings of this study were combined with recent results from Kao *et al* (Table 4). With the addition of Rotterdam Study results the evidence for association between rs16847548 strengthened, irrespective of the outcome definition used (original HR=1.31, P=0.0024; joint HR=1.26,

Table 3 *Hazard ratio's for classical SCD risk factors by phenotype definition in the total Rotterdam Study population with QT interval measurements and without bundle branch block, atrial fibrillation or QRS duration over 120 milliseconds (n=5,661). Presented are hazard ratio's per QT interval quintile adjuste for age, sex and RR interval (inverse heart rate) and hazard ratio for sex adjusted for age.*

	Cases, n	All SCD, HR (95% CI)	Cases, n	Witnessed SCD, HR (95% CI)
QT quintile 1	36	1.00 (ref)	16	1.00 (ref)
QT quintile 2	38	1.38 (0.85–2.25)	21	2.01 (1.00–4.03)
QT quintile 3	34	1.48 (0.86–2.56)	21	2.64 (1.24–5.64)
QT quintile 4	46	2.20 (1.18–4.09)	27	4.42 (1.87–10.44)
Male	75	1.00 (ref)	50	1.00 (ref)
Female	79	0.43 (0.31–0.60)	35	0.29 (0.19–0.46)

P=0.0011 for all SCD and HR=1.33, *P*=0.00019 for witnessed SCD). For rs12537209, the significance level only increased when we used the restricted witnessed SCD definition (original HR=0.57, *P*=0.0035; joint HR=0.69, *P*=0.0075 for all SCD and HR=0.60, *P*=0.0018, for witnessed SCD).

Table 4 *Joint analysis of Rotterdam Study results and ARIC plus CHS results. All results are based on sex and age adjusted allelic models.*

	rs12567209 Allelic		**Dominant**		**rs16847548 Allelic**	
ARIC* + CHS†						
HR	0.57		0.53		1.31	
Ln(HR)‡	-0.562		-0.635		0.270	
Se§	0.192		0.201		0.089	
P	0.0035		0.0015		0.0024	
RS°	All	Witnessed	All	Witnessed	All	Witnessed
HR	0.85	0.69	0.80	0.70	1.18	1.40
Ln(HR)	-0.165	-0.378	-0.229	-0.364	0.165	0.339
Se	0.204	0.311	0.217	0.318	0.117	0.153
P	0.42	0.23	0.29	0.25	0.16	0.027
Joint						
HR	0.69	0.60	0.64	0.57	1.26	1.33
Ln(HR)	-0.374	-0.511	-0.448	-0.558	0.232	0.287
Se	0.140	0.164	0.147	0.170	0.071	0.077
P	0.0074	0.0018	0.0024	0.0010	0.0011	0.00019

* ARIC: Atherosclerosis Risk in Communities Study
† CHS: Cardiovascular Health Study
‡ Ln(HR): natural log of the hazard ratio = beta estimate from Cox regression model
§ SE: standard error of the regression coefficient
° RS: Rotterdam Study

DISCUSSION

Main finding

In the present study, additional evidence is presented that common variation at the *NOS1AP* locus is associated with increased SCD risk. Adjustment for potential confounders, such as cardiovascular co-morbidities or risk factors did not change the results. Overall in the Rotterdam Study sample, no significant effects for association between common variants and SCD were observed. However, when we restricted the case definition to only witnessed SCD, significant effects for rs12143842 and its proxy rs16847549 were seen despite the reduced number of cases. Combination of the results with recently published data from the Cardiovascular Health Study and Atherosclerosis Risk in Communities study by Kao *et al.*[18] resulted in a strengthened overall significance of the association between rs16847548 and SCD, regardless of the phenotype definition. For rs12567209, the incremental effect on significance was dependent on the outcome definition. When a stricter outcome definition was used, this resulted in strengthened statistical significance.

NOS1AP and SCD

The mechanism by which *NOS1AP* influences QT interval and/or SCD risk remains to be fully elucidated. *NOS1AP* encodes the CAPON protein that is expressed in the heart and interacts with Nitric Oxide Synthase 1 (NOS1) enhancing nitric oxide (NO) production[31]. NOS1 and NO play an important role in regulating cardiac physiology such as calcium turnover and adrenergic response[32,33]. Overexpression of CAPON leads to inhibition of the L-type calcium channels and an increase in the delayed rectifier potassium current resulting in hastening of the action potential[31]. It remains to be proven whether the influence of common variation in *NOS1AP* on this physiological mechanism is also the mechanism for the increased risk of SCD. In the original report by Kao *et al*, association of the common variants with SCD was independent of the effect on QT interval duration. We observed that for rs12143842, a proxy for the variant reported by Kao *et al* rs16847548, the effect on SCD was attenuated upon inclusion of QT interval in the model. For rs12567209 (which has not been associated with QT interval duration) adjustment for QT interval resulted in a slightly stronger effect of the common variant on SCD, however no significant results were observed in the Rotterdam Study sample alone. From the existing data it is difficult to say whether these genetic variants exert their effect on SCD through modulation of the QT interval or via other mechanisms. Both in the presented and previously described sample, the QT interval was measured well before the event. Additionally, the common variants might very well not be the causal SNPs. As indicated by Kao *et al*, the effect measures of these SNPs on QT interval and SCD might differ due to unequal linkage to the causal SNP, leaving residual association between the SNP and SCD after adjustment for QT interval. The fact that we observed attenuation of the effect on SCD for rs12143842 could be due to stronger correlation between this variant and the causal SNP.

Methodological considerations

Advantages of our study are the large community based sample size, prospective ascertainment of risk factors and active surveillance for SCD events over a relatively long duration of follow-up. Extensive information was available with respect to SCD events, facilitating careful case adjudication. The homogeneity of the Rotterdam Study population and quality control procedures identifying population outliers result in a low chance of population substructure underlying the findings. In prior GWAS analysis of QT interval, no indication of population stratification was observed in the Rotterdam Study[15].

A limitation is the possibility of alternative causes of abrupt death, especially at increasing age. Specifically, for an unwitnessed abrupt death alternative causes of sudden death might lead to misclassification of the outcome due to less extensive information for adjudication. Since case ascertainment was blinded to genotype, this misclassification is non-differential and would lead to bias towards the null hypothesis of finding no association. Therefore, a reduction in misclassification of the outcome by excluding unwitnessed SCD could explain the observation of stronger associations for witnessed SCD. This is supported by the behavior of risk estimates for classically established SCD risk factors. Finally, with the available cases the power for replicating the findings for rs16847548 and rs12567209 was 68% and 92%, respectively, under the assumption there is no case misclassification within the Rotterdam Study.

CONCLUSION

This study provided additional evidence for an association between common genetic variation within *NOS1AP* and sudden cardiac death. The strength of association depended on the strictness of the phenotype definition used. The mechanism through which this effect is exerted remains to be elucidated.

REFERENCES

1. Myerburg, R.J. and Castellanos, A. (2006) Emerging paradigms of the epidemiology and demographics of sudden cardiac arrest. *Heart Rhythm*, 3, 235-239.
2. Friedlander, Y., Siscovick, D.S., Arbogast, P., Psaty, B.M., Weinmann, S., Lemaitre, R.N., Raghunathan, T.E. and Cobb, L.A. (2002) Sudden death and myocardial infarction in first degree relatives as predictors of primary cardiac arrest. *Atherosclerosis,* 162, 211-216.
3. Friedlander, Y., Siscovick, D.S., Weinmann, S., Austin, M.A., Psaty, B.M., Lemaitre, R.N., Arbogast, P., Raghunathan, T.E. and Cobb, L.A. (1998) Family history as a risk factor for primary cardiac arrest. *Circulation*, 97, 155-160.
4. Jouven, X., Desnos, M., Guerot, C. and Ducimetiere, P. (1999) Predicting sudden death in the population: the Paris Prospective Study I. *Circulation*, 99, 1978-1983.
5. Kaikkonen, K.S., Kortelainen, M.L., Linna, E. and Huikuri, H.V. (2006) Family history and the risk of sudden cardiac death as a manifestation of an acute coronary event. *Circulation*, 114, 1462-1467.
6. Hong, Y., Rautaharju, P.M., Hopkins, P.N., Arnett, D.K., Djousse, L., Pankow, J.S., Sholinsky, P., Rao, D.C. and Province, M.A. (2001) Familial aggregation of QT-interval variability in a general population: results from the NHLBI Family Heart Study. *Clin Genet,* 59, 171-177.
7. Newton-Cheh, C., Larson, M.G., Corey, D.C., Benjamin, E.J., Herbert, A.G., Levy, D., D'Agostino, R.B. and O'Donnell, C.J. (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Heart Rhythm*, 2, 277-284.
8. Straus, S.M., Kors, J.A., De Bruin, M.L., van der Hooft, C.S., Hofman, A., Heeringa, J., Deckers, J.W., Kingma, J.H., Sturkenboom, M.C., Stricker, B.H. *et al.* (2006) Prolonged QTC interval and risk of sudden cardiac death in a population of older adults. *J Am Coll Cardiol*, 47, 362-367.
9. Aarnoudse, A.J., Newton-Cheh, C., de Bakker, P.I., Straus, S.M., Kors, J.A., Hofman, A., Uitterlinden, A.G., Witteman, J.C. and Stricker, B.H. (2007) Common *NOS1AP* variants are associated with a prolonged QTC interval in the Rotterdam Study. Circulation, 116, 10-16.
10. Arking, D.E., Pfeufer, A., Post, W., Kao, W.H., Newton-Cheh, C., Ikeda, M., West, K., Kashuk, C., Akyol, M., Perz, S. *et al.* (2006) A common genetic variant in the NOS1 regulator *NOS1AP* modulates cardiac repolarization. *Nat Genet,* 38, 644-651.
11. Post, W., Shen, H., Damcott, C., Arking, D.E., Kao, W.H., Sack, P.A., Ryan, K.A., Chakravarti, A., Mitchell, B.D. and Shuldiner, A.R. (2007) Associations between genetic variants in the *NOS1AP* (CAPON) gene and cardiac repolarization in the old order Amish. *Hum Hered*, 64, 214-219.
12. Raitakari, O.T., Blom-Nyholm, J., Koskinen, T.A., Kahonen, M., Viikari, J.S. and Lehtimaki, T. (2008) Common variation in *NOS1AP* and KCNH2 genes and QT interval duration in young adults. The Cardiovascular Risk in Young Finns Study. *Ann Med*, 1-8.
13. Tobin, M.D., Kahonen, M., Braund, P., Nieminen, T., Hajat, C., Tomaszewski, M., Viik, J., Lehtinen, R., Ng, G.A., Macfarlane, P.W. *et al.* (2008) Gender and effects of a common genetic variant in the NOS1 regulator *NOS1AP* on cardiac repolarization in 3761 individuals from two independent populations. *Int J Epidemiol*, 37, 1132-1141.
14. Eijgelsheim, M., Aarnoudse, A.L., Rivadeneira, F., Kors, J.A., Witteman, J.C., Hofman, A., van Duijn, C.M., Uitterlinden, A.G. and Stricker, B.H. (2009) Identification of a common variant at the *NOS1AP* locus strongly associated to QT-interval duration. *Hum Mol Genet*, 18, 347-357.
15. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M., de Bakker, P.I., Yin, X., Estrada, K., Bis, J.C., Marciante, K., Rivadeneira, F., Noseworthy, P.A. *et al.* (2009) Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet*, 41, 399-406.

16. Nolte, I.M., Wallace, C., Newhouse, S.J., Waggott, D., Fu, J., Soranzo, N., Gwilliam, R., Deloukas, P., Savelieva, I., Zheng, D. *et al.* (2009) Common genetic variation near the phospholamban gene is associated with cardiac repolarisation: meta-analysis of three genome-wide association studies. *PLoS One*, 4, e6138.
17. Pfeufer, A., Sanna, S., Arking, D.E., Muller, M., Gateva, V., Fuchsberger, C., Ehret, G.B., Orru, M., Pattaro, C., Kottgen, A. *et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet*, 41, 407-414.
18. Kao, W.H., Arking, D.E., Post, W., Rea, T.D., Sotoodehnia, N., Prineas, R.J., Bishe, B., Doan, B.Q., Boerwinkle, E., Psaty, B.M. *et al.* (2009) Genetic Variations in Nitric Oxide Synthase 1 Adaptor Protein Are Associated With Sudden Cardiac Death in US White Community-Based Populations. *Circulation*.
19. Bleumink, G.S., Knetsch, A.M., Sturkenboom, M.C., Straus, S.M., Hofman, A., Deckers, J.W., Witteman, J.C. and Stricker, B.H. (2004) Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure The Rotterdam Study. *Eur Heart J*, 25, 1614-1619.
20. Bots, M.L., Hoes, A.W., Koudstaal, P.J., Hofman, A. and Grobbee, D.E. (1997) Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study. *Circulation*, 96, 1432-1437.
21. Dehghan, A., Kardys, I., de Maat, M.P., Uitterlinden, A.G., Sijbrands, E.J., Bootsma, A.H., Stijnen, T., Hofman, A., Schram, M.T. and Witteman, J.C. (2007) Genetic variation, C-reactive protein levels, and incidence of diabetes. *Diabetes*, 56, 872-878.
22. Mosterd, A., Hoes, A.W., de Bruyne, M.C., Deckers, J.W., Linker, D.T., Hofman, A. and Grobbee, D.E. (1999) Prevalence of heart failure and left ventricular dysfunction in the general population; The Rotterdam Study. *Eur Heart J,* 20, 447-455.
23. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22, 819-829.
24. Hofman, A., Grobbee, D.E., de Jong, P.T. and van den Ouweland, F.A. (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*, 7, 403-422.
25. Priori, S.G., Aliot, E., Blomstrom-Lundqvist, C., Bossaert, L., Breithardt, G., Brugada, P., Camm, A.J., Cappato, R., Cobbe, S.M., Di Mario, C. *et al.* (2001) Task Force on Sudden Cardiac Death of the European Society of Cardiology. *Eur Heart J*, 22, 1374-1450.
26. de Bruyne, M.C., Kors, J.A., Hoes, A.W., Kruijssen, D.A., Deckers, J.W., Grosfeld, M., van Herpen, G., Grobbee, D.E. and van Bemmel, J.H. (1997) Diagnostic interpretation of electrocardiograms in population-based research: computer program research physicians, or cardiologists? *J Clin Epidemiol*, 50, 947-952.
27. Kors, J.A. and van Herpen, G. (2009) Methodology of QT-interval measurement in the modular ECG analysis system (MEANS). *Ann Noninvasive Electrocardiol*, 14 Suppl 1, S48-53.
28. van Bemmel, J.H., Kors, J.A. and van Herpen, G. (1990) Methodology of the modular ECG analysis system MEANS. *Methods Inf Med*, 29, 346-353.
29. Li, Y. and Abecasis, G.R. (2006) Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet*, S79, 2290.
30. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*, 17, R122-128.
31. Chang, K.C., Barth, A.S., Sasano, T., Kizana, E., Kashiwakura, Y., Zhang, Y., Foster, D.B. and Marban, E. (2008) CAPON modulates cardiac repolarization via neuronal nitric oxide synthase signaling in the heart. *Proc Natl Acad Sci USA*, 105, 4477-4482.
32. Barouch, L.A., Harrison, R.W., Skaf, M.W., Rosas, G.O., Cappola, T.P., Kobeissi, Z.A., Hobai, I.A., Lemmon, C.A., Burnett, A.L., O'Rourke, B. *et al.* (2002) Nitric oxide regulates the heart by spatial confinement of nitric oxide synthase isoforms. *Nature*, 416, 337-339.
33. Burkard, N., Rokita, A.G., Kaufmann, S.G., Hallhuber, M., Wu, R., Hu, K., Hofmann, U., Bonz, A., Frantz, S., Cartwright, E.J. *et al.* (2007) Conditional neuronal nitric oxide synthase overexpression impairs myocardial contractility. *Circ Res*, 100, e32-44.

PART 3

Genome-wide association studies on ECG derived traits

3.1 Common variants at ten loci influence QT interval duration in the QTGEN Study

Newton-Cheh C*, Eijgelsheim M*, Rice KM*, de Bakker PI*, Yin X, Estrada K, Bis JC, Marciante K, Rivadeneira F, Noseworthy PA, Sotoodehnia N, Smith NL, Rotter JI, Kors JA, Witteman JC, Hofman A, Heckbert SR, O'Donnell CJ, Uitterlinden AG, Psaty BM, Lumley T#, Larson MG#, Stricker BH#

Nat Genet. 2009 Apr;41(4):399-406.

* and # denote equal author contributions

ABSTRACT

QT interval duration reflecting myocardial repolarization on the electrocardiogram is a heritable risk factor for sudden cardiac death and drug-induced arrhythmias. We conducted a meta-analysis of 3 genome-wide association studies in 13,685 individuals of European ancestry from the Framingham Heart Study, the Rotterdam Study and the Cardiovascular Health Study. We observed associations at $P<5\times10^{-8}$ with variants in *NOS1AP*, *KCNQ1*, *KCNE1*, *KCNH2* and *SCN5A*, known to be involved in myocardial repolarization and Mendelian Long QT Syndromes. Associations at five novel loci included 16q21 near *NDRG4* and *GINS3*, 6q22 near *PLN*, 1p36 near *RNF207*, 16p13 near *LITAF* and 17q12 near *LIG3* and *RIFFL*. Collectively, the 14 independent variants at these 10 loci explain 5.4-6.5% of variation in QT interval. Identifying the causal variants and defining their impact on myocardial repolarization may add incrementally to the prevention of sudden cardiac death and drug-induced arrhythmias.

INTRODUCTION

Sudden cardiac death (SCD) and drug-induced ventricular arrhythmia, a major barrier to drug development, are poorly predicted[1]. Prolongation of electrocardiographic QT interval duration, a measure of myocardial repolarization time, is a risk factor for drug-induced arrhythmias and SCD. Continuous QT interval duration is heritable ($h^2 \approx 0.35$) and has multiple environmental and genetic contributors[2]. Its genetic determinants in populations are poorly characterized[3]. Congenital Long and Short QT Syndromes of ventricular arrhythmias and SCD due to extremes of QT interval duration are often caused by mutations with large effect sizes, commonly in ion channels involved in myocardial repolarization. These mutations are typically private to specific families and individually explain little of the population variation in QT interval duration or SCD risk[4]. The few common variants in candidate genes associated with QT interval duration thus far reported[5-9] leave much of its heritability unexplained. Genome-wide association studies covering the majority of common variation in the human genome can be used in large samples to identify genetic variants that typically confer modest effect sizes for quantitative complex traits such as QT interval duration. We completed a meta-analysis of three genome-wide association studies of QT interval duration in 13,685 self-identified white individuals of European ancestry drawn from three prospective cohort studies: the Framingham Heart Study (FHS, n = 7,650), the Rotterdam Study (RS, n = 4,606) and the first and second rounds of genotyping in the Cardiovascular Health Study (CHS, n = 1,429)[10]. We used genotypes from the Affymetrix 500K chip and 50K gene-centered MIP, the Illumina 550K and Illumina 370CNV arrays, respectively, to impute genotypes for 2,543,686 autosomal SNPs with reference to HapMap CEU linkage disequilibrium patterns.

METHODS

Study samples

The QTGEN consortium includes European-derived samples from three cohorts. The Framingham Heart Study (FHS) is a community-based, longitudinal cohort study comprising three generations of individuals in multigenerational pedigrees and additional unrelated individuals. The current study included individuals from Generation 1 (11th examination), Generation 2 (1st examination) and Generation 3 (1st examination)[11-13]. The Rotterdam Study (RS) is a prospective population-based cohort study of chronic diseases begun in 1990[14,15]. The current RS study sample included data from one of 4 examination cycles at which the first eligible electrocardiogram was available for each individual. The Cardiovascular Health Study (CHS) is a prospective, cohort study of risk factors for heart disease and stroke begun in 1989 and included 4,925 self-described white participants[16]. The CHS study sample used in this analysis included self-identified whites from the first two rounds of genotyping in a nested case-cohort study of myocardial infarction. Data on QT interval and covariates came from the baseline examination, at which prevalent cardiovascular disease was an exclusion criterion in the parent case-cohort study. All studies were approved by local institutional review boards and written informed consent was given.

Individuals were excluded for bundle branch block or QRS duration >120 milliseconds, atrial fibrillation or flutter, pacemaker activity, or use of a QT-altering medication (not applied to FHS Generation 3). After exclusions there were 7,650 FHS, 4,606 RS and 1,429 CHS individuals with phenotype and genotype data who contributed to genotype-phenotype association analyses.

QT measurement methods

In FHS, paper electrocardiograms were scanned and digital caliper measurements were made using proprietary software. In the Rotterdam Study, digital measurements of the QT interval were made using the Modular ECG Analysis System (MEANS)[17]. In the Cardiovascular Health Study, the electrocardiograms were recorded on MAC PC-DT ECG recorder (Marquette Electronics Inc, Milwaukee, WI, USA) machines and measurements of QT interval made using the Marquette 12SL algorithm (see *Supplementary Methods* for full details).

Phenotype modeling

The overall strategy involved linear regression to adjust QT interval for effects of age, sex, and RR interval (inverse heart rate) and residuals were used in genotype-phenotype association testing. In FHS, cohort-, sex- and cardiac cycle-specific regression models were created to adjust for age and RR interval. Residuals from these regression models were standardized to mean 0, standard deviation (SD) 1 and then averaged across up to 4 cardiac cycles. The averaged residuals were then restandardized to mean 0, SD 1 in cohort- and sex-specific samples. These residuals were then used in genotype-phenotype association testing. In RS, sex-specific regression models were constructed to adjust for RR interval and age and generate residuals. In CHS, the adjustment for age, sex, RR interval and study site was performed in the genotype-phenotype association step (see below).

Genotyping

In FHS, genotyping was performed using the Affymetrix 500K GeneChip array, called using the BRLMM algorithm[18], and a custom-designed gene-centric 50K MIP. In RS, genotyping was performed using the Infinium II HumanHap550K Genotyping BeadChip version 3. In CHS, genotyping was performed using the Illumina 370CNV BeadChip system. Associations of poorly imputed SNPs were validated by re-genotyping FHS samples using Sequenom and RS samples using Taqman. See *Supplementary Methods* for cohort-specific genotyping details including filters.

Imputation

We imputed estimated allele dosage, defined as the expected number of copies of the minor allele (a fractional value between 0 and 2), of all autosomal SNPs using MACH[19] (HapMap CEU release 22, build 36) in FHS and RS and using BIMBAM[20] (release 21a, build 35) in CHS (see *Supplementary Methods* for details).

Genotype-phenotype association method

In FHS, standardized QT residuals were tested for association with imputed allele dosage under an additive genetic model using the linear mixed effects model of the kinship package in R to account for relatedness[21,22]. In the Rotterdam Study, QT residuals were tested for association using MACH2QTL, which uses dosage value (0–2, continuous) as a predictor in a linear regression framework[19]. In CHS, QT interval was linearly regressed on age, sex, clinic, RR-interval and SNP dosage using R[21]. The regression was weighted to reflect case-cohort sampling probabilities.

Meta-analysis

The minor allele from HapMap CEU genotypes was used to define the coded allele in all analyses, regardless of frequency in individual cohorts. For an A/T SNP, the T allele is the 'coded allele' under the following coding: AA=0, AT=1, TT=2. To implement genomic control, the lambda value was used to correct the standard error as follows, SE_corrected=SE*sqrt(lambda). Each effect estimate (beta) was standardized to the standard deviation of the cohort-specific adjusted residuals to put all results on the SD scale. The ratio of the observed to the expected variance of the imputed allele frequency was used as the quality metric for the imputation of a given SNP. To account for the difference in power, and thus the interpretation of resulting p-values for each SNP, we created a variable N_effective which discounted the total sample size by imputation quality as follows: N_effective= N*(observed/expected variance). We used inverse variance-weighted fixed effects meta-analysis of the beta estimates from linear regression as the primary meta-analysis method. Weighting by the square root of the N_effective resulted in very similar -log(p-value) for results with $P < 0.01$ (r=0.9980). The scripts developed for this project are freely available at http://www.broad.mit.edu/~debakker/meta.html. We *a priori* declared results significant at $\alpha = 5\times10^{-8}$, based on estimates adjusting for all common variant tests in the human genome of European ancestry individuals for a target genome-wide $P < 0.05$[23].

Exchange with QTSCD consortium

We submitted to the QTSCD consortium a list of our top SNP associations (one SNP per signal of association) and received QTSCD results for those SNPs. A reciprocal exchange was performed. We performed meta-analysis of the QTGEN meta-analysis results (n=13,685) with the QTSCD results (n=15,854) using inverse variance weights.

QT genotype score

A QT genotype score was calculated using the effect estimates from the QTGEN meta-analysis for the coded allele of each of 14 SNPs. The score, on the standard deviation scale, was calculated as follows for each individual: QT genotype score = $beta_1$*allele_copy_number$_1$ + $beta_2$*allele_copy_number$_2$ + … $beta_{14}$*allele_copy_number$_{14}$.

RESULTS

The mean ages of the individuals in the FHS, RS and CHS cohorts were 40, 69, and 73 years, respectively. Additional clinical characteristics are shown in Table 1. QT interval measures were adjusted for age and RR-interval (inverse heart rate) using cohort- and sex-specific linear regression, and the standardized residuals served as the phenotype for the association analysis. We started with a set of SNPs passing study-specific quality control filters: in FHS, 378,163 SNPs from the Affymetrix 500K chip + 50K gene-centered MIP; in RS, 512,349 SNPs from the Illumina 550K array; and in CHS, 332,946 SNPs from the Illumina 370CNV array. We imputed genotypes with reference to phased chromosomes from HapMap CEU (see Methods)[20]. After quality control filtering, we used genotypes from up to 2,543,686 SNPs to test for association in cohort-specific analyses. Genomic control was used to adjust for test-statistic inflation[24], which was minimal, with λ_{gc} ranging from 1.02 to 1.04. Cohort-specific quantile-quantile plots of p-value distributions are shown in Supplementary Figure 1. Using inverse variance weights, we combined genomic-controlled association results under an additive model from the three cohorts in fixed effects meta-analysis (overall $\lambda_{gc} = 1.012$). Nine loci were independently associated at a genome-wide $P < 5\times10^{-8}$ (Table 2, Figure 1, Figure 2). An additional locus was borderline significant ($P = 8\times10^{-8}$), but was externally validated. Five of the ten associated loci are related to the myocardial repolarization genes previously known to be associated with QT interval duration in the general population or in Mendelian conditions: *NOS1AP*, *KCNQ1*, *KCNH2*, *KCNE1* and *SCN5A*. We observed an excess number of associations compared with the expectation under the null. For a nominal $P < 10^{-5}$ we observed 568 associations compared with 25 expected under the null hypothesis ($P << 10^{-4}$, Supplementary Table 1). This finding suggests that among the many false positives at less stringent statistical thresholds additional truly associated variants may exist.

Table 1 *Clinical characteristics by cohort and by sex.*

	Framingham Heart Study		**Rotterdam Study**		**Cardiovascular Health Study**	
	men n = 3,440	women n = 4,210	men n = 1,854	women n = 2,752	men n = 802	women n = 627
Age (yearS)	40 (10)	40 (11)	68 (8)	69 (9)	73 (6)	73 (5)
Body mass index	27 (4)	25 (5)	26 (3)	27 (4)	26 (3)	26 (5)
Hypertension	23%	14%	34%	41%	54%	60%
Diabetes	2.3%	1.6%	10%	10%	16%	11%
Raw QT (msec)	390 (35)	391 (38)	398 (29)	399 (29)	417 (35)	413 (31)
Heart rate (beats/ minute)	67 (13)	72 (14)	68 (12)	71 (12)	66 (11)	69 (10)
RR interval (msec)	943 (167)	880 (164)	904 (155)	865 (140)	979 (167)	928 (140)
QTc (msec)*	404 (22)	419 (22)	421 (23)	431 (22)	423 (19)	431 (22)
Standard deviation of QT residuals (msec)†	20.6	20.8	18.3	17.9	17.3	17.3

* Bazett's correction for heart rate: QTc = QT/sqrt (RR interval).

† Residuals are from sex-specific linear regression models adjusting for age and RR interval.

We had the opportunity to compare our top results with the QTSCD consortium, which included 15,854 individuals of European ancestry[25]. All associations but one were confirmed at 2-sided $P < 0.05$ (Table 2).

Genes known to be involved in myocardial repolarization

At the *NOS1AP* locus, we observed the strongest association in the genome for rs12143842, 6kb 5' of *NOS1AP*, with 0.21 SD QT increase per minor allele copy (minor allele frequency, MAF = 0.26, $P = 8\times10^{-46}$, Table 2, Figure 2A). All results are shown on the standard deviation scale (1 SD ≈ 17.5 msec). Two additional independent signals were observed at rs12029454 (MAF 0.15) and rs16857031 (MAF 0.14) in intron 2 and intron 1, respectively, all with r^2 to each other <0.05 in HapMap and with $P<0.05$ when entered into a single regression model in FHS and RS (CHS with a smaller sample is underpowered, Supplementary Table 2). We have previously reported

Table 2 *SNPs with evidence for independent association at 10 loci with $P < 5\times10^{-8}$. A SNP at the LIG3 locus met our significance threshold in an interim analysis and was confirmed in the QTSCD consortium study. Chromosomal positions and coded alleles are given relative to forward strand of NCBI build 36. Effect sizes are shown as beta estimates from linear regression models for increasing copy of the coded allele and are on the standard deviation scale (1 SD ≈ 17.5 msec). A beta estimate of 0.08 SD is equivalent to a change in QT interval of 1.4 msec and an effect of 0.48 SD is equivalent to an 8.4 msec change. The effective sample size reflects the power relative to the total sample size of 13,685 with imputed data resulting from variation in imputation quality (see Methods). Selected genes from each locus are shown for reference. Results using the same coded allele from the QTSCD study (reported separately) and meta-analysis of the QTGEN and QTSCD study using inverse variance weighting are shown (n ≤ 29,539). Chr = chromosome. SE = standard error.*

						QTGEN	
SNP	**Chr**	**Function/gene**	**Other genes within 500kb at novel loci**	**Coded allele**	**Allele frequency**	**Effective sample size**	**Beta estimate**
rs12143842	1q	upstream *NOS1AP*		T	0.26	13,241	0.21
rs12029454	1q	intron *NOS1AP*		A	0.15	12,172	0.21
rs16857031	1q	intron *NOS1AP*		G	0.14	13,154	0.19
rs2074238	11p	intron *KCNQ1*		T	0.06	2,888	-0.47
rs37062	16q	intron *CNOT1*	*GINS3, NDRG4, SLC38A7, GOT2*	G	0.24	13,440	-0.12
rs11756438	6q	intron *c6orf204*	*SLC35F1, PLN, ASF1A*	A	0.47	12,707	0.09
rs12576239	11p	intron *KCNQ1*		T	0.13	13,211	0.12
rs846111	1p	3' UTR *RNF207*	*NPHP4, CHDS, ACOT7, PLEKHG5, KLH21*	C	0.28	6,480	0.12
rs4725982	7q	downstream *KCNH2*		T	0.22	13,706	0.09
rs8049607	16p	upstream *LITAF*	*CLEC16A, SNN, ZC3H7A, TNFRSF17*	T	0.49	10,543	0.08
rs1805128	21q	missense *KCNE1*		A	0.010	7,644	0.48
rs12053903	3p	intron *SCN5A*		C	0.34	13,491	-0.08
rs2074518	17q	intron *LIG3*	*RFFL*	T	0.46	13,488	-0.07
rs2968864	7q	downstream *KCNH2*		C	0.25	12,932	-0.08

association in FHS and RS samples of rs10494366 at *NOS1AP* ($P = 5\times10^{-30}$ in the current report) and this association has been widely replicated[5,26-29]. This SNP is not significant in models adjusting for the 3 SNPs identified in the current study ($P > 0.05$) and it shows some degree of correlation to each of the 3 SNPs (to rs12043842 $r^2 = 0.46$-0.47 in FHS and RS and $r^2 = 0.11$ in HapMap CEU; to rs12029454 $r^2 = 0.17$ in RS and HapMap CEU; to rs16857031 $r^2 = 0.17$ in HapMap CEU)[30]. We conclude that there are three independent signals at the locus and that rs10494366 captures the association signal from at least one of these 3 SNPs.

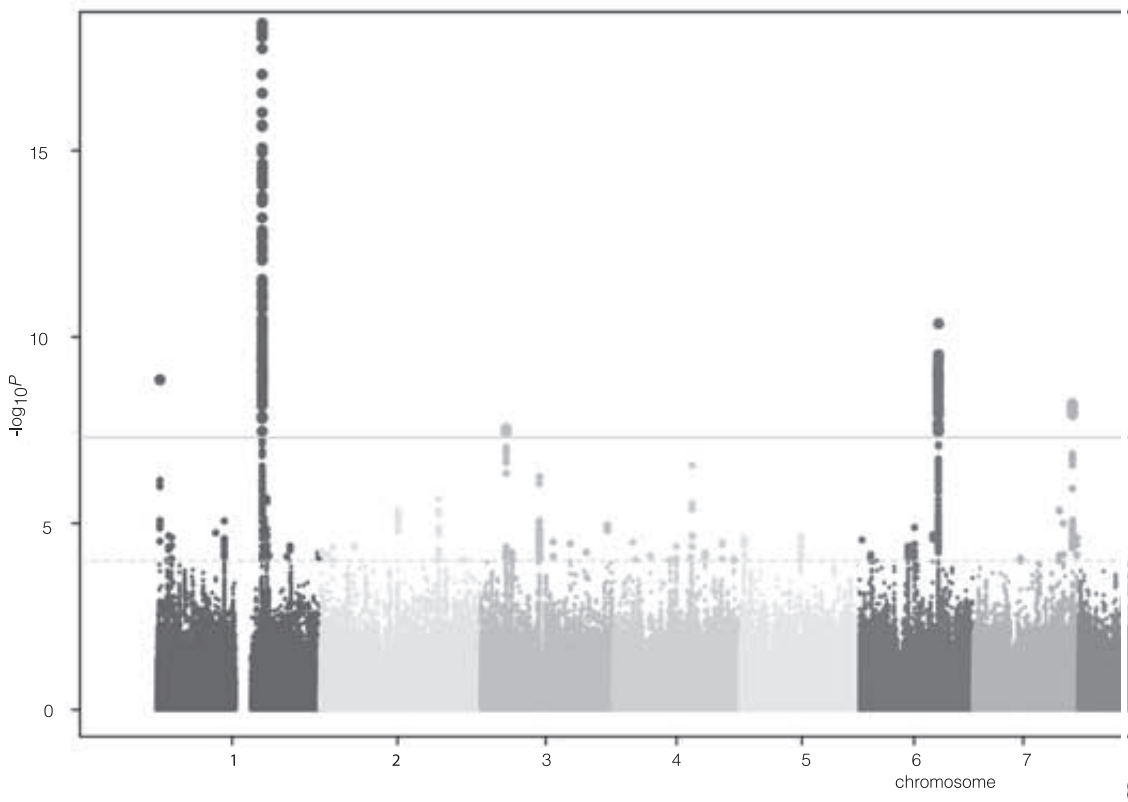
We identified two common variants in intron 1 of *KCNQ1* that were associated with QT interval duration (Table 2, Figure 2B). Rare mutations in *KCNQ1*, a potassium channel involved in myocardial repolarization, have been associated with Long QT Syndrome type 1 and Short QT Syndrome type 2[4]. In this meta-analysis, SNP rs2074238 (MAF = 0.06) was associated with 0.47 SD shorter QT interval for each minor allele ($P = 3\times10^{-16}$) and rs12576239 (MAF = 0.13) was associated with 0.12 SD longer QT interval for each minor allele ($P = 2\times10^{-10}$). The two SNPs were independently associated with QT in models that included both SNPs ($P = 6\times10^{-5}$, $P = 1\times10^{-4}$, respectively in FHS and $P = 3\times10^{-10}$, $P = 0.03$ in RS, Supplementary Table 2). Coupled with the low correlation of the two SNPs (HapMap CEU r^2=0.009, FHS r^2=0.014, RS r^2=0.011) these findings support two independent association signals at the locus. Pfeufer *et al.* previously reported association with QT interval of rs757092 (MAF=0.38) which lies 3kb away from rs12576239 in intron 1 and to which it is partially correlated ($r^2 = 0.31$ HapMap CEU) and 14kb away from rs2074238 to which it is not correlated ($r^2 = 0.005$ HapMap CEU)[9]. We did not find supportive evidence of association of rs757092, which was well imputed, with QT interval in QTGEN ($P = 0.11$).

The common SNP rs4725982, 3' of *KCNH2*, was associated with QT interval (+0.09 SD/minor allele, MAF=0.22, $P = 6\times10^{-9}$, Table 2, Figure 2F). A second SNP at *KCNH2*, rs2968864 was associated with shorter QT interval duration for increasing minor allele count, but did not reach our

		QTSCD			Meta-analysis QTGEN + QTSCD		
SE	***P*-value**	**Beta estimate**	**SE**	***P*-value**	**Beta estimate**	**SE**	***P*-value**
0.02	8×10^{-46}	0.16	0.01	5×10^{-36}	0.18	0.01	2×10^{-78}
0.02	6×10^{-28}	0.15	0.02	3×10^{-20}	0.17	0.01	3×10^{-45}
0.02	3×10^{-23}	0.12	0.02	1×10^{-14}	0.15	0.01	1×10^{-34}
0.06	3×10^{-16}	-0.33	0.14	0.02	-0.45	0.05	3×10^{-17}
0.02	3×10^{-15}	-0.09	0.01	5×10^{-12}	-0.10	0.01	3×10^{-25}
0.01	4×10^{-11}	0.08	0.01	2×10^{-12}	0.08	0.01	5×10^{-22}
0.02	2×10^{-10}	0.08	0.02	3×10^{-7}	0.10	0.01	1×10^{-15}
0.02	1×10^{-9}	0.08	0.01	4×10^{-9}	0.10	0.01	1×10^{-16}
0.02	6×10^{-9}	0.08	0.01	1×10^{-8}	0.09	0.01	5×10^{-16}
0.01	2×10^{-8}	0.07	0.01	4×10^{-8}	0.07	0.01	5×10^{-15}
0.09	2×10^{-8}	-0.06	0.04	0.16	0.05	0.04	0.22
0.01	3×10^{-8}	-0.06	0.01	6×10^{-8}	-0.07	0.01	1×10^{-14}
0.01	8×10^{-8}	-0.05	0.01	7×10^{-6}	-0.06	0.01	6×10^{-12}
0.02	1×10^{-7}	-0.08	0.01	1×10^{-9}	-0.08	0.01	8×10^{-16}

pre-specified genome-wide significance threshold for unselected genetic variants in the QTGEN samples (-0.08 SD/minor allele, MAF=0.25, $P = 1\times10^{-7}$, Table 2, Figure 2F). Rare mutations in *KCNH2*, a potassium channel involved in myocardial repolarization and drug-induced arrhythmias, are known to underlie congenital Long QT Syndrome type 2 and Short QT Syndrome type 1[4]. The two SNPs were significant or nearly so when entered into a single regression model ($P = 4\times10^{-3}$, $P = 3\times10^{-4}$, respectively in FHS and $P = 4\times10^{-3}$, $P = 0.16$, respectively in RS, Supplementary Table 2). Coupled with the low correlation between the SNPs in HapMap CEU (r^2=0.09), the two SNPs thus appear to represent independent signals of association. The missense variant rs1805123 (K897T) has been associated with QT interval in most studies, including our own[6,8,9,31], and is perfectly correlated with rs2968864 (r^2=1.0 in FHS, data not shown), which is thus not a novel finding[8]. An intronic SNP has been previously reported by Pfeufer *et al.* to be associated with QT interval (rs3815459), is poorly correlated with the currently reported rs4725982 or rs2968864/rs1805123 variants (r^2 = 0.08, r^2 = 0.08, respectively in KORA, personal communication, Arne

Figure 1 *QT interval association results for 2,543,686 imputed SNPs in 13,685 individuals from 3 cohorts. Results are shown on the -log10(P) scale and are truncated at -log10(P) = 18 for display purposes. The solid bar corresponds to the genome-wide significance threshold of 5×10^{-8}.*



Pfeufer) and could not be imputed because it is not represented in HapMap. Another variant previously reported by us (rs3807375)[8] has limited correlation with rs2968864 (r^2=0.21 HapMap CEU) and rs4725982 (r^2 = 0.39 HapMap CEU) and is not significant in models containing rs2968864 and rs4725982, suggesting that it does not represent an independent signal of association.

SNP rs1805128 was associated with QT interval duration (+0.48 SD/minor allele, MAF = 0.01, $P = 2\times10^{-8}$, Table 2, Figure 2H). This SNP encodes a change from aspartate to asparagine at amino acid 85 (D85N) in *KCNE1*, a potassium channel involved in myocardial repolarization in which rare mutations result in Long QT Syndrome type 5[4]. D85N is poorly covered by the fixed genotyping arrays used here, but was included on the supplemental Affymetrix 50K array used in FHS, for which results are presented. Association of rs1805128 has been reported by Gouas *et al.* with extremes of QT interval duration in 398 individuals from the DESIR cohort ($P = 0.02$)[7], and by us in 4,487 CHS participants ($P = 0.003$)[33], and more recently in 5,043 individuals from the Health2000 study ($P = 4\times10^{-11}$)[33]. Importantly, the D85N variant has also been related to drug-induced arrhythmia[34,35] and Long QT Syndrome[36].

In the published literature, no common variants in *SCN5A* have been convincingly associated with QT interval duration in European-derived individuals. We observed an association of rs12053903 in intron 27 of *SCN5A* with QT interval (-0.08 SD/minor allele, MAF = 0.34,

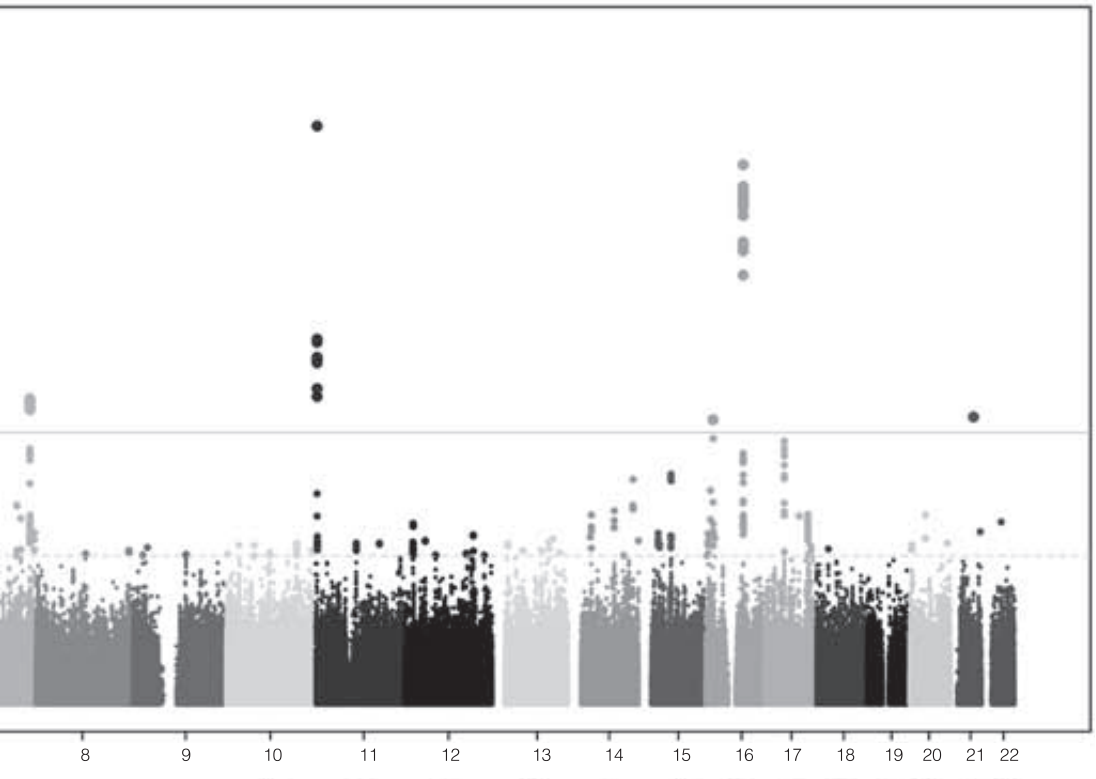
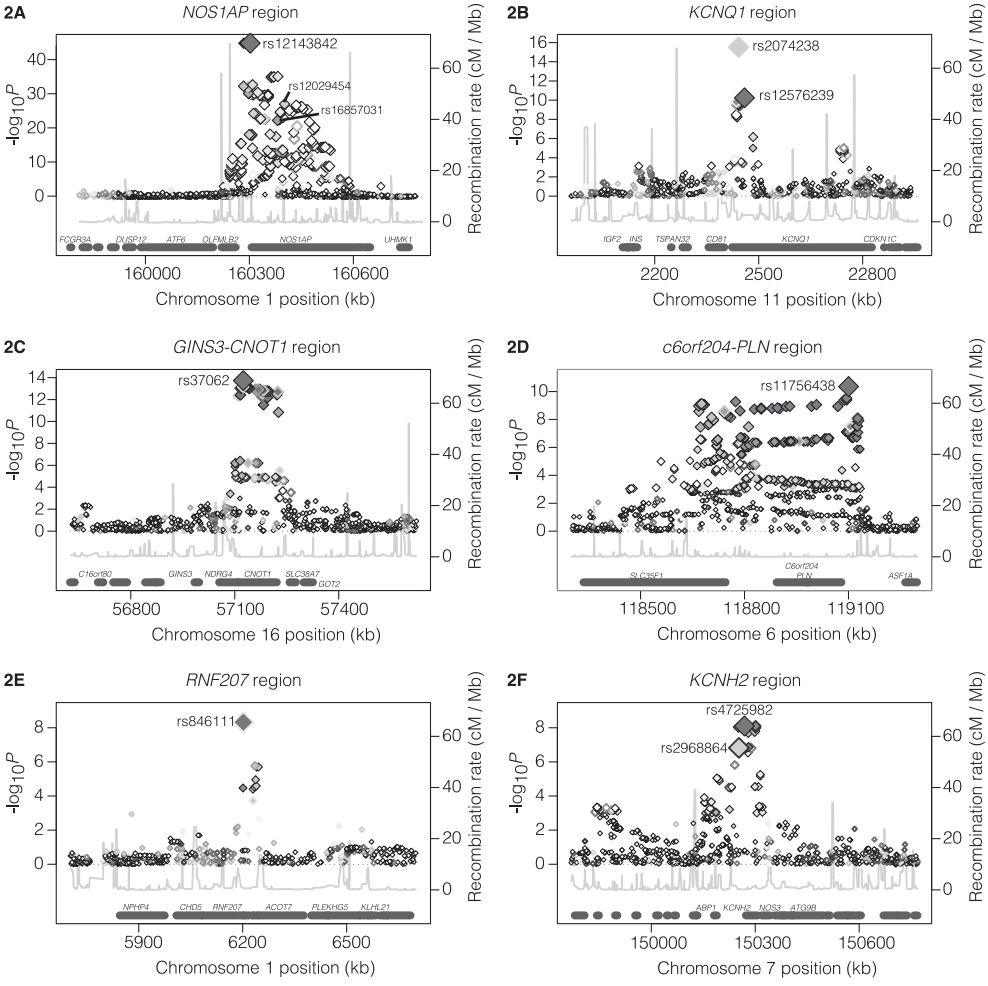
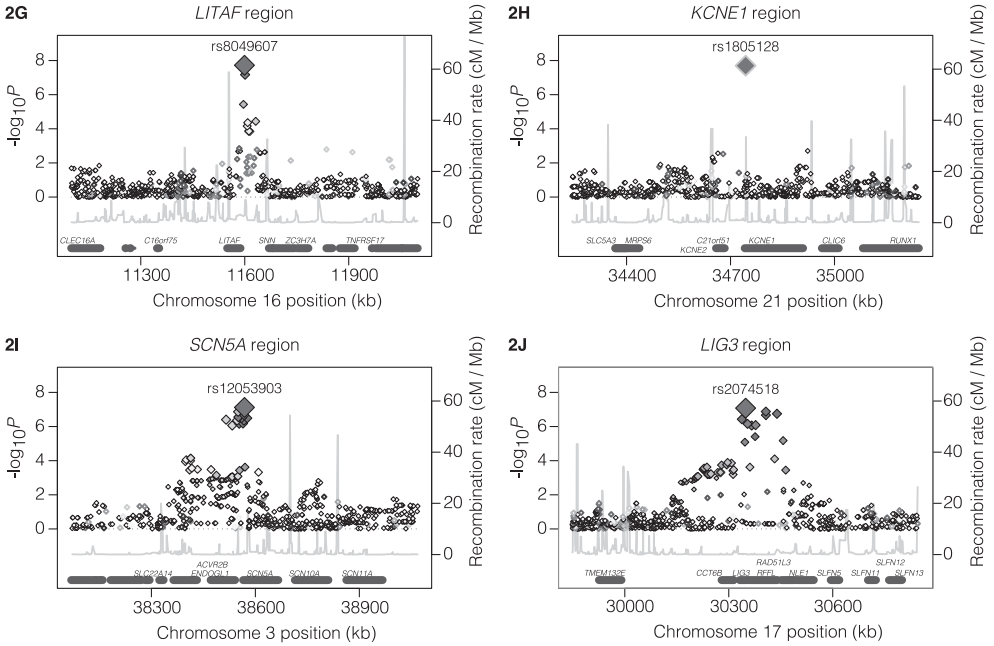


Figure 2 *Regional association plots for one megabase surrounding each associated locus. Statistical significance of associated SNPs at each locus are shown on the -log(P) scale as a function of chromosomal position (NCBI Build 36). The primary associated SNP at each locus is shown in black. The correlation of the primary SNP to other SNPs at the locus is shown on a scale from minimal (white) to maximal (black). The quality of imputation as assessed by the observed/expected variance on allele dosage is represented by the darkness of the diamond outline ranging from maximal (black) to minimal (light gray). Estimated recombination rates from HapMap and RefSeq annotations are shown. The loci shown include: (A) 1q23.3 including NOS1AP with three independent associations, rs12143842, rs12029454 and rs16857031, (B) 11p15.5 including KCNQ1 with two independent associations, rs2074238 and rs12576329, (C) 16q21 including GINS3, NDRG4 and CNOT1, (D) 6q22.31 including c6orf204 and PLN, (E) 1p36.31 including RNF207, (F) 7q36.1 including KCNH2 with two independent associations, rs4725982 and rs2968864, (G) 16p13.3 including LITAF, (H) 21q22.12 including KCNE1, (I) 3p22.2 including SCN5A and (J) 17q12 including LIG3 and RFFL.*





$P = 3\times10^{-8}$, Table 2, Figure 2I). Rare mutations in *SCN5A*, the cardiac sodium channel, result in Long QT Syndrome type 3[37]. A common missense variant in *SCN5A*, S1102Y, is associated with QT prolongation, ventricular arrhythmias and sudden cardiac death in African Americans [37,38], but is nearly monomorphic (MAF <0.01) in individuals of European ancestry.

The finding of 9 associated common variants in 5 genes known to influence myocardial repolarization and cardiac arrhythmias at the top of our list of results, 8 achieving a stringent genome-wide significance threshold in our meta-analysis of three independent cohorts, gave us confidence in the validity of the five novel loci that exceeded this threshold.

QT interval associations with novel loci

The first novel locus on chromosome 16q21 near *NDRG4, SETD6, CNOT1, SLC38A7* and *GINS3* included SNP rs37062 (MAF = 0.24) which falls in intron 40 of *CNOT1*, a regulator of RNA transcription, and was associated with 0.12 SD reduced QT per minor allele ($P = 3\times10^{-15}$, Table 2, Figure 2C). None of the nearby genes is known to modulate myocardial repolarization in humans, although recent experiments in zebrafish suggest potential candidates at the locus. Milan *et al* tested zebrafish mutants generated by Amsterdam *et al*. in an insertional mutagenesis screen[39] for altered response to challenge with dofetilide, a QT interval prolonging medication used in humans that prolongs cardiomyocyte action potential duration in humans and zebrafish. They found that a mutant with an insertion in intron1 of GINS complex subunit 3 (*GINS3*) was resistant to the QT-prolonging effects of exposure to dofetilide[40]. The GINS complex is involved in the

establishment of DNA replication forks. In humans, *GINS3* falls near the 16q21 interval associated with QT interval in our meta-analysis (127kb from rs37062). In addition, a recent report on *NDRG4* (N-myc downstream-regulated gene family member 4, 56kb from rs37062) in zebrafish observed expression restricted to the central nervous system and the heart starting at 24 hours post-fertilization[41]. Morpholino knockdown of *NDRG4* was associated with hypoplastic hearts with pericardial edema, dilated atria, looping defects and slower heart rates compared to controls. Zebrafish respond to exposure to QT-prolonging drugs with heart rate slowing[42], although this may be a non-specific finding in the *NDRG4* morphants. While we cannot exclude a source of the association in the many other genes at this locus, *GINS3* and *NDRG4* are promising candidates for further work.

The second novel locus is on chromosome 6q22.31. SNP rs11756438 (MAF = 0.47) lies in an intron of a predicted gene of unknown function *c6orf204* and near *SLC35F1* and *PLN* and was associated with 0.09 SD higher QT interval per minor allele ($P = 4\times10^{-11}$, Table 2, Figure 2D). Interestingly, *PLN* (122kb away from this SNP) encodes phospholamban, an inhibitor of cardiac sarcoplasmic reticulum Ca^{++}-ATPase (SERCA2a). Phospholamban knockout mice show enhanced myocardial contractility in response to beta adrenergic agonists[43]. Cardiomyocyte dysregulation of Ca^{++} handling due to increased phospholamban activity has been linked to dilated cardiomyopathy and heart failure in mouse models[44] and in a human family with cardiomyopathy and ventricular tachycardia[45]. While more work will be required to localize the source of the signal of association reported here, it is interesting to note that *NOS1AP* activates neuronal nitric oxide synthase 1[5], a regulator of calcium cycling in the sarcoplasmic reticulum, and that rare variants in *CACNA1C*, a subunit of the L-type voltage-dependent calcium channel, cause congenital Long QT Syndrome[46]. These observations suggest a unifying hypothesis that genetic variation influencing calcium cycling in cardiac myocytes influences repolarization and, when altered, contributes to arrhythmogenesis.

The third novel locus was on chromosome 1p36.31 near several genes including *CHD5, RPL22, RNF207, ICMT, HES3, GPR153*, and *ACOT7*. The top SNP at the locus, rs846111 (MAF = 0.28), lies in the 3' untranslated region of *RNF207* and was associated with 0.12 SD higher QT interval for each copy of the minor allele ($P = 1\times10^{-9}$, Table 2, Figure 2e). *RNF207*, which encodes ring finger protein 207, is of unknown function, in a family of molecules involved generally in protein-protein interaction and ubiquitination.

The fourth novel locus was on chromosome 16p13.3 upstream of *LITAF*, encoding lipopolysaccharide induced tumor necrosis factor, which has no known relationship to myocardial repolarization but missense mutations in this gene have been related to Charcot-Marie-Tooth, a hereditary motor and sensory neuropathy[47]. The top SNP at the locus, rs8049607 (MAF = 0.49), was associated with 0.08 SD higher QT interval for each minor allele copy ($P = 2\times10^{-8}$, Table 2, Figure 2g).

The fifth novel locus was on chromosome 17q12 near *LIG3* and *RFFL*. The top SNP at the locus, rs2074518 (MAF = 0.46) in intron 11 of *LIG3*, was associated with 0.07 lower QT interval for each minor allele copy ($P = 8\times10^{-8}$, Table 2, Figure 2j). While the result did not achieve genome-wide significance in our data alone, replication was observed in the QTSCD consortium ($P = 7\times10^{-6}$,

joint $P = 6\times10^{-12}$). *LIG3* encodes DNA ligase III, is involved in DNA base-excision repair and is not an obvious candidate to modulate myocardial repolarization. The nearby gene *RFFL* encodes the rififylin protein and is involved in the endocytic recycling compartment.

Estimates of the coverage by the imputed SNPs of SNPs found in HapMap CEU at each of the novel loci are shown in Supplementary Table 3.

Technical validation of poorly imputed SNPs and secondary analysis

Because the imputation quality of individual SNPs varied due to variation in coverage of SNPs by fixed genotyping arrays, we directly genotyped 3 sentinel SNPs in the entire RS sample and in a subset of the FHS sample (total n≈7,000), as well as three additional SNPs only in the FHS subset. For example, rs2074238 in intron 1 of *KCNQ1* had an effective sample size (see Methods) in 6,975 individuals examined of only 2,359 when accounting for the relatively low imputation quality in Framingham (observed/expected variance = 0.10) and Rotterdam (0.45). The SNP was filtered out in the CHS analysis by the imputation QC thresholds. We compared the significance of imputed association results in the directly genotyped subsample to that of the direct genotyping results. Direct genotyping confirmed the association of rs2074238 with QT interval with a substantially stronger significance, consistent with the rise in effective sample size from 2,359 to 6,975 individuals ($P = 1\times10^{-13}$ imputed vs $P = 6\times10^{-23}$ directly genotyped, Table 3). The appropriate filters to be applied to imputed genotypes of varying quality are a matter of debate, but it was certainly encouraging that no association based on imputed results failed to be supported by directly genotyped SNP results (Table 3). Including poorly imputed variants in analyses may be valuable even though they have substantially reduced power to detect truly associated variants.

Table 3 *QT interval association results of directly genotyped SNPs compared to imputed SNPs. Shown are meta-analysis of genotype-phenotype association results using imputed and directly genotyped SNPs in 1) a subset of the Framingham sample and the entire Rotterdam Study (n ≤ 6,975) or 2) the Framingham subset only (n ≤ 2,566). Three SNPs were genotyped only in the Framingham Heart Study (and did not specifically have low imputation quality). Effects are shown on the standard deviation (SD) scale. For SNPs that were less well imputed (small effective sample size) the increase in significance tracks with the fall in standard error and rise in effective sample size. N_effective is the sample size (N)* (observed/expected variance) [see Methods].*

	Imputed SNP genotype association				Directly genotyped SNP association				Sample
SNP	**Beta (SD)**	**SE**	***P***	**N_effective**	**Beta (SD)**	**SE**	***P***	**N**	
rs2074238	-0.47	0.06	$1x10^{-13}$	2,359	-0.31	0.03	$6x10^{-23}$	6,975	FHS+RS
rs846111	0.09	0.03	$4x10^{-4}$	3,709	0.10	0.02	$2x10^{-7}$	6,825	FHS+RS
rs8049607	0.09	0.02	$1x10^{-6}$	5,964	0.08	0.02	$4x10^{-6}$	6,921	FHS+RS
rs12576239	0.13	0.04	$2x10^{-3}$	2,560	0.13	0.04	$1x10^{-3}$	2,566	FHS only
rs4725982	0.10	0.03	$3x10^{-3}$	2,556	0.11	0.03	$1x10^{-3}$	2,556	FHS only
rs12053903	-0.04	0.03	0.17	2,385	-0.04	0.03	0.24	2,465	FHS only

Because of the strong effect of sex on QT interval variation, which explains approximately 5% of its variability, we tested for but observed no significant interaction of sex with the SNP-QT associations.

DISCUSSION

Confirmation of association in an independent sample

Examination of results for 14 SNPs at 10 loci identified by the QTGEN consortium in the QTSCD consortium data, demonstrated strong confirmation of results for 12 of the 14 associations (Table 2)[25]. One SNP that only weakly replicated is the less common *KCNQ1* SNP (rs2074238), which is poorly imputed from Affymetrix arrays used in QTSCD and is likely due to low power given the strong association in our data ($P = 6\times10^{-23}$ upon direct genotyping in 6,975 individuals). A SNP that did not replicate is the low frequency (MAF = 0.01) missense SNP rs1805128 (D85N) in *KCNE1*, which is poorly imputed from Affymetrix genotypes and is well replicated in external studies with direct genotyping[7,32,33]. Both SNPs thus had limited power to be replicated due to genotype imprecision from poor imputation quality. Additionally, genetic variants reaching genome-wide significance in the QTSCD consortium were strongly confirmed in the QTGEN consortium, including rs10919071 at a locus containing *ATP1B1* (QTGEN $P = 4\times10^{-5}$) and rs17779747 at a locus containing *KCNJ2*, a Long QT Syndrome gene (QTGEN $P = 4\times10^{-5}$).

Population impact of 14 variants associated in QTGEN

In summary, the QTGEN meta-analysis of three genome-wide association studies detected nine common variants at five known candidate genes and an additional five common variants at novel loci not previously recognized to modulate myocardial repolarization. In total, these variants explain a substantial proportion of variation in QT interval (Framingham 5.4%, Rotterdam 6.5%, CHS 2.3%, Supplementary Table 2). These variants in aggregate explain more of QT interval variation than any other covariate (excluding heart rate) including female sex, a known risk factor for QT prolongation and drug-induced arrhythmia.

To assess the potential clinical impact of the genetic variants examined here, we constructed a QT genotype score using the allele copy number and the effect estimates for the 14 SNPs from our meta-analysis and tested the score in the FHS and RS samples (the larger samples). The top quintile of QT genotype score was associated with 9.7 msec and 12.4 msec higher Bazett-corrected QTc (the heart rate correction used in clinical settings) compared to the bottom quintile in FHS and RS samples ($P = 5\times10^{-28}$, $P = 1\times10^{-31}$, respectively). A prolonged QTc ≥450 msec in men and ≥470 msec in women has previously been shown to be associated with 2.5-fold increased hazard of sudden cardiac death in the Rotterdam Study[1]. The top quintile of QT genotype score was associated with odds ratios for prolonged QTc of 2.6 and 3.1 in FHS and RS ($P = 3\times10^{-5}$, $P = 6\times10^{-7}$, respectively, Supplementary Table 4). The finding that the top 20% of genotype score in the population has a QTc increase compared to the bottom 20% in excess of the QT-prolonging effect of some drugs causing arrhythmias (as little as 8 msec), and that one of the SNPs (D85N *KCNE1*) is associated with the congenital Long QT Syndrome[36] and drug-induced

arrhythmias[34,35], further supports the clinical relevance of the genetic variants identified in the current report. Tests of the hypothesis that these variants, individually or in aggregate, contribute to risk of sudden cardiac death or drug-induced arrhythmias will require additional work in well-powered samples.

Although genetic effects have been thought to be weaker at older ages, striking associations were identified among these populations of middle aged and older adults. This may be expected for common variants of modest effects that elude negative selection and modulate traits in which environmental factors play only a modest role.

Additional fine mapping with direct genotyping to refine the signal of association and to identify the specific genes involved will be required. As illustrated by the number of common variants in Long and Short QT Syndrome genes, the spectrum of allele frequencies and effect sizes for the variants at many genes ranges from rare variants of strong effect underlying Mendelian forms of disease to less common variants with intermediate effects to highly polymorphic variants with comparatively modest effects. Certainly, this study will have missed variants that have even more modest effects, that are poorly captured by fixed genotyping arrays including those in the minor allele frequency range from 0.5-5%, and that due to random sampling variation failed to rise to the top of our results but might in an equally sized independent sample. Resequencing of each gene will be needed to fully characterize the allelic architecture of QT variation in the general population as well as its relevance to the approximately 25% of LQTS families without recognized mutations in known genes and the great majority of those who die of sudden cardiac death in the general population without recognized genetic risks.

REFERENCES

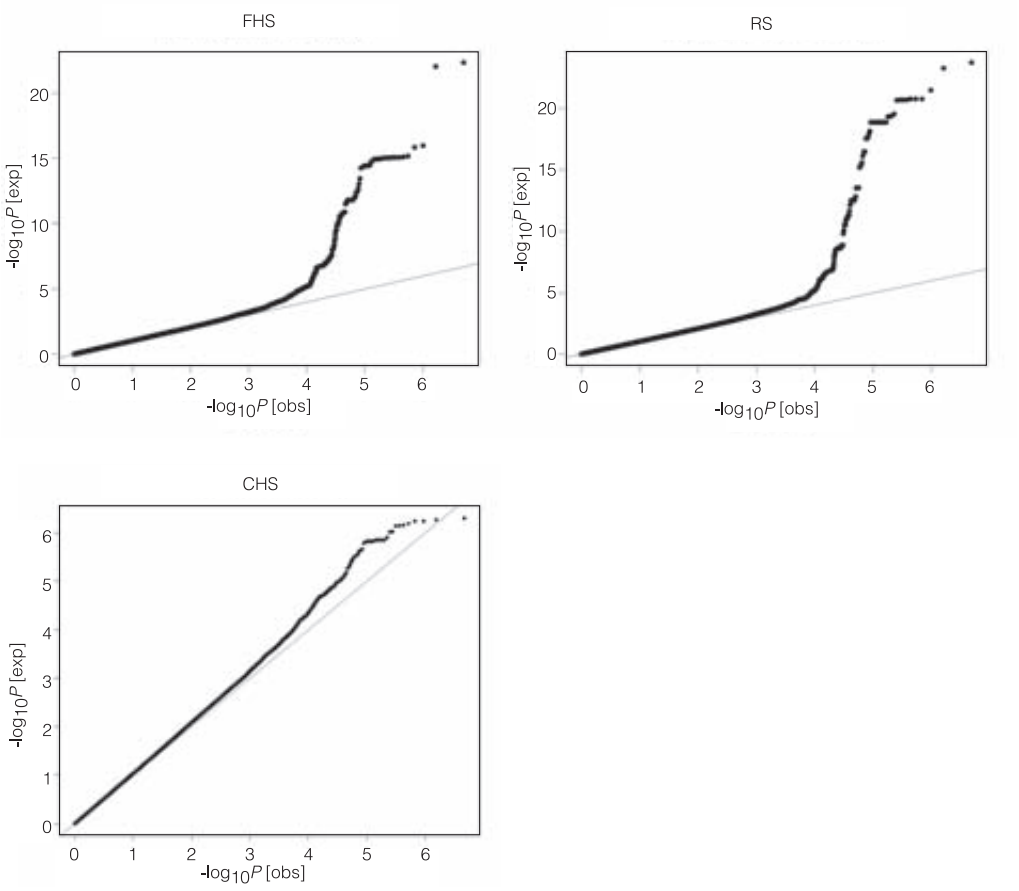
1. Straus, S.M., Kors, J.A., De Bruin, M.L., van der Hooft, C.S., Hofman, A., Heeringa, J., Deckers, J.W., Kingma, J.H., Sturkenboom, M.C., Stricker, B.H. *et al.* (2006) Prolonged QTC interval and risk of sudden cardiac death in a population of older adults. *J Am Coll Cardiol*, 47, 362-367.
2. Newton-Cheh, C., Larson, M.G., Corey, D.C., Benjamin, E.J., Herbert, A.G., Levy, D., D'Agostino, R.B. and O'Donnell, C.J. (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Heart Rhythm*, 2, 277-284.
3. Newton-Cheh, C. and Shah, R. (2007) Genetic determinants of QT interval variation and sudden cardiac death. *Curr Opin Genet Dev*, 17, 213-221.
4. Splawski, I., Shen, J., Timothy, K.W., Lehmann, M.H., Priori, S., Robinson, J.L., Moss, A.J., Schwartz, P.J., Towbin, J.A., Vincent, G.M. *et al.* (2000) Spectrum of mutations in long-QT syndrome genes. *KVLQT1, HERG, SCN5A, KCNE1*, and *KCNE2*. *Circulation*, 102, 1178-1185.
5. Arking, D.E., Pfeufer, A., Post, W., Kao, W.H., Newton-Cheh, C., Ikeda, M., West, K., Kashuk, C., Akyol, M., Perz, S. *et al.* (2006) A common genetic variant in the *NOS1* regulator *NOS1AP* modulates cardiac repolarization. *Nat Gene*t, 38, 644-651.
6. Bezzina, C.R., Verkerk, A.O., Busjahn, A., Jeron, A., Erdmann, J., Koopmann, T.T., Bhuiyan, Z.A., Wilders, R., Mannens, M.M., Tan, H.L. *et al.* (2003) A common polymorphism in *KCNH2* (HERG) hastens cardiac repolarization. C*ardiovasc Res*, 59, 27-36.
7. Gouas, L., Nicaud, V., Berthet, M., Forhan, A., Tiret, L., Balkau, B. and Guicheney, P. (2005) Association of *KCNQ1*, *KCNE1, KCNH2* and *SCN5A* polymorphisms with QTC interval length in a healthy population. *Eur J Hum Genet*, 13, 1213-1222.
8. Newton-Cheh, C., Guo, C.Y., Larson, M.G., Musone, S.L., Surti, A., Camargo, A.L., Drake, J.A., Benjamin, E.J., Levy, D., D'Agostino, R.B., Sr. *et al.* (2007) Common genetic variation in *KCNH2* is associated with QT interval duration: the Framingham Heart Study. *Circulation*, 116, 1128-1136.

9. Pfeufer, A., Jalilzadeh, S., Perz, S., Mueller, J.C., Hinterseer, M., Illig, T., Akyol, M., Huth, C., Schopfer-Wendels, A., Kuch, B. *et al.* (2005) Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. *Circ Res*, 96, 693-701.
10. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., Boerwinkle, E. *et al.* (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts, 2, 73-80.
11. Dawber, T.R., Meadors, G.F. and Moore, F.E., Jr. (1951) Epidemiological approaches to heart disease: the Framingham Study. *American journal of public health and the nation's health*, 41, 279-281.
12. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Sr., Fox, C.S., Larson, M.G., Murabito, J.M. *et al.* (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*, 165, 1328-1335.
13. Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J. and Castelli, W.P. (1979) An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*, 110, 281-290.
14. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22, 819-829.
15. Hofman, A., Grobbee, D.E., de Jong, P.T. and van den Ouweland, F.A. (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*, 7, 403-422.
16. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A. *et al.* (1991) The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*, 1, 263-276.
17. van Bemmel, J.H., Kors, J.A. and van Herpen, G. (1990) Methodology of the modular ECG analysis system MEANS. *Methods Inf Med*, 29, 346-353.
18. Rabbee, N. and Speed, T.P. (2006) A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics, 22, 7-12.
19. Li, Y. and Abecasis, G.R. (2006) Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet*, S79, 2290.
20. Servin, B. and Stephens, M. (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS.Genet*, 3, e114-.
21. R Developmental Core Team (2007) A language and environment for statistical computing. R Foundation for Statistical Computing.
22. Therneau, T. (2008) kinship: mixed effects Cox models, sparse matrices, and modeling data from large pedigrees. R package version 1.1.0-19 ed.
23. Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32, 381-385.
24. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, 55, 997-1004.
25. Pfeufer, A., Sanna, S., Arking, D.E., Muller, M., Gateva, V., Fuchsberger, C., Ehret, G.B., Orru, M., Pattaro, C., Kottgen, A. *et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet*, 41, 407-414.
26. Aarnoudse, A.J., Newton-Cheh, C., de Bakker, P.I., Straus, S.M., Kors, J.A., Hofman, A., Uitterlinden, A.G., Witteman, J.C. and Stricker, B.H. (2007) Common *NOS1AP* variants are associated with a prolonged QTC interval in the Rotterdam Study. *Circulation*, 116, 10-16.
27. Lehtinen, A.B., Newton-Cheh, C., Ziegler, J.T., Langefeld, C.D., Freedman, B.I., Daniel, K.R., Herrington, D.M. and Bowden, D.W. (2008) Association of *NOS1AP* genetic variants with QT interval duration in families from the Diabetes Heart Study. *Diabetes*, 57, 1108-1114.
28. Post, W., Shen, H., Damcott, C., Arking, D.E., Kao, W.H., Sack, P.A., Ryan, K.A., Chakravarti, A., Mitchell, B.D. and Shuldiner, A.R. (2007) Associations between genetic variants in the *NOS1AP* (CAPON) gene and cardiac repolarization in the old order Amish. *Hum Hered*, 64, 214-219.
29. Tobin, M.D., Kahonen, M., Braund, P., Nieminen, T., Hajat, C., Tomaszewski, M., Viik, J., Lehtinen, R., Ng, G.A., Macfarlane, P.W. *et al.* (2008) Gender and effects of a common genetic variant in the *NOS1* regulator *NOS1AP* on cardiac repolarization in 3761 individuals from two independent populations. Int J Epidemiol, 37, 1132-1141.
30. Eijgelsheim, M., Aarnoudse, A.L., Rivadeneira, F., Kors, J.A., Witteman, J.C., Hofman, A., van Duijn, C.M., Uitterlinden, A.G. and Stricker, B.H. (2009) Identification of a common variant at the *NOS1AP* locus strongly associated to QT-interval duration. *Hum Mol Genet*, 18, 347-357.

31. Pietila, E., Fodstad, H., Niskasaari, E., Laitinen, P.P., Swan, H., Savolainen, M., Kesaniemi, Y.A., Kontula, K. and Huikuri, H.V. (2002) Association between HERG K897T polymorphism and QT interval in middle-aged Finnish women. *J Am Coll Cardiol*, 40, 511-514.
32. Sotoodehnia, N., Vatta, M., Lemaitre, R., Rautaharju, P., Durda, P., Towbin, J.A., Friedlander, Y., Tracy, R., Manolio, T.A., Burke, G.L. *et al.* (2005) *KCNE1* Gene D85N Variant, QT Interval, and Risk Of Mortality. Circulation, 111, e231.
33. Marjamaa, A., Newton-Cheh, C., Porthan, K., Reunanen, A., Lahermo, P., Vaananen, H., Jula, A., Karanko, H., Swan, H., Toivonen, L. *et al.* (2009) Common candidate gene variants are associated with QT interval duration in the general population. *Journal of internal medicine*, 265, 448-458.
34. Paulussen, A.D., Gilissen, R.A., Armstrong, M., Doevendans, P.A., Verhasselt, P., Smeets, H.J., Schulze-Bahr, E., Haverkamp, W., Breithardt, G., Cohen, N. *et al.* (2004) Genetic variations of *KCNQ1, KCNH2, SCN5A, KCNE1*, and *KCNE2* in drug-induced long QT syndrome patients. *J Mol Med*, 82, 182-188.
35. Wei, J., Yang, I.C.H., Tapper, A.R., Murray, K.T., Viswanathan, P., Rudy, Y., Bennett, P.B., Norris, K., Balser, J.R., Roden, D.M. *et al.* (1999) *KCNE1* polymorphism confers risk of drug-induced long QT syndrome by altering kinetic properties of I-Ks potassium channels. *Circulation*, 100, 495-495.
36. Salisbury, B.A., Judson, R.S., Pungliya, M., Carr, J., Qi, M., Zareba, W., Robinson, J.L., Moss, A.J., Will, M.L., Tester, D.J. *et al.* (2006) The single nucleotide polymorphism D85N-*KCNE1* is associated with both congenital and drug-induced long QT. *Heart Rhythm.*, 3, S98-S98.
37. Splawski, I., Timothy, K.W., Tateyama, M., Clancy, C.E., Malhotra, A., Beggs, A.H., Cappuccio, F.P., Sagnella, G.A., Kass, R.S. and Keating, M.T. (2002) Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. *Science*, 297, 1333-1336.
38. Burke, A., Creighton, W., Mont, E., Li, L., Hogan, S., Kutys, R., Fowler, D. and Virmani, R. (2005) Role of *SCN5A* Y1102 polymorphism in sudden cardiac death in blacks. *Circulation*, 112, 798-802.
39. Amsterdam, A., Nissen, R.M., Sun, Z., Swindell, E.C., Farrington, S. and Hopkins, N. (2004) Identification of 315 genes essential for early zebrafish development. *Proc Natl Acad Sci USA*, 101, 12792-12797.
40. Milan, D.J., Kim, A.M., Winterfield, J.R., Jones, I.L., Pfeufer, A., Sanna, S., Arking, D.E., Amsterdam, A.H., Sabeh, K.M., Mably, J.D. *et al.* (2009) Drug-sensitized zebrafish screen identifies multiple genes, including *GINS3*, as regulators of myocardial repolarization. *Circulation*, 120, 553-559.
41. Qu, X., Jia, H., Garrity, D.M., Tompkins, K., Batts, L., Appel, B., Zhong, T.P. and Baldwin, H.S. (2008) ndrg4 is required for normal myocyte proliferation during early cardiac development in zebrafish. *Dev Biol.*, 317(2), 486-496
42. Milan, D.J., Peterson, T.A., Ruskin, J.N., Peterson, R.T. and MacRae, C.A. (2003) Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation*, 107, 1355-1358.
43. Luo, W., Grupp, I.L., Harrer, J., Ponniah, S., Grupp, G., Duffy, J.J., Doetschman, T. and Kranias, E.G. (1994) Targeted ablation of the phospholamban gene is associated with markedly enhanced myocardial contractility and loss of beta-agonist stimulation. *Circ Res*, 75, 401-409.
44. Schmitt, J.P., Kamisago, M., Asahi, M., Li, G.H., Ahmad, F., Mende, U., Kranias, E.G., MacLennan, D.H., Seidman, J.G. and Seidman, C.E. (2003) Dilated cardiomyopathy and heart failure caused by a mutation in phospholamban. *Science*, 299, 1410-1413.
45. Haghighi, K., Kolokathis, F., Gramolini, A.O., Waggoner, J.R., Pater, L., Lynch, R.A., Fan, G.C., Tsiapras, D., Parekh, R.R., Dorn, G.W. *et al.* (2006) A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. *Proc.Natl.Acad.Sci.USA*, 103, 1388-1393.
46. Splawski, I., Timothy, K.W., Decher, N., Kumar, P., Sachse, F.B., Beggs, A.H., Sanguinetti, M.C. and Keating, M.T. (2005) Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc.Natl.Acad.Sci.USA*, 102, 8089-8096.
47. Street, V.A., Bennett, C.L., Goldy, J.D., Shirk, A.J., Kleopa, K.A., Tempel, B.L., Lipe, H.P., Scherer, S.S., Bird, T.D. and Chance, P.F. (2003) Mutation of a putative protein degradation gene LITAF/SIMPLE in Charcot-Marie-Tooth disease 1C. *Neurology*, 60, 22-26.

3.1 Supplement

Supplementary Figure 1 *Shown are quantile-quantile plots of -log(p-value) of QT interval association tests in Framingham Heart Study (FHS), Rotterdam Study (RS), and Cardiovascular Health Study (CHS).*



Supplementary Table 1 *P-value distributions from meta-analysis of 2,543,686 imputed SNPs in 13,685 individuals, binned by -log$_{10}$(P). The ratio of observed to expected SNPs exceeding different thresholds is shown.*

-log10(*P*)	<2	<3	<4	<5	<6	<7	<8
observed	27,702	4,260	1,197	568	459	346	304
expected	25,437	2,544	254	25	2.5	0.25	0.03
ratio	1.1	1.7	4.7	22	180	1,360	11,951

Supplementary Table 2 *Cohort-specific association results in multivariable regression models including 14 SNPs at 10 loci.*

	Framingham Heart Study			Rotterdam Study			Cardiovascular Health Study		
SNP	**Beta (SD)**	**SE**	***P***	**Beta (SD)**	**SE**	***P***	**Beta (SD)**	**SE**	***P***
rs12143842	0.15	0.02	$8x10^{-12}$	0.19	0.03	$9x10^{-14}$	0.17	0.07	$9x10^{-3}$
rs12029454	0.11	0.03	$4x10^{-5}$	0.08	0.03	$7x10^{-3}$	-0.02	0.14	$9x10^{-1}$
rs16857031	0.07	0.03	$4x10^{-3}$	0.17	0.03	$6x10^{-8}$	0.24	0.08	$3x10^{-3}$
rs2074238	-0.47	0.12	$6x10^{-5}$	-0.41	0.06	$3x10^{-10}$	-0.66	0.55	$3x10^{-1}$
rs37062	-0.11	0.02	$2x10^{-8}$	-0.13	0.02	$5x10-^{8}$	-0.13	0.05	$8x10^{-3}$
rs11756438	0.07	0.02	$4x10^{-5}$	0.11	0.02	$6x10^{-8}$	0.12	0.06	$6x10^{-2}$
rs12576239	0.10	0.03	$1x10^{-4}$	0.07	0.03	$3x10^{-2}$	0.12	0.08	$1x10^{-1}$
rs846111	0.13	0.03	$6x10^{-6}$	0.10	0.03	$3x10^{-4}$	0.25	0.08	$4x10^{-3}$
rs4725982	0.06	0.02	$4x10^{-3}$	0.08	0.03	$4x10^{-3}$	0.12	0.06	$5x10^{-2}$
rs8049607	0.10	0.02	$3x10^{-6}$	0.08	0.02	$1x10^{-4}$	-0.01	0.05	$9x10^{-1}$
rs1805128	0.47	0.08	$2x10^{-8}$	0.06	0.06	$3x10^{-1}$	0.79	0.84	$4x10^{-1}$
rs12053903	-0.09	0.02	$3x10^{-6}$	-0.05	0.02	$2x10^{-2}$	-0.12	0.05	$2x10^{-2}$
rs2074518	-0.06	0.02	$2x10^{-4}$	-0.06	0.02	$2x10^{-3}$	-0.09	0.06	$9x10^{-2}$
rs2968864	-0.08	0.02	$3x10^{-4}$	-0.03	0.02	$2x10^{-1}$	-0.12	0.05	$3x10^{-2}$
	Model r^2 = 5.4%			Model r^2 = 6.5%			Model r^2 = 2.3%		

Supplementary Table 3 *Coverage statistics for novel loci. For purposes of coverage estimation, we defined the associated interval at each of the five novel loci to be the genomic span surrounding the top association signal(s) bounded by SNPs with r^2 to each independent signal ≥0.20, including all intervening SNPs. Shown are the genomic span of these loci, the mean imputation quality score (ranging from 0 to 1) and the proportion of SNPs in the interval captured at an imputation quality score >0.50 and >0.80. As can be seen, coverage at four of these novel loci is quite good (>90% of all SNPs captured at imputation quality >0.50) but the locus containing RNF207 is less well covered.*

Chr	Gene	Start	Stop	Genomic span (kb)	Mean	>0.50	>0.80
16q	*CNOT1*	57085908	57257853	172	0.89	0.93	0.86
6q	*c6orf204*	118634581	119134543	500	0.95	0.99	0.95
1p	*RNF207*	6201001	6245523	45	0.63	0.71	0.43
16p	*LITAF*	11574706	11631856	57	0.70	0.97	0.27
17q	*LIG3*	30084616	30492053	407	0.92	0.96	0.89

Supplementary Table 4 *Quintiles of QT interval score and odds ratios for prolonged QTc. Shown are the odds ratios, 95% confidence intervals and the p-values for each quintile of QT interval score relative to the lowest quintile (q1).*

	Rotterdam Study			Framingham Heart Study		
	OR	**95% CI**	***P***	**OR**	**95% CI**	***P***
q1	ref			ref		
q2	1.64	1.01-2.65	0.04	1.24	0.76-2.05	0.39
q3	1.95	1.22-3.11	0.005	1.91	1.21-3.03	0.006
q4	2.26	1.43-3.57	$5x10^{-4}$	1.52	0.94-2.46	0.09
q5	3.08	1.98-4.78	$6x10^{-7}$	2.55	1.64-3.98	$3x10^{-5}$

SUPPLEMENTARY METHODS

QT measurement methods

In FHS, paper electrocardiograms recorded on Marquette machines were scanned and digital caliper measurements were made using proprietary software (eResearchTechnology, generations 1 and 2) or using Rigel 1.7.2. (AMPS, LLC, New York, NY, USA, generation 3). The QT interval was measured from the beginning of the QRS to the end of the T wave (or the nadir of the T-U when U waves were present) in two cardiac cycles from lead II, one cycle from lead V2 and one cycle from lead V5[1]. The correlation of a single cycle QT measure between two readers was 0.77 and the coefficient of variation was 2.6%, as previously published[1]. To increase precision, the average of measures from 4 cardiac cycles was used in analyses.

In the Rotterdam Study, electrocardiograms were measured on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the QT interval were made using the Modular ECG Analysis System (MEANS)[2]. The operation of the waveform recognition algorithms has been described and validated extensively[2,3]. The MEANS program determines common QRS onset and T offset for all 12 leads together on one representative averaged beat in a 10-second recording by use of template matching techniques. A common QRS onset and T offset are determined over all 12 leads in the representative complex. The presence of U waves or low T waves in individual leads then becomes less relevant. In the Common Standards for Quantitative Electrocardiography (CSE) study, in which different ECG computer programs were compared with a group of expert cardiologists, the measurement performance of MEANS ranked among the best[3].

In the Cardiovascular Health Study, the electrocardiograms were recorded on MAC PC-DT ECG recorder (Marquette Electronics Inc, Milwaukee, WI, USA) machines and measurements of QT interval made using the Marquette 12SL algorithm, which measures the QT as a global interval measured from the median complex derived from the cardiac cycles occurring in 10 seconds. QT measures were compared to values generated using the Dalhousie Program by the EPICORE reading center[4], which measures a global interval from a complex obtained with selective averaging of all normally conducted complexes. 12SL values with QT within 40 msec of the Dalhousie Program were accepted for use (97.4%). If the QT interval differed between the two programs by ≥40 msec (only 2.6% of ECGs), then the value closer to the median rate-corrected QT was used.

Genotyping

In FHS, genotyping was performed by Affymetrix (Santa Clara, CA, USA) using the Affymetrix 500K GeneChip array and a custom-designed gene-centric 50K MIP. Affymetrix 500K genotypes were called using the BRLMM algorithm[5]. In FHS, the following exclusions were applied to exclude individuals with call rate ≤ 97%, per subject heterozygosity ± 5SD away from mean, or excess Mendelian errors resulting in 8,481 individuals with genotype regardless of phenotype and then to exclude SNPs with HWE $P < 10^{-6}$ (15,586), call rate ≤ 97% (64,511), mishap $P < 10^{-9}$ (45,361), Mendel errors >100 (4,857), minor allele frequency <0.01 (67,269), incompatible strand with HapMap genotypes (release 22, n = 2) and SNPs not present on HapMap (13,394), resulting in a set of 378,163 SNPs to be used in imputation.

In RS, genotyping was performed by the genetic laboratory of the Department of Internal Medicine, Erasmus Medical Center, Rotterdam using the Infinium II HumanHap550K Genotyping BeadChip version 3 (Illumina, San Diego, CA, USA). The Illumina 550K BeadChip array was genotyped in all participants of the original Rotterdam Study cohort with proper quality DNA samples (n = 6,449). Intensity files were analyzed using the BeadStudio Genotyping Module software v.3.1.14. A no-call threshold of 0.15 was applied to a custom-generated cluster file derived from the Illumina-provided cluster file (based on the cluster definitions applied to the HapMap CEPH cohort). In the custom-cluster file 2,308 SNPs with GenCall scores < 0.90 were visually checked by two observers and manually re-clustered or zeroed accordingly. Poorly performing samples with low call rate and 10th percentile GenCall score were excluded prior to calling genotypes. Any samples with a call rate below 97.5% (n = 209), excess autosomal heterozygosity > 0.336 ~FDR < 0.1% (n = 21), mismatch between called and phenotypic sex (n = 36), or if there were outliers identified by the IBS clustering analysis clustering > 3 standard deviations away from the population mean (n = 102) or IBS probabilities > 97% (n = 129) were excluded from the analysis. In total, 5,974 genotyped samples were available after exclusions.

In CHS, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system. Genotypes were called using the Illumina BeadStudio software as above. Samples were excluded from analysis for sex mismatch, discordance with prior genotyping, or call rate < 95%. SNPs were excluded from analysis for HWE $P < 10^{-5}$; SNPs with call rates < 95% were manually reclustered using the Illumina software. All three studies make use of genotypes that were not specifically generated to examine the genetic basis of QT interval duration.

Associations of poorly imputed SNPs were validated by re-genotyping FHS samples at the Broad Institute using the Sequenom platform (San Diego, CA, USA) and RS samples at the Erasmus Medical Center using Taqman MGB platform using Assays-by-Design (Applied Biosystems, Foster City, CA, USA).

Imputation

In FHS, model parameters were estimated using MACH[6] v1.0.15 (using flags – rounds 100, –greedy) in 200 unrelated FHS individuals, prioritized for high call rate (≥98.9%), low Mendel error rates and non-outlier status in EIGENSTRAT[7] principal components analysis. With these

model parameters, we used MACH[6] to impute allele dosage, defined as the expected number of copies of the minor allele (a fractional value between 0 and 2), of all autosomal SNPs on HapMap CEU based on phased chromosomes of release 22, build 36.

In RS, the following exclusions were applied to identify 512,349 SNPs to be used for imputation: HWE $P < 10^{-6}$, call rate ≤ 98% and minor allele frequency < 0.01. In total 49,117 SNPs were excluded. For setting model parameters 200 random subjects were selected and used for every chromosome to estimate error and recombination rate (with flags – greedy, – rounds 100). Imputation was then performed to impute genotypes oriented to the positive strand of the human genome reference sequence for all autosomal SNPs in HapMap CEU using release 22, build 36. For each SNP in each individual, imputation results are reported as an allele dosage.

In CHS, the following exclusions were applied to identify a final set of 332,946 SNPs: call rate >95%, HWE $P > 10^{-5}$, 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios). Imputation was performed using BIMBAM[8] v0.95 with reference to HapMap CEU using release 21a, build 35 using one round of imputations and the default expectation-maximization warm-ups and runs. SNPs were excluded for variance on the allele dosage ≤0.01.

REFERENCES

1. Newton-Cheh, C., Larson, M.G., Corey, D.C., Benjamin, E.J., Herbert, A.G., Levy, D., D'Agostino, R.B. and O'Donnell, C.J. (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Heart Rhythm*, 2, 277-284.
2. van Bemmel, J.H., Kors, J.A. and van Herpen, G. (1990) Methodology of the modular ECG analysis system MEANS. *Methods Inf Med*, 29, 346-353.
3. Willems, J.L., Abreu-Lima, C., Arnaud, P., van Bemmel, J.H., Brohet, C., Degani, R., Denis, B., Gehring, J., Graham, I., van Herpen, G. *et al.* (1991) The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*, 325, 1767-1773.
4. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A. *et al.* (1991) The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*, 1, 263-276.
5. Rabbee, N. and Speed, T.P. (2006) A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics, 22, 7-12.
6. Li, Y. and Abecasis, G.R. (2006) Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet*, S79, 2290.
7. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat.Genet.*, 38, 904-909.
8. Servin, B. and Stephens, M. (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS.Genet*, 3, e114-.

3.2 Genome-wide association analysis identifies multiple loci related to resting heart rate

Eijgelsheim M*, Newton-Cheh C* , Sotoodehnia N*, de Bakker PI*, Müller M#, Morrison AC#, Smith AV#, Isaacs A#, Sanna S#, Dörr M#, Navarro P#, Fuchsberger C#, Nolte IM#, de Geus EJ#, Estrada K, Hwang SJ, Bis JC, Rückert IM, Alonso A, Launer LJ, Hottenga JJ, Rivadeneira F, Noseworthy PA, Rice KM, Perz S, Arking DE, Spector TD, Kors JA, Aulchenko YS, Tarasov KV, Homuth G, Wild SH, Marroni F, Gieger C, Licht CM, Prineas RJ, Hofman A, Rotter JI, Hicks AA, Ernst F, Najjar SS, Wright AF, Peters A, Fox ER, Oostra BA, Kroemer HK, Couper D, Völzke H, Campbell H, Meitinger T, Uda M, Witteman JC, Psaty BM, Wichmann HE, Harris TB, Kääb S, Siscovick DS, Jamshidi Y, Uitterlinden AG, Folsom AR, Larson MG, Wilson JF, Penninx BW, Snieder H, Pramstaller PP, van Duijn CM, Lakatta EG, Felix SB, Gudnason V, Pfeufer A, Heckbert SR, Stricker BH, Boerwinkle E, O'Donnell CJ.

Hum Mol Genet. 2010 Oct 1; 19(19): 3885-3894.

* and # denote equal author contributions

ABSTRACT

Higher resting heart rate is associated with increased cardiovascular disease and mortality risk. Though heritable factors play a substantial role in population variation, little is known about specific genetic determinants. This knowledge can impact clinical care by identifying novel factors that influence pathologic heart rate states, modulate heart rate through cardiac structure and function or by improving our understanding of the physiology of heart rate regulation. To identify common genetic variants associated with heart rate we performed a meta-analysis of fifteen genome-wide association studies, including 38,991 subjects of European ancestry, estimating the association between age-, sex-, and body mass-adjusted RR interval (inverse heart rate) and ~2.5 million markers. Results with $P<5\times10^{-8}$ were considered genome-wide significant. We constructed regression models with multiple markers to assess whether results at less stringent thresholds were likely to be truly associated with RR interval. We identified six novel associations with resting heart rate at six loci: 6q22 near *GJA1*, 14q12 near *MYH7*, 12p12 near *SOX5*, *c12orf67*, *BCAT1*, *LRMP* and *CASC1*, 6q22 near *SLC35F1*, *PLN* and *c6orf204*, 7q22 near *SLC12A9*, *Ufsp1* and 11q12 near *FADS1*. Associations at 6q22 400kb away from *GJA1*, at 14q12 *MYH6* and at 1q32 near *CD34* identified in previously published GWAS were confirmed. In aggregate, these variants explain ~0.7% of RR interval variance. A multi-variant regression model including 20 variants with $P<10^{-5}$ increased explained variance to 1.6% suggesting that some loci falling short of genome-wide significance are likely truly associated. Future research is warranted to elucidate underlying mechanisms that may impact clinical care.

INTRODUCTION

Higher resting heart rate is associated with increased risk of cardiovascular disease[1], cardiovascular mortality[2,3], including sudden death[4], and all-cause mortality independent of traditional risk factors[5-9]. However, it is not known whether heart rate directly impacts mortality or merely reflects unrecognized subclinical disease[1,7,8]. Recently it was shown that physical exercise reduces the increased cardiovascular mortality risk associated with higher heart rate[3]. This suggests that heart rate is a clinically relevant and potentially modifiable risk factor.

Heart rate is a complex trait, determined by multiple environmental, genetic and other endogenous factors. Heredity plays a substantial role in the inter-individual variation of resting heart rate, accounting for 26 to 32% of heart rate variation in prior studies[10-13]. Twin studies report even higher heritability estimates up to 55 to 63%[14,15]. Candidate gene approaches have identified multiple loci associated with heart rate[12,13,16-20], but the results have been inconsistent and difficult to replicate. Genome-wide genotyping arrays of single nucleotide polymorphisms (SNPs) assay common variation in the human genome and can identify genetic variants with modest influences on a complex trait such as heart rate, as shown by two recent genome-wide association studies (GWAS) that identified common variation at or near *MYH6*, *GJA1* and *CD34* associated with heart rate[21,22]. These chromosomal loci identified in an unbiased genome-wide study may represent novel risk factors for cardiovascular disease outcomes. This knowledge may also have impact on clinical care 1) by identifying novel factors that cause pathologic heart rate states (such as sick sinus syndrome or other arrhythmia's), 2) by identifying factors that influence cardiac structure or function (e.g. stroke volume) and thereby modulate heart rate (since cardiac output = heart rate x stroke volume), or 3) by improving our understanding of the physiologic basis of heart rate regulation. Altogether this will generate insights into the underlying mechanisms of heart rate as a well established, but poorly understood, risk indicator for cardiovascular disease and mortality.

To identify additional genetic determinants of heart rate, we performed a meta-analysis of genome-wide association studies of resting heart rate, measured as the RR interval on the electrocardiogram (ECG), in 38,991 individuals of European ancestry derived from fifteen studies in the RRGEN consortium.

METHODS

Study participants

The RRGEN sample consisted of subjects from the five participating studies in the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium[23] – comprised of the Age, Gene, Environment Susceptibility (AGES) Study, the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS) and the Rotterdam Study (RS) – as well as the Cooperative Health Research in the Region Augsburg (KORA) study, the SardiNIA study, the Study of Health in Pomerania (SHIP), TwinsUK, Netherlands Study of Depression and Anxiety (NESDA) and three population isolate studies in the European

Special Populations Network (EUROSPAN), the Erasmus Rucphen Family (ERF), the South Tyrolean Micro-Isolate (MICROS) and the Orkney Complex Disease Study (ORCADES). For RRGEN both the baseline Rotterdam Study (RS-I) and first extended cohort (RS-II) were used[24,25]. KORA subjects in RRGEN were drawn from the F3 and S4 cohorts.

Individuals were excluded if they were non-Caucasian, had atrial fibrillation, a second or third degree atrio-ventricular block, a pacemaker, a diagnosis of prevalent myocardial infarction or prevalent heart failure, used beta-adrenergic blocking agents, non-dihydropyridine calcium antagonists or digoxin, or had a heart rate below 50 (RR interval>1200 milliseconds) or above 100 beats per minute (RR interval<600 milliseconds).

To adhere to STrenghtening the reporting of Genetic Associations studies (STREGA) statement guidelines[26] we include an online supplement with additional information (Supplementary Material). All studies have approval from their institutional review committee, and the subjects of all cohorts provided written informed consent. A more detailed description of cohorts is given in the Supplementary Material.

RR interval measurement methods

All cohorts recorded 12-lead ECGs from which the RR interval (which equals the inverse heart rate) was measured. For cohort-specific details on RR interval measurement methods please refer to the Supplementary Material.

Genotyping and imputation

Affymetrix and Illumina arrays were used for genotyping. Using genotype information generated on these platforms, all cohorts imputed genotypes for a common set of ~2.5 million autosomal SNPs based on linkage disequilibrium patterns observed in HapMap CEU reference samples (Utah residents of Northern and Western European descent). The genetic trait analyzed was the imputed allele dosage, a fractional value between 0 and 2, reflecting the estimated number of allele copies of a SNP for each subject. A more detailed description is given in Supplementary Table 1.

Statistical methods

Resting RR interval was adjusted for age, sex and body mass index. For each SNP, we tested the genotype for association with covariate adjusted-RR interval under an additive genetic model using linear regression models. We then conducted a fixed-effects, inverse variance weighted meta-analysis using beta estimates and standard errors from each of the cohorts with applying genomic control on a per study basis and additionally post meta-analysis ($P_{\text{meta-gc}}$). Genomic control refers to the correction made to the test statistics to account for any inflation of the test statistic distribution, which can result from unaccounted population substructure or other technical biases[27,28]. We mapped all SNPs to dbSNP build 129, resulting in a unique set of 2,650,552 autosomal SNPs, after confirming consistency of the coded allele across all studies. Scripts used for this meta-analysis are available online. (http://www.broadinstitute.org/~debakker/meta.html)

Genome-wide significance was defined as $P<5\times10^{-8}$, based on the estimated multiple testing burden for all common variants in populations of European ancestry[29]. To identify independent signals reaching the genome-wide significance threshold within a locus, genome-wide association meta-analysis results were aggregated into bins by index SNP at a linkage disequilibrium r^2 threshold of 0.1, such that all results within a given bin were correlated to the index SNP at $r^2\geq0.1$ but to any index SNP in other bins at $r^2<0.1$. Four index SNPs at two loci were subsequently analyzed in conditional regression models (n<=33,846) to assess statistical independence.

Finally, we adopted the polygenic regression modelling approach as recently described by Purcell *et al.*[30] and implemented in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink) to estimate the genetic variance explained by associated loci at progressively less stringent *P*-value thresholds within the RS-II sample. The outcome of this analysis is a *P*-value threshold at which the explained variance is maximized. The list of loci included in the score yielding the maximum explained variance will include non-genome-wide loci that in aggregate contribute to the model's performance and indicate that additional true positive signals are likely to be present within that list. For this analysis, we removed the RS-I and RS-II data from the discovery meta-analysis to remove the risk of correlation between discovery and validation sample and overestimation of the explained variance. Subsequently, we obtained a list of independent signals to be included in the model based on their statistical significance from the meta-analysis without RS-I or RS-II (PLINK – clump option, $r^2\geq0.05$, 1 megabase window)[31]. We summarized variation across associated loci (using significance thresholds from $P<5\times10^{-8}$ to $P<0.05$) into quantitative scores per individual. The score was then used as a predictor in a linear regression analysis in the RS-II sample (n=1,589) and the resulting R^2 is reported as the measure for explained variance for each *P*-value threshold.

RESULTS

There were 38,991 individuals available for genotype-phenotype association analysis after exclusions. Subject characteristics are shown in Table 1. While cohort-specific quantile-quantile plots of *P*-value distributions approximated expectations under the null, the meta-analysis of all results showed a clear excess of low *P*-values (Supplementary Figure 1). Study specific genomic inflation lambda values ranged from 0.98 to 1.05, suggesting that population stratification or other technical artifacts were minimal (Supplementary Table 1).

Meta-analysis of results from all studies resulted in 156 SNPs reaching the pre-specified genome-wide statistical significance level ($P<5\times10^{-8}$) before applying post-meta-analysis genomic control. In total, nine independent genome-wide significant signals were observed across seven chromosomal loci (Table 2, Figure 1), with all but one locus harboring multiple SNPs reaching the genome-wide significance threshold. (Figure 2A-E).

Genome-wide significant associations were observed at 6q22 nearest to *GJA1*, 14q12 near *MYH6* and *MYH7*; at 12p12 near *SOX5*, *C12orf67*, *BCAT1*, *LRMP* and *CASC1*; at 6q22 near *SLC35F1*, *6orf204* and *PLN* (>3Mb away from 6q22 *GJA1*); at 7q22 near *SLC12A9*; at 11q12 near *FADS1* and at 1q32 near *CD34*. The genomic inflation lambda value of the meta-analysis was 1.05. After

Table 1 *Baseline characteristics of samples included by cohort*

	Sample size	Male, n (%)	Age, mean (SD*), y	Body mass index, mean, (SD*), kg/m^2	Heart rate (SD*), beats/ minute	RR interval (SD*), msec†	SD* of RR–residual‡
AGES	1,651	622 (37.7)	75.9 (5.5)	26.8 (4.4)	68.3 (10.2)	897.1 (131.0)	128.5
ARIC	6,308	2,855 (45.3)	53.9 (5.6)	26.7 (4.7)	67.3 (9.0)	907.6 (118.2)	116.1
CHS	2,544	951 (37.4)	72.2 (5.3)	26.2 (4.4)	65.7 (9.2)	930.1 (123.6)	121.3
ERF	1,275	508 (39.8)	47.1 (14.0)	26.5 (4.5)	64.5 (9.2)	948.5 (128.8)	126.9
FHS	7,243	3,305(45.6)	40.2(10.5)	26.1(5.0)	69.3(11.1)	888.2(139.2)	121.8
KORA F3	995	480 (48.2)	60.0 (10.1)	27.3 (4.4)	65.6 (9.7)	933.8 (130.8)	128.1
KORA S4	1,398	654 (46.8)	52.8 (8.7)	27.3 (4.4)	66.4 (9.4)	921.1 (125.6)	122.5
MICROS	919	399 (44.4)	44.8 (16.0)	25.6 (4.8)	68.8 (11.7)	897.0 (151.4)	92.0
NESDA	1,456	437 (30.0)	39.8 (12.2)	25.06 (4.7)	68.1 (9.6)	898.7 (125.4)	125.1
ORCADES	546	240 (44.0)	52.6 (14.9)	27.6 (4.9)	62.5 (8.1)	975.2 (119.6)	118.5
RS-I	3,781	1,441 (38.1)	68.5 (8.6)	26.1 (3.6)	71.2 (10.2)	860.6 (126.4)	124.1
RS-II	1,589	695 (43.7)	64.8 (7.4)	27.0 (4.0)	70.3 (10.1)	871.1 (123.9)	122.5
SardiNIA	3,977	1,678 (42.1)	42.9 (17.3)	25.3 (4.7)	64.5 (10.1)	907.4 (130.0)	127.4
SHIP	2,582	1,260 (48.8)	46.8 (15.7)	26.8 (4.7)	72.1 (11.4)	852.7 (134.4)	133.5
TwinsUK	2,727	117 (4.3)	51.7 (12.5)	25.7 (4.4)	67.1 (9.6)	911.5 (126.3)	125.5

* SD=standard deviation
† msec=milliseconds
‡ RR-residual=Residuals are from linear regression models adjusting for age, sex and body mass index.

applying post-meta-analysis genomic control, the signal at 1q34 near *CD34* and a second independent signal at 6q22 *GJA1* lost genome-wide significance.

The strongest association was observed for an intergenic SNP 370kb upstream of *GJA1* at 6q22 rs9398652 (minor allele frequency [MAF]=0.10) with 12.6 msec shorter RR interval per minor A allele, which is equivalent to a 0.95 beats/minute (bpm) higher heart rate based on the baseline mean heart rate of 66.8 bpm, across all studies, (P=8.0×10^{-16}, $P_{meta\text{-}gc}$=3.8×10^{-15}, Table 2 and Figure 2A). Cho *et al.* observed genome-wide significant association between rs12110693 and pulse rate in an Asian population based GWAS[21]. This SNP is in perfect linkage disequilibrium (r^2=1) with rs939862 in both Caucasian and Asian HapMap reference populations and thus reflects the same signal 370kb from *GJA1*. A second SNP only 8kb away from *GJA1* also reached genome-wide significance pre-genomic control, rs11154022 (MAF=0.33) with 5.8 msec longer RR interval (0.43 bpm lower heart rate) per A allele, (P=3.5×10^{-8}, $P_{meta\text{-}gc}$=7.2×10^{-8} Figure 2B, Table 2). This SNP had low correlation with rs9398652 (r^2=0.006 in HapMap CEU), suggesting a novel independent association signal. Since HapMap is limited to 90 subjects we assessed the linkage disequilibrium in our data. All observed r^2 values ranged between 0.0001 and 0.004, which is lower than seen in the HapMap CEU reference population. Conditional analysis confirmed that these two signals are independent with $P_{conditional}$=2.4×10^{-11} and $P_{conditional}$=3.3×10^{-8}, respectively for rs9398652 and rs1154022 in the subset (n<=33,846) used for this analysis.

The second locus with two signals in low correlation reaching genome-wide significance is located on chromosome 14q12. The strongest association at this locus was observed for rs452036

Table 2 *Association analyses results for independent index SNPs from loci with P<5×10⁻⁸ in the meta analysis.*

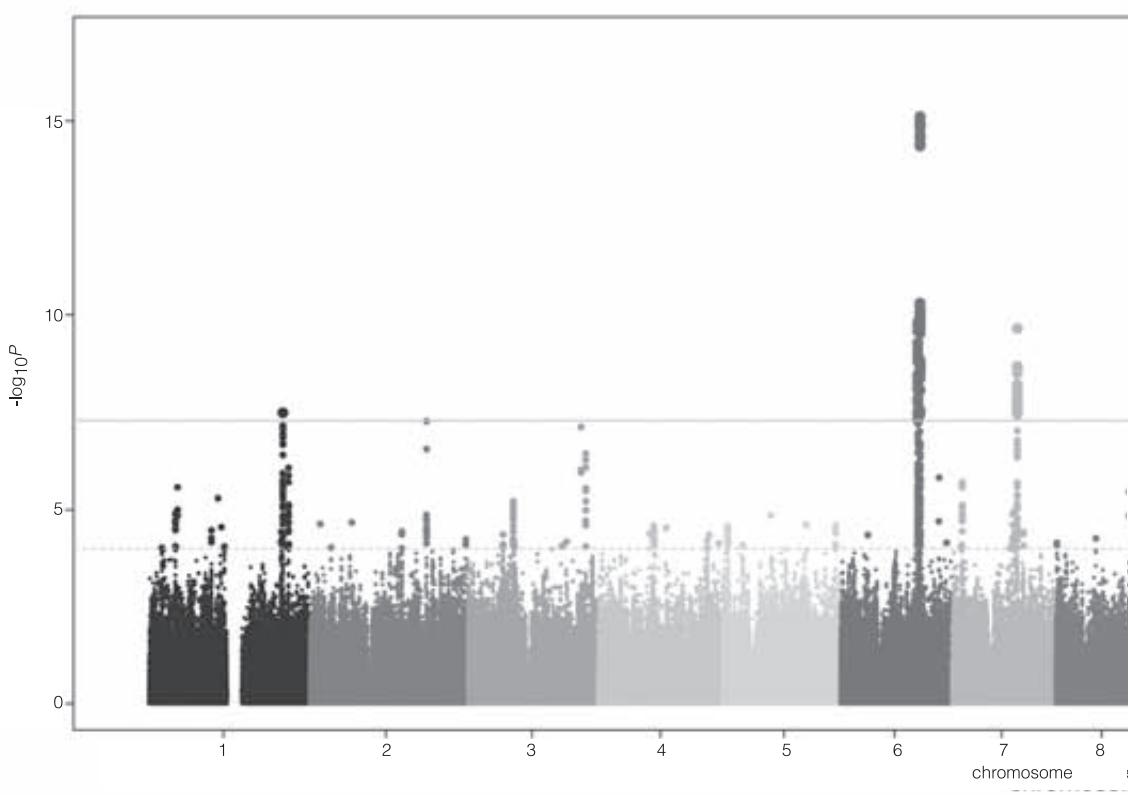
Chr†	Basepair position (kb)	SNP	Correlation to index SNP*	Function/ Gene	Coded/ Non-Coded allele	Allele frequency	Effective sample	Effect Estimate	SE‡	2-sided *P*	2-sided $P_{meta-gc}$
6q22	122,187	rs9398652	-	Intergenic, 400kb from *GJA1*	A / C	0.10	37,050	-12.6	1.56	$7.7x10^{-16}$	$3.8x10^{-15}$
6q22	121,790	rs11154022	0.006	Intergenic, 8kb from *GJA1*	A / G	0.33	31,676	5.8	1.05	$3.5x10^{-8}$	$7.2x10^{-8}$
14q12	22,935	rs452036	-	Intronic *MYH6*	A / G	0.36	34,640	-7.8	1.00	$8.1x10^{-15}$	$3.8x10^{-14}$
14q12	22,931	rs365990	0.96	Non-synonymous coding *MYH6* (Ala-1101-Val)	G / A	0.37	32,627	-7.7	1.02	$5.4x10^{-14}$	$2.1x10^{-13}$
14q12	23,046	rs223116	0.08	Intergenic, nearest to *MYH7, NDNG*	A / G	0.24	26,899	-7.4	1.30	$1.1x10^{-8}$	$2.5x10^{-8}$
12p12	24,662	rs17287293	-	Intergenic	G / A	0.15	37,988	8.6	1.31	$5.7x10^{-11}$	$1.6x10^{-10}$
6q22	118,680	rs281868	-	Intronic *SLC35F1*	G / A	0.50	32,109	-6.3	0.99	$1.5x10^{-10}$	$4.3x10^{-10}$
7q22	100,291	rs314370	-	Intronic *SLC12A9*	C / T	0.19	35,170	-7.6	1.21	$2.3x10^{-10}$	$6.1x10^{-10}$
7q22	100,324	rs12666989	0.88	Non-synonymous coding *UfSp1* (Leu-41-Val)	C / T	0.18	35,750	-7.0	1.21	$9.4x10^{-9}$	$2.1x10^{-8}$
11q12	61,327	rs174547	-	Intronic *FADS1*	C / T	0.33	34,907	-6.2	1.01	$8.2x10^{-10}$	$2.1x10^{-9}$
1q32	206,195	rs2745967	-	Intergenic near *CD34*	G / A	0.37	34,913	5.4	0.98	$3.2x10^{-8}$	$6.6x10^{-8}$

Chromosomal positions and coded alleles are given relative to forward strand of NCBI build 36. Effect sizes (on the millisecond scale) are shown as beta estimates from linear regression models for each additional copy of the coded allele. The effective sample size reflects the imputation quality-adjusted sample size. Final column shows the *P*-value from inverse-variance weighted meta-analyses.
† Chr = chromosome
* CEU HapMap population linkage disequilibrium r^2 values to the index SNP
‡ SE = standard error

located in intron 19 of *MYH6* (MAF =0.36, 7.8 msec shorter RR interval (0.58 bpm higher heart rate) per C allele, $P=8.1.\times10^{-15}$, $P_{meta-gc}=3.8\times10^{-14}$, Figure 2C). This replicates the finding by Holm *et al.* who previously described an association between rs452036 and heart rate[22]. The non-synonymous coding variant rs365990, which results in an amino acid change at position 1101 (Alanine>Valine) of the *MYH6* gene product, is a possible functional variant since it showed strong correlation with ($r^2=0.96$ in HapMap CEU) and association results are indistinguishable from rs452036 (Table 2). A second SNP located near *MYH7* rs223116, (MAF=0.24), associated with a 7.4 msec shorter RR interval per A allele (0.55 bpm higher heart rate, $P=1.1\times10^{-8}$, $P_{meta-gc}=2.5\times10^{-8}$, Figure 2D) was in low correlation with rs452036 ($r^2=0.08$ in HapMap CEU) and reflects a novel association. We observed r^2 values similar to HapMap CEU with values ranging from 0.03 to 0.08. Conditional analysis confirmed the presence of two independent signals with $P_{conditional}=7.9\times10^{-14}$ and $P_{conditional}=3.7\times10^{-4}$ for rs452036 and rs223116, respectively.

For the other five loci only a single signal of association met our genome-wide significance threshold, meaning all other genome-wide significant SNPs have an r^2>0.1 in HapMap CEU to the most significant SNP at that locus. For these loci the minor allele frequencies of the index SNPs ranged from 15 to 50%, effect sizes ranged between 5.4 and 8.6 milliseconds and $P_{meta\text{-}gc}$ ranged between 2.1×10^{-9} and 1.6×10^{-10}. (Table 2, Figure 2C to 2H) The *CD34* locus lost genome-wide significance upon meta-analytic genomic-control but replicates the association reported by Cho *et al*[21]. Of these five loci only the index SNP at 7q22 *SLC12A9* shows strong correlation with a non-synonymous coding SNP. This coding SNP (rs12666989, r^2=0.88 to rs314370 in HapMap CEU) results in a Leucine>Valine substitution at amino acid position 41 in the *UfSp1* gene product.

Figure 1 *RR interval association results for ~2.5 million imputed autosomal SNPs in 38,991 individuals from 15 cohorts. Results are shown on the -log$_{10}$(P) scale (Y-axis). The X-axis depicts chromosomal position. The gray horizontal line corresponds to the genome-wide significance threshold of P=5×10^{-8}.*



Additionally, an association result that just missed our genome-wide significance threshold (P=5.2×10^{-8}, $P_{meta\text{-}gc}$=1.1×10^{-7}) was observed for a non-synonymous coding SNP in *CCDC141* at 2q31. This SNP, rs17362588 (MAF=0.12, 8.3 msec shorter RR interval per A allele), results in an amino acid substitution at position 360 (Tryptophan> Arginine) of the encoded protein.

Evidence for additional causal loci not reaching genome-wide significance

We used polygenic modelling methods to quantify the genetic variance explained and to indicate if loci falling short of genome-wide significance are likely to harbour additional variants that influence heart rate. The explained variance of resting heart rate in RS-II (first extended Rotterdam Study cohort; n=1,589) was ~0.7% when the score was calculated based on genome-wide significant signals. Inclusion of 20 independent variants with P<1×10^{-5} resulted in the maximal proportion of explained variance of 1.6% (Supplementary Figure 2). These 20 variants included signals from the 7 genome-wide significant loci and signals from an additional 12 loci, including several loci with genes of potential cardiac relevance (see Supplementary Table 2 for full list).

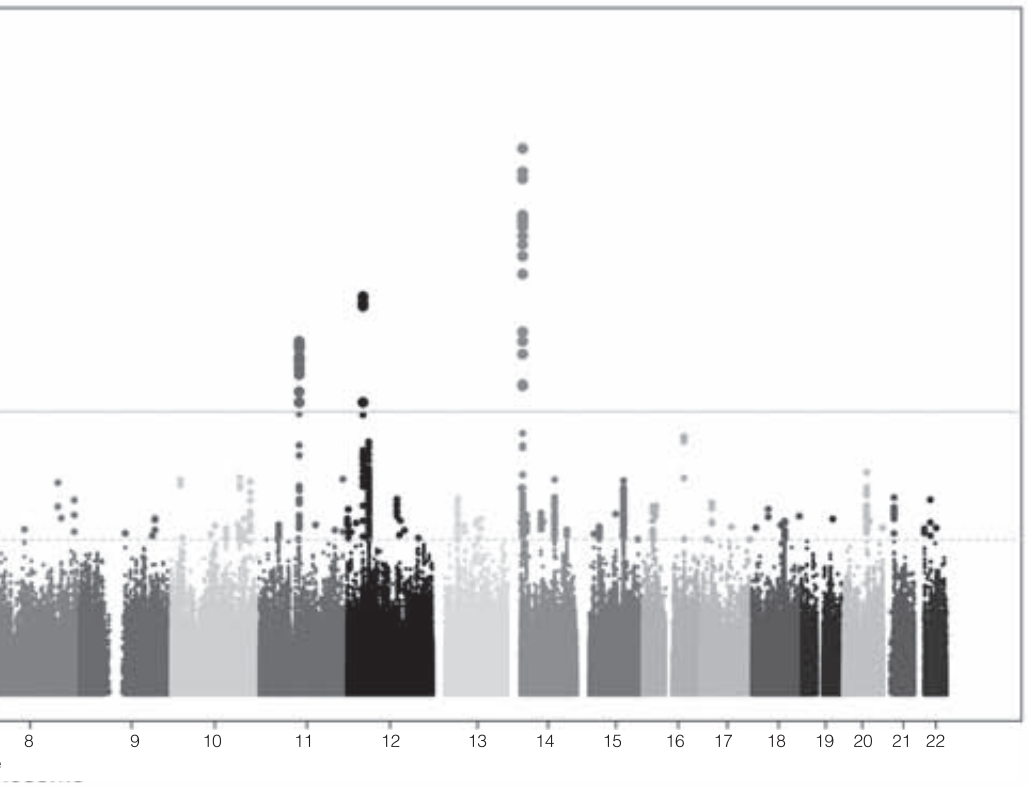
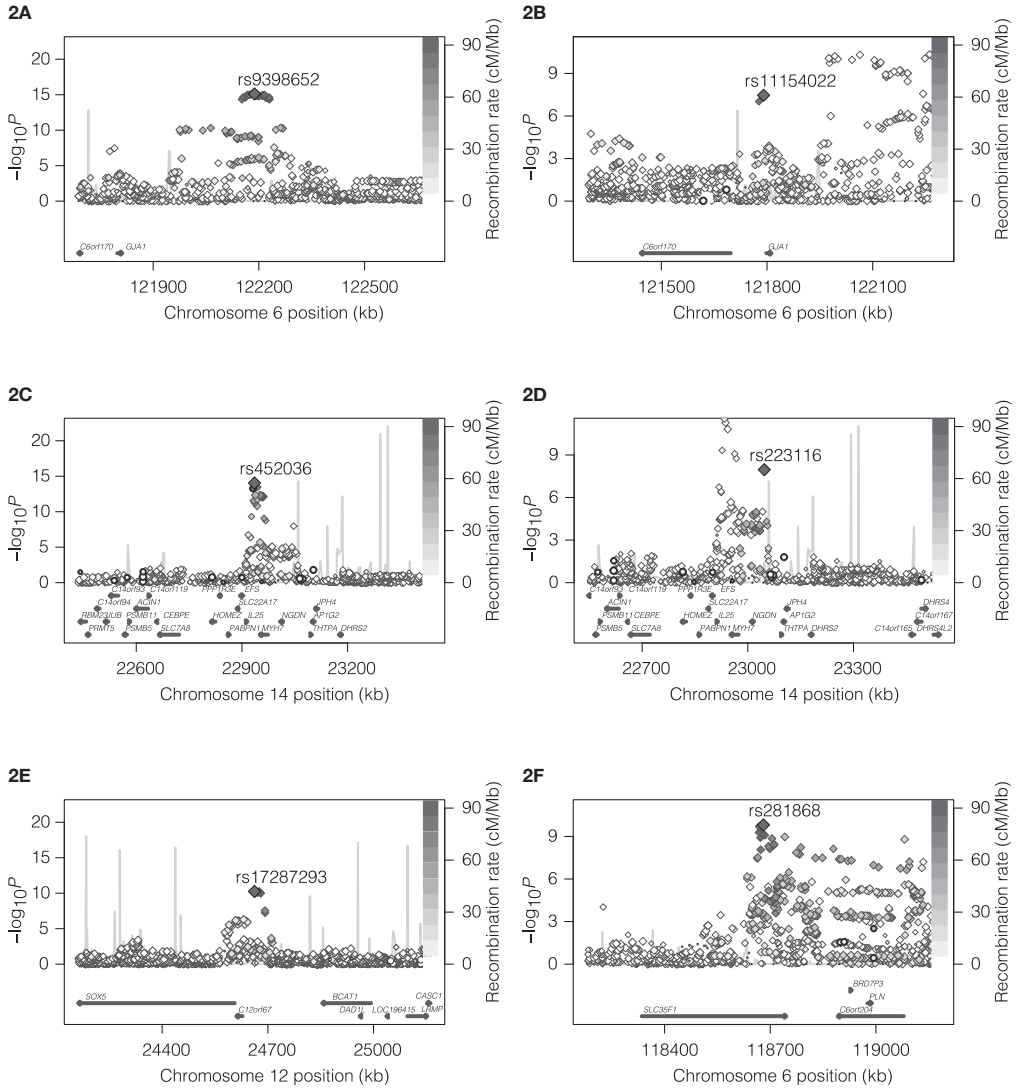
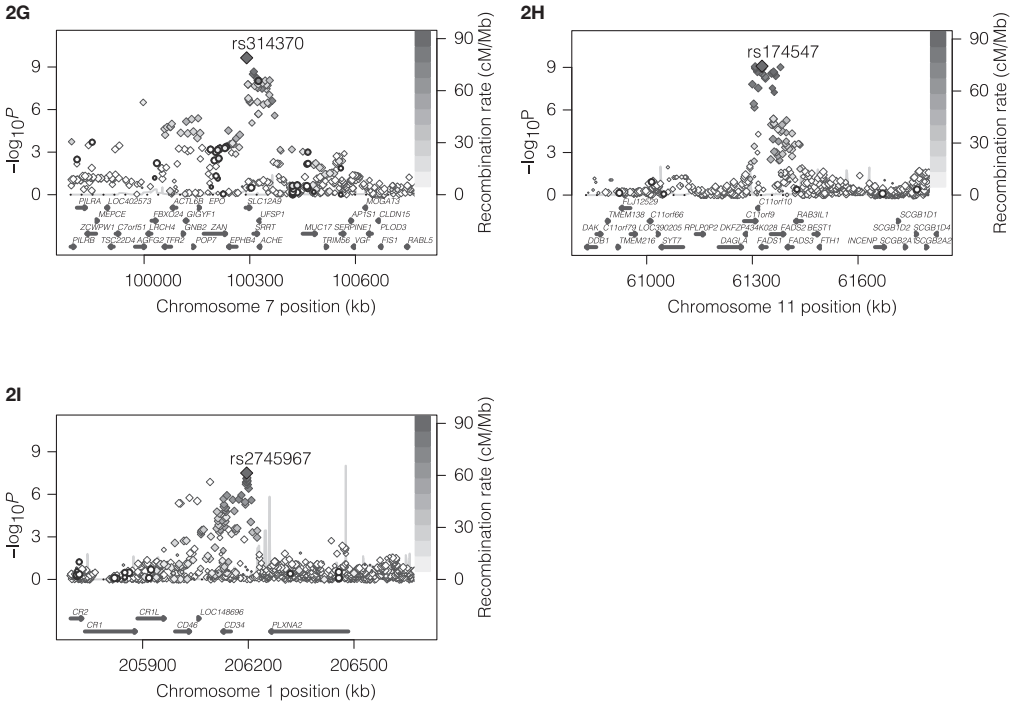


Figure 2 *Regional association plots covering one megabase surrounding the index SNP. Statistical significance of SNPs is shown on the –log$_{10}$(P) scale as a function of chromosomal position. The primary SNP is annotated by rs-number. Correlation of each plotted SNP to the primary SNP is indicated on a color scale from white (minimal correlation) to black (maximal correlation). Non-synonymous coding SNPs resulting in amino acid changes in the encoded protein are plotted as grey circles. Estimated recombination rates from HapMap (grey lines) and RefSeq gene annotations (arrows) are shown.*





DISCUSSION

The application of genome-wide association methods in a large sample of subjects of European ancestry identified common variants at multiple loci associated with inter-individual variation in resting heart rate. Genetic determinants of heart rate could alter the function of the sinus node (the dominant pacemaker in the normal heart) either directly through altered pacemaking activity[32], or indirectly through sympathetic or parasympathetic inputs to the heart. We were therefore encouraged to find that genes at some of the loci identified here are cardiac ion channels or their regulatory proteins. Besides a direct effect on sinus node function, effects on cardiac structure – either developmental or through remodeling – and function could underlie the observed associations.

The most strongly related locus on chromosome region 6q22 included two independent association signals 8kb and 370kb away from *GJA1*. The latter finding is in line with a recent GWAS that described an association between rs12110693 and pulse rate in an Asian sample[21]. rs9398652 described in the present study is in perfect linkage disequilibrium (r^2=1) with rs12110693 in both the Caucasian and Asian HapMap reference populations and thus reflects the same signal 370kb from *GJA1*. *GJA1* encodes CX43, a connexin family protein and a major component of cardiac gap junction, crucial in electrical coupling of myocytes[33]. Mutations in *GJA1* cause Mendelian inher-

ited hypoplastic left heart syndrome[34]. To our knowledge, a role for Cx43 in pacemaker function in the adult sinus node has not been reported.

The 14q12 *MYH6* locus was also associated with resting heart rate in the RRGEN study. Its gene product, the myosin heavy chain-6 protein, is a component of the hexameric myosin protein. Mutations in *MYH6* have been related to Mendelian forms of hypertrophic cardiomyopathy[35], atrial septal defect[36] and dilated cardiomyopathy[37]. The index SNP was strongly correlated with an amino acid-altering common variant in *MYH6*. The association of the coding SNP with heart rate raises the possibility that it is the causal variant and *MYH6* the causal gene. This finding replicates the result of Holm *et al*. who performed a GWAS on heart rate in an Icelandic population based sample [22]. We do report a novel independent signal located near *MYH7*. Future research is warranted to define the allelic architecture of this locus in relation to heart rate. Of note, cardiac specific microRNA's encoded within intronic regions of *MYH6* (miR-208a) and *MYH7* (miR-208b) have regulatory effects on cardiac conduction[38,39].

The locus on chromosome 12p12 includes several genes, but without a clear candidate for association with heart rate in the associated interval. The closest genes are *BCAT1*, which encodes a cytosolic form of the branched-chain amino acid transaminase enzyme that catalyses transamination of branched-chain amino acids to their respective alpha-keto acids essential for cell growth and protein synthesis and *SOX5*, a member of the SOX family[40] and of incompletely understood function. SOX genes play a major role in cell fate modulation through transcriptional activity but without a clear cardiac role for *SOX5*[41]. The variant at 12p12 we describe here, rs17287293, is a perfect proxy for rs11047543 which was associated with PR interval (reflecting atrial depolarization duration and atrioventricular nodal conduction time) in a GWAS[42], independent of heart rate, suggesting pleiotropic electrophysiological effects.

The second locus on 6q22 is located near *SLC35F1* and *PLN*, which encodes phospholamban. This locus is located more than 3Mb away from the loci near *GJA1* on 6q22. We have previously described common variation at this locus to be associated with heart rate corrected QT interval[43-45]. The index SNP reported here was in moderate linkage disequilibrium with two SNPs associated with heart rate-adjusted QT interval duration (rs11756438[44], r^2=0.43 in HapMap CEU, RR interval $P=6.5\times10^{-6}$; and rs11970286[43], r^2=0.58 in HapMap CEU, RR interval $P=2.7\times10^{-8}$), resulting in overlapping signals in independent phenotypes. In addition to higher heart rate and longer QT interval, this locus has also been associated with decreased end-diastolic left ventricular diameter in the EchoGen Study[46]. In humans, the sinoatrial node shows comparable expression of phospholamban compared with atrial myocytes[32]. However, basal cAMP-mediated, Protein Kinase A (PKA)-dependent phosphorylation of phospholamban is elevated in sinoatrial nodal pacemaker cells compared with other cardiac cell types[47]. Basal PKA-dependent phosphorylation is obligatory for spontaneous basal pacemaking activity and graded changes in the phosphorylation of phospholamban cause graded changes of the pacemaker cell basal heart rate[47,48]. The location of the top SNP at this locus in an intron of *SLC35F1*, which is mainly expressed in the brain[49], raises the possibility that causal variation influencing this gene in fact underlies the association at the locus. However, the observation that heart rate, QT interval and left ventricular structure are all associated with genetic variation at this locus makes phospholamban the best

candidate in light of its role in excitation-contraction coupling and intracellular calcium signalling, critical to action potential development in both the sinoatrial node and ventricles.

A locus without a clear candidate to explain its strong association with resting heart rate is located on chromosome 7q22 with the index SNP in *SLC12A9* encoding a cation-Cl^- co-transporter-interacting protein. However, this SNP was strongly correlated with genome-significant SNPs in *TRIP6* encoding a thyroid receptor-interacting protein, *ACHE* which encodes acetylcholinesterase and a non-synonymous-coding SNP in *UfSp1*. This gene encodes an ubiquitin-fold modifier protease[50,51].

The exact same signal that we report from the *FADS1* locus on 11q12 was previously associated with cholesterol levels[52-54] and fatty acid metabolism[55,56]. The direct product of the reaction catalyzed by *FADS1* is arachidonyl-CoA, which has been shown to release Ca^{2+} from the sarcoplasmic reticulum[57].

The 1q32 *CD34/C1ORF132* locus has no clear potential mechanism through which it is related to heart rate and it lost genome-wide significance after applying post-meta-analysis genomic control. However, it is likely to be truly associated with pulse rate since it was previously identified in a non-Caucasian sample[21], and our study thus represents a replication of this finding in a different ethnic group.

The 2q31 *CCDC141* locus just missed our genome-wide significance threshold, but the most significant SNP is a non-synonymous coding variant within *CCDC141*, a coiled coil domain containing protein of unknown function. Interestingly, the nearby *TTN* gene encodes Titin, which is expressed in cardiac and skeletal myocytes[58] and plays a key role in muscle assembly, force transmission and maintenance of resting tension[59].

In additional analysis we have shown that based on the results of this GWAS the explained variance could be increased to 1.6% with the inclusion of additional signals with $P<10^{-5}$. The goal of this analysis was to describe if additional variants associated with heart rate are likely to be present. Although true-positives are amongst these loci, we cannot discriminate them from the false-positives. We did observe that the loci contributing to the maximal explained variance include loci of specific interest. Three loci that stand out are 3q26 near *GNB4*, 12p13 *CACNA1C* and 14q11 near *PRKD1*. *GNB4* encodes Gβ4 known to influence G-protein-activated inwardly rectifying K+ channels (GIRK) that play an important role in heartbeat regulation[60-63] and are activated on binding of acetylcholine to the muscarinic M2-receptor present in the sinoatrial node[32,63]. *CACNA1C* encodes the alpha-1 subunit of the voltage-dependent calcium channel and is related to Timothy's syndrome including, among other traits, prolonged QT interval, high arrhythmia risk and with a single observation of in utero bradycardia[64]. Lastly, the *PRKD1* gene product is relevant for calcium/calmodulin dependent kinase affecting cardiac remodeling and contraction[65]. Validation studies in independent samples have to indicate whether these loci are truly associated with heart rate.

Strengths and limitations

The large sample derived from several population-based cohort studies allowed us to identify common variants with modest effects. In addition, these cohort studies have extensive data on covariates and disease status that allowed us to harmonize exclusion criteria and phenotype modeling prior to analyzing the data.

Additional signals may have been missed due to random sampling variation, restriction to autosomal SNPs, poor coverage of certain genomic regions or rare alleles by the genotyping platforms used or a lack of power to detect even smaller effects. Lastly, heart rate is a very dynamic trait. Strong environmental influences like chronic physical activity or training as well as variability in the time at rest or posture at the moment of measurements and other factors, such as anaemia or anxiety, which we have not accounted for would add noise to the phenotype, which would be expected to bias our study toward the null but not toward the false inference of association.

CONCLUSION

RRGEN identified nine signals at seven loci at which common genetic variation is associated with resting heart rate. Six of these signals at 6 loci are novel while the other three signals replicate previous findings from GWAS. Several of these loci include genes encoding ion channel regulator proteins with known involvement in heart rate regulation, or proteins with cardiovascular relevant functions and known associated Mendelian disorders. These variants may impact clinical care by identifying novel factors that influence pathologic heart rate states, are relevant for cardiac structure and thereby modulate heart rate, or by improving our understanding of the physiologic basis of heart rate regulation.

REFERENCES

1. Fox, K., Borer, J.S., Camm, A.J., Danchin, N., Ferrari, R., Lopez Sendon, J.L., Steg, P.G., Tardif, J.C., Tavazzi, L. and Tendera, M. (2007) Resting heart rate in cardiovascular disease. *J Am Coll Cardiol*, 50, 823-830.
2. Greenland, P., Daviglus, M.L., Dyer, A.R., Liu, K., Huang, C.F., Goldberger, J.J. and Stamler, J. (1999) Resting heart rate is a risk factor for cardiovascular and noncardiovascular mortality: the Chicago Heart Association Detection Project in Industry. *Am J Epidemiol*, 149, 853-862.
3. Nauman, J., Nilsen, T.I., Wisloff, U. and Vatten, L.J. (2010) Combined effect of resting heart rate and physical activity on ischaemic heart disease: mortality follow-up in a population study (the HUNT study, Norway). *Journal of epidemiology and community health*, 64, 175-181.
4. Jouven, X., Zureik, M., Desnos, M., Guerot, C. and Ducimetiere, P. (2001) Resting heart rate as a predictive risk factor for sudden death in middle-aged men. *Cardiovasc Res*, 50, 373-378.
5. Kannel, W.B., Kannel, C., Paffenbarger, R.S., Jr. and Cupples, L.A. (1987) Heart rate and cardiovascular mortality: the Framingham Study. *Am Heart J*, 113, 1489-1494.
6. Kristal-Boneh, E., Silber, H., Harari, G. and Froom, P. (2000) The association of resting heart rate with cardiovascular, cancer and all-cause mortality. Eight year follow-up of 3527 male Israeli employees (the CORDIS Study). *Eur Heart J*, 21, 116-124.
7. Chang, M., Havlik, R.J., Corti, M.C., Chaves, P.H., Fried, L.P. and Guralnik, J.M. (2003) Relation of heart rate at rest and mortality in the Women's Health and Aging Study. *Am J Cardiol*, 92, 1294-1299.
8. Jouven, X., Empana, J.P., Escolano, S., Buyck, J.F., Tafflet, M., Desnos, M. and Ducimetiere, P. (2009) Relation of heart rate at rest and long-term (>20 years) death rate in initially healthy middle-aged men. *Am J Cardiol*, 103, 279-283.

9. Hsia, J., Larson, J.C., Ockene, J.K., Sarto, G.E., Allison, M.A., Hendrix, S.L., Robinson, J.G., LaCroix, A.Z., Manson, J.E. and Women's Health Initiative Research, G. (2009) Resting heart rate as a low tech predictor of coronary events in women: prospective cohort study. *Bmj*, 338, b219.
10. Russell, M.W., Law, I., Sholinsky, P. and Fabsitz, R.R. (1998) Heritability of ECG measurements in adult male twins. *J Electrocardiol*, 30 Suppl, 64-68.
11. Singh, J.P., Larson, M.G., O'Donnell, C.J., Tsuji, H., Evans, J.C. and Levy, D. (1999) Heritability of heart rate variability: the Framingham Heart Study. *Circulation*, 99, 2251-2254.
12. Martin, L.J., Comuzzie, A.G., Sonnenberg, G.E., Myklebust, J., James, R., Marks, J., Blangero, J. and Kissebah, A.H. (2004) Major quantitative trait locus for resting heart rate maps to a region on chromosome 4. *Hypertension*, 43, 1146-1151.
13. Laramie, J.M., Wilk, J.B., Hunt, S.C., Ellison, R.C., Chakravarti, A., Boerwinkle, E. and Myers, R.H. (2006) Evidence for a gene influencing heart rate on chromosome 5p13-14 in a meta-analysis of genome-wide scans from the NHLBI Family Blood Pressure Program. *BMC Med Genet*, 7, 17.
14. Dalageorgou, C., Ge, D., Jamshidi, Y., Nolte, I.M., Riese, H., Savelieva, I., Carter, N.D., Spector, T.D. and Snieder, H. (2008) Heritability of QT interval: how much is explained by genes for resting heart rate? *J Cardiovasc Electrophysiol*, 19, 386-391.
15. De Geus, E.J., Kupper, N., Boomsma, D.I. and Snieder, H. (2007) Bivariate genetic modeling of cardiovascular stress reactivity: does stress uncover genetic variance? *Psychosom Med*, 69, 356-364.
16. Wilk, J.B., Myers, R.H., Zhang, Y., Lewis, C.E., Atwood, L., Hopkins, P.N. and Ellison, R.C. (2002) Evidence for a gene influencing heart rate on chromosome 4 among hypertensives. *Hum Genet*, 111, 207-213.
17. Ranade, K., Jorgenson, E., Sheu, W.H., Pei, D., Hsiung, C.A., Chiang, F.T., Chen, Y.D., Pratt, R., Olshen, R.A., Curb, D. *et al.* (2002) A polymorphism in the beta1 adrenergic receptor is associated with resting heart rate. *Am J Hum Genet*, 70, 935-942.
18. An, P., Rice, T., Rankinen, T., Leon, A.S., Skinner, J.S., Wilmore, J.H., Bouchard, C. and Rao, D.C. (2006) Genome-wide scan to identify quantitative trait loci for baseline resting heart rate and its response to endurance exercise training: the HERITAGE Family Study. *Int J Sports Med*, 27, 31-36.
19. Wilk, J.B., Myers, R.H., Pankow, J.S., Hunt, S.C., Leppert, M.F., Freedman, B.I., Province, M.A. and Ellison, R.C. (2006) Adrenergic receptor polymorphisms associated with resting heart rate: the HyperGEN Study. *Ann Hum Genet*, 70, 566-573.
20. Wilton, S.B., Anderson, T.J., Parboosingh, J., Bridge, P.J., Exner, D.V., Forrest, D. and Duff, H.J. (2008) Polymorphisms in multiple genes are associated with resting heart rate in a stepwise allele-dependent manner. *Heart Rhythm*, 5, 694-700.
21. Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M. *et al.* (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet*, 41, 527-534.
22. Holm, H., Gudbjartsson, D.F., Arnar, D.O., Thorleifsson, G., Thorgeirsson, G., Stefansdottir, H., Gudjonsson, S.A., Jonasdottir, A., Mathiesen, E.B., Njolstad, I. *et al.* (2010) Several common variants modulate heart rate, PR interval and QRS duration. *Nat Genet*, 42, 117-122.
23. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., Boerwinkle, E. *et al.* (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts, 2, 73-80.
24. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22, 819-829.
25. Hofman, A., Breteler, M.M., van Duijn, C.M., Janssen, H.L., Krestin, G.P., Kuipers, E.J., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. *et al.* (2009) The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol*, 24, 553-572.
26. Little, J., Higgins, J.P., Ioannidis, J.P., Moher, D., Gagnon, F., von Elm, E., Khoury, M.J., Cohen, B., Davey-Smith, G., Grimshaw, J. *et al.* (2009) Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Eur J Epidemiol*, 24, 37-55.
27. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*, 17, R122-128.

28. Devlin, B., Roeder, K. and Wasserman, L. (2001) Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol*, 60, 155-166.
29. Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32, 381-385.
30. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460, 748-752.
31. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559-575.
32. Chandler, N.J., Greener, I.D., Tellez, J.O., Inada, S., Musa, H., Molenaar, P., Difrancesco, D., Baruscotti, M., Longhi, R., Anderson, R.H. *et al.* (2009) Molecular architecture of the human sinus node: insights into the function of the cardiac pacemaker. *Circulation*, 119, 1562-1575.
33. Rohr, S. (2004) Role of gap junctions in the propagation of the cardiac action potential. Cardiovasc Res, 62, 309-322.
34. Dasgupta, C., Martinez, A.M., Zuppan, C.W., Shah, M.M., Bailey, L.L. and Fletcher, W.H. (2001) Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). *Mutat Res*, 479, 173-186.
35. Niimura, H., Patton, K.K., McKenna, W.J., Soults, J., Maron, B.J., Seidman, J.G. and Seidman, C.E. (2002) Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation*, 105, 446-451.
36. Ching, Y.H., Ghosh, T.K., Cross, S.J., Packham, E.A., Honeyman, L., Loughna, S., Robinson, T.E., Dearlove, A.M., Ribas, G., Bonser, A.J. *et al.* (2005) Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet*, 37, 423-428.
37. Carniel, E., Taylor, M.R., Sinagra, G., Di Lenarda, A., Ku, L., Fain, P.R., Boucek, M.M., Cavanaugh, J., Miocic, S., Slavov, D. *et al.* (2005) Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation*, 112, 54-59.
38. Callis, T.E., Pandya, K., Seok, H.Y., Tang, R.H., Tatsuguchi, M., Huang, Z.P., Chen, J.F., Deng, Z., Gunn, B., Shumate, J. *et al.* (2009) MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest*, 119, 2772-2786.
39. van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J. and Olson, E.N. (2007) Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science*, 316, 575-579.
40. Wunderle, V.M., Critcher, R., Ashworth, A. and Goodfellow, P.N. (1996) Cloning and characterization of SOX5, a new member of the human SOX gene family. *Genomics*, 36, 354-358.
41. Lefebvre, V. (2009) The SOXD transcription factors – SOX5, SOX6, and SOX13 – are key cell fate modulators. *The international journal of biochemistry & cell biology*, 42, 429-432.
42. Pfeufer, A., van Noord, C., Marciante, K.D., Arking, D.E., Larson, M.G., Smith, A.V., Tarasov, K.V., Muller, M., Sotoodehnia, N., Sinner, M.F. *et al.* Genome-wide association study of PR interval. *Nat Genet*, 42, 153-159.
43. Pfeufer, A., Sanna, S., Arking, D.E., Muller, M., Gateva, V., Fuchsberger, C., Ehret, G.B., Orru, M., Pattaro, C., Kottgen, A. *et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet*, 41, 407-414.
44. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M., de Bakker, P.I., Yin, X., Estrada, K., Bis, J.C., Marciante, K., Rivadeneira, F., Noseworthy, P.A. *et al.* (2009) Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet*, 41, 399-406.
45. Nolte, I.M., Wallace, C., Newhouse, S.J., Waggott, D., Fu, J., Soranzo, N., Gwilliam, R., Deloukas, P., Savelieva, I., Zheng, D. *et al.* (2009) Common genetic variation near the phospholamban gene is associated with cardiac repolarisation: meta-analysis of three genome-wide association studies. *PLoS One*, 4, e6138.
46. Vasan, R.S., Glazer, N.L., Felix, J.F., Lieb, W., Wild, P.S., Felix, S.B., Watzinger, N., Larson, M.G., Smith, N.L., Dehghan, A. *et al.* (2009) Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *Jama*, 302, 168-178.
47. Vinogradova, T.M., Sirenko, S., Lyashkov, A.E., Younes, A., Li, Y., Zhu, W., Yang, D., Ruknudin, A.M., Spurgeon, H. and Lakatta, E.G. (2008) Constitutive phosphodiesterase activity restricts spontaneous beating rate of cardiac pacemaker cells by suppressing local Ca2+ releases. *Circ Res*, 102, 761-769.
48. Vinogradova, T.M., Lyashkov, A.E., Zhu, W., Ruknudin, A.M., Sirenko, S., Yang, D., Deo, S., Barlow, M., Johnson, S., Caffrey, J.L. *et al.* (2006) High basal protein kinase A-dependent phosphorylation drives rhythmic internal Ca^{2+} store oscillations and spontaneous beating of cardiac pacemaker cells. *Circ Res*, 98, 505-514.

49. Nishimura, M., Suzuki, S., Satoh, T. and Naito, S. (2009) Tissue-specific mRNA expression profiles of human solute carrier 35 transporters. *Drug metabolism and pharmacokinetics*, 24, 91-99.
50. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M. and Boerwinkle, E. (2008) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circulation Cardiovascular Genetics*, in press.
51. Kang, S.H., Kim, G.R., Seong, M., Baek, S.H., Seol, J.H., Bang, O.S., Ovaa, H., Tatsumi, K., Komatsu, M., Tanaka, K. *et al.* (2007) Two novel ubiquitin-fold modifier 1 (Ufm1)-specific proteases, UfSP1 and UfSP2. *J Biol Chem*, 282, 5256-5262.
52. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T. *et al.* (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*, 41, 47-55.
53. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*, 41, 56-65.
54. Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M. *et al.* (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet*, 41, 35-46.
55. Gieger, C., Geistlinger, L., Altmaier, E., Hrabe de Angelis, M., Kronenberg, F., Meitinger, T., Mewes, H.W., Wichmann, H.E., Weinberger, K.M., Adamski, J. *et al.* (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*, 4, e1000282.
56. Schaeffer, L., Gohlke, H., Muller, M., Heid, I.M., Palmer, L.J., Kompauer, I., Demmelmair, H., Illig, T., Koletzko, B. and Heinrich, J. (2006) Common genetic variants of the FADS1 FADS2 gene cluster and their reconstructed haplotypes are associated with the fatty acid composition in phospholipids. *Hum Mol Genet*, 15, 1745-1756.
57. Dettbarn, C. and Palade, P. (1993) Arachidonic acid-induced Ca^{2+} release from isolated sarcoplasmic reticulum. *Biochem Pharmacol*, 45, 1301-1309.
58. Bang, M.L., Centner, T., Fornoff, F., Geach, A.J., Gotthardt, M., McNabb, M., Witt, C.C., Labeit, D., Gregorio, C.C., Granzier, H. *et al.* (2001) The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circ Res*, 89, 1065-1072.
59. Itoh-Satoh, M., Hayashi, T., Nishi, H., Koga, Y., Arimura, T., Koyanagi, T., Takahashi, M., Hohda, S., Ueda, K., Nouchi, T. *et al.* (2002) Titin mutations as the molecular basis for dilated cardiomyopathy. *Biochem Biophys Res Commun*, 291, 385-393.
60. Kubo, Y., Baldwin, T.J., Jan, Y.N. and Jan, L.Y. (1993) Primary structure and functional expression of a mouse inward rectifier potassium channel. *Nature*, 362, 127-133.
61. Fleischmann, B.K., Duan, Y., Fan, Y., Schoneberg, T., Ehlich, A., Lenka, N., Viatchenko-Karpinski, S., Pott, L., Hescheler, J. and Fakler, B. (2004) Differential subunit composition of the G protein-activated inward-rectifier potassium channel during cardiac development. *J Clin Invest*, 114, 994-1001.
62. Rosskopf, D., Nikula, C., Manthey, I., Joisten, M., Frey, U., Kohnen, S. and Siffert, W. (2003) The human G protein beta4 subunit: gene structure, expression, Ggamma and effector interaction. *FEBS Lett*, 544, 27-32.
63. Ruiz-Velasco, V., Ikeda, S.R. and Puhl, H.L. (2002) Cloning, tissue distribution, and functional expression of the human G protein beta 4-subunit. *Physiol Genomics*, 8, 41-50.
64. Splawski, I., Timothy, K.W., Decher, N., Kumar, P., Sachse, F.B., Beggs, A.H., Sanguinetti, M.C. and Keating, M.T. (2005) Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc Natl Acad Sci USA*, 102, 8089-8096; discussion 8086-8088.
65. Avkiran, M., Rowland, A.J., Cuello, F. and Haworth, R.S. (2008) Protein kinase d in the cardiovascular system: emerging roles in health and disease. *Circ Res*, 102, 157-163.

3.2 Supplement

SUPPLEMENTARY METHODS

Cohort descriptions

For the description of the CHARGE consortium and participating cohorts we refer to the CHARGE design paper[1]. In short, the AGES – Reykjavik Study represents a sample from the population-based Reykjavik Study, originally comprising 31,795 individuals. Between 2002 and 2006 the AGES – Reykjavik Study 5,764 survivors of the original cohort were re-examined[2]. The ARIC study is a population-based prospective cohort study, sponsored by the National Heart, Lung, and Blood Institute, and included 15,792 individuals aged 45 to 64 years at baseline from 4 US communities[3]. CHS is a National Heart, Lung, and Blood Institute sponsored population-based cohort study in adults over 65 years of age conducted at 4 field centers. Originally CHS 5,201 predominantly white individuals were enrolled in 1989-1990[4]. The FHS started in 1948 with the recruitment of an original cohort of 5,209 men and women aged 28 to 62 years of age[5]. In 1971, children and spouses of children of the original cohort were enrolled[6]. The third generation cohort enrollment started in 2002 and comprises 4,095 children of offspring cohort participants[7]. The RS is a prospective population-based cohort study comprising 7,983 subjects aged 55 years or older (RS-I)[8]. In 2000-2001, an additional 3,011 individuals aged 55 years or older (RS-II) were recruited[9].

In addition to cohorts in the CHARGE consortium, additional cohorts participated and are discussed below. EUROSPAN consists of five isolated population cohorts. For the current effort we used ECG measurements that were available in three of the participating studies, namely ERF, MICROS and ORCADES. The studied populations originate from Rucphen, (The Netherlands), South Tyrol (Italy) and the Orkney Islands (Scotland, UK), respectively. The Erasmus Rucphen Study (ERF) is derived from a recent genetic isolate in the southwest Netherlands. This population was founded in the middle of the 18th century by approximately 150 individuals and was isolated until the last few decades. Twenty couples living in the region in the 19th century were chosen. These couples parented a minimum of 6 children, each of whom was baptized between 1880 and 1900 in the community church. All living descendants of these pairs (as well as their spouses), ascertained on the basis of municipal and baptismal records, were traced and invited to participate. The MICROS study is part of the genomic health care program 'GenNova' and was carried out in three villages of the Val Venosta on the populations of Stelvio, Vallelunga and Martello. This study was an extensive survey carried out in South Tyrol (Italy) in the period

2001-2003. An extensive description of the study is available elsewhere[10]. Briefly, study participants were volunteers from three isolated villages located in the Italian Alps, bordering with Austria and Switzerland. Due to geographical, historical and political reasons, the entire region experienced prolonged isolation from surrounding populations. The Orkney Complex Disease Study (ORCADES) is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney. Genetic diversity in this population is decreased compared to Mainland Scotland, consistent with the high levels of endogamy historically. Data for participants aged 18-88 years, from a subgroup of ten islands, were used for this analysis. The KORA Study is a series of population-based epidemiological surveys of persons living in or near the city of Augsburg, Germany. All survey participants are residents of German nationality identified through the registration office and between 25 and 74 years old at the time of enrollment. Survey S3 was conducted between 1994 and 1995 and survey S4 between 1999 and 2001. KORA F3, a follow up examination at 10 years of follow up after S3, occurred in 2004 and 2005[11]. The Netherlands Study of Depression and Anxiety (NESDA) is an ongoing 8-year longitudinal cohort study to examine (predictors of) the long-term course of depression and anxiety disorders. The rationales, methods and recruitment strategy have been described elsewhere[12]. Subjects were recruited in various settings (community, general practices, mental health organizations) and include subjects with and without depression and anxiety disorders. The SardiNIA study is a population based study, that recruited and phenotyped 6,148 individuals, male and female, ages 14-102 years, from a cluster of four towns in the Lanusei Valley of Sardinia[13]. For the GWA scans a total of 4,305 related individuals were examined[14]. Genotyped individuals had four Sardinian grandparents and were selected without regard to their phenotypes. The Study of Health in Pomerania (SHIP) is a longitudinal population-based cohort study in West Pomerania, a region in the northeast of Germany with a total population of 212,157 inhabitants[15]. A two-stage cluster sampling method adopted from the WHO MONICA Project Augsburg, Germany yielded 12 five-year age strata for both genders, each including 292 individuals. For the baseline cohort, a sample of 6,267 eligible subjects aged 20 to 79 years was drawn from population registries where all German citizens are registered. The final study population compromised 4,310 subjects (response proportions 69%). The Twins UK Registry comprises unselected, mostly female volunteers ascertained from the general population through national media campaigns in the UK[16]. Means and ranges of quantitative phenotypes in Twins UK were similar to an age-matched singleton sample from the general population[17]. Zygosity was determined by standardized questionnaire and confirmed by DNA fingerprinting. Written informed consent was obtained from all participants before they entered the studies, which were approved by the local research ethics committee.

Phenotype measurement

In ARIC, RR interval was measured automatically from 12-lead ECGs performed at baseline. The study ECGs were recorded using MAC PC ECG machines (Marquette Electronics, Milwaukee, Wisconsin) in all four clinical centers. All ECGs were visually inspected for technical errors and inadequate quality. ECGs were initially processed in a central laboratory at the EPICORE Center (University of Alberta, Edmonton, Alberta, Canada) and during later phases of the study at the

EPICARE Center (Wake Forest University, Winston-Salem, North Carolina). Initial ECG processing was done by the Dalhousie ECG program, and processing was later repeated with the 2001 version of the GE Marquette 12 SL program (GE Marquette, Milwaukee, Wisconsin).

In AGES, RR interval duration was automatically measured from 12-lead electrocardiograms using the Marquette 12 SL analysis program (General Electric Marquette Medical Division, Milwaukee, Wisconsin, USA).

In CHS, electrocardiograms were recorded using MAC PC ECG machines (Marquette Electronics, Milwaukee, Wisconsin) in all clinical centers. ECGs were initially processed in a central laboratory at the EPICORE Center (University of Alberta, Edmonton, Alberta, Canada) and during later phases of the study, at the EPICARE Center (Wake Forest University, Winston-Salem, North Carolina). All ECGs were visually inspected for technical errors and inadequate quality. RR interval was calculated from heart rate using the baseline ECG for eligible subjects. Initial ECG processing was done by the Dalhousie ECG program, and processing was later repeated with the 2001 version of the GE Marquette 12 SL program (GE Marquette, Milwaukee, Wisconsin).

In FHS, paper electrocardiograms were scanned and digital caliper measurements were made using proprietary software (eResearchTechnology, generations 1 and 2) or using Rigel 1.7.2. (AMPS, LLC, New York, NY, USA, generation 3). The RR interval between QRS complexes from consecutive beats in sinus rhythm was measured, after excluding all premature atrial or ventricular beats. The RR interval trait examined was the average RR interval from up to 4 cardiac cycles.

In KORA F3 and S4 12-lead resting electrocardiograms were recorded with digital recording systems (KORA F3: Mortara Portrait, Mortara Inc., Milwaukee, USA; KORA S4: Hörmann Bioset 9000, Hörmann Medizinelektronik, Germany). The Mortara Portrait determines RR interval by the proprietary XL-ECG algorithm which has not been published but has shown to be in good accordance with other published electrocardiogram measurement algorithms[18]; RR interval in KORA S4 (Hörmann Bioset) was determined using the Hannover EKG analysis software (HES-Version 3.22-12) by computerized analysis of all leads and all cycles of a 10 second interval as described earlier[19,20]. In the Hannover algorithm RR intervals are taken as the intervals between the reference points detected in adjacent QRS complexes. The mean RR interval was computed, after exclusion of RR intervals that immediately precede and follow any premature ventricular complex. Only ECGs classified as 'appropriate technical quality' according to visual inspection were used. The ECG examinations in both studies were performed according to a standard protocol, after ten minutes resting in supine position.

In MICROS 12-lead resting ECGs were recorded using a digital recording system (Mortara Portrait, Mortara, Milwaukee, WI, USA). The Mortara Portrait determines RR interval by the proprietary XL-ECG algorithm. In brief, computerized analysis of an averaged cycle computed from all leads and all cycles of the 10 second recording after exclusion of ectopic beats was performed. RR intervals were determined as the intervals between the reference points detected in adjacent QRS complexes. The median RR interval was computed, after exclusion of RR intervals that immediately precede and follow any premature ventricular complex.

In NESDA, all respondents received an average physiological recording of 100 minutes, per-

formed with the 'Vrije Universiteit Ambulatory Monitoring System' [21]. The VU-AMS is a light-weight ambulatory device that records the electrocardiogram (ECG) from three electrodes in lead II configuration and changes in thorax impedance (dZ) from four electrodes placed at chest and back of the subjects. From the R-waves in the ECG signal the inter beat intervals (IBI) were computed online and the IBI time series was visually checked for ectopic beats or missed R-waves offline. For the GWA mean heart rate was computed across valid IBIS across ten minutes of quiet supine rest at the start of the recording period.

In ORCADES digital 10 second ECGS were taken after 10 minutes supine rest, using a PC link with RR interval calculated using CardioView software (NUMED cardiac diagnostics, Sheffield, UK).

In the RS-I and RS-II, electrocardiograms were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the RR intervals were made using the Modular ECG Analysis System (MEANS)[22]. The MEANS program locates the QRS complexes and determines a stable reference point in each complex. The QRS detector of MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the QRS complexes among the other parts of the signal. RR intervals are taken as the intervals between the reference points in adjacent QRS complexes. The median RR interval was computed, after exclusion of RR intervals that immediately precede and follow any premature ventricular complex. SHIP and ERF both use the same ECG analysis methods as the Rotterdam Study.

In SardiNIA the RR interval was calculated from heart rate that was measured from the electrocardiogram (measured on Cardiette 600 machines) during physical examination.

In TwinsUK, ECG data were available on 3,043 individuals before exlcusions. Two thousand seven hundred twenty six had automated measurements of the RR interval by the Cardiofax ECG-9020K (Nihon Kohden UK Ltd., Middlesex, UK) and 317 were scored manually using a high-resolution digitizing board (GTCO CalComp Peripherals, USA). The dataset for analyses consequently included 2,727 individuals, of which 1980 were DZ twins (i.e. 990 pairs) and 747 singletons. These singletons included 474 MZ twins of which the mean RR interval of both twins was used to optimize information.

Supplementary Table 1 *Summary methods per cohort.*

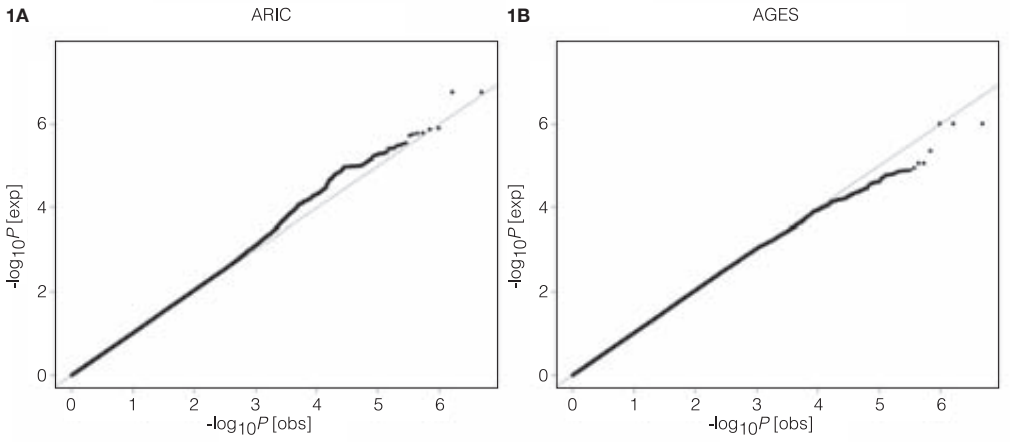
	Array	Calling algorithm	SNP call rate filter	HWE p-value filter	MAF	Imputation software	NCBI Build	Statistical Analysis	Number of SNPs	Lambda
AGES	Illumina 370CNV	BeadStudio	<97%	<10-6	<1%	Mach v1.0.16	Build 36	ProbABEL[23]	2,532,729	1.012
ARIC	Affymetrix 6.0	Birdseed	<95%	<10-5	<1%	Mach v1.0.16[24]	Build 35	ProbABEL	2,557,232	1.017
CHS	Illumina 370CNV	BeadStudio	<97%	<10-5	<1%	BimBam v0.99[25]	Build 36	R[26]	2,333,043	1.043
ERF	Illumina 300, Affymetrix 250K	Beadstudio	<98%	<10 6	<1%	Mach v1.0.16	Build 36	ProbABEL	2,543,887	1.027
FHS	Affymetrix 500K + gene-centric 50K MIP	BRLMM[27]	<97%	<10-6	<1%	Mach v1.0.15	Build 36	Kinship package in R[28]	2,540,128	1.016
KORA F3	Affymetrix 500K	BRLMM	<93%	<10-5	<5%	Mach v1.0.10	Build 35	Mach2QTL[24]	2,557,252	0.997
KORA S4	Affymetrix 6.0	Birdseed	<93%	<10-5	<5%	Mach v1.0.16	Build 36	Mach2QTL	2,543,887	0.995
MICROS	Illumina 300	Beadstudio	<98%	<10 6	<1%	Mach v1.0.16	Build 36	ProbABEL	2,543,887	0.979
NESDA	Affymetrix 500K	PERLEGEN	<95%	none	<1%	IMPUTE v0.5.0[29]	Build 36	SNPTEST	2,081,096	1.012
ORCADES	Illumina 300	Beadstudio	<97%	<10 5	<1%	Mach v1.0.15	Build 36	ProbABEL	2,543,887	1.055
RS-I	Illumina 550K	BeadStudio	<98%	<10-6	<1%	Mach v1.0.15	Build 36	Mach2QTL	2,543,887	1.016
RS-II	Illumina 550K	Genome-Studio	<98%	<10-6	<1%	Mach v1.0.16	Build 36	Mach2QTL	2,543,887	1.015
SardiNIA	Affymetrix 10K, Affymetrix 500K	BRLMM	<90%	<10-3 (10k), <10-6 (500k)	<5%	Mach	Build 35	MERLIN[30]	2,252,228	1.027
SHIP	Affymetrix 6.0	Birdseed	none	none	none	IMPUTE v0.5.0	Build 36	SNPTEST	2,748,910	1.007
TwinsUK	Illumina 300K Duo, Illumina 300, Illumina 550K, Illumina 610K	Beadstudio	<95%	<10 4	<1%	IMPUTE v0.5.0	Build 36	SNPTEST	2,237,157	1.003

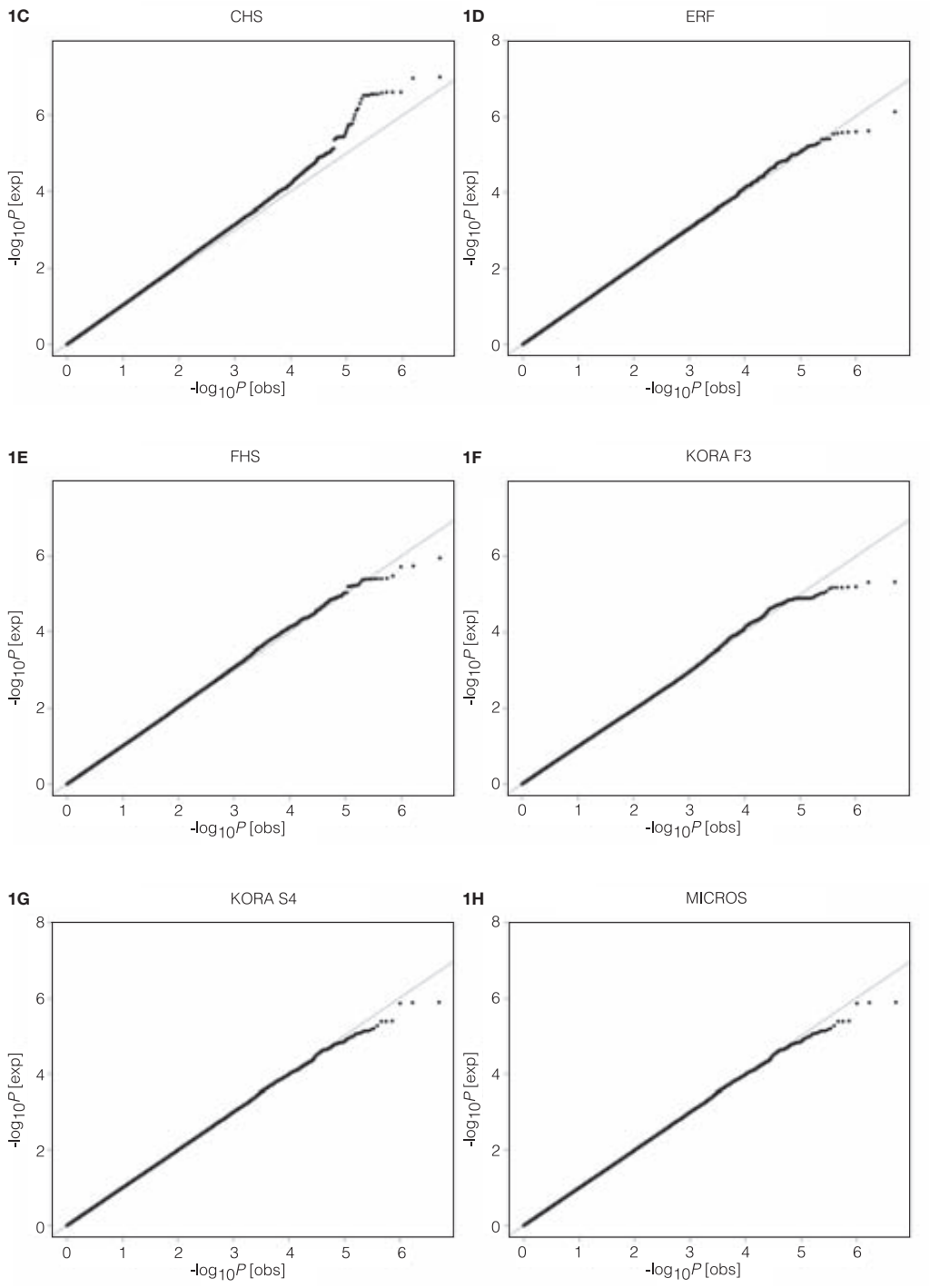
Supplementary Table 2 *rs-numbers and annotation of quality controlled genotyped index.*

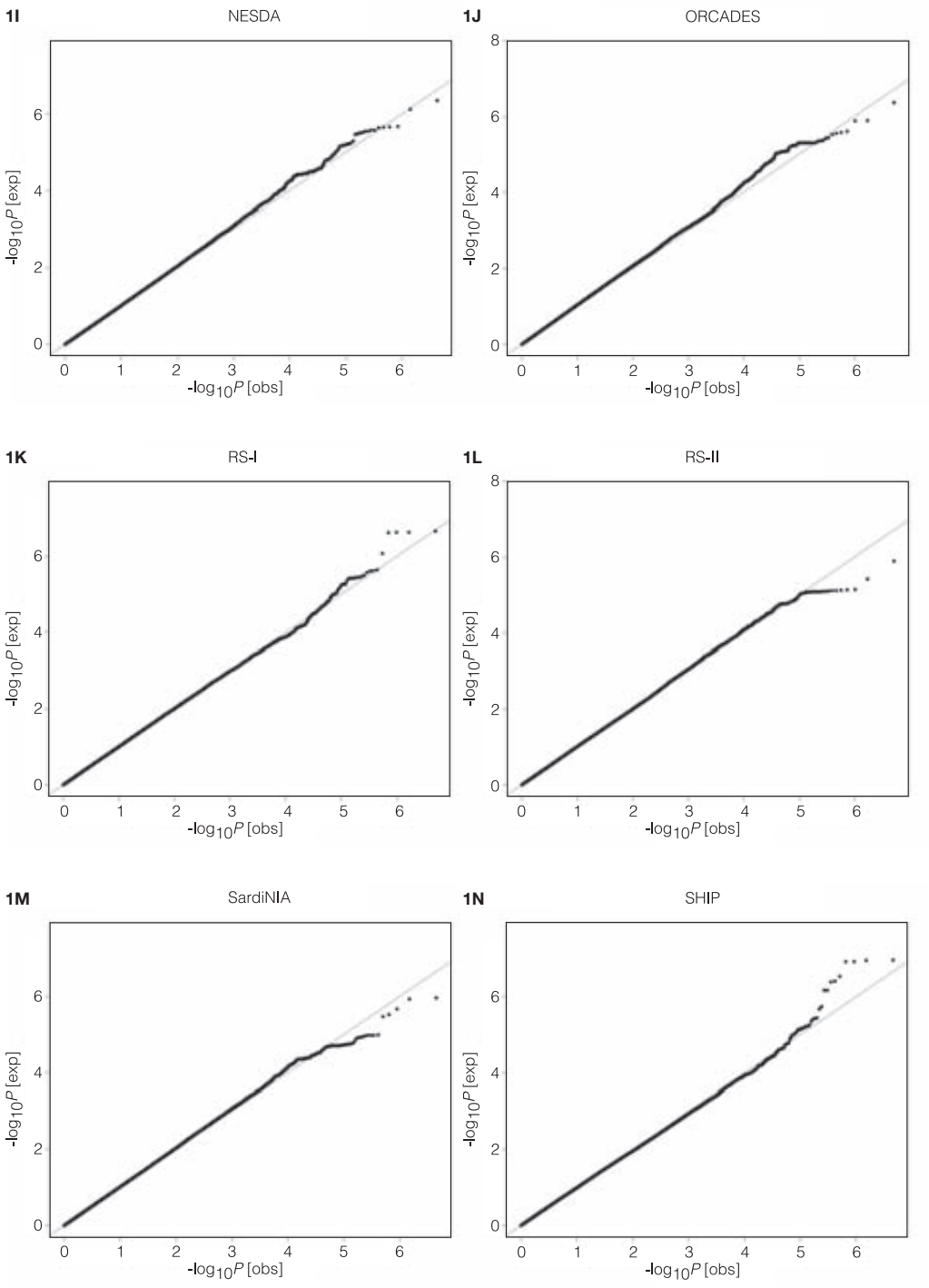
SNP	CHR	POS	Nearest Gene annotation
rs272564	1p34	44,784,860	*c1orf164*
rs11579530	1q32	206,197,041	Intergenic, near ***CD34***
rs16845015	2q21	141,430,306	*LRP1B*
rs1873164	2q31	287,738,258	Upstream *CCDC141*
rs4894803	3q26	173,282,950	*FNDC3B*
rs7612445	3q26	180,655,287	Upstream *GNB4*
rs281868	6q22	118,684,736	***SLC35F1***
rs2269579	6q22	122,198,112	Intergenic, ***GJA1***
rs314370	7q22	100,291,144	***SLC12A9***
rs17099385	10q25	121,680,302	Intergenic, *SEC23IP*
rs174546	11q12	61,326,406	3' UTR ***FADS1*** / ***FADS3***
rs2238018	12p13	2,048,922	*CACNA1C*
rs4246224	12p12	24,675,406	Intergenic, ***SOX5*** / ***BCAT1***
rs2200155	12p11	33,626,202	Intergenic, *SYT10*
rs2887596	12q14	76,678,888	Intergenic, *NAV3*
rs4981691	14q11	28,738,258	Intergenic, *PRKD1*
rs365990	14q12	22,931,651	Non synonymous coding ***MYH6***
rs223116	14q12	23,046,850	Intergenic, ***MYH7*** / *NGDN* / *ZFHX2*
rs2883661	16p12	19,899,685	Intergenic, *GPRC5B*
rs1364215	16q22	63,455,370	Intergenic, *CDH11*

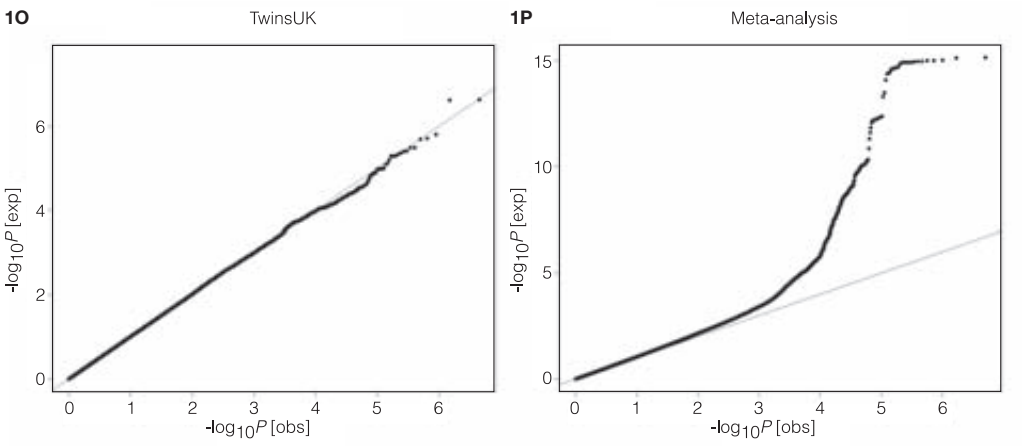
SNPs in multi-variant polygenic model that in aggregate resulted in the maximal explained variance.
CHR=chromosome, POS=position
Genome-wide loci from main analysis are printed in bold

Supplementary Figure 1 *Quantile – Quantile plots of observed/expected $-\log_{10}(P)$ per cohort and meta-analysis.*

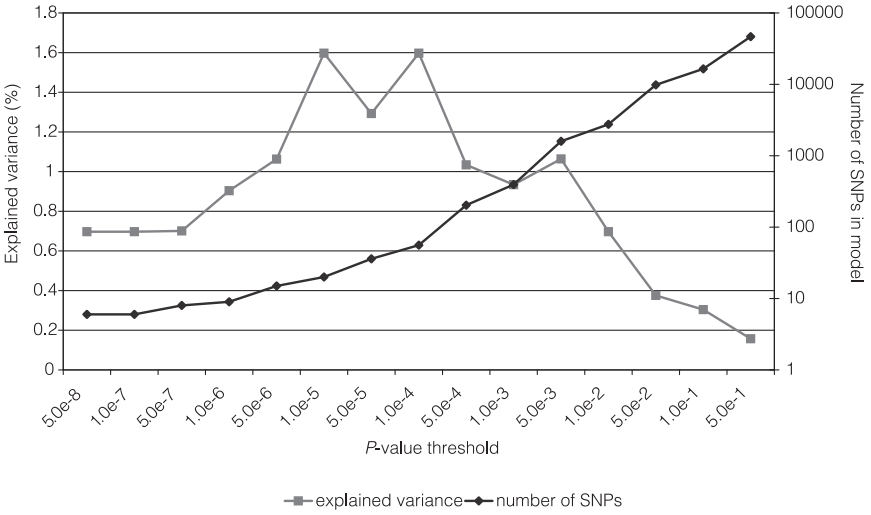



1C
CHS
1D
ERF
1E
FHS
1F
KORA F3
1G
KORA S4
1H
MICROS
-log10P [exp]
-log10P [obs]


1I
NESDA
1J
ORCADES
1K
RS-I
1L
RS-II
1M
SardiNIA
1N
SHIP
-log10P [exp]
-log10P [obs]



Supplementary Figure 2 *Graphical presentation of the multiple variant modeling approach estimating explained variance in the RS-II sample.*



X-axis: *P*-value thresholds below which all independent signals meeting this threshold were included in the genotype score included in the regression model to estimate the explained variance in the RS-II sample. Left Y-axis: explained variance in the RS-II sample for any given score at the different *P*-value thresholds. Right Y-axis: number of independent SNPs included in the model for a given *P*-value threshold.

REFERENCES

1. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., Boerwinkle, E. *et al.* (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts, 2, 73-80.
2. Harris, T.B., Launer, L.J., Eiriksdottir, G., Kjartansson, O., Jonsson, P.V., Sigurdsson, G., Thorgeirsson, G., Aspelund, T., Garcia, M.E., Cotch, M.F. *et al.* (2007) Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *American journal of epidemiology*, 165, 1076-1087.
3. (1989) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*, 129, 687-702.
4. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A. *et al.* (1991) The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*, 1, 263-276.
5. Dawber, T.R., Meadors, G.F. and Moore, F.E., Jr. (1951) Epidemiological approaches to heart disease: the Framingham Study. *American journal of public health and the nation's health*, 41, 279-281.
6. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. and Castelli, W.P. (1975) The Framingham Offspring Study. Design and preliminary data. *Preventive medicine*, 4, 518-525.
7. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Sr., Fox, C.S., Larson, M.G., Murabito, J.M. *et al.* (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*, 165, 1328-1335.
8. Hofman, A., Grobbee, D.E., de Jong, P.T. and van den Ouweland, F.A. (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*, 7, 403-422.
9. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22, 819-829.
10. Pattaro, C., Marroni, F., Riegler, A., Mascalzoni, D., Pichler, I., Volpato, C.B., Dal Cero, U., De Grandi, A., Egger, C., Eisendle, A. *et al.* (2007) The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. *BMC Med Genet*, 8, 29.
11. Wichmann, H.E., Gieger, C. and Illig, T. (2005) KORA-gen-resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, 67 Suppl 1, S26-30.
12. Penninx, B.W., Beekman, A.T., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W., Assendelft, W.J. *et al.* (2008) The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International journal of methods in psychiatric research*, 17, 121-140.
13. Pilia, G., Chen, W.M., Scuteri, A., Orru, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P. *et al.* (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet*, 2, e132.
14. Scuteri, A., Sanna, S., Chen, W.M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orru, M., Usala, G. *et al.* (2007) Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet*, 3, e115.
15. John, U., Greiner, B., Hensel, E., Ludemann, J., Piek, M., Sauer, S., Adam, C., Born, G., Alte, D., Greiser, E. *et al.* (2001) Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Praventivmed*, 46, 186-194.
16. Spector, T.D. and Williams, F.M. (2006) The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet*, 9, 899-906.
17. Andrew, T., Hart, D.J., Snieder, H., de Lange, M., Spector, T.D. and MacGregor, A.J. (2001) Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res*, 4, 464-477.
18. Michelucci, A., Mortara, D., Lazzeri, C., Barletta, G., Capalbo, A., Badia, T., Del Bene, R., Bano, C., Gensini, G.F. and Franchi, F. (2002) Simultaneous assessment of electrocardiographic parameters for risk stratification: validation in healthy subjects. *Ital Heart J*, 3, 308-317.
19. Pfeufer, A., Jalilzadeh, S., Perz, S., Mueller, J.C., Hinterseer, M., Illig, T., Akyol, M., Huth, C., Schopfer-Wendels, A., Kuch, B. *et al.* (2005) Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. *Circ Res*, 96, 693-701.
20. Zywietz, C., Borovsky, D., Gotsch, G. and Joseph, G. (1990) Methodology of ECG interpretation in the Hannover program. *Methods Inf Med*, 29, 375-385.

21. Licht, C.M., de Geus, E.J., Zitman, F.G., Hoogendijk, W.J., van Dyck, R. and Penninx, B.W. (2008) Association between major depressive disorder and heart rate variability in the Netherlands Study of Depression and Anxiety (NESDA). *Arch Gen Psychiatry*, 65, 1358-1367.
22. van Bemmel, J.H., Kors, J.A. and van Herpen, G. (1990) Methodology of the modular ECG analysis system MEANS. *Methods Inf Med*, 29, 346-353.
23. Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23, 1294-1296.
24. Li, Y. and Abecasis, G.R. (2006) Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet*, S79, 2290.
25. Guan, Y. and Stephens, M. (2008) Practical issues in imputation-based association mapping. *PLoS Genet*, 4, e1000279.
26. R Developmental Core Team (2007) A language and environment for statistical computing. R Foundation for Statistical Computing.
27. Rabbee, N. and Speed, T.P. (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics*, 22, 7-12.
28. Therneau, T. (2008) kinship: mixed effects Cox models, sparse matrices, and modeling data from large pedigrees. R package version 1.1.0-19 ed.
29. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 39, 906-913.
30. Abecasis, G.R., Cherny, S.S., Cookson, W.O. and Cardon, L.R. (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*, 30, 97-101.

3.3 Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction

Nona Sotoodehnia*, Aaron Isaacs*, Paul I.W. de Bakker*, Marcus Dörr*, Christopher Newton-Cheh*, Ilja M. Nolte*, Pim van der Harst*, Martina Müller*, Mark Eijgelsheim*, Alvaro Alonso*, Andrew A. Hicks*, Sandosh Padmanabhan*, Caroline Hayward*, Albert Vernon Smith*, Ozren Polasek*, Steven Giovannone*, Jingyuan Fu*, Jared W. Magnani, Kristin D. Marciante, Arne Pfeufer, Sina A. Gharib, Alexander Teumer, Man Li, Joshua C. Bis, Fernando Rivadeneira, Thor Aspelund, Anna Köttgen, Toby Johnson, Kenneth Rice, Mark P.S. Sie, Amanda Ying Wang, Norman Klopp, Christian Fuchsberger, Sarah H. Wild, Irene Mateo Leach, Karol Estrada, Uwe Völker, Alan F. Wright, Folkert W. Asselbergs, Jiaxiang Qu, Aravinda Chakravarti, Moritz F. Sinner, Jan A. Kors, Astrid Petersmann, Tamara B. Harris, Elsayed Z. Soliman, Patricia B. Munroe, Bruce M. Psaty, Ben A. Oostra, L. Adrienne Cupples , Siegfried Perz, Rudolf A. de Boer, André G. Uitterlinden, Henry Völzke, Timothy D. Spector, Fang-Yu Liu, Eric Boerwinkle, Anna F. Dominiczak, Jerome I. Rotter, Gé van Herpen, Daniel Levy, H.-Erich Wichmann, Wiek H. van Gilst, Jacqueline C.M. Witteman, Heyo K. Kroemer, W.H. Linda Kao, Susan R. Heckbert, Thomas Meitinger, Albert Hofman, Harry Campbell, Aaron R. Folsom, Dirk J. van Veldhuisen, Christine Schwienbacher, Christopher J. O'Donnell, Claudia Beu Volpato, Mark J. Caulfield, John M. Connell, Lenore Launer, Xiaowen Lu, Lude Franke, Rudolf S.N. Fehrmann, Gerard te Meerman, Harry J.M. Groen, Rinse K. Weersma, Leonard H. van den Berg, Cisca Wijmenga, Roel A. Ophoff, Gerjan Navis, Igor Rudan, Harold Snieder*, James F. Wilson*, Peter P. Pramstaller*, David S. Siscovick*, Thomas J. Wang*, Vilmundur Gudnason*, Cornelia M. van Duijn*, Stephan B. Felix*, Glenn I. Fishman*, Yalda Jamshidi*, Bruno H Ch Stricker*, Nilesh J. Samani*, Stefan Kääb*, Dan E. Arking*

Nat Genet. 2010 Dec; 42(12): 1068-1076

* denote equal author contributions

ABSTRACT

The QRS interval, from the beginning of the Q wave to the end of the S wave on an electrocardiogram, reflects ventricular depolarization and conduction time and is a risk factor for mortality, sudden death and heart failure. We performed a genome-wide association meta-analysis in 40,407 individuals of European descent from 14 studies, with further genotyping in 7,170 additional Europeans, and we identified 22 loci associated with QRS duration ($P < 5\times10^{-8}$). These loci map in or near genes in pathways with established roles in ventricular conduction such as sodium channels, transcription factors and calcium-handling proteins, but also point to previously unidentified biologic processes, such as kinase inhibitors and genes related to tumorigenesis. We demonstrate that *SCN10A*, a candidate gene at the most significantly associated locus in this study, is expressed in the mouse ventricular conduction system, and treatment with a selective *SCN10A* blocker prolongs QRS duration. These findings extend our current knowledge of ventricular depolarization and conduction.

INTRODUCTION

The electrocardiographic QRS interval reflects ventricular depolarization, and its duration is a function of electrophysiological properties within the His-Purkinje system and the ventricular myocardium. A diseased ventricular conduction system can lead to life-threatening bradyarrhythmias, such as heart block, and tachyarrhythmias, such as ventricular fibrillation. Longer QRS duration is a predictor of mortality and sudden death in the general population and in cohorts with hypertension and coronary artery disease[1-3]. In a population-based study, prolonged baseline QRS was associated with incident heart failure[4].

Twin and family studies suggest a genetic contribution to QRS duration, with heritability estimates of up to 40%[5,6]. Prior candidate gene and smaller genome-wide studies identified a limited number of loci associated with QRS duration, supporting the hypothesis of the contribution of common genetic variation to QRS duration[7-9]. To identify additional loci and highlight physiologic processes associated with ventricular conduction, we performed a meta-analysis of 14 genome-wide association studies (GWAS) of QRS duration in a total of 40,407 individuals of European descent, where we adjusted the analyses for age, sex, height, and body mass index after appropriate sample exclusions (Methods). After an initial discovery phase, we further genotyped selected variants representing nine loci with *P*-values ranging from 1×10^{-6} to 5×10^{-9} in an additional cohort of 7,170 European individuals.

METHODS

Participating Studies

Details of the 15 participating studies are available in the Supplementary Note.

Phenotype modeling

We excluded individuals of non-European ancestry and those with QRS duration longer than 120 ms, which is often due to acquired left or right bundle branch block. We also excluded individuals with characteristics that may influence QRS duration, including a history of prior myocardial infarction or heart failure, atrial fibrillation on the ECG, pacemaker, Wolff-Parkinson-White syndrome, or use of class I and class III antiarrhythmic medications at the time of ECG acquisition. Covariates measured at baseline include age, gender, study site or cohort (if relevant), height and body-mass index.

GWAS genotyping and imputations

Either Affymetrix or Illumina arrays were used for genotyping (Supplementary Table 1b). Each study performed filtering of both individuals and SNPs to ensure robustness for genetic analysis (Supplementary Table 1b). Each study utilized the genotypes generated with these platforms to impute genotypes for approximately 2.5 million autosomal SNPs based on LD patterns observed in the HapMap CEU samples. Imputed genotypes were coded as dosages, fractional values between 0 and 2 reflecting the estimated number of copies of a given allele for a given SNP for each

individual. Most studies used a hidden Markov model as implemented in the MACH software. In CHS, imputation was performed using BIMBAM.

Extension genotyping

To extend our analyses, we genotyped nine SNP variants representing nine loci with *P*-values ranging from 1×10^{-6} to 5×10^{-9} in an additional group of 7,170 individuals in the PREVEND study. Of the nine SNPs, four represented loci with *P*-values between 5×10^{-8} to 5×10^{-9} in the discovery phase (Table 1). The remaining five index SNPs (rs1733724 near *DKK1*; rs1662342 in a *MYL12A*

Table 1 *Significant loci at $P<5\times10^{-08}$ in combined GWAS and candidate SNP meta-analysis. In each locus at least one marker exceeds the genome-wide significance threshold of $P<5\times10^{-8}$. At locus 1, six signals were identified ($r^2<0.05$) that exceeded genome-wide threshold. In a multiSNP model that included all 6 SNPs, there was evidence that at least 4 of these SNPs were independently associated with QRS duration. The bolded allele is the coded allele. Beta values (β) estimate the difference in QRS interval in milliseconds per copy of the coded allele, adjusted for the covariates in the model. Chr, chromosome; AF, coded allele frequency; SE, standard error; GC, genomic control adjusted; UTR, untranslated region. AF is an average weighted by study size.*

Locus	Chr	Index SNP	Coded/Non-coded Allele	AF	GWAS β	GWAS SE_{GC}	GWAS P_{GC}	I^2
1	3	rs6801957	**T**/C	0.41	0.77	0.07	1.10×10^{-28}	45.3
	3	rs9851724	**C**/T	0.33	-0.66	0.07	1.91×10^{-20}	57.1
	3	rs10865879	**T**/C	0.41	0.77	0.07	1.10×10^{-28}	53.6
	3	rs11710077	**T**/A	0.21	-0.84	0.09	5.74×10^{-22}	23.8
	3	rs11708996	**C**/G	0.16	0.79	0.10	1.26×10^{-16}	0.0
	3	rs2051211	**G**/A	0.26	-0.44	0.08	1.57×10^{-08}	0.0
2	6	rs9470361	**A**/G	0.25	0.87	0.08	3.00×10^{-27}	14.6
3	6	rs11153730	**C**/T	0.49	0.59	0.07	1.26×10^{-18}	5.3
4	1	rs9436640	**G**/T	0.46	-0.59	0.07	4.57×10^{-18}	51.2
5	5	rs13165478	**A**/G	0.36	-0.55	0.07	7.36×10^{-14}	64.6
6	7	rs1362212	**A**/G	0.18	0.69	0.09	1.12×10^{-13}	0.0
7	14	rs11848785	**G**/A	0.27	-0.50	0.08	1.04×10^{-10}	0.0
8	12	rs883079	**C**/T	0.29	0.49	0.08	1.33×10^{-10}	8.3
9	12	rs10850409	**A**/G	0.27	-0.49	0.08	3.06×10^{-10}	0.0
10	10	rs7342028	**T**/G	0.27	0.48	0.08	4.95×10^{-10}	0.0
11	18	rs991014	**T**/C	0.42	0.42	0.07	6.20×10^{-10}	0.0
12	2	rs17020136	**C**/T	0.21	0.51	0.08	1.90×10^{-9}	0.0
13	3	rs4687718	**A**/G	0.14	-0.63	0.11	6.25×10^{-9}	0.0
14	2	rs7562790	**G**/T	0.40	0.39	0.07	8.22×10^{-9}	0.0
15	1	rs17391905	**G**/T	0.05	-1.35	0.23	8.72×10^{-9}	4.0
16	17	rs9912468	**G**/C	0.43	0.39	0.07	1.06×10^{-8}	28.2
17	7	rs7784776	**G**/A	0.43	0.39	0.07	1.42×10^{-8}	0.0
18	1	rs4074536	**C**/T	0.29	-0.42	0.07	2.36×10^{-8}	0.5
19	13	rs1886512	**A**/T	0.37	-0.40	0.07	4.31×10^{-8}	0.0
20	3	rs2242285	**A**/G	0.42	0.37	0.07	4.79×10^{-8}	35.4
21	10	rs1733724	**A**/G	0.25	0.49	0.09	1.26×10^{-7}	0.0
22	17	rs17608766	**C**/T	0.16	0.53	0.10	3.71×10^{-7}	13.8

intron; rs17608766, intronic in *GOSR2*; rs17362588, missense variant in *CCDC141*; rs2848901, intronic in *FHOD3*) had *P*-values ranging from 1×10^{-6} to 5×10^{-8}, and needed additional statistical evidence in favor of the alternative hypothesis that they represented true associations. The SNPs were genotyped using TaqMan Allelic Discrimination Assays (ABI, Foster City, CA, USA).

Statistical Methods

Associations between QRS duration and SNPs were tested using linear regression models under the assumption of an additive (allelic trend) model of genotypic effect. These models were adjusted for age, gender, height, BMI, and study site (as appropriate). In family-based cohorts, linear mixed modeling was implemented to control for relatedness among samples[10]. A genomic control correction factor (λ_{GC}), calculated from all imputed SNPs, was applied on a per-study basis to account for cryptic population sub-structure and other potential biases[11].

The regression results were meta-analyzed using inverse variance weighted fixed-effects models[12]. We conducted the meta-analysis by three independent analysts using three different software packages, MANTEL, METABEL, and METAL. All results were extremely concordant,

Prevend β	Prevend *P*	$P_{Overall}$	Multi SNP β	Multi SNP *P*	Nearest Gene	SNP Annotation
-	-	1.10×10^{-28}	0.54	3.43×10^{-14}	*SCN10A*	intron
-	-	1.91×10^{-20}	-0.60	5.78×10^{-16}	*SCN10A/ SCN5A*	intergenic
-	-	1.10×10^{-28}	0.33	1.67×10^{-04}	*SCN5A/EXOG*	intergenic
-	-	5.74×10^{-22}	-0.44	1.33×10^{-06}	*SCN5A*	intron
-	-	1.26×10^{-16}	0.47	7.23×10^{-06}	*SCN5A*	intron
-	-	1.57×10^{-08}	-0.18	3.71×10^{-02}	*EXOG*	intron
-	-	3.00×10^{-27}	-	-	*CDKN1A*	intergenic
-	-	1.26×10^{-18}	-	-	*C6orf204/ SLC35F1/PLN/ BRD7P3*	intergenic
-	-	4.57×10^{-18}	-	-	*NFIA*	intron
-	-	7.36×10^{-14}	-	-	*HAND1/SAP30L*	intergenic
-	-	1.12×10^{-13}	-	-	*TBX20*	intergenic
-	-	1.04×10^{-10}	-	-	*SIPA1L1*	intron
-	-	1.33×10^{-10}	-	-	*TBX5*	3'-UTR
-	-	3.06×10^{-10}	-	-	*TBX3*	intergenic
-	-	4.95×10^{-10}	-	-	*VTI1A*	intron
-	-	6.20×10^{-10}	-	-	*SETBP1*	intron
-	-	1.90×10^{-9}	-	-	*HEATR5B/STRN*	intron
-	-	6.25×10^{-9}	-	-	*TKT/PRKCD/ CACNA1D*	intron
-	-	8.22×10^{-9}	-	-	*CRIM1*	intron
-1.17	0.005	3.26×10^{-10}	-	-	*C1orf185/RNF11/ CDKN2C/FAF1*	intergenic
-	-	1.06×10^{-8}	-	-	*PRKCA*	intron
0.36	0.015	1.28×10^{-9}	-	-	*IGFBP3*	intergenic
-	-	2.36×10^{-8}	-	-	*CASQ2*	missense
-0.28	0.047	1.27×10^{-8}	-	-	*KLF12*	intron
0.29	0.040	1.09×10^{-8}	-	-	*LRIG1/SLC25A26*	intron
0.34	0.035	3.05×10^{-8}	-	-	*DKK1*	intergenic
0.92	4.7×10^{-5}	4.75×10^{-10}	-	-	*GOSR2*	intron, 3'-UTR

reflecting a robust analysis. To be conservative, we subsequently corrected all *P*-values from the meta-analysis for the overall inflation (λ_{GC}=1.059). Results were considered statistically significant at a *P*-value of 5×10^{-8} after inflation correction, a figure that reflects the estimated testing burden of one million independent SNPs in samples of European ancestry[13]. Regions harboring association signals were visualized using SNAP[14].

To discern independent SNPs in regions with multiple genome-wide significant 'hits', we used an LD-binning procedure as implemented in PLINK. Starting with the most significant result as the index SNP, all surrounding SNPs within 500 kb (regardless of *P*-value) with a very liberal pairwise r^2>0.05 were 'clumped' with the index SNP, using LD patterns from HapMap-CEU (release 27). The procedure was repeated until all SNPs found membership in a clump. Thus, all index SNPs are, by definition, in very weak LD (if at all) with one another (pairwise r^2<0.05), and are, as such, suggestive of independent signals of association.

We then evaluated a multivariate regression model based on 28 index SNPs that reached genome-wide significance in the discovery meta-analysis to test which index SNPs represented truly independent signals. We set the significance threshold for claiming independence based on the estimated number of uncorrelated tests we performed across the 20 genetic loci (that contain the 28 'independent' index SNPs). At these 20 loci, there are 21,551 SNPs surrounding the index SNPs (within 500 kb) in HapMap-CEU, but after correcting for LD we arrived at ~1,563 tests, which corresponds to a threshold of $P<3.2\times10^{-5}$. Through meta-analysis of the multivariate *P*-values across the participating cohorts with the largest sample size, we found significant evidence for four independent SNPs at the *SCN10A-SCN5A* locus (Table 1), but not elsewhere.

To test for an association with QRS greater than 120 ms, we calculated a SNP score for each individual in the RS and ARIC studies by adding up the number of QRS-prolonging alleles from the dosage counts and weighting by the β estimate from the meta-analysis. Logistic regression modeling was then performed with QRS greater than 120 ms as the dependent variable (dichotomous trait), which was regressed on the score, adjusting for age, sex, height, BMI and study site.

Because QRS interval is strongly influenced by sex, and inherited conduction defects can show a pronounced influence with aging, we explored whether the 23 genetic associations identified by our discovery GWAS meta-analysis (index SNPs at each of the 20 QRS-associated loci plus three additional independent SNPs at locus 1) differed by sex or age. For interactions with sex, we performed analyses in each cohort including an interaction term for sex × genotype, and then we used inverse variance weighting to meta-analyze the interaction terms. For interactions with age, each cohort performed separate analyses for each SNP stratified by decade, and then we performed regression analyses to assess the effect of age on the magnitude of the genetic effect.

eQTL analysis

We used genomics data from 1,240 PAXgene whole blood samples[15]. The samples were expression profiled on an Illumina HT-12 platform and genotyped using either an Illumina Hap370 or 610-Quad platform. Ungenotyped SNPs were imputed with the IMPUTE software. We

applied a 500kb window (250kb on each side) around each of the 25 QRS SNPs and tested the cis expression-genotype association using SNPTEST.

One hundred forty-two independent probes were examined at 22 QRS loci (comprising 25 index SNPs), resulting in a total of 198 probe-SNP pairs tested. Bonferroni adjustment was applied for the tested probe-SNP pairs ($P < 2.5 \times 10^{-4}$ was deemed significant). The eQTLs were checked for possible polymorphisms within the probe region. 1000 Genomes Project data (April 2009 release) was used to assess LD between the detected eSNP and the SNPs located within the probe region. If $r^2 > 0.05$ between an eSNP and the SNPs in the probe region, the eQTL was deemed a false positive assuming the 'probe SNP' caused differential hybridization.

Bioinformatics

All identified QRS loci were chosen to generate a gene product interaction network using Ingenuity Pathway Analysis software[16]. For locus 1, two genes were independently associated with QRS duration (*SCN5A-SCN10A*) and both were included in the network. For loci 3, 5, 12, 13, 15, 21 or 22, where it was difficult to discern to which of several genes the association signal might map, several genes (listed in Table 1 for each of the loci) were included in the model. Of these seven loci, three (loci 13, 15 and 21) had two members each map to the network. We limited the relational network to known direct relationships (for example, protein-protein interaction, phosphorylation or physical binding). To ensure a spanning network, Ingenuity algorithmically incorporated additional nodes that interact with the QRS-associated loci. However, we limited these 'linkers' such that the relational distance among the QRS-associated loci was no more than one linker node. In some cases, this resulted in a genome-wide significant locus to remain unconnected to the network (for example, *GOSR2, CRIM1* and *NFIA*). We systematically searched *PubMed* to find direct gene product relationships among the network nodes to complement the original network. Relationships highlighted in this network have been compiled from a range of experimental conditions, organisms and tissue types, and therefore may not correspond to actual pathways in the human heart. For the functional enrichment analysis, we used two independent online tools, the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.7) and GOTree Machine (GOTM), to identify Gene Ontology categories overrepresented among the QRS-associated loci relative to the entire human genome background[17,18]. A modified version of Fisher's Exact test (DAVID) and the hypergeometric test (GOTM) were used to determine the probability that a functional category is enriched compared to that expected from random chance.

Mouse model

Quantitative real-time PCR

Enhanced green fluorescent protein $(EGFP)^+$ Purkinje cells and $EGFP^-$ ventricular myocytes were isolated from neonatal *CNTN2-EGFP* transgenic mice by fluorescence-activated cell sorting[19,20]. Quantitative PCR was performed user primers for *SCN10A* and ribosomal S26 as an internal reference using the $2\text{-}\Delta\Delta CT$ method[21].

Electrophysiologic testing

Telemetric devices (DSI) were implanted into adult CD1/129SI/SV1MJ mice and recordings were obtained before and 30 min after intraperitoneal injection of the Nav1.8 antagonist A-803467 (100 mg/kg)[22]. Intracardiac recordings were obtained using an open-chest model under isoflurane anesthesia (1.5% v/v) using an octapolar catheter (EPR-800, Millar Instruments) placed in the right jugular vein.

RESULTS

Meta-analysis of genome-wide association results

We conducted meta-analyses for approximately 2.5 million single nucleotide polymorphisms (SNPs) in 40,407 individuals of European ancestry from 14 GWAS (Supplementary Tables 1A and 1B). Overall, 612 variants in 20 loci exceeded our genome-wide significance *P*-value threshold of 5×10^{-8} after adjusting for modest genomic inflation ($\lambda_{GC} = 1.059$) (Figure 1 and Supplementary Figure 1). The loci associated with QRS interval duration are detailed in Table 1 and Supplementary Figure 2, with the index SNP (representing the most significant association) labeled for each independent signal.

Across the genome, the most significant association for QRS interval duration (locus 1) was on chromosome 3p22 (Figure 2A), where we identified six potentially independent association signals based on the linkage disequilibrium (LD) patterns in HapMap-CEU (pairwise r^2 among index SNPs<0.05). In conditional analyses where all six SNPs were included in the same regres-

Figure 1 *Manhattan plot. Manhattan plot showing the association of SNPs with QRS interval duration in a GWAS of 40,407 individuals. The dashed horizontal line marks the threshold for genome-wide significance ($P = 5\times10^{-8}$). Twenty loci (labeled) reached genome-wide significance. Two additional loci, GOSR2 and DKK1, reached significance after genotyping of select SNPs in an additional sample of 7,170 individuals (see Results).*

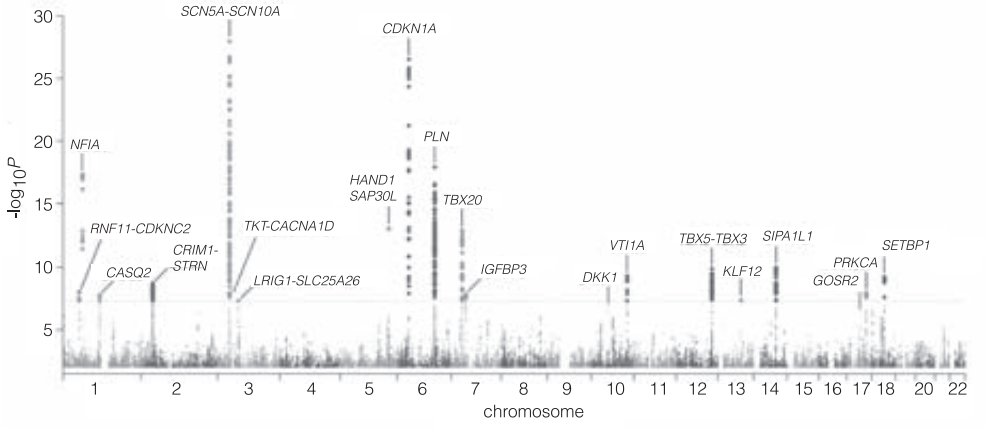
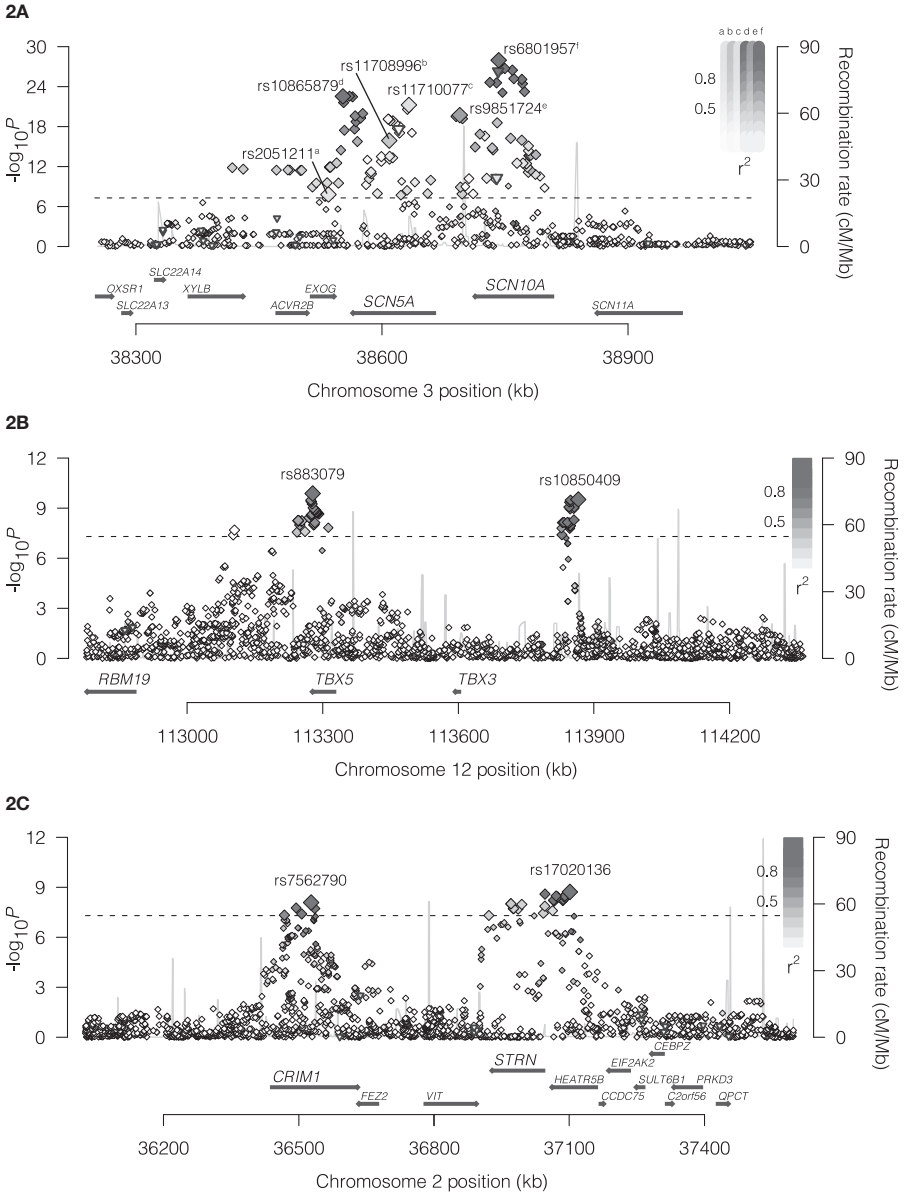


Figure 2 *Association plots for select loci. Each SNP is plotted with respect to its chromosomal location (x axis) and its P value (y axis on the left). The tall spikes indicate the recombination rate (y axis on the right) at that region of the chromosome. The black-outlined triangles indicate coding region SNPs. (a) Locus 1 (SCN5A-SCN10A) on chromosome 3. The six index signals are named with their rs numbers and highlighted in different grey scales (a-f). Other SNPs in linkage disequilibrium with the index SNP are denoted in the same scale. Color saturation indicates the degree of correlation with the index SNP. (b) Locus 8 (TBX5) and locus 9 (TBX3) on chromosome 12. (c) Locus 12 (HEATR5B-STRN) and locus 14 (CRIM1) on chromosome 2.*



sion model, there was compelling evidence that at least four SNPs from this region were independently associated with QRS duration (Table 1). Two of these associations were in or near *SCN10A*, a voltage-gated sodium channel gene. Variation at this locus was recently associated with QRS duration in two GWAS. The top SNP identified in those two studies, rs6795970, was in strong LD with our top signal, rs6801975 (r^2=0.93)[8,9]. Two additional signals were identified in *SCN5A*, a sodium channel gene adjacent to *SCN10A* (Table 1).

The second most significant locus (locus 2) was on chromosome 6p21 near *CDKN1A*, a cyclin dependent kinase inhibitor. The *CDKN1A* locus was recently associated with QRS interval duration in an Icelandic population[9]. The index SNP in the prior report, rs1321311, is in strong LD with our top signal, rs9470361 (r^2 = 0.88). *CDKN2C*, which encodes another cyclin-dependent kinase inhibitor, is located in locus 15, which encompasses several other genes, including *C10rf185*, *RNF11* and *FAF1*.

Locus 3 on chromosome 6q22 contains the *PLN-SLC35F1-C6orf204-BRD7P3* cluster of genes. *PLN* encodes phospholamban, a key regulator of sarcoplasmic reticulum calcium reuptake. Significant associations were found in several other regions harboring calcium-handling genes, including locus 12 (*STRN-HEATR5B*), locus 16 (*PRKCA*) and locus 18 (*CASQ2*).

Locus 4 mapped to an intronic SNP in *NFIA*, which encodes a transcription factor. Several other significant loci also mapped in or near genes encoding transcription factors, including locus 5 (*HAND1*), locus 6 (*TBX20*), locus 8 (*TBX5*), locus 9 (*TBX3*) and locus 19 (*KLF12*). Common variation in *TBX5* was recently associated with QRS duration[9]. The index signal in the prior report, rs3825214, was in moderate LD with our top signal, rs883079 (r^2 = 0.67).

Additional regions identified include locus 7 (*SIPA1L1*), locus 10 (*VTI1A*), locus 11 (*SETBP1*), locus 13 (*TKT-CACNA1D-PRKCD*), locus 14 (*CRIM1*), locus 17 (the nearest gene, *IGFBP3*, is 660 kb away) and locus 20 (*LRIG1*).

Collectively, the identified index SNPs across these 20 loci explained approximately 5.7% (± 2.3% (s.d.)) of the observed variance in QRS duration, consistent with a polygenic model in which each of the discovered variants exerts only a modest effect on QRS interval. None of these index SNPs showed a significant interaction with sex or age after Bonferroni correction. We observed moderate levels of heterogeneity of the effect ($25<I^2<75$) for several index SNPs (Table 1). However, only *HAND1/SAP30L* showed significant evidence of heterogeneity using Cochran's Q test corrected for 23 independent genome-wide variants (Cochran's P = 0.005).

Extension of findings in an additional 7,170 individuals

Based on the discovery meta-analysis, we selected the index SNPs at four loci (loci 15, 17, 19 and 20) with P values ranging between $P = 5 \times 10^{-8}$ and $P = 5 \times 10^{-9}$ and from all five loci with P values ranging from $P = 1 \times 10^{-6}$ to $P = 5 \times 10^{-8}$ (Online Methods) for genotyping in an additional 7,170 European individuals in order to boost the study's power. In a joint analysis combining all 47,577 individuals, the significance for the four loci with P values between $P = 5 \times 10^{-8}$ and $P = 5 \times 10^{-9}$ increased, indicating these represent true positive associations (Table 1). The joint analysis also provided further evidence for two other loci (locus 21 near *DKK1* and locus 22 tagged by an intronic SNP in *GOSR2*) that reached genome-wide significance, bringing the total

number of significant loci to 22, with 25 independently associated index SNPs (Table 1). The index SNP (rs1733724) in *DKK1* was previously associated with QRS duration in an Icelandic population[9].

Association with conduction defect

Based on this series of QRS associations, we sought to test the hypothesis that QRS-prolonging alleles, on average, increase risk of ventricular conduction defects. To address this question, we calculated a risk score in each individual by adding up the number of QRS-prolonging alleles identified in this study weighted by the observed effect sizes (β estimates) from the final meta-analysis. In an independent set of 519 individuals from the Atherosclerosis Risk in Communities (ARIC) and Rotterdam (RS) studies with bundle branch block or nonspecific prolongation of QRS interval (QRS>120 milliseconds (ms)) compared with those individuals with normal conduction (N = 12,804), we found evidence that the cumulative burden of QRS-prolonging alleles is associated with risk of ventricular conduction defect ($P = 0.004$). This result was largely driven by those individuals with nonspecific intraventricular conduction defects, as opposed to those with left or right bundle branch block (Supplementary Table 2A,B). Similar results were observed using an unweighted genotype risk score.

Putative functional variants

Of the 612 genome-wide significant SNPs, one in *SCN5A* (rs1805124, His558Arg, $P = 2.4 \times 10^{-18}$), two in *SCN10A* (rs12632942, Leu1092Pro, $P = 5.1 \times 10^{-11}$, and rs6795970, Ala1073Val, $P = 5 \times 10^{-27}$), one in *C6orf204* near *PLN* (rs3734381, Ser137Gly, $P = 1.1 \times 10^{-10}$) and one in *CASQ2* (index SNP rs4074536, Thr66Ala, $P = 2.4 \times 10^{-8}$) were non-synonymous (Figure 2 and Supplementary Figure 2). The PolyPhen-2 program predicts all five of these variants to be benign, which is consistent with small-effect associations: each copy of the minor allele was associated with cross-sectional differences in QRS duration of less than 1 ms.

The 25 index SNPs (from Table 1) were subsequently tested for association with gene cis-expression levels in 1,240 PAXgene whole blood samples[15]. Four cis-eQTLs were detected after stringent Bonferroni correction. The most notable eQTLs were observed for probes in exonic regions of *TKT* (rs4687718, $P = 5.87 \times 10^{-70}$) and *CDKN1A* (rs9470361, $P = 1.41 \times 10^{-10}$) and in an intronic probe for *C6orf204* near *PLN* (rs11153730, $P = 1.54 \times 10^{-10}$) (Supplementary Figure 3). We additionally assessed cis regulation for all HapMap SNPs for these three loci (± 250 kb around the SNPs). The top eSNPs for *TKT* (rs9821134) and *C6orf204* (rs11970286) were in moderate to high LD ($r^2 = 0.47$ and $r^2 = 0.91$, respectively) with the top QRS signals at these loci. However, the top eSNP for *CDKN1A*, rs735013, was only weakly correlated with the QRS index SNP rs9470361 ($r^2 = 0.089$). In a conditional analysis that included both *CDKN1A* locus SNPs in the regression model, both rs735013 and rs9470361 remained independently associated with expression levels ($P = 1.7 \times 10^{-9}$ and $P = 2.3 \times 10^{-5}$, respectively). Additionally, rs735013 itself was marginally associated with QRS duration (coded allele frequency = 0.39, $\beta = 0.33$ ms (s.e. = 0.07 ms), $P = 2.4 \times 10^{-6}$). Whether these associations in whole blood samples will be similar to associations in cardiac myocytes and conduction tissue deserves further investigation.

Pleiotropic effects of variants associated with QRS duration and other ECG measurements

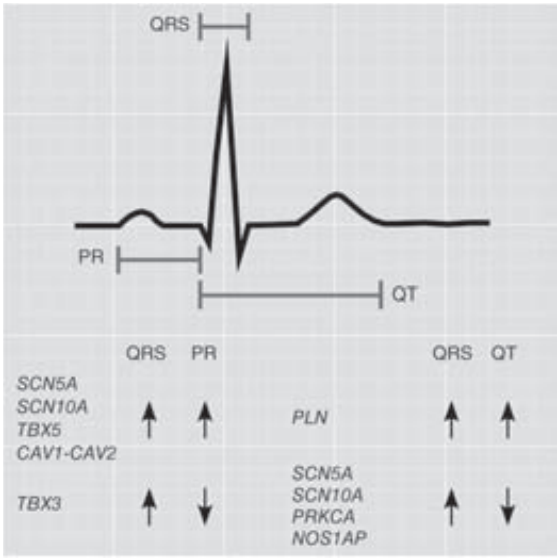
To explore the shared genetic underpinnings between atrial and ventricular depolarization and conduction (as measured by PR and QRS intervals) as well as ventricular depolarization and repolarization (QRS and QT intervals), we examined the effects of previously published PR and QT SNPs with respect to QRS interval. Several QRS loci were previously associated with PR or QT intervals, including *PLN, TBX5-TBX3* and *SCN5A-SCN10A*, the last of which is associated with all three traits (Supplementary Table 3). We also tested nine PR SNPs and 16 QT SNPs for their effect on QRS duration (Supplementary Table 3)[23-25]. Our results suggest roles for *CAV1/2* (rs3807989, $P = 5.8 \times 10^{-6}$) and *NOS1AP* (rs12143842, $P = 1.3 \times 10^{-4}$) in QRS duration. Indeed *CAV1/2* was recently associated with QRS interval[9].

QRS duration is positively correlated with both PR interval ($r = 0.09$) and QT interval ($r = 0.44$)[9]. To test if these relationships are also observed genetically, we compared the directionality of the association of SNPs at the published PR and QT loci with those for QRS duration. Generally, the effects of SNPs on PR interval were positively correlated with their effects on QRS duration ($r = 0.53$). With the exception of *TBX3*, the loci influencing both PR and QRS intervals (*SCN5A, SCN10A, TBX5* and *CAV1-CAV2*) do so in a concordant fashion (that is, variants that prolong PR also prolong QRS duration) (Figure 3 and Supplementary Tables 3A and 3B). By contrast, although QT and QRS are positively correlated at the population level, the effects of SNPs on QT interval were marginally negatively correlated with their effects on QRS interval ($r = -0.08$). Of the index SNPs at the four loci significantly associated with both QT and QRS interval (*SCN5A-SCN10A, PRKCA, NOS1AP* and *PLN*), only the *PLN* locus SNPs showed effects in the same direction (Figure 3 and Supplementary Tables 3A and 3B).

Bioinformatic network analysis of QRS-associated loci

To examine the relationships between genetic loci associated with QRS duration, we developed an *in silico* relational network linking the loci based on published direct gene product interactions obtained from curated databases (Supplementary Figure 4)[16]. Most loci meeting genome-wide significance mapped to this network after a minimum number of 'linker' nodes were incorporated to create a spanning network. This analysis provides a graphical overview of the interconnections among QRS-associated genetic loci and highlights both known and putative molecular mechanisms regulating ventricular conduction (see below for further discussion). Several of the 'linker' nodes incorporated in the network, such as *GJA1* (encoding connexin 43), *NEDD4, KCNMA1* and *RYR2*, are known modulators of cardiac electrical activity. Functional enrichment analysis of the QRS-associated network nodes (that is, loci with $P < 5 \times 10^{-8}$) using two independent software tools revealed that programs involved in heart development were highly overrepresented (P value range: $P = 5.8 \times 10^{-6}$ to $P = 9.6 \times 10^{-5}$)[17,18].

Figure 3 *Pleiotropic associations of PR, QRS, and QT loci. Electrocardiographic tracing delineating the PR, QRS and QT intervals. PR and QRS intervals reflect myocardial depolarization and conduction time through the atria and down the atrioventricular node (PR) and throughout the ventricle (QRS) and are weakly positively correlated (r = 0.09). The majority of loci that influence both PR and QRS (SCN5A, SCN10A, TBX5 and CAV1-2) do so in a concordant fashion (meaning variants that prolong PR duration also prolong QRS duration). The notable exception is a region on chromosome 12, where variants in the TBX5 locus have a concordant effect, whereas those in nearby TBX3 have a discordant effect. By contrast, although QRS (ventricular depolarization) and QT (ventricular repolarization) are moderately positively correlated, the majority of loci (SCN5A, SCN10A, PRKCA and NOS1AP) that influence both phenotypes do so in a discordant fashion (meaning variants that prolong the QRS interval shorten the QT interval). The exception is the locus at PLN, where the variants have a concordant effect.*



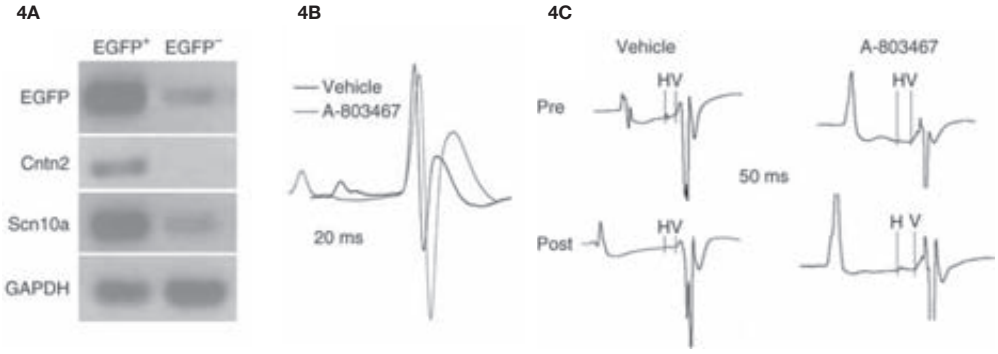
SCN10A function in a mouse model

We undertook functional studies to determine whether our most significant locus was associated with ventricular conduction in mice. Transcriptional profiling suggests that *Scn10a* mRNA, which encodes the $Na_v1.8$ sodium channel, is expressed in the ventricular myocardium and at higher levels in the specialized conduction system[19]. These data were confirmed and extended by quantitative PCR (Figure 4A), demonstrating a 25.7-fold ± 1.1-fold (s.e.) enrichment of *Scn10a* $Na_v1.8$ in Purkinje cells compared to working ventricular myocytes (n = 3 for each cell type; $P = 0.002$).

Telemetric electrocardiographic recordings (lead II position) were obtained in conscious mice treated with A-803467, a potent *Scn10a* $Na_v1.8$ antagonist, which blocks $Na_v1.8$ 100 times more potently than $Na_v1.5$ with the doses used[26]. These studies showed a significant increase in

QRS duration (11.6 ms (± 2.6 ms (s.e.)) to 14.5 ms (± 0.54 ms), $n = 7$, $P < 0.001$), whereas treatment with vehicle alone was without effect (11.4 ms (± 0.29 ms) to 11.9 ms (± 0.42 ms), $n = 7$, P value was not significant). The PR interval was also increased in drug-treated mice, from 31.4 ms (± 0.98 ms) to 42.5 ms (± 3.3 ms), $n = 7$, $P < 0.01$), whereas treatment with vehicle alone resulted in no significant change (32.6 ms (± 1.0 ms) to 33.4 ms (± 0.69 ms), $n = 7$, P value was not significant) (Figure 4B). To further delineate the site of ventricular conduction slowing, we performed intra-cardiac recordings from mice treated with A-803467. These studies confirmed the significant increase in QRS duration (from 12.26 ms (± 0.62 ms) to 14.56 ms (± 0.58 ms), $n = 7$, $P = 0.015$), whereas treatment with vehicle alone was without significant effect (12.39 ms (± 0.52 ms) to 13.65 ms (± 0.97 ms), $n = 5$, P value was not significant). A-803467 treatment resulted in a 35.7% ± 1.2% increase in HV interval (from 9.33 ms (± 0.74 ms) to 12.67 ms (± 1.06 ms), $P = 0.009$), whereas treatment with vehicle alone was without significant effect (10.67 ms (± 0.83 ms) to 11.17 ms (± 1.10 ms), P value was not significant) (Figure 4C). Taken together, these data indicate that the QRS prolongation may primarily reflect conduction slowing in the specialized ventricular conduction system.

Figure 4 *Expression and function of Scn10a in the murine heart. (a) Neonatal ventricular myocytes from Cntn2-eGFP BAC transgenic mice were fluorescence-activated cell sorted and eGFP⁺ and eGFP⁻ pools were analyzed by RT-PCR. Transcripts encoding eGFP, Cntn2 and Scn10a were highly enriched in the eGFP⁺ fraction. Quantitative RT-PCR demonstrated 25.7-fold enrichment of Scn10a $Na_v1.8$. (b) Representative telemetric electrocardiographic recordings (lead II configuration) obtained 30 min after administration of vehicle alone (black tracing) or the Scn10a Nav1.8 antagonist A-803467 (gray tracing). The two tracings are aligned at the onset of the QRS wave, and both PR interval and QRS interval prolongation were observed in drug-treated mice. (c) Representative intracardiac recordings showing HV intervals obtained before (Pre) and after (Post) administration of vehicle or A-803467. Significant HV prolongation was observed in drug-treated mice.*



DISCUSSION

Our meta-analysis of 14 GWAS consisting of 40,407 individuals of European descent with additional genotyping in 7,170 Europeans yielded genome-wide significant associations of QRS duration with common variants in 22 loci. Variations in four of these loci (locus 1, *SCN5A-SCN10A*; locus 2, *CDKN1A*; locus 8, *TBX5*; and locus 21, *DKK1*) were previously associated with QRS duration in smaller independent studies using both candidate gene and genome-wide approaches[7-9]. The 22 loci include genes in a number of interconnected pathways, including some previously known to be involved in cardiac conduction, such as sodium channels, calcium-handling proteins and transcription factors, as well as previously unidentified processes not known to be involved in cardiac electrophysiology, such as kinase inhibitors, growth factor-related genes and others.

The electrocardiographic QRS interval reflects ventricular depolarization and conduction time. Ventricular myocyte depolarization occurs through cardiac membrane excitatory inward currents mediated by voltage-gated sodium channels[27]. The primary determinants of conduction velocity are the magnitude of excitatory inward currents flowing through these sodium channels, the extent of cell-to-cell communication through gap junction–connexin coupling, and cell and tissue architecture and morphology[27]. Multiple pathways suggested in this study determine or modulate these key components of ventricular depolarization and conduction. Candidate genes in these pathways are briefly discussed in Box 1.

Our strongest association signal (locus 1) mapped in or near two voltage-gated sodium channel genes: *SCN5A* and *SCN10A*. *SCN5A* encodes the cardiac $Na_v1.5$ sodium channel and is well known for its role in cardiac conduction, and other cardiovascular and electrophysiologic phenotypes[28,29]. *SCN10A* encodes the $Na_v1.8$ sodium channel. We provide new data demonstrating that the *SCN10A* transcript and product is preferentially expressed in the mouse His-Purkinje system compared with the ventricular myocardium, and that $Na_v1.8$ channel blockers result in QRS and HV interval prolongation, indicative of a slowing of impulse propagation in the specialized ventricular conduction system and a delayed activation of the ventricular myocardium. Notably, a recent study reported shortening of the PR interval in Scn10a knockdown mice and concluded that $Na_v1.8$ prolongs cardiac conduction and that rs6795970, encoding the A1073V variant, is a gain-of-function allele[8]. Alternatively, the more rapid conduction they observed in the knockdown mice could reflect compensatory upregulation of *TTX*-sensitive currents, a phenomenon previously observed in $Na_v1.8$-deficient DRG neurons[30].

We, and others, demonstrated previously that, in addition to their association with QRS duration, variants in *SCN5A* and *SCN10A* are associated with atrial conduction (PR interval) and myocardial repolarization (QT interval), as well as atrial and ventricular fibrillation[8,9,25]. These results emphasize the crucial role played by these genes in cardiac conduction and the generation of arrhythmias.

Calcium regulation is integral to impulse propagation, modulating cellular electrophysiology including sodium channel and gap junction function, as well as tissue architecture[28,31,32]. Several of the loci associated with QRS duration contain genes directly related to calcium processes. As depicted in Supplementary Figure 4 and detailed in Box 1, these genes encode interrelated

proteins that influence Ca^{2+} signaling (*PLN* in locus 3; *PRKCA* in locus 16; and *CASQ2* in locus 18) and downstream effects (*STRN* in locus 12).

Transcription factors regulating embryonic electrophysiologic development are critical for the integrity of impulse conduction[33]. We identified six transcription factors (*TBX3* in locus 9; *TBX5* in locus 8; *TBX20* in locus 6; *HAND1* in locus 5; *NFIA* in locus 4; and *KLF12* in locus 19) in loci associated with QRS duration. Several of these transcription factors affect cardiac morphogenesis and may influence conduction by altering cellular and tissue architecture. Notably, they may also have direct electrophysiologic consequences by modifying factors involved in impulse conduction. For example, *HAND1* and T-box factors regulate *GJA5* (encoding connexin 40) and/or *GJA1* (encoding connexin 43), and *TBX5* binds to the *ATP2A2* (also known as *SERCA2A*) promoter[34].

Our study suggests a number of processes and pathways not previously known to be involved in cardiac electrophysiology, including cyclin-dependent kinase inhibitors and genes related to tumorigenesis and cellular transformation. How these previously unidentified processes influence QRS duration remains to be defined.

In pleiotropic analyses, most variants influencing both PR and QRS interval, with the exception of *TBX3*, were concordant in effect direction, consistent with the known shared physiologic processes underlying the two traits: depolarization and conduction time in the sino-atrial node, atria and atrioventricular node (PR interval) and depolarization and conduction time in the ventricles (QRS interval). By contrast, although QRS (ventricular depolarization) and QT (ventricular repolarization) are moderately positively correlated, most loci influencing both traits showed discordant effect directions (with the exception of the *PLN* locus). Investigating the physiologic foundations for these concordant and discordant PR-QRS and QT-QRS relationships could be particularly informative for elucidating the mechanisms by which these loci influence cardiac depolarization, conduction and repolarization.

Several limitations of our study should be considered. First, although we have identified 22 loci significantly associated with QRS duration, the broad nature of LD among common variants generally precludes an unambiguous identification of the culprit variant or of the functional gene. For several genes (*SCN5A, SCN10A*, *C6orf204* and CASQ2), there are common coding SNPs in high LD with the index SNP, which may lend some support for a functional role for these genes. Furthermore, our expression analysis in blood revealed very strong *cis*-eQTL associations for *TKT* and *CDKN1A*, lending additional support to these genes as functional candidates. It would be desirable to perform similar eQTL analyses based on expression data in myocardial cells or in conduction tissue. For our top signal in *SCN10A*, a gene which until recently was not known to be expressed in the heart, our functional work in mice confirms that *SCN10A* is involved in ventricular depolarization and conduction. Further fine mapping is needed at all 22 loci to conclusively test all genetic variation (rare and common) for a role in QRS modulation.

To minimize the potential for confounding due to population substructure, we limited our analyses to individuals of European descent, a population for which we could assemble the largest number of samples. At the individual study level, the GWAS showed very little evidence for gross stratification (genomic inflation factor, λ_{GC}, values ranged from 1.00 to 1.05). However, one of our QRS loci, mapping to *HAND1-SAP30L*, showed evidence of heterogeneity. In genetic

association studies, heterogeneity can be due to sampling error, differences in phenotypic measurement, differences in LD structure between populations, technical artifacts, or genuine biological heterogeneity, but it would be difficult to conclude on the basis of our data here which of these is the most likely explanation[35].

Our study underscores the power of a large genome-wide association study to extend prior biological understanding of cardiac ventricular conduction. Better understanding of the complex biologic pathways and molecular genetics associated with cardiac conduction and QRS duration may offer insight into the molecular basis underlying the pathogenesis of conduction abnormalities that can result in increased risk of sudden death, heart failure and cardiac mortality.

Box 1 *Noteworthy genes within loci associated with QRS duration.*

Of the 22 loci identified, common variants in four loci (*SCN5A-SCN10A, CDKN1A, TBX5* and *DKK1*) were previously associated with QRS duration in genetic association studies. Mutations in two loci (*SCN5A* and *TBX5*) lead to inherited syndromes associated with conduction disease. Animal experiments show a role for several additional loci (*HAND1, TBX3* and *TBX5*) in cardiac ventricular conduction, as detailed below. The remainder are new QRS loci, and their role in cardiac conduction remains to be elucidated.

1 Cardiac sodium channel genes:

• ***SCN5A* (locus 1):** *SCN5A* encodes the cardiac $Na_v1.5$ sodium channel and is well known to influence cardiac conduction, as well as other cardiovascular and electrophysiologic phenotypes[28,29].

• ***SCN10A* (locus 1):** *SCN10A* encodes the $Na_v1.8$ sodium channel, present in both ventricular myocardium and conduction fibers. Selective *SCN10A* blocker prolongs QRS interval.

2 Calcium-handling proteins:

• ***CASQ2* (locus 18):** *CASQ2* regulates opening of the ryanodine receptor (encoded by *RYR2*)[36,37]. Cellular depolarization through sodium channels triggers calcium influx through L-type calcium channels, which in turn provokes *RYR2*-mediated calcium release from the sarcoplasmic reticulum. *CASQ2* mutations have been associated with catecholaminergic polymorphic ventricular tachycardia[38,39].

• ***PLN* (locus 3):** Calcium uptake into the sarcoplasmic reticulum by SERCA2a is regulated by *PLN* (encoding cardiac phospholamban)[40]. The phosphorylation state of *PLN* is dependent on signaling pathways involving phosphatases and kinases, including that encoded by *PRKCA*[41]. We previously demonstrated that this locus is associated with both cardiac electrical properties (QT interval duration and heart rate) and size (left ventricular end diastolic dimension) in genome-wide association analyses[23,24,41,42].

• ***PRKCA* (locus 16):** Protein kinase C alpha activity affects dephosphorylation of the sarcoplasmic reticulum Ca^{2+} ATPase-2 (SERCA-2) pump inhibitory protein phospholamban (*PLN*) and alters sarcoplasmic reticulum Ca^{2+} loading and the Ca^{2+} transient[43].

• ***STRN* (locus 12):** Striatin is a Ca^{2+}-calmodulin binding protein that directly binds to caveolin scaffolding protein. Striatin has recently been implicated in a canine model of arrhythmogenic right ventricular cardiomyopathy[44,45].

3 Transcription factors:

• ***TBX3* (locus 9) and *TBX5* (locus 8):** *TBX3* and *TBX5* encode transcription factors found in the cardiac conduction system. *TBX5* (activator) competes with *TBX3* (repressor) for the regulation of working myocardial genes such as *GJA1*[46,47]. Common variations near *TBX3* and *TBX5* were associated with PR and QRS durations[9,25]. Mutations in *TBX3* and *TBX5* have been associated with rare inherited syndromes manifested by an array of defects including ventricular structural and/or conduction defects.

• ***TBX20* (locus 6):** *TBX20* demarcates the left and right ventricles[48] and mutations in *TBX20* have been implicated in multiple structural defects in mouse and human models[49,50].

• ***HAND1* (locus 5):** *HAND1* encodes a transcription factor essential to cardiac morphogenesis[51], with a mutation in this gene having been identified in human hearts with septal defects[52]. Overexpression of *HAND1* in the adult mouse heart leads to loss of connexin 43 (encoded by *GJA1*) expression, QRS prolongation and predisposition to ventricular arrhythmia[53].

• ***NFIA* (locus 4) and *KLF12* (locus 19):** Little is known about the role of Nuclear Factor One (NFIA) and Kruppel like protein 12 (KLF12) in cardiac tissue development.

4 Cyclin dependent kinase inhibitors:

• ***CDKN1A* (locus 2):** *CDKN1A* is a negative regulator of cell cycle entry into G2/M phase and is upregulated by *ERBB2* activation. *ERBB2*, encoding a member of the EGF-receptor family of tyrosine kinases, is essential for proper heart development, and its ligand neuregulin-1 promotes formation of the murine cardiac conduction system[54]. Furthermore, *ERBB2* can modulate gap junction assembly and alter appropriate phosphorylation of connexin 43 in glial cells[55]. In addition, *CDKN1A* is upregulated by *PRKCA* (locus 16)[56].

• ***CDKN2C* (locus 15):** A member of the family of cyclin-dependent kinase inhibitors that prevent the activation of the CDK kinases, thus functioning as a cell-growth regulator that controls cell cycle G1 progression.

5 Other pathways:

• ***CRIM1* (locus 14):** *CRIM1*, encoding a cell-surface transmembrane protein that may bind to various members of the TGF-beta superfamily of ligands, is expressed in mouse and human cardiac tissues[57,58]. *CRIM1* interacts with bone morphogenetic proteins, which induce the expression of *CDKN1A* (p21)[57,59].

• ***LRIG1* (locus 20):** *LRIG1* is upregulated in malignancies. It negatively regulates the proto-oncogenic, tyrosine kinase receptor family *ERBB2*[60].

• ***SETBP1* (locus 11):** *SETBP1* encodes a ubiquitously expressed protein that binds to *SET*[61]. The *SETBP1-SET* interaction has been hypothesized to be a component in tumor development.

• ***TKT* (locus 13):** Transketolase (TKT) is a ubiquitous enzyme used in multiple metabolic pathways, including the pentose phosphate pathway[62].

• ***DKK1* (locus 21):** *DKK1*, implicated in several tumors, inhibits the Wnt signaling pathway[63]. Wnt signaling is an important modulator of connexin43-dependent intercellular coupling in the heart[64]. In cardiac tissue, it has an embryologic role with regard to axial development[65].

• ***SIPA1L1* (locus 7):** *SIPA1L1* appears to play a role in non-canonical Wnt signaling and contributes to development[66].

REFERENCES

1. Oikarinen, L., Nieminen, M.S., Viitasalo, M., Toivonen, L., Jern, S., Dahlof, B., Devereux, R.B. and Okin, P.M. (2004) QRS duration and QT interval predict mortality in hypertensive patients with left ventricular hypertrophy: the Losartan Intervention for Endpoint Reduction in Hypertension Study. *Hypertension*, 43, 1029-1034.
2. Desai, A.D., Yaw, T.S., Yamazaki, T., Kaykha, A., Chun, S. and Froelicher, V.F. (2006) Prognostic Significance of Quantitative QRS Duration. *Am J Med*, 119, 600-606.
3. Elhendy, A., Hammill, S.C., Mahoney, D.W. and Pellikka, P.A. (2005) Relation of QRS duration on the surface 12-lead electrocardiogram with mortality in patients with known or suspected coronary artery disease. *Am J Cardiol*, 96, 1082-1088.
4. Dhingra, R., Pencina, M.J., Wang, T.J., Nam, B.H., Benjamin, E.J., Levy, D., Larson, M.G., Kannel, W.B., D'Agostino, R.B., Sr. and Vasan, R.S. (2006) Electrocardiographic QRS duration and the risk of congestive heart failure: the Framingham Heart Study. *Hypertension*, 47, 861-867.
5. Busjahn, A., Knoblauch, H., Faulhaber, H.D., Boeckel, T., Rosenthal, M., Uhlmann, R., Hoehe, M., Schuster, H. and Luft, F.C. (1999) QT interval is linked to 2 long-QT syndrome loci in normal subjects. *Circulation*, 99, 3161-3164.
6. Hanson, B., Tuna, N., Bouchard, T., Heston, L., Eckert, E., Lykken, D., Segal, N. and Rich, S. (1989) Genetic factors in the electrocardiogram and heart rate of twins reared apart and together. *Am J Cardiol*, 63, 606-609.
7. Bezzina, C.R., Shimizu, W., Yang, P., Koopmann, T.T., Tanck, M.W., Miyamoto, Y., Kamakura, S., Roden, D.M. and Wilde, A.A. (2006) Common sodium channel promoter haplotype in asian subjects underlies variability in cardiac conduction. *Circulation*, 113, 338-344.
8. Chambers, J.C., Zhao, J., Terracciano, C.M., Bezzina, C.R., Zhang, W., Kaba, R., Navaratnarajah, M., Lotlikar, A., Sehmi, J.S., Kooner, M.K. *et al.* (2010) Genetic variation in SCN10A influences cardiac conduction. *Nat Genet*, 42, 149-152.
9. Holm, H., Gudbjartsson, D.F., Arnar, D.O., Thorleifsson, G., Thorgeirsson, G., Stefansdottir, H., Gudjonsson, S.A., Jonasdottir, A., Mathiesen, E.B., Njolstad, I. *et al.* (2010) Several common variants modulate heart rate, PR interval and QRS duration. *Nat Genet*, 42, 117-122.
10. Chen, W.M. and Abecasis, G.R. (2007) Family-based association tests for genomewide association scans. *Am J Hum Genet*, 81, 913-926.
11. Devlin, B., Roeder, K. and Wasserman, L. (2001) Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol*, 60, 155-166.
12. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*, 17, R122-128.
13. Peer, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32, 381-385.
14. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, 24, 2938-2939.
15. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A. *et al.* (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*, 42, 295-302.
16. Calvano, S.E., Xiao, W., Richards, D.R., Felciano, R.M., Baker, H.V., Cho, R.J., Chen, R.O., Brownstein, B.H., Cobb, J.P., Tschoeke, S.K. *et al.* (2005) A network-based analysis of systemic inflammation in humans. *Nature*, 437, 1032-1037.
17. Zhang, B., Schmoyer, D., Kirov, S. and Snoddy, J. (2004) GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *BMC bioinformatics*, 5, 16.
18. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4, 44-57.
19. Pallante, B.A., Giovannone, S., Fang-Yu, L., Zhang, J., Liu, N., Kang, G., Dun, W., Boyden, P.A. and Fishman, G.I. Contactin-2 expression in the cardiac Purkinje fiber network. *Circ Arrhythm Electrophysiol*, 3, 186-194.
20. Sreejit, P., Kumar, S. and Verma, R.S. (2008) An improved protocol for primary culture of cardiomyocyte from neonatal mice. *In vitro cellular & developmental biology*, 44, 45-50.
21. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* (San Diego, Calif, 25, 402-408.

22. Lee, P., Morley, G., Huang, Q., Fischer, A., Seiler, S., Horner, J.W., Factor, S., Vaidya, D., Jalife, J. and Fishman, G.I. (1998) Conditional lineage ablation to model human diseases. *Proc Natl Acad Sci USA*, 95, 11371-11376.
23. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M., de Bakker, P.I., Yin, X., Estrada, K., Bis, J.C., Marciante, K., Rivadeneira, F., Noseworthy, P.A. *et al.* (2009) Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet*, 41, 399-406.
24. Pfeufer, A., Sanna, S., Arking, D.E., Muller, M., Gateva, V., Fuchsberger, C., Ehret, G.B., Orru, M., Pattaro, C., Kottgen, A. *et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet*, 41, 407-414.
25. Pfeufer, A., van Noord, C., Marciante, K.D., Arking, D.E., Larson, M.G., Smith, A.V., Tarasov, K.V., Muller, M., Sotoodehnia, N., Sinner, M.F. *et al.* (2010) Genome-wide association study of PR interval. *Nat Genet*, 42, 153-159.
26. Jarvis, M.F., Honore, P., Shieh, C.C., Chapman, M., Joshi, S., Zhang, X.F., Kort, M., Carroll, W., Marron, B., Atkinson, R. *et al.* (2007) A-803467, a potent and selective Nav1.8 sodium channel blocker, attenuates neuropathic and inflammatory pain in the rat. *Proc Natl Acad Sci USA*, 104, 8520-8525.
27. Desplantez, T., Dupont, E., Severs, N.J. and Weingart, R. (2007) Gap junction channels and cardiac impulse propagation. *J Membr Biol*, 218, 13-28.
28. Abriel, H. Cardiac sodium channel Na(v)1.5 and interacting proteins: Physiology and pathophysiology. *J Mol Cell Cardiol*, 48, 2-11.
29. Remme, C.A., Wilde, A.A. and Bezzina, C.R. (2008) Cardiac sodium channel overlap syndromes: different faces of SCN5A mutations. *Trends in cardiovascular medicine*, 18, 78-87.
30. Akopian, A.N., Souslova, V., England, S., Okuse, K., Ogata, N., Ure, J., Smith, A., Kerr, B.J., McMahon, S.B., Boyce, S. *et al.* (1999) The tetrodotoxin-resistant sodium channel SNS has a specialized function in pain pathways. *Nature neuroscience*, 2, 541-548.
31. Saimi, Y. and Kung, C. (2002) Calmodulin as an ion channel subunit. Annual review of physiology, 64, 289-311.
32. Potet, F., Chagot, B., Anghelescu, M., Viswanathan, P.C., Stepanovic, S.Z., Kupershmidt, S., Chazin, W.J. and Balser, J.R. (2009) Functional Interactions between Distinct Sodium Channel Cytoplasmic Domains through the Action of Calmodulin. *J Biol Chem*, 284, 8846-8854.
33. Wolf, C.M. and Berul, C.I. (2006) Inherited conduction system abnormalities – one group of diseases, many genes. *Journal of cardiovascular electrophysiology*, 17, 446-455.
34. Zhu, Y., Gramolini, A.O., Walsh, M.A., Zhou, Y.Q., Slorach, C., Friedberg, M.K., Takeuchi, J.K., Sun, H., Henkelman, R.M., Backx, P.H. *et al.* (2008) TBX5-dependent pathway regulating diastolic function in congenital heart disease. *Proc Natl Acad Sci USA*, 105, 5519-5524.
35. Lebrec, J.J., Stijnen, T. and van Houwelingen, H.C. (2010) Dealing with heterogeneity between cohorts in genome-wide SNP association studies. *Statistical applications in genetics and molecular biology*, 9, Article 8.
36. Wei, L., Hanna, A.D., Beard, N.A. and Dulhunty, A.F. (2009) Unique isoform-specific properties of calsequestrin in the heart and skeletal muscle. *Cell calcium*, 45, 474-484.
37. Terentyev, D., Nori, A., Santoro, M., Viatchenko-Karpinski, S., Kubalova, Z., Gyorke, I., Terentyeva, R., Vedamoorthyrao, S., Blom, N.A., Valle, G. *et al.* (2006) Abnormal interactions of calsequestrin with the ryanodine receptor calcium release channel complex linked to exercise-induced sudden cardiac death. *Circ Res*, 98, 1151-1158.
38. Priori, S.G., Napolitano, C., Memmi, M., Colombi, B., Drago, F., Gasparini, M., DeSimone, L., Coltorti, F., Bloise, R., Keegan, R. *et al.* (2002) Clinical and molecular characterization of patients with catecholaminergic polymorphic ventricular tachycardia. *Circulation*, 106, 69-74.
39. Postma, A.V., Denjoy, I., Hoorntje, T.M., Lupoglazoff, J.M., Da Costa, A., Sebillon, P., Mannens, M.M., Wilde, A.A. and Guicheney, P. (2002) Absence of calsequestrin 2 causes severe forms of catecholaminergic polymorphic ventricular tachycardia. *Circulation research*, 91, e21-26.
40. Wang, Y. and Goldhaber, J.I. (2004) Return of calcium: manipulating intracellular calcium to prevent cardiac pathologies. *Proc Natl Acad Sci USA*, 101, 5697-5698.
41. Vasan, R.S., Glazer, N.L., Felix, J.F., Lieb, W., Wild, P.S., Felix, S.B., Watzinger, N., Larson, M.G., Smith, N.L., Dehghan, A. *et al.* (2009) Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *Jama*, 302, 168-178.
42. Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P.I., Muller, M., Morrison, A.C., Smith, A.V., Isaacs, A., Sanna, S., Dorr, M. *et al.* (2010) Genome-wide association analysis identifies multiple loci related with resting heart rate. *Hum Mol Genet*, 19, 3885-3894.

43. Braz, J.C., Gregory, K., Pathak, A., Zhao, W., Sahin, B., Klevitsky, R., Kimball, T.F., Lorenz, J.N., Nairn, A.C., Liggett, S.B. *et al.* (2004) PKC-alpha regulates cardiac contractility and propensity toward heart failure. *Nature medicine*, 10, 248-254.
44. Baillat, G., Moqrich, A., Castets, F., Baude, A., Bailly, Y., Benmerah, A. and Monneron, A. (2001) Molecular cloning and characterization of phocein, a protein found from the Golgi complex to dendritic spines. *Molecular biology of the cell*, 12, 663-673.
45. Meurs, K.M., Mauceli, E., Lahmers, S., Acland, G.M., White, S.N. and Lindblad-Toh, K. (2010) Genome-wide association identifies a deletion in the 3' untranslated region of striatin in a canine model of arrhythmogenic right ventricular cardiomyopathy. *Human genetics*, 128, 315-324.
46. Boogerd, K.J., Wong, L.Y., Christoffels, V.M., Klarenbeek, M., Ruijter, J.M., Moorman, A.F. and Barnett, P. (2008) MSX1 and MSX2 are functional interacting partners of T-box factors in the regulation of Connexin43. *Cardiovasc Res*, 78, 485-493.
47. Hoogaars, W.M., Tessari, A., Moorman, A.F., de Boer, P.A., Hagoort, J., Soufan, A.T., Campione, M. and Christoffels, V.M. (2004) The transcriptional repressor TBX3 delineates the developing central conduction system of the heart. *Cardiovasc Res*, 62, 489-499.
48. Singh, R., Horsthuis, T., Farin, H.F., Grieskamp, T., Norden, J., Petry, M., Wakker, V., Moorman, A.F., Christoffels, V.M. and Kispert, A. (2009) TBX20 interacts with smads to confine tbx2 expression to the atrioventricular canal. *Circ Res*, 105, 442-452.
49. Posch, M.G., Gramlich, M., Sunde, M., Schmitt, K.R., Lee, S.H., Richter, S., Kersten, A., Perrot, A., Panek, A.N., Al Khatib, I.H. *et al.* (2009) A gain-of-function TBX20 mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects. *Journal of medical genetics*, 47, 230-235.
50. Bakker, M.L., Boukens, B.J., Mommersteeg, M.T., Brons, J.F., Wakker, V., Moorman, A.F. and Christoffels, V.M. (2008) Transcription factor TBX3 is required for the specification of the atrioventricular conduction system. *Circ Res*, 102, 1340-1349.
51. Riley, P., Anson-Cartwright, L. and Cross, J.C. (1998) The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nat Genet*, 18, 271-275.
52. Reamon-Buettner, S.M., Ciribilli, Y., Traverso, I., Kuhls, B., Inga, A. and Borlak, J. (2009) A functional genetic study identifies HAND1 mutations in septation defects of the human heart. *Hum Mol Genet*, 18, 3567-3578.
53. Breckenridge, R.A., Zuberi, Z., Gomes, J., Orford, R., Dupays, L., Felkin, L.E., Clark, J.E., Magee, A.I., Ehler, E., Birks, E.J. *et al.* (2009) Overexpression of the transcription factor Hand1 causes predisposition towards arrhythmia in mice. *J Mol Cell Cardiol*, 47, 133-141.
54. Rentschler, S., Zander, J., Meyers, K., France, D., Levine, R., Porter, G., Rivkees, S.A., Morley, G.E. and Fishman, G.I. (2002) Neuregulin-1 promotes formation of the murine cardiac conduction system. *Proc Natl Acad Sci USA*, 99, 10464-10469.
55. Hofer, A., Saez, J.C., Chang, C.C., Trosko, J.E., Spray, D.C. and Dermietzel, R. (1996) c-erbB2/neu transfection induces gap junctional communication incompetence in glial cells. *J Neurosci*, 16, 4311-4321.
56. Besson, A. and Yong, V.W. (2000) Involvement of p21(WAF1/Cip1) in protein kinase C alpha-induced cell cycle progression. *Molecular and cellular biology*, 20, 4580-4590.
57. Wilkinson, L., Kolle, G., Wen, D., Piper, M., Scott, J. and Little, M. (2003) CRIM1 regulates the rate of processing and delivery of bone morphogenetic proteins to the cell surface. *J Biol Chem*, 278, 34181-34188.
58. Kolle, G., Georgas, K., Holmes, G.P., Little, M.H. and Yamada, T. (2000) CRIM1, a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogenesis. *Mechanisms of development*, 90, 181-193.
59. Pardali, K., Kowanetz, M., Heldin, C.H. and Moustakas, A. (2005) Smad pathway-specific transcriptional regulation of the cell cycle inhibitor p21(WAF1/cip1). *Journal of cellular physiology*, 204, 260-272.
60. Laederich, M.B., Funes-Duran, M., Yen, L., Ingalla, E., Wu, X., Carraway, K.L., 3rd and Sweeney, C. (2004) The leucine-rich repeat protein LRIG1 is a negative regulator of ErbB family receptor tyrosine kinases. *J Biol Chem*, 279, 47050-47056.
61. Minakuchi, M., Kakazu, N., Gorrin-Rivas, M.J., Abe, T., Copeland, T.D., Ueda, K. and Adachi, Y. (2001) Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. *European journal of biochemistry/FEBS*, 268, 1340-1351.
62. Zhao, J. and Zhong, C.J. (2009) A review on research progress of transketolase. Neuroscience bulletin, 25, 94-99.

63. Fedi, P., Bafico, A., Nieto Soria, A., Burgess, W.H., Miki, T., Bottaro, D.P., Kraus, M.H. and Aaronson, S.A. (1999) Isolation and biochemical characterization of the human Dkk-1 homologue, a novel inhibitor of mammalian Wnt signaling. *J Biol Chem*, 274, 19465-19472.
64. Ai, Z., Fischer, A., Spray, D.C., Brown, A.M. and Fishman, G.I. (2000) Wnt-1 regulation of connexin43 in cardiac myocytes. *J Clin Invest*, 105, 161-171.
65. Korol, O., Gupta, R.W. and Mercola, M. (2008) A novel activity of the Dickkopf-1 amino terminal domain promotes axial and heart development independently of canonical Wnt inhibition. *Dev Biol*, 324, 131-138.
66. Tsai, I.C., Amack, J.D., Gao, Z.H., Band, V., Yost, H.J. and Virshup, D.M. (2007) A Wnt-CK1varepsilon-Rap1 pathway regulates gastrulation by modulating SIPA1L1, a Rap GTPase activating protein. *Developmental cell*, 12, 335-347.

3.3 Supplement

SUPPLEMENTARY METHODS

Details on participating studies

Fourteen genome-wide association studies (GWAS) consisting of individuals of European descent from Europe and the United States contributed to the discovery phase of this study. To extend our analyses, we genotyped select variants representing nine loci in an additional cohort (PREVEND). All studies received approval from the appropriate institutional review committees, and the subjects in each cohort provided written informed consent.

AGES

The Age, Gene/Environment Susceptibility (AGES) Reykjavik Study was initiated to examine genetic susceptibility and gene/environment interaction as these contribute to phenotypes common in old age, and represents a continuation of the Reykjavik Study cohort begun in 1967 and is comprised of 5776 randomly recruited survivors from the original cohort. QRS interval duration was automatically measured from 12-lead electrocardiograms using the Marquette 12 SL analysis program (General Electric Marquette Medical Division, Milwaukee, Wisconsin, USA).

ARIC

The Atherosclerosis Risk in Communities study *(http://www.cscc.unc.edu/aric/)* includes 15,792 men and women from four communities in the United States (Jackson, Mississippi; Forsyth County, North Carolina; Washington County, Maryland; suburbs of Minneapolis, Minnesota) enrolled in 1987–1989 and prospectively followed. ECGs were recorded at baseline using MAC PC ECG machines (Marquette Electronics) and processed initially by the Dalhousie ECG program in a central laboratory at the EPICORE Center (University of Alberta). Processing was later repeated for the present study using the GE Marquette 12-SL program (2001 version) at the EPICARE Center (Wake Forest University). All ECGs were visually inspected for technical errors and inadequate quality. QRS interval was measured automatically from baseline ECGs.

BRIGHT

The MRC BRIGHT study *(http://www.brightstudy.ac.uk/)* comprises 2000 severely hypertensive probands ascertained from families with multiplex affected sibships or as parent-offspring trios. Case ascertainment and phenotyping has been described previously. Briefly, cases have BP

readings ≥150/100 mmHg based on one reading or ≥145/95 mmHg based on the mean of three readings. Twelve-lead ECG recordings (Siemens-Sicard 440; *http://www.brightstudy.ac.uk/info/sopo4.html*), which produces an automated measurement of the QRS interval, were available for all subjects. All data were transferred from each recruitment centre by electronic modem to electrophysiologists from the West of Scotland Primary Prevention Study (Professor Peter MacFarlane) for central reporting. All individuals included in the analysis were of white British ancestry (up to level of grandparents).

CHS

The Cardiovascular Health Study *(http://www.chs-nhlbi.org)* is a prospective, longitudinal cohort study of risk factors for cardiovascular disease in the elderly, was begun in 1989 and included 4,925 self-described White participants. People 65 years of age or older were recruited from four field centers in the United States. The CHS study sample used in this analysis includes participants without clinically-recognized cardiovascular disease at baseline who described their race as White, consented to genetic testing, and had DNA available for genotyping. Study electrocardiograms were recorded using MAC PC ECG machines (Marquette Electronics, Milwaukee, Wisconsin) in all clinical centers. ECGs were initially processed in a central laboratory at the EPICORE Center (University of Alberta, Edmonton, Alberta, Canada) and during later phases of the study, at the EPICARE Center (Wake Forest University, Winston-Salem, North Carolina). All ECGs were visually inspected for technical errors and inadequate quality. QRS interval was measured using the baseline ECG for eligible subjects. Initial ECG processing was done by the Dalhousie ECG program, and processing was later repeated with the 2001 version of the GE Marquette 12-SL program (GE Marquette, Milwaukee, Wisconsin).

ERF

The Erasmus Rucphen Family study is comprised of a family-based cohort embedded in the Genetic Research in Isolated Populations (GRIP) program in the southwest of the Netherlands. The aim of this program is to identify genetic risk factors for the development of complex disorders. In ERF, twenty-two families that had a minimum of five children baptized in the community church between 1850 and 1900 were identified with the help of detailed genealogical records. All living descendants of these couples, and their spouses, were invited to take part in the study. Comprehensive interviews, questionnaires, and examinations were completed at a research center in the area; approximately 3,200 individuals participated. Examinations included 12 lead ECG measurements. Electrocardiograms were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the QRS intervals were made using the Modular ECG Analysis System (MEANS). The QRS detector of MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the QRS complexes among the other parts of the signal. Data collection started in June 2002 and was completed in February 2005. In the current analyses, 1,466 participants for whom complete phenotypic, genotypic and genealogical information was available were studied.

FHS

The Framingham Heart Study *(http://www.framinghamheartstudy.org/)* is a community-based, longitudinal cohort study comprising three generations of individuals in multigenerational pedigrees and additional unrelated individuals. The current study included individuals from Generation 1 (11th examination), Generation 2 (1st examination) and Generation 3 (1st examination). In FHS, paper electrocardiograms recorded on Marquette machines were scanned and digital caliper measurements were made using proprietary software (eResearchTechnology, generations 1 and 2) or using Rigel 1.7.2. (AMPS, LLC, New York, NY, USA, generation 3). The QRS duration was measured from the Q-onset to S-offset in two cardiac cycles from lead II and averaged.

KORA F3 and S4

The KORA study is a series of independent population-based epidemiological surveys of participants living in the city of Augsburg, Southern Germany, or the two adjacent counties. All survey participants are residents of German nationality identified through the registration office and aged between 25 and 74 years at recruitment. The baseline survey KORA S3 was conducted in the years 1994/95 and KORA S4 in 1999-2001. 3,006 participants from KORA S3 were re-examined in a 10-year follow-up (KORA F3) in the years 2004/05. Genomewide data for the analysis of the length of the QRS interval is available for random subsets of 1,644 persons from KORA F3 and 1,814 study participants from KORA S4. In both studies, 12-lead resting electrocardiograms were recorded with digital recording systems (F3: Mortara Portrait, Mortara Inc., Milwaukee, USA, S4: Hörmann Bioset 9000, Hörmann Medizinelektronik, Germany).

KORCULA

The KORCULA study sampled Croatians from the Adriatic island of Korcula, between the ages of 18 and 88. The fieldwork was performed in 2007 in the eastern part of the island, targeting healthy volunteers from the town of Korčula and the villages of Lumbarda, Žrnovo and Račišće. Mortara ELI 350 was used in ECG recording.

MICROS

The MICROS study *(http://www.biomedcentral.com/1471-2350/8/29)* is part of the genomic health care program 'GenNova' and was carried out in three villages of the Val Venosta on the populations of Stelvio, Vallelunga and Martello. This study was an extensive survey carried out in South Tyrol (Italy) in the period 2001-2003. Study participants were volunteers from three isolated villages located in the Italian Alps, in a German-speaking region bordering with Austria and Switzerland. Due to geographical, historical and political reasons, the entire region experienced a prolonged period of isolation from surrounding populations. Genotyping was performed on just under 1,400 participants with 1,334 available for analysis after data cleaning. Information on participants' health status was collected through a standardized questionnaire and clinical examinations, including digitized ECG measurements (Mortara Portrait, Mortara Inc., Milwaukee, USA). Individuals with identified U-waves were excluded from analysis. The Mortara portrait

determines QRS complex by a proprietary algorithm (Michelucci 2002). Laboratory data were obtained from standard blood analyses.

ORCADES

The Orkney Complex Disease Study (ORCADES) is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney. Genetic diversity in this population is decreased compared to Mainland Scotland, consistent with high levels of endogamy historically. Participants included here were aged 18-92 years and came from a subgroup of ten islands. The Cardioview ECG device was used in the phenotyping.

ROTTERDAM STUDY (RS1 and RS2)

The Rotterdam Study is a prospective population-based cohort study comprising 7,983 subjects aged 55 years or older (RS-I), which started in 1990. In 2000-2001, an additional 3,011 individuals aged 55 years or older were recruited (RS-II) *(http://www.erasmus-epidemiology.nl /research/ergo.htm)*. In the RS-I and RS-II, electrocardiograms were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the QRS intervals were made using the Modular ECG Analysis System (MEANS). The QRS detector of MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the QRS complexes among the other parts of the signal.

SHIP

The Study of Health in Pomerania *(http://ship.community-medicine.de)* is a longitudinal population-based cohort study in West Pomerania, a region in the northeast of Germany. From the total population comprising 212,157 inhabitants in 1995, a two-stage stratified cluster sample of adults aged 20 to 79 years was drawn. From the net sample of 6,265 eligible subjects, 4,308 subjects (2192 women) of Caucasian origin participated in the baseline examination, SHIP-0 (response 68.8%). For the present analyses both electrocardiographic and genotyping data were available from 2,985 participants of the SHIP baseline cohort without exclusion criteria. QRS intervals in SHIP were measured from digitally stored electrocardiograms (Personal 120LD, Esaote, Genova, Italy) using MEANS according to the method described above for the RS.

SPLIT

The SPLIT study samples Croatians from the town of Split, between the ages 18 and 85. The sampling started in 2008, and continues throughout 2010. Mortara ELI 350 was used in ECG recording.

TWINSUK

The Twins UK Registry *(http://www.twinsuk.ac.uk)* comprises unselected, mostly female volunteers ascertained from the general population through national media campaigns in the UK. Means and ranges of quantitative phenotypes in Twins UK were similar to an age-matched singleton sample from the general population. Zygosity was determined by standardized ques-

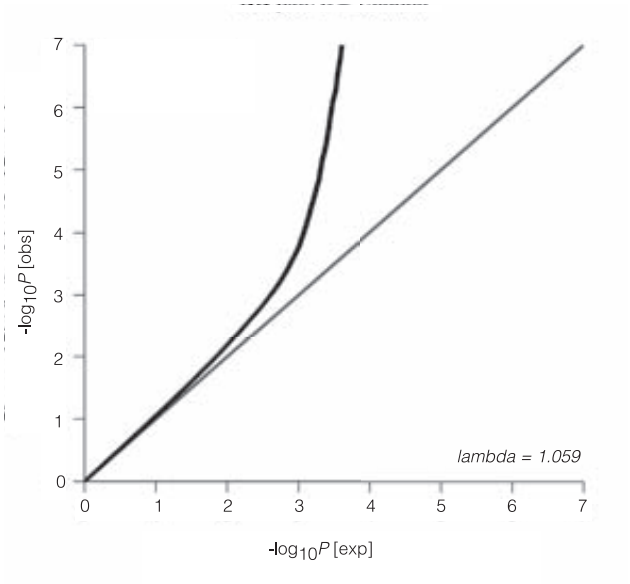
tionnaire and confirmed by DNA fingerprinting. QRS duration data were available on 2,726 of these individuals measured automatically by the Cardiofax ECG-9020K (Nihon Kohden UK Ltd., Middlesex, UK).

PREVEND

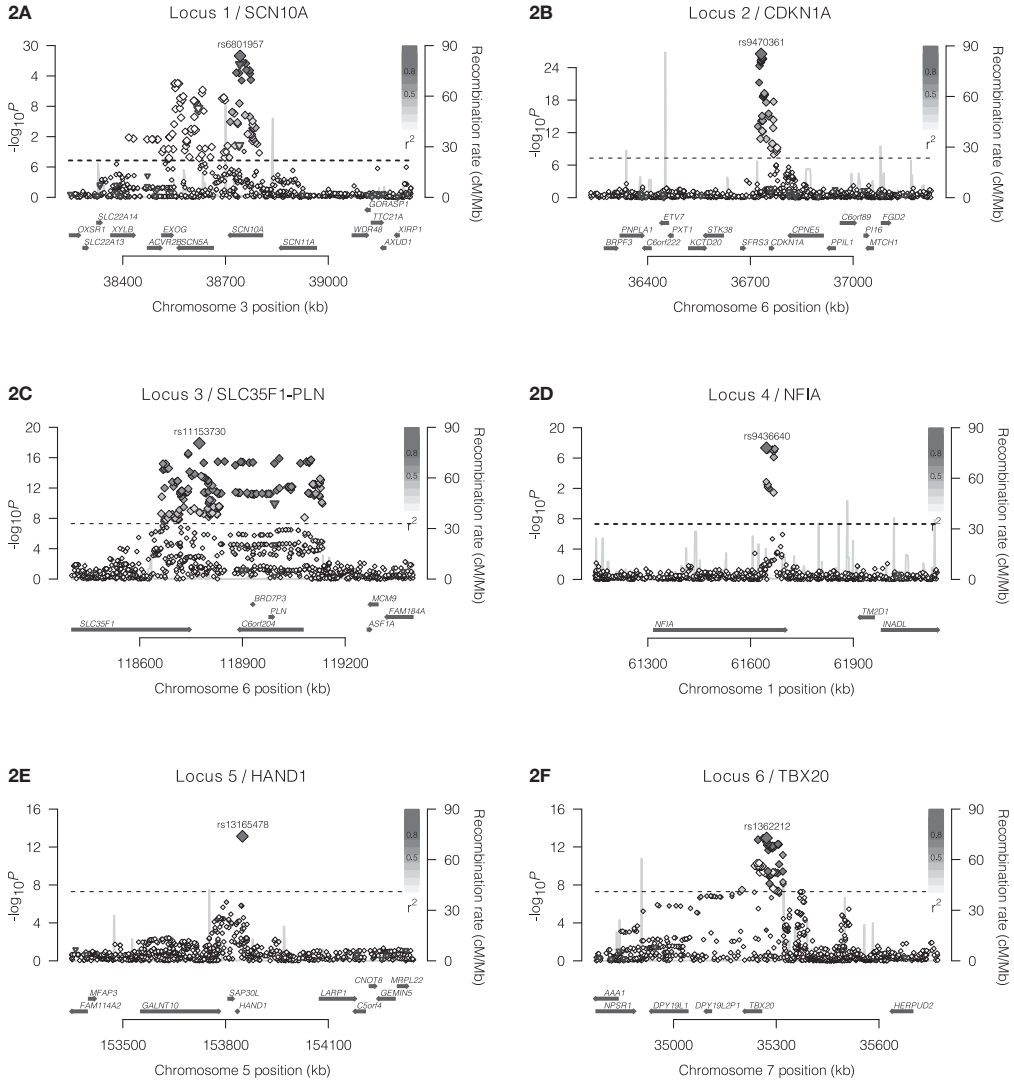
The Prevention of REnal and Vascular ENd stage Disease (PREVEND) study is an ongoing prospective study investigating the natural course of increased levels of urinary albumin excretion and its relation to renal and cardiovascular disease. Inhabitants 28 to 75 years of age (n=85,421) in the city of Groningen, The Netherlands, were asked to complete a short questionnaire, 47% responded, and individuals were then selected with a urinary albumin concentration of at least 10 mg/L (n = 7,768) and a randomly selected control group with a urinary albumin concentration less than 10 mg/L (n = 3,395). Details of the protocol have been described elsewhere *(www.prevend.org)*. Standard 12-lead electrocardiograms were recorded with CardioPerfect equipment (Cardio Control; currently Welch Allyn, Delft, The Netherlands) and digital measurements of the QRS intervals were made using the Modular ECG Analysis System (MEANS). The QRS detector of MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the QRS complexes among the other parts of the signal.

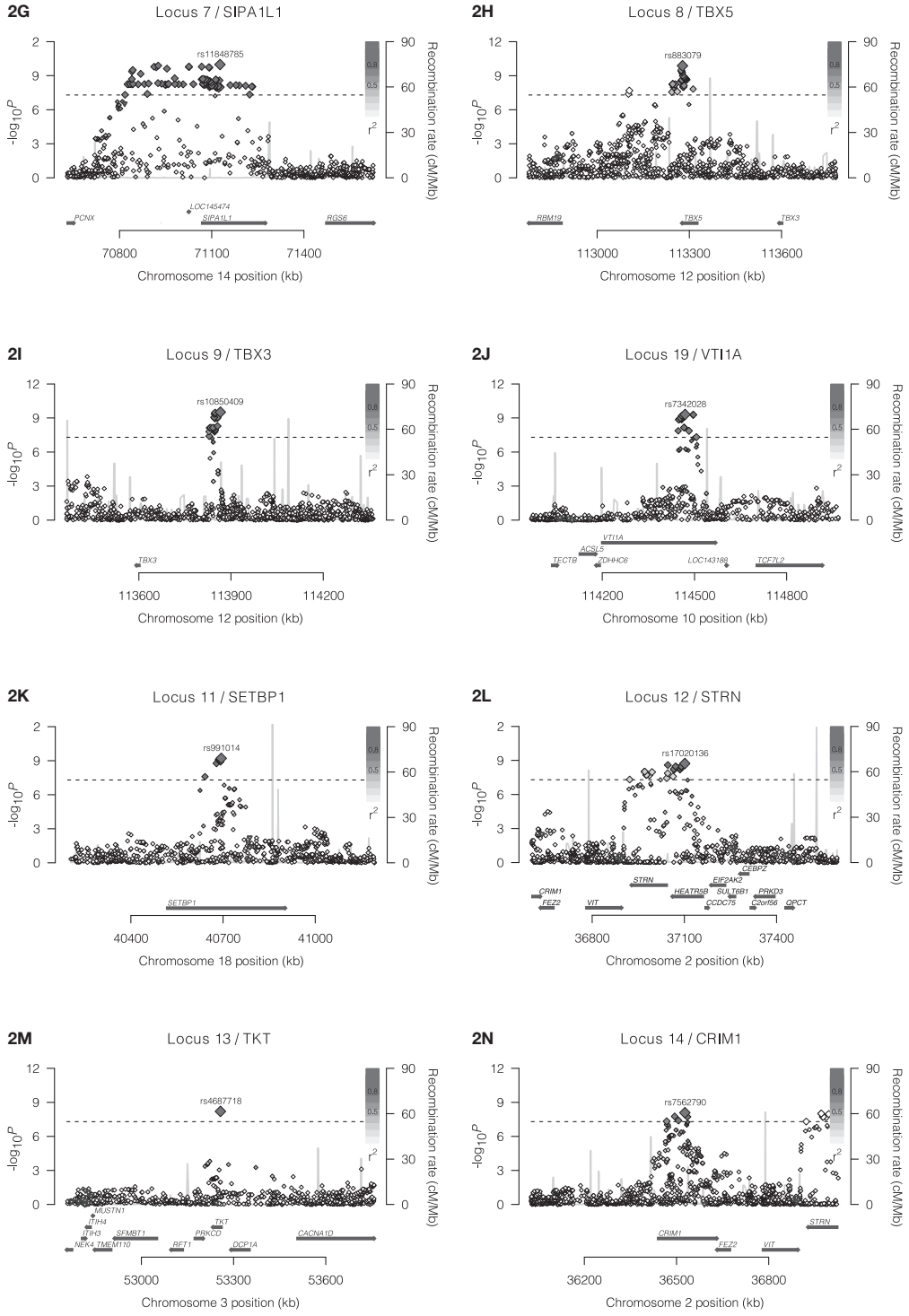
SUPPLEMENTARY FIGURES

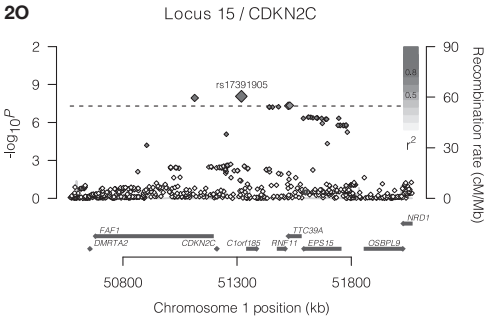
Supplementary Figure 1 *The quantile-quantile (Q-Q) plots demonstrate robust behavior in the bulk of the distribution (lower-left corner) (consistent with a modest λ_{gc} of 1.05). In the tail of the distribution, we observe a departure away from the null hypothesis, presumably due to the presence of true associations.*

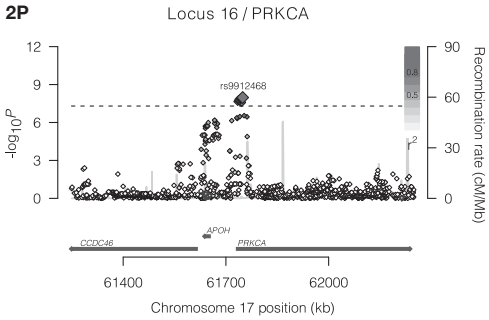


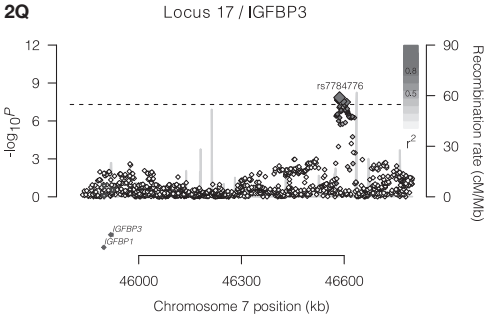
Supplementary Figure 2 *Regional association plots. Association results at each significantly associated locus. Loci are displayed in the order listed in Table 1. Each SNP is plotted with respect to its chromosomal location (x-axis) and its P-value (y-axis on the left). Each panel spans approximately ±500 kb around each index SNP and has known gene transcripts annotated at the bottom. The SNPs are colored according to their degree of linkage disequilibrium (r^2) with the index variant which is highlighted with a diamond and displayed by rs number and significance level achieved in the meta-analysis. The triangles indicate coding region SNPs. The tall spikes indicate the recombination rate (y-axis on the right) at that region of the chromosome.*

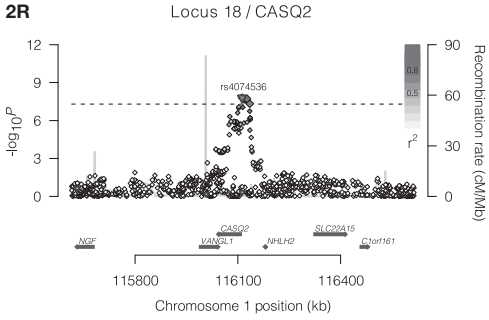


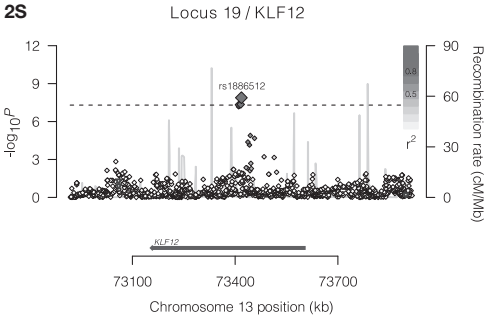

2G
Locus 7 / SIPA1L1
rs11848785
PCNX
LOC145474
SIPA1L1
RGS6
70800
71100
71400
Chromosome 14 position (kb)
2H
Locus 8 / TBX5
rs883079
RBM19
TBX5
TBX3
113000
113300
113600
Chromosome 12 position (kb)
2I
Locus 9 / TBX3
rs10850409
TBX3
113600
113900
114200
Chromosome 12 position (kb)
2J
Locus 19 / VTI1A
rs7342028
VTI1A
ACSL5
TECTB
ZDHHC6
LOC143188
TCF7L2
114200
114500
114800
Chromosome 10 position (kb)
2K
Locus 11 / SETBP1
rs991014
SETBP1
40400
40700
41000
Chromosome 18 position (kb)
2L
Locus 12 / STRN
rs17020136
CEBPZ
STRN
EIF2AK2
CRIM1
HEATR5B
SULT6B1
PRKD3
FEZ2
VIT
CCDC75
C2orf56
QPCT
36800
37100
37400
Chromosome 2 position (kb)
2M
Locus 13 / TKT
rs4687718
MUSTN1
TIH4
TKT
TIH3
SEMBT1
PRKCD
CACNA1D
NEK4
TMEM110
RFT1
DCP1A
53000
53300
53600
Chromosome 3 position (kb)
2N
Locus 14 / CRIM1
rs7562790
STRN
CRIM1
FEZ2
VIT
36200
36500
36800
Chromosome 2 position (kb)
-log10P
Recombination rate (cM/Mb)
r2

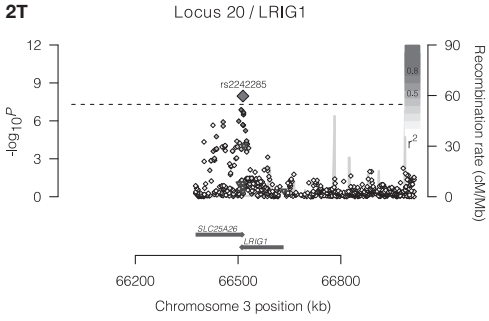

2O
Locus 15 / CDKN2C
rs17391905
-log10P
Recombination rate (cM/Mb)
r2
FAF1
DMRTA2
CDKN2C
C1orf185
RNF11
EPS15
TTC39A
OSBPL9
NRD1
50800
51300
51800
Chromosome 1 position (kb)

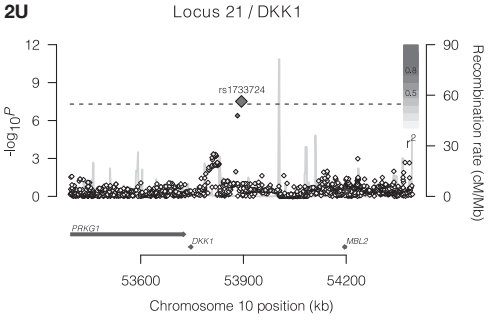

2P
Locus 16 / PRKCA
rs9912468
-log10P
Recombination rate (cM/Mb)
r2
CCDC46
APOH
PRKCA
61400
61700
62000
Chromosome 17 position (kb)

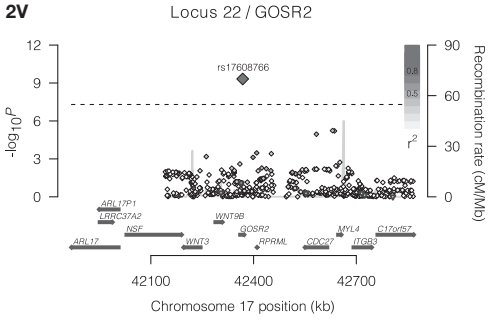

2Q
Locus 17 / IGFBP3
rs7784776
-log10P
Recombination rate (cM/Mb)
r2
IGFBP3
IGFBP1
46000
46300
46600
Chromosome 7 position (kb)


2R
Locus 18 / CASQ2
rs4074536
-log10P
Recombination rate (cM/Mb)
r2
NGF
CASQ2
VANGL1
NHLH2
SLC22A15
C1orf161
115800
116100
116400
Chromosome 1 position (kb)

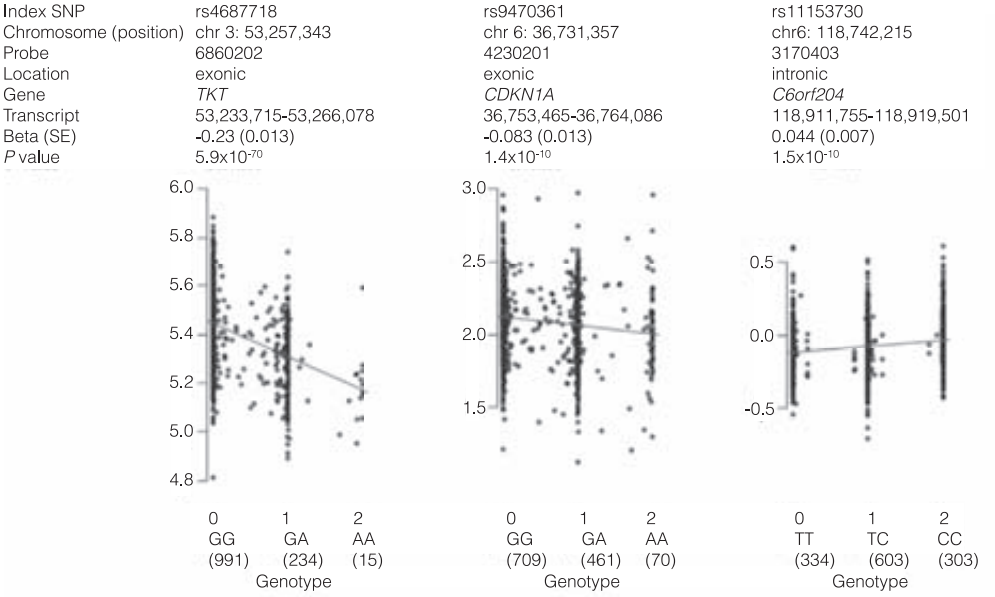

2S
Locus 19 / KLF12
rs1886512
-log10P
Recombination rate (cM/Mb)
r2
KLF12
73100
73400
73700
Chromosome 13 position (kb)


2T
Locus 20 / LRIG1
rs2242285
-log10P
Recombination rate (cM/Mb)
r2
SLC25A26
LRIG1
66200
66500
66800
Chromosome 3 position (kb)

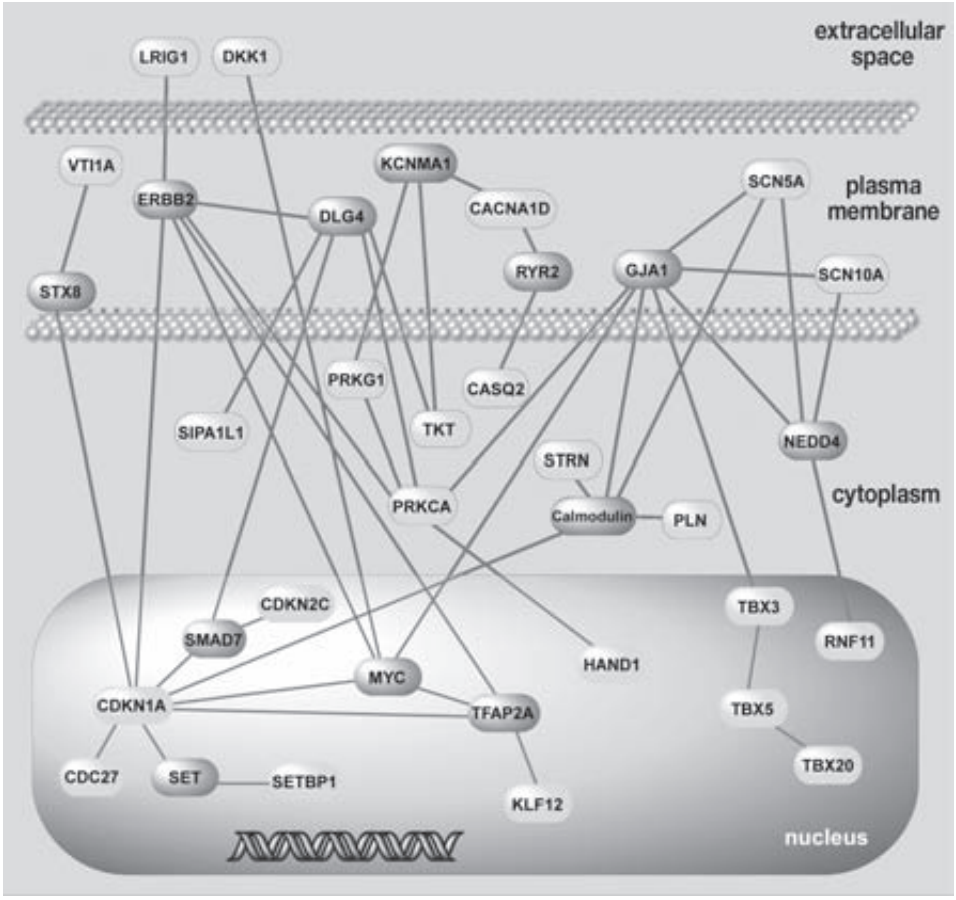

2U
Locus 21 / DKK1
rs1733724
-log10P
Recombination rate (cM/Mb)
r2
PRKG1
DKK1
MBL2
53600
53900
54200
Chromosome 10 position (kb)


2V
Locus 22 / GOSR2
rs17608766
-log10P
Recombination rate (cM/Mb)
r2
ARL17P1
LRRC37A2
ARL17
NSF
WNT9B
WNT3
GOSR2
RPRML
MYL4
CDC27
C17orf57
ITGB3
42100
42400
42700
Chromosome 17 position (kb)

Supplementary Figure 3 *Cis expression-genotype association analysis. The most striking cis eQTLs were observed for probes in exonic regions of TKT (rs4687718, P=5.87×10^{-70}) and CDKN1A (rs9470361, P=1.41×10^{-10}) and an intronic probe for C6orf204 near PLN (rs11153730, P=1.54×10^{-10}). The y-axis indicates normalized expression data and the x-axis indicates the dosage genotype values. NCBI genomic build 36 was used in probe numbering.*



Supplementary Figure 4 *Network map. In silico relational network linking the loci associated with QRS interval duration. Most loci meeting genome-wide significance mapped to this network (shown in light grey). For loci where either multiple genes were independently associated with QRS interval duration (SCN5A and SCN10A in locus 1) or where it was difficult to discern to which of several genes the association signal might map (loci 3, 5, 12, 13, 15, 21, or 22), several genes (listed in Table 1 for each of the loci) were included in the model. Of these seven latter loci, three (loci 13, 15, and 21) had 2 gene members map to the network. All interactions depicted in this relational network represent direct gene product interactions obtained from curated databases. To ensure that the interactome spanned across the maximum number of QRS-associated genes, several nodes (shown in dark grey) were added to the network based on the strength of their connectivity with the original loci. Linker nodes were added only if they connected to a minimum of two network nodes, without bias in regards to function. The minimum number of linkers required to connect network nodes was selected. Our network analysis shows that many of the genetic loci associated with QRS duration interact with each other and are likely to be functionally linked, although the relevance of these relationships in the human heart needs to be experimentally assessed.*



SUPPLEMENTARY TABLES

Supplementary Table 1A *Study participant characteristics.*

Characteristic	AGES	ARIC	BRIGHT	CHS	SPLIT	KORCULA	ERF	FHS
N, Participants with ECG and genotype data	3188	9013	1566	3271	433	428	1591	7950
N, Participants after exclusion	2251	8085	1302	2845	395	378	1466	7499
Sex, women, %	64.0	54.4	63.0	62.8	63.5	62.5	59.5	54.1
Age, yars, mean	76.0	54.0	58.8	72.1	49.3	54.5	47.8	39.2
Age, years, range	66-95	44-66	21-89	65-94	18-85	18-88	18-83	19-79
QRS interval, ms, mean	90.4	96.2	92.9	88.3	96.1	95.9	97.1	87.2
QRS interval, ms, range	60-120	61-120	66-118	56-120	70-120	76-119	68-120	59-120
Height, cm, mean	166.1	168.6	170.0	164.3	171.2	168.0	166.5	168.9
BMI, kg/m2, mean	27.0	26.8	27.4	26.2	26.7	28.0	26.7	26.2
Hypertension, %	77.8	24.1	100	51.9	25.2	28.8	15	8.3
Diabetes mellitus, %	10.4	7.6	0.1	11.2	3.6	6.1	2.8	1.6
Heart rate, bpm, mean	66.6	66.5	63.0	64.7	65.7	65.8	63.1	68.0

Supplementary Table 1A continued

Characteristic	KORA S4*	KORA F3*	MICROS	ORCADES	RS 1*	RS 2*	SHIP	TwinsUK	PREVEND**
N, Participants with ECG and genotype data	1814	1644	1244	719	5974	2157	3548	2687	7500
N, Participants after exclusion	1654	1393	1061	690	4081	1838	2985	2484	7170
Sex, women, %	52.5	51.7	57.8	54.9	62.9	57.9	52.6	95.0	53.0
Age, years, mean	53.5	61.4	44.2	53.3	68.3	64.8	48.1	51.3	48.7
Age, years, range	25-74	35-79	18-87	18-92	55-101	55-95	20-81	17-83	28-75
QRS interval, ms, mean	91.5	92.4	94.3	90.0	96.6	97.5	97.1	87.7	96.2
QRS interval, ms, range	64-120	62-120	69-120	60-120	64-120	70-120	60-120	60-120	50-20
Height, cm, mean	167.6	167.1	166.3	167.3	166.7	168.2	169.1	163.0	173.0
BMI, kg/m^2, mean	27.6	27.9	25.3	27.6	26.3	27.3	27.0	25.7	26.0
Hypertension, %	16.6	41.6	15.5	24.9	51.8	58.5	49.5	16.4	31.1
Diabetes mellitus, %	3.0	8.8	3.1	2.7	8.6	9.3	6.3	1.5	3.2
Heart rate, bpm, mean	64.9	64.1	68.0	60.7	70.2	69.7	72.0	66.5	69.0

*The KORA and RS studies both have two separate cohorts.

**PREVEND study participants were used for the candidate SNP extension genotyping only. All other studies were included in the GWAS meta-analysis.

Supplementary Table 1B *Study genome-wide genotyping characteristics.*

Characteristic	AGES	ARIC	BRIGHT	CHS	SPLIT	KORCULA	ERF	FHS
Array	Illumina CNV370	Affy 6.0	Affy 500K	Illumina CNV370	Illumina CNV370	Illumina CNV370	Illumin 318K, 370K, Affy 250K	Affy 500K, 50K MIP
Genotype calling software	Bead Studio	Birdseed	CHIAMO	Bead Studio	Bead Studio	Bead Studio	BeadStudio	BRLMM
SNP call rate exclusion	<97%	<95%	<95%	<95%	<98%	<98%	<98%	<=97%
SNP MAF exclusion	<0.01	<1%	<1%	<1%	<1%	<1%	NA	<0.01
P HWE exclusion	$<10x10^{-6}$	$<10x10^{-5}$	$<10x10^{-7}$	$<10x10^{-5}$	$<10x10^{-6}$	$<10x10^{-6}$	$<10x10^{-6}$	$<10x10^{-6}$
Imputation software	Mach1 v1.0.16	Mach1 v1.0.16	IMPUTE	BIMBAM	Mach v1.0.15	Mach v1.0.15	Mach v1.0.15	Mach v1.0.15
NCBI Build for imputation	Build 36	Build 35	Build 35	Build 36	Build 36	Build 36	Build 36	Build 36
GWAS statistical analysis	ProbABEL, R	Mach2QTL + plink	SNPTEST	R	GeneABEL, ProbABEL, R	GeneABEL, ProbABEL, R	GeneABEL, ProbABEL	R
Related individuals?	No	No	No	No	Yes	Yes	Yes	Yes
Familial adjustment method	N/A	N/A	N/A	N/A	Mmscore in ProbABEL	Mmscore in ProbABEL	Mmscore in ProbABEL	Kinship package in R
Genomic control factor (λ)	1.01	1.01	1.00	1.03	1.02	1.03	1.01	1.03

Supplementary Table 1B continued

Characteristic	KORA S4	KORA F3	MICROS	ORCADES	RS 1	RS 2	SHIP	TwinsUK
Array	Affy 6.0	Affy 500K	Illumina HumHap 300v2	Illumina CNV370 & Illumina HumHap 300v2	Illumina550K	Illumina550K Duo, 610KQuad	Affy 6.0	Illumina Hap300 Duo, Hap300, Hap550, Hap610
Genotype calling software	Birdseed	BRLMM	BeadStudio	Bead Studio	BeadStudio	Genome-Studio	Birdseed	Illuminus
SNP call rate exclusion	<93%	<95%	<98%	<98%	<98%	<98%	None	<95%
SNP MAF exclusion	<1%	<1%	<1%	<1%	<1%	<1%	None	<1%
P HWE exclusion	$<10x10^{-5}$	$<10x10^{-5}$	$<10x10^{-6}$	$<10x10^{-6}$	$<10x10^{-6}$	$<10x10^{-6}$	None	$<10x10^{-4}$
Imputation software	Mach1 v1.0.16	Mach1 v1.010	Mach v1.0.16	Mach 1.0 ML	Mach v1.0.15	Mach v1.0.16	Impute v0.5.0	Impute v0.3.2
NCBI Build for imputation	Build 36	Build 35	Build 36	Build 36	Build 36	Build 36	Build 36	Build 36
GWAS statistical analysis	ProbABEL v0.1-2	ProbABEL v0.1-2	ProbABEL	GeneABEL, ProbABEL, R	Mach2QTL as implemented in GRIMP	Mach2QTL as implemented in GRIMP	SNPTEST v.1.1.5	SNPTEST v.1.1.4
Related individuals?	No	No	Yes	Yes	No	No	No	Yes
Familial adjustment method	N/A	N/A	Mmscore in ProbABEL	Mmscore in ProbABEL	N/A	N/A	N/A	Huber-White robust variance estimation in R
Genomic control factor (λ)	1.01	1.01	1.00	1.00	1.01	1.02	1.04	1.02

Supplementary Table 2B *Mean QRS duration and sample sizes for individuals stratified by QRS >120 ms and specific ventricular conduction defects*

	QRS≤120 ms (mean±sd)	QRS>120 ms (mean±sd)	LBBB (mean±sd)	RBBB (mean±sd)	NIVCD (mean±sd)
ARIC	7996 (96.2±9.3)	213 (138.3±16.2)	26 (157.7±11.7)	62 (148.5±13.4)	125 (129.8±11.3)
Rotterdam	4769 (96.9±10.6)	306 (143.5±17.2)	81 (157.8±12.7)	107 (148.8±14.9)	118 (129.0±8.9)

Excludes individuals with prevalent heart failure or myocardial infarction. sd, standard deviation; LBBB, left bundle branch block; RBBB, right bundle branch block; NIVCD, non-specific intraventricular conduction defect.

Supplementary Table 2B *Effects of a weighted genotype risk score on QRS >120 ms and stratified on specific ventricular conduction defects*

	QRS >120ms		LBBB		RBBB		NIVCD	
	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
ARIC	**1.12 (1.04-1.22)**	**0.003**	1.11 (0.89-1.40)	0.34	1.00 (0.86-1.16)	0.98	**1.20 (1.08-1.33)**	**0.0006**
Rotterdam	1.04 (0.97-1.12)	0.21	1.00 (0.88-1.13)	0.97	1.02 (0.91-1.14)	0.73	1.11 (0.99-1.23)	0.07
Combined	**1.08 (1.02-1.13)**	**0.004**	1.02 (0.92-1.14)	0.67	1.01 (0.93-1.11)	0.79	**1.15 (1.07-1.25)**	**0.0002**

Excludes individuals with prevalent heart failure or myocardial infarction. **Bold** indicates significant results ($P < 0.05$). LBBB, left bundle branch block; RBBB, right bundle branch block; NIVCD, non-specific intraventricular conduction defect; OR, odds ratio; CI, confidence interval.

Supplementary Table 3A *Effect of QRS duration hits on PR interval and QT interval*

Locus	Nearest Gene	Index SNP	Chr	Position	Coded/Non-coded Allele	QRS β	QRS SE	PR β
1	*SCN10A*	rs6801957	3	38,742,319	**T**/C	0.77	0.07	**3.79**
	SCN10A	rs9851724	3	38,694,939	**C**/T	-0.66	0.07	**-1.70**
	SCN5A	rs11710077	3	38,632,903	**T**/A	-0.84	0.09	**-1.80**
	SCN5A	rs11708996	3	38,608,927	**C**/G	0.79	0.10	**3.04**
2	*CDKN1A*	rs9470361	6	36,731,357	**A**/G	0.87	0.08	0.74
3	*C6orf204/ SLC35F1/PLN/ BRD7P3*	rs11153730	6	118,774,215	**C**/T	0.59	0.07	-0.56
4	*NFIA*	rs9436640	1	61,585,698	**G**/T	-0.59	0.07	0.39
5	*HAND1/SAP30L*	rs13165478	5	153,849,233	**A**/G	-0.55	0.07	0.39
6	*TBX20*	rs1362212	7	35,078,546	**A**/G	0.69	0.09	0.50
7	*SIPA1L1*	rs11848785	14	71,127,108	**G**/A	-0.50	0.08	0.66
8	*TBX5*	rs883079	12	113,255,960	**C**/T	0.49	0.08	**1.15**
9	*TBX3*	rs10850409	12	113,844,460	**A**/G	-0.49	0.08	**1.70**
10	*VTI1A*	rs7342028	10	114,469,252	**T**/G	0.48	0.08	0.20
11	*SETBP1*	rs991014	18	40,693,884	**T**/C	0.42	0.07	0.37
12	*HEATR5B/STRN*	rs17020136	2	37,159,666	**C**/T	0.51	0.08	-0.34
13	*TKT/CACNA1D/ PRKCD*	rs4687718	3	53,257,343	**A**/G	-0.63	0.11	-0.27
14	*CRIM1*	rs7562790	2	36,585,206	**G**/T	0.39	0.07	-0.32
15	*C1orf185/RNF11/ CDKN2C/FAF1*	rs17391905	1	51,258,161	**G**/T	-1.35	0.23	**-3.01**
16	*PRKCA*	rs9912468	17	61,748,819	**G**/C	0.39	0.07	0.39
17	*IGFBP3*	rs7784776	7	46,393,385	**G**/A	0.39	0.07	0.17
18	*CASQ2*	rs4074536	1	116,023,009	**C**/T	-0.42	0.07	0.32
19	*KLF12*	rs1886512	13	73,418,187	**A**/T	-0.40	0.07	-0.40
20	*LRIG1/SLC25A26*	rs2242285	3	66,514,292	**A**/G	0.37	0.07	0.55
21	*DKK1*	rs1733724	10	53,893,983	**A**/G	0.49	0.09	0.03
22	*GOSR2*	rs17608766	17	42,368,270	**C**/T	0.53	0.10	0.48

QT interval results are drawn from the QTSCD study. **Bold** indicates significant SNPs after Bonferroni correction for the number of SNPs tested. The bolded allele is the coded allele. Effect size (β) is reported in milliseconds (ms) per copy of the coded allele. Chr, chromosome; AF, coded allele frequency; SE, standard error.

	PR SE	PR *P*	QT β	QT SE	QT *P*
	0.21	**1.80×10^{-73}**	**-0.67**	**0.20**	**1.05×10^{-3}**
	0.22	**7.98×10^{-15}**	**0.95**	**0.21**	**6.66×10^{-6}**
	0.26	**3.18×10^{-12}**	**0.92**	**0.24**	**1.34×10^{-4}**
	0.29	**6.00×10^{-26}**	**-0.93**	**0.28**	**7.78×10^{-4}**
	0.24	2.01×10^{-3}	-0.64	0.24	6.64×10^{-3}
	0.20	6.20×10^{-3}	**1.61**	**0.20**	**5.19×10^{-16}**
	0.20	0.06	-0.44	0.20	0.024
	0.22	0.07	-0.27	0.21	0.2
	0.27	0.07	-0.17	0.27	0.53
	0.23	4.26×10^{-3}	-0.09	0.22	0.67
	0.23	**9.08×10^{-7}**	0.42	0.22	0.06
	0.23	**3.72×10^{-13}**	-0.33	0.23	0.15
	0.23	0.39	-0.22	0.22	0.33
	0.21	0.07	0.06	0.20	0.78
	0.26	0.18	0.43	0.25	0.08
	0.32	0.40	0.11	0.30	0.71
	0.21	0.12	0.20	0.20	0.31
	0.71	**2.09×10^{-5}**	-0.38	0.70	0.59
	0.21	0.06	**-0.92**	**0.20**	**3.66×10^{-6}**
	0.21	0.41	0.15	0.20	0.46
	0.23	0.16	-0.63	0.22	4.47×10^{-3}
	0.22	0.06	0.28	0.22	0.19
	0.21	8.27×10^{-3}	0.07	0.20	0.73
	0.29	0.92	0.84	0.28	2.46×10^{-3}
	0.30	0.12	0.88	0.29	2.86×10^{-3}

Supplementary Table 3B *Effect of PR and QT interval SNPs on QRS duration*

Trait	Locus	Index SNP	Chr	Position	Coded-Non-coded Allele	Trait β	Trait SE	QRS β	QRS SE	QRS *P*
PR interval	*SCN10A*	rs6800541	3	38,749,836	**C**/T	**3.77**	**0.21**	**0.74**	**0.07**	**$5.85x10^{-29}$**
	SCN5A	rs11708996	3	38,608,927	**C**/G	**3.04**	**0.29**	**0.79**	**0.09**	**$1.66x10^{-17}$**
	TBX5-TBX3	rs1896312	12	113,830,807	**C**/T	**1.95**	**0.23**	**-0.44**	**0.07**	**$2.63x10^{-9}$**
	CAV1-CAV2	rs3807989	7	115,973,477	**A**/G	**2.30**	**0.21**	**0.30**	**0.07**	**$5.84x10^{-6}$**
	MEIS1	rs11897119	2	66,625,504	**C**/T	1.36	0.21	0.10	0.07	0.12
	NKX2-5	rs251253	5	172,412,942	**C**/T	-1.49	0.21	0.10	0.07	0.13
	SOX5	rs11047543	12	24,679,606	**A**/G	-2.09	0.29	0.10	0.09	0.29
	ARHGAP24	rs7692808	4	86,860,173	**A**/G	-2.01	0.22	-0.04	0.07	0.60
	WNT11	rs4944092	11	75,587,267	**G**/A	-1.19	0.22	0.04	0.07	0.60
QT interval	*SCN5A*	rs11129795	3	38,568,397	**A**/G	**-1.27**	**0.23**	**0.78**	**0.08**	**$1.95x10^{-24}$**
	PLN	rs11970286	6	118,787,067	**T**/C	**1.64**	**0.20**	**0.55**	**0.07**	**$7.07x10^{-17}$**
	PLN	rs12210810	6	118,759,897	**C**/G	**-3.13**	**0.43**	**-0.75**	**0.15**	**$4.08x10^{-7}$**
	NOS1AP	rs12143842	1	160,300,514	**T**/C	**2.88**	**0.23**	**-0.29**	**0.08**	**$1.25x10^{-4}$**
	ATP1B1	rs10919071	1	167,366,107	**G**/A	-2.05	0.29	-0.21	0.10	0.03
	LIG3	rs2074518	17	30,356,290	**T**/C	-1.23*	0.18*	-0.10	0.07	0.12
	KCNJ2	rs17779747	17	66,006,587	**T**/G	-1.16	0.21	-0.10	0.07	0.15
	KCNE1	rs1805128	21	34,743,550	**T**/C	4.03*	1.58*	-0.29	0.22	0.19
	KCNH2	rs4725982	7	150,268,796	**T**/C	1.58*	0.35*	-0.10	0.08	0.24
	NOS1AP	rs4657178	1	160,477,234	**T**/C	2.19	0.22	-0.08	0.07	0.27
	KCNQ1	rs12296050	11	2,445,918	**T**/C	1.44	0.25	-0.06	0.08	0.45
	NDRG4	rs7188697	16	57,179,679	**G**/A	-1.66	0.23	-0.06	0.08	0.46
	KCNH2	rs2968863	7	150,254,070	**T**/C	-1.35	0.23	0.05	0.08	0.55
	KCNQ1	rs2074238	11	2,441,379	**T**/C	-8.22*	1.05*	0.18	0.34	0.59
	RNF207	rs846111	1	6,201,957	**C**/G	1.49	0.25	-0.04	0.09	0.66
	LITAF	rs8049607	16	11,599,254	**T**/C	1.25	0.22	-0.01	0.07	0.88

QT results are drawn from the QTSCD study, unless otherwise noted. PR results are from.13 **Bold** indicates significant SNPs after Bonferroni correction for the number of SNPs tested. The bolded allele is the coded allele. Effect size (β) is reported in milliseconds (ms) per copy of the coded allele. Chr, chromosome; MAF, minor allele frequency; SE, standard error.

*Genome-wide significant results ($P<5x10^{-8}$) are drawn from the QTGEN study, and standardized beta estimates and SE were converted to ms using SD=17.5 ms.

PART 4

Genome-wide association studies of correlated traits

4 Meta-analysis of genome-wide association studies of correlated traits

Mark Eijgelsheim, Karol Estrada, Christopher Newton-Cheh, André G Uitterlinden, Bruno HCh Stricker, Paul IW de Bakker

Manuscript in preparation

ABSTRACT

Univariate genome-wide association studies meta-analyzed by pooling data from several studies in a collaborative effort are frequently used to identify loci for complex traits. Often multiple correlated quantitative endophenotypes are analyzed without taking benefit of the mutual trait correlation. Multivariate analysis is the solution to make benefit of the underlying correlation structure and to increase the statistical power. In addition, this method will also be more powerful for identifying loci having effects on multiple uncorrelated traits (i.e. pleiotropy). However, this approach is computationally intensive and requires individual subject level information on genotype and phenotype. We sought a method to combine single trait genome-wide association results of correlated traits, and so minimizing the analysis burden but utilizing the additional information that can be derived from the trait correlation. Here we describe a simple approach using z-score test statistic and Fisher's method based meta-analyis and present the results of a simulation study to evaluate its performance. We show that the proposed methods have a power advantage just like bivariate regression analysis in identifying associated genetic loci at several trait correlation levels when compared to single trait analysis. The correlation between bivariate regression analysis results and our proposed application of Fisher's method is very high (Pearson correlation coefficient = 0.996) with slightly more conservative *p*-values for the proposed approximation. In conclusion, the proposed method can be applied without prior knowledge on trait correlation structure in an efficient way using standard computer hardware and software and does not require access to individual level genotype or phenotype data.

INTRODUCTION

It is common practice in genetic association studies to test multiple endophenotypic traits. Common complex disorders are quantitative traits that reflect polygenic liability[1]. For example, our group studies multiple correlated and uncorrelated endophenotypes for sudden cardiac death and arrhythmia that are believed to be close to biology, easy to measure and readily available from body surface electrocardiogram recordings, i.e. QT interval (cardiac repolarization duration), PR interval (interval between atrial and ventricular depolarization), QRS duration (cardiac depolarization) and RR interval (inverse heart rate). The results of several of these individual trait GWAS showed shared quantitative trait loci (QTL) in addition to QTL's private to the single traits[2-7]. The single trait approach ignores information provided by cross-covariance and multivariate analysis methods have been implemented in e.g. PLINK[8], making multivariate genome-wide association analysis feasible that will increase power in case of correlation between traits. Because also uncorrelated traits can share common genetic background, combination of information on different uncorrelated traits can improve power to identify shared associations (QTL's with pleiotropic effects). However, multivariate analysis methods have a major disadvantage, namely that all genome-wide association analysis and consortium meta-analyses have to be redone for every possible combination of traits one wants to study. This comes with a huge burden on computational and personnel resources.

We tested the hypothesis that combining two individual trait genome-wide association (meta-) analysis results would be a feasible and less demanding alternative to bivariate analysis. We here describe the results of a comparison between bivariate analysis and meta-analysis of individual trait results of simulated data showing we can obtain similar results in a more efficient way. This gain in efficiency comes from the requirement of only aggregate result level data for the individual trait and not individual subject level information on genotype and phenotype. Furthermore, the minimal required computational resource is limited to standard up-to-date hardware and software. The approach described can be applied to both correlated and uncorrelated traits without prior knowledge on trait correlation.

METHODS

Proposed method

We propose a method that consists of three steps to combine results of univariate genome-wide association analysis of two individual traits taking the trait correlation into account. The first step consists of performing bidirectional Z-score meta-analysis. Hereto, perform a Z-score meta-analysis on the two original individual trait result sets (further named $Z_{meta\text{-}same}$) to identify loci with same direction effects (formula 1) and a second Z-score meta-analysis after flipping the sign (direction of effect) for all genetic variants in one of the two result sets to identify opposite direction effects ($Z_{meta\text{-}opp}$) (formula 2). Both traits should be weighted equally (i.e. $\omega_1 = \omega_2$).

(1) $Z_{meta-same} = \omega_1 * Z_{Trait1} + \omega_2 * Z_{Trait2}$

(2) $Z_{meta-opp} = \omega_1 * Z_{Trait1} - \omega_2 * Z_{Trait2}$

For each of the two resulting Z-score combinations the inflation of the test statistic distribution, lambda, is calculated by dividing the observed median chi-square value (i.e. Z-score²) over the expected median chi-square under the null hypothesis of no association (i.e. 0.4549364) (formula 3).

(3) $$\lambda = \frac{observed\ [median(Z_{meta}{}^2)]}{expected\ [median(\chi^2)]}$$

A positive trait correlation will result in inflation of the $Z_{meta-same}$ meta-analysis and deflation of the $Z_{meta-opp}$ meta-analysis.

The second step consists of calculating the in-/deflation adjusted test statistic for each Z-score combination by dividing the observed chi-square value by the lambda value (formula 4) and subsequently the corresponding $P_{\lambda\text{-adjusted}}$-values can be derived.

(4) $$\chi^2_{meta-adjusted} = \frac{\chi^2_{meta-observed}}{\lambda}$$

It is important to note for step 3 that the resulting Z-score based meta-analysis test statistic (and thus $P_{\lambda\text{-adjusted}}$) distributions are uncorrelated due to the sign inversion. The analysis can be stopped at this point (see Results and Discussion below) resulting in two *P*-values for each genetic variant representing the evidence for same and opposite directions of effect, respectively. We will refer to this as bidirectional Z-score meta-analysis.

The third and final step consists of combining the two independent $P_{\lambda\text{-adjusted}}$ distributions using Fisher's method for combining independent *P*-values (formula 5)[9].

(5) $$\chi^2_{2k} = -2\Sigma_{i=1}^{k} \log_e(p_i)$$

The resulting chi-square value distribution with four degrees of freedom (two times the number of traits) provides the overall association evidence for each genetic variant taking underlying correlation and directionality of the effects into account. Throughout the paper we will refer to this as Fisher's method.

Simulations

To asses the similarity of the proposed methods we simulated three sets of two correlated standardized traits (mean = 0, standard deviation = 1) at correlation levels of 0.2, 0.4 and 0.6, respectively, for 50,000 subjects using the rmvnorm() function in R v2.11.0. In addition, 50,000 independent variants, with a uniform allele frequency distribution between 0.05 and 0.95, for 50,000 subjects were simulated under the null hypothesis of no association using PLINK v1.07[10].

Subsequently we forced in true positive associations. Both trait 1 and trait 2 were enriched with both associations private to the trait as well as a common set of pleiotropic variants (Table 1). Each trait was enriched with one association explaining 0.5% of the original trait variance, 40 variants each explaining 0.05% of the original trait variance and 200 variants each explaining 0.01% of the original trait variance, all private to one trait. In addition, pleiotropic effects were introduced by forcing associations between variants and both traits with the same parameters. This resulted in a total of 723 forced associations. The sign of the effect was randomly allocated so the pleiotropic simulated effects will have same or opposite direction effects on the two correlated traits.

The simulated datasets were analyzed with standard univariate and bivariate analysis methods as well as the proposed bidirectional Z-score and Fisher methods. Multivariate analysis performed using PLINK v1.06-b[8] was considered the gold standard and used as a reference. Single trait analysis was performed with PLINK v1.07[10]. We assessed the similarity of results between bivariate analysis, the bidirectional Z-score meta-analysis and the Fisher based approximation.

Table 1 *Overview of simulated effects (applied at all three trait correlation levels)*

	Explained proportion of original trait variance	No. of forced associations
Trait 1 specific effects	0.5%	1
	0.05%	40
	0.001%	200
Trait 2 specific effects	0.5%	1
	0.05%	40
	0.01%	200
Pleiotropic effect (effect simulated on both trait 1 and trait 2)	0.5%	1
	0.05%	40
	0.01%	200
Total per trait		482
Total both traits		723

RESULTS

Analysis of simulated data sets under the null

Under the null model with no genotype-phenotype associations all results showed normal behaviour. (Supplementary Figures 1A-D). As expected the two bidirectional Z-score meta-analysis test statistic distributions were uncorrelated (r=0). Both the multivariate association analysis as well as Fisher's method showed expected *P*-value distributions without inflation or false positives. The Pearson correlation between the multivariate model and Fisher's method was 0.996 at all simulated trait correlation levels. Figure 1A graphically represents the similarity between bivariate analysis and Fisher's based approximation at a trait correlation of 0.6.

Figure 1A *Correlation plot for bivariate analysis versus Fisher under the null model of no genotype – phenotype association.*

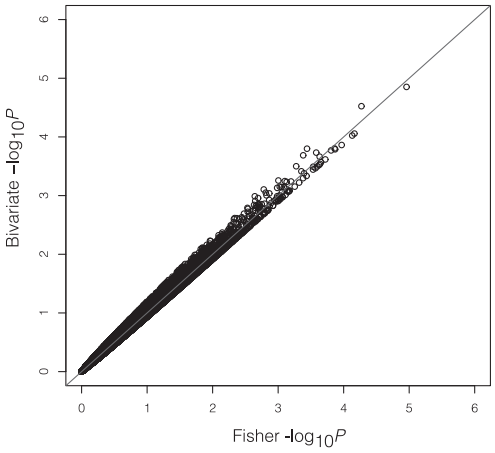
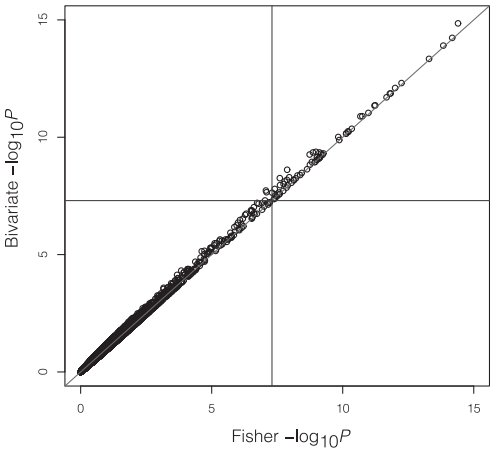


Figure 1B *Correlation plot bivariate analysis versus Fisher for the simulated dataset with forced positive control associations. Axes are truncated at -log10(p) = 15 for display purposes.*



We then generated the univariate and multivariate analysis results with the positive control associations for each of the correlated trait sets. The phenotype correlation between the two traits was slightly attenuated to 0.55, 0.37 and 0.19, respectively, as a result of forcing in the positive control associations (Table 2). In Figure 2 we report the number of markers reaching the genome-wide significance level commonly used in GWA studies, $P<5\times10^{-8}$, for univariate analysis, bivariate analysis, bidirectional Z-score analysis (using a stricter threshold of $P<2.5\times10^{-8}$) and Fisher's method at all correlation levels. Bivariate analysis is superior over univariate analysis in identifying QTL's at all correlation levels. The Fisher based *P*-values are highly correlated with bivariate association results (Pearson correlation = 0.996) at all trait correlation levels but is in general more conservative resulting in slightly less true positive QTL's (Figure 1B and Figure 2).

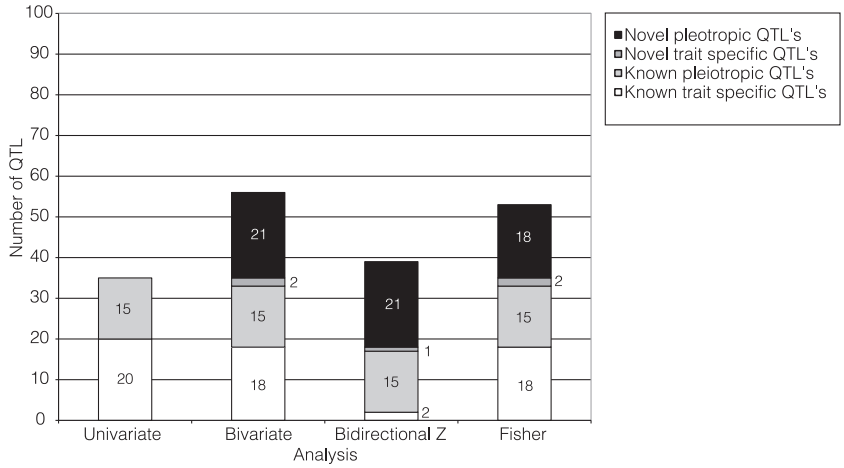
Tabel 2 *Lambda value of bidirectional Z-score meta-analysis under situation of no genotype-phenotype association and after spiking positive controls*

	Null		**Spiked**	
Directionality	*R**	*Lambda†*	*R**	*Lambda†*
Same	0.2	1.22	0.19	1.22
Opposite		0.81		0.84
Same	0.4	1.42	0.37	1.41
Opposite		0.60		0.66
Same	0.6	1.64	0.55	1.60
Opposite		0.39		0.46

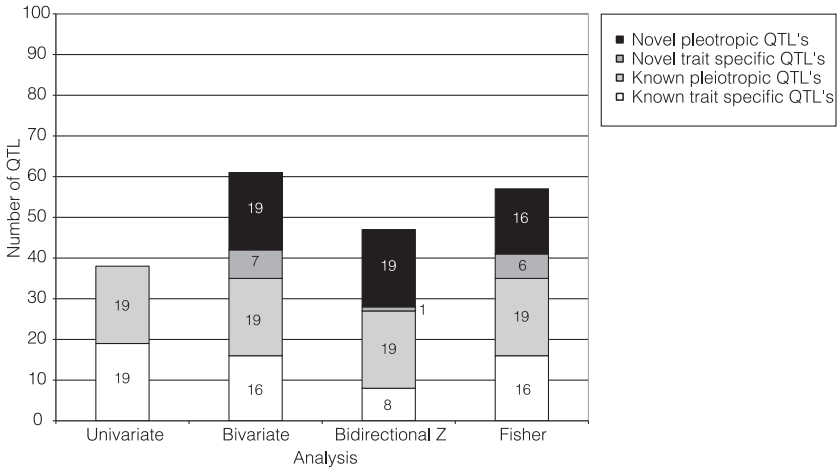
*Trait correlation, † Inflation/deflation of bidirectional Z-score meta-analysis test statistic distributions

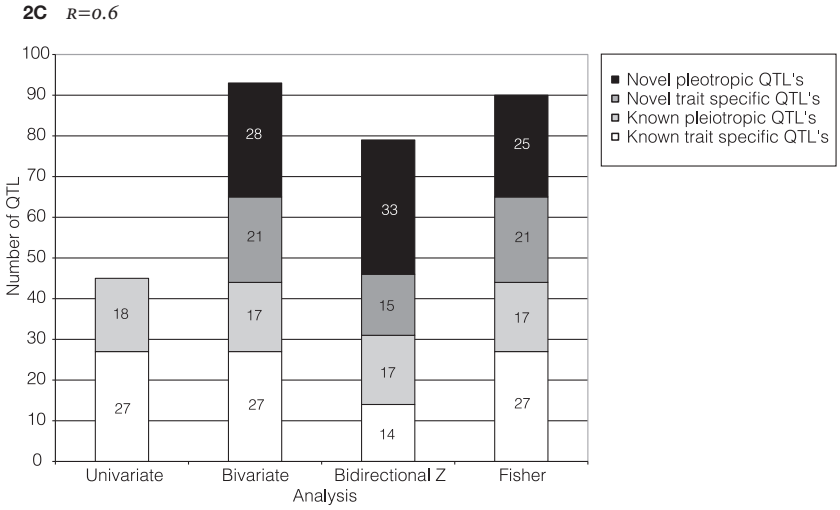
Figure 2 *The number of markers reaching the genome-wide significance level commonly used in GWA studies, $P<5\times10^{-8}$, for univariate analysis, bivariate analysis, bidirectional Z-score analysis (using a stricter threshold of 2.5×10^{-8}) and Fisher's method at A (R=0.2), B (R=0.4) and C (R=0.6).*

2A *R=0.2*



2B *R=0.4*





Novel pleiotropic QTL: genetic variant simulated to be associated with both trait 1 AND trait 2 and identified with a multivariate method, not univariate association analysis; **Novel trait specific QTL**: genetic variant simulated to be associated with both trait 1 OR trait 2 and identified with a multivariate method, not univariate association analysis; **Known pleiotropic QTL**: genetic variant simulated to be associated with both trait 1 AND trait 2 and identified with univariate association analysis. If also identified with a multivariate method not considered novel; **Novel trait specific QTL**: genetic variant simulated to be associated with both trait 1 OR trait 2 and identified with univariate association analysis. If also identified with a multivariate method not considered novel.

When the bidirectional Z-score based meta-analysis were considered as the final result, similar results were observed regarding the number of novel identified loci (at $P<2.5\times10^{-8}$) despite the low correlation with bivariate analysis (Pearson r=0.66 and 0.65 for same and opposite Z-score directionality, respectively). Since we expected that the QTL's identified with bidirectional Z-score based meta-analysis did not overlap completely with the QTL's identified by the Fisher method we looked into more detail in the properties of the QTL's identified by either the bidirectional Z-score meta analysis or the Fisher method but not both. Several examples for traits with a correlation of 0.6 are given in Supplemental table 2. In general the Z-score based meta-analysis identifies more pleiotropic effects while the Fisher approximation (and bivariate analysis) identified more QTL's private to one trait, as can be seen in Figure 2.

DISCUSSION

The proposed methods for combining univariate GWAS results facilitate large scale multivariate analysis without the requirement of re-analysis of individual data or knowledge of the correlation structure. Using easy implementable methods we can obtain results similar to multivariate analysis in an efficient way. We showed that the obtained test statistic distribution does not suffer from false positive results. We also showed that the proposed methods have increased power to detect both private and pleiotropic QTL's of two traits at different correlation levels and that the obtained results have a high correlation with standard bivariate analysis. In addition,

the intermediate results from this method can be used. The bidirectional Z-score meta-analysis has greater power to identify QTL's with independent effects on both traits when compared to Fisher's combination. Based on the expected underlying genetic effects or specific interests it can be motivated to regard the results of the bidirectional Z score meta-analysis as the outcome of interest and omit the final Fisher combination to come to the overall test-statistic.

A few issues should be considered. It is not unlikely that for two traits the numbers of subjects contributing to the GWAS's are not equal. Bivariate analysis in PLINK handles this with either case wise deletion or mean imputation (i.e. missing phenotypes are replaced with the sample mean). We performed simulation analysis for an unequal sample size scenario in which the results for trait 2 were based on half of the sample contributing to trait 1. The results were compared with bivariate analysis results using mean imputation. With maintenance of equal weighting in the SZS analysis the correlation between bivariate analysis and our approximation was 0.996 (Supplemental Figure 3).

In addition, the described methods have some practical advantages. In contrast to currently available software the proposed approximation can be used to estimate bivariate association signals for imputed genotypes (i.e. genotype dosage information). However, one should be careful interpreting the results for markers with a low effective sample size for either trait. A low effective sample size is the result of poor imputation of that marker resulting in uncertainty at the individual level on the true genotype. In the described methods all test statistic results are assumed to be the result of a consistent contribution by a fixed number of individuals per trait. A stringent filter on effective sample size or imputation quality should be advised when the method is applied on imputation based results. A practical limitation of the proposed methods is the applicability to two traits only. It is not possible to perform the bidirectional Z-score meta-analysis to obtain independent test statistics that can be easily combined using Fisher's method on more than two traits.

In conclusion, we present efficient methods for bivariate analysis that can increase power of QTL detection using aggregate level GWAS results data.

REFERENCES

1. Plomin, R., Haworth, C.M. and Davis, O.S. (2009) Common disorders are quantitative traits. *Nat Rev Genet*, 10, 872-878.
2. Chambers, J.C., Zhao, J., Terracciano, C.M., Bezzina, C.R., Zhang, W., Kaba, R., Navaratnarajah, M., Lotlikar, A., Sehmi, J.S., Kooner, M.K. *et al.* Genetic variation in SCN10A influences cardiac conduction. *Nat Genet*, 42, 149-152.
3. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M., de Bakker, P.I., Yin, X., Estrada, K., Bis, J.C., Marciante, K., Rivadeneira, F., Noseworthy, P.A. *et al.* (2009) Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet*, 41, 399-406.
4. Pfeufer, A., Sanna, S., Arking, D.E., Muller, M., Gateva, V., Fuchsberger, C., Ehret, G.B., Orru, M., Pattaro, C., Kottgen, A. *et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Gene*t, 41, 407-414.
5. Pfeufer, A., van Noord, C., Marciante, K.D., Arking, D.E., Larson, M.G., Smith, A.V., Tarasov, K.V., Muller, M., Sotoodehnia, N., Sinner, M.F. *et al.* (2010) Genome-wide association study of PR interval. *Nat Genet*, 42, 153-159.

6. Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P.I., Muller, M., Morrison, A.C., Smith, A.V., Isaacs, A., Sanna, S., Dorr, M. *et al.* (2010) Genome-wide association analysis identifies multiple loci related with resting heart rate. *Hum Mol Genet*, 19, 3885-3894.
7. Sotoodehnia, N. Isaacs, A. de Bakker, P.I. Dorr, M. Newton-Cheh, C. Nolte, I.M. van der Harst, P. Muller, M. Eijgelsheim, M. Alonso, A. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet*, 42, 1068-1076.
8. Ferreira, M.A. and Purcell, S.M. (2009) A multivariate test of association. *Bioinformatics*, 25, 132-133.
9. Fisher, R.A. (1932) *Statistical Methods for Research Workers.* 4th ed. Oliver and Boyd, London.
10. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559-575.

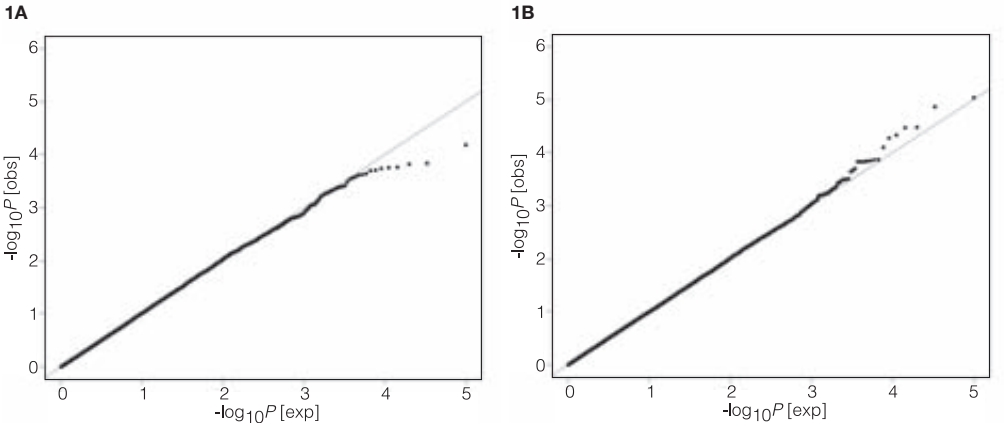
4 Supplement

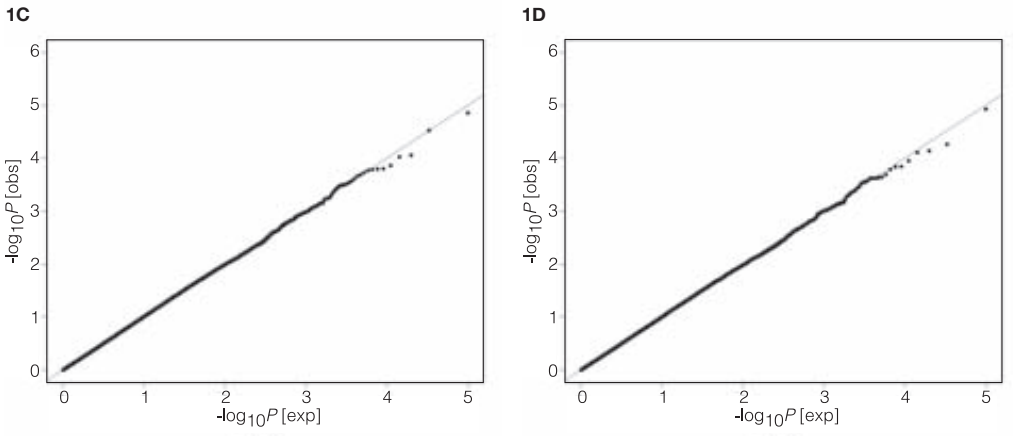
Supplemental Table 1 *Examples of method specific QTLs identified in addition to univariate analysis at r=0.6.*

Identified by	Properties	Example						
		Proportion variance explained*	Univariate *P*-value		Bidirectional Z-score *P*-value		Fisher *P*-value	Bivariate *P*-value
			Trait 1	Trait 2	Same	Opposite		
Fisher	Single trait effects	Trait 1: 2e-7 (0) Trait 2: 5e-4 (5e-4)	$1.8x10^{-7}$	$6.2x10^{-7}$	$4.6x10^{-3}$	$2.8x10^{-7}$	$2.8x10^{-8}$	$2.8x10e^{-8}$
Bidirectional Z	Pleiotropic	Same direction Trait 1: 5.8e-4 (1e-4) Trait 2: 4.4e-4 (1e-4)	$8.1x10^{-8}$	$2.8x10^{-6}$	$2.0x10^{-8}$	$4.8x10^{-1}$	$1.8x10^{-7}$	$6.3x10^{-8}$
		Opposite direction Trait 1: 1.2e-4 (1e-4) Trait 2: 1.7e-4 (1e-4)	$1.6x10^{-2}$	$3.3x10^{-3}$	$7.7x10^{-1}$	$1.2x10^{-8}$	$1.3x10^{-7}$	$2.9x10^{-7}$

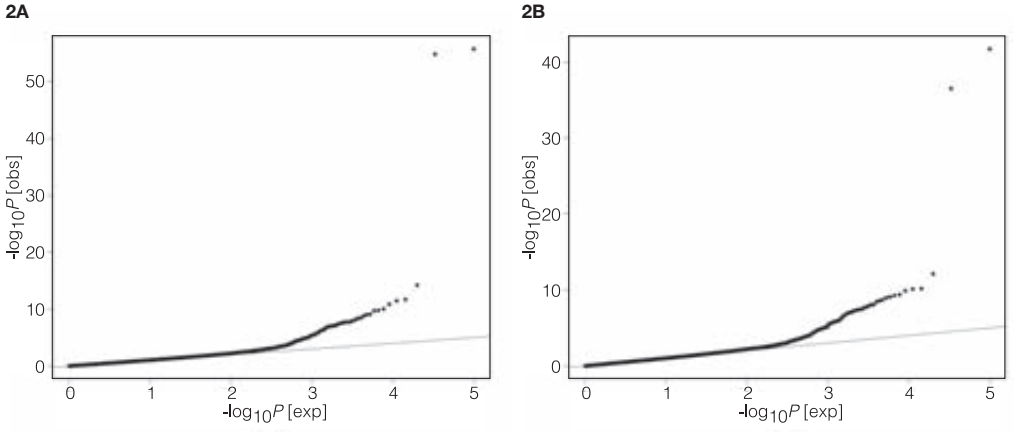
* = explained proportion of trait variance: final trait variance after forcing in all associations (original trait variance = effect simulated)

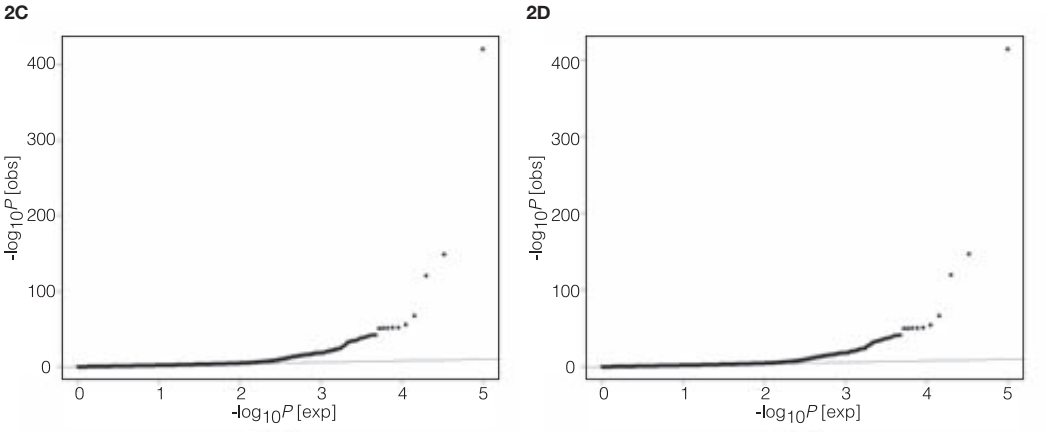
Supplementary Figure 1A-D *Quantile-Quantile plots showing observed versus the expected P-value distribution under the null model of no genotype-phenotype association for univariate analysis of (A) trait 1, (B) trait 2, (C) bivariate analysis and (D) Fisher method. Plots are presented for traits with a Pearson correlation coefficient of 0.6.*



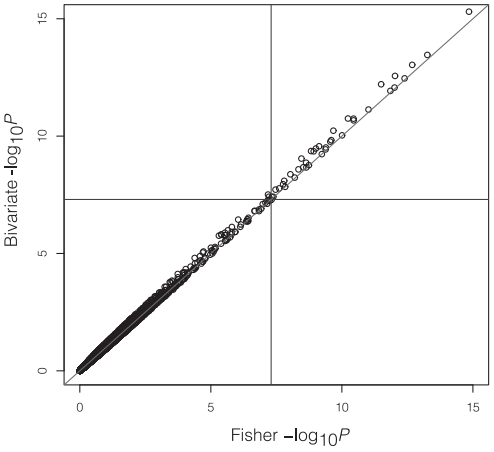


Supplementary Figure 2A-D *Quantile-Quantile plots showing observed versus the expected P-value distribution with simulated positive control type-phenotype associations for univariate analysis of (A) trait 1, (B) trait 2, (C) bivariate analysis and (D) Fisher method. Plots are presented for traits with a correlation coefficient of 0.6.*





Supplementary Figure 3 *Correlation plot. Bivariate analysis versus Fisher's method for the simulated dataset with forced positive control associations with unequal sample size (ratio of 1:2). Axes are truncated at -10log(P) = 15 for display purposes.*



PART 5

Common genetic variation and effect modification

5 Common variation in the *CACNA1C* gene modifies the effect of diltiazem on heart rate

Mark Eijgelsheim, Eline M. Rodenburg, Jan A. Kors, Jacqueline C. M. Witteman, Albert Hofman, André G. Uitterlinden, Bruno H. Ch. Stricker

Manuscript in preparation

ABSTRACT

Calcium antagonists block the L-type calcium channels by binding to the α1C subunit, which is encoded by the *CACNA1C* gene. Clinically, only verapamil and diltiazem depress sinus node function and atrio-ventricular conduction. Recently, rs2238018 located in *CACNA1C* was associated with RR interval (inverse heart rate). The objective of this study was to investigate whether the effect of diltiazem and verapamil on RR interval is modified by rs2238018. Hereto, we included 18,063 ECGs from 8,967 participants of the Rotterdam Study who were successfully genotyped and after exclusion of ECGs with left or right bundle branch block, RR interval <600 milliseconds (>100 beats per minute) or >1500 milliseconds (<40 beats per minute), or absence of P-waves. The study outcome was the RR interval within genotype strata with diltiazem and verapamil use compared to no use. Interaction was formally tested by the inclusion of a drug * genotype interaction term. Age, sex, heart failure, myocardial infarction, hypertension, body mass index and concomitant use of beta-blocker, digoxin and anti-arrhythmics were used as covariates in a repeated measurement analysis. Use of diltiazem (292 exposed ECGs) was associated with a 35.4 millisecond longer RR interval (95%CI: 21.1-49.6). The effect of diltiazem was potentiated in rs2238018 allele carriers with RR interval prolongation up to 74.4 (95%CI: 47.5-107.3) and 77.2 (95%CI: 4.1-150.3) milliseconds for heterozygous and homozygous variant allele carriers, respectively (*P*-value interaction=0.003). For verapamil (98 exposed ECGs) a similar but non-significant trend was observed. In conclusion, we demonstrated that *CACNA1C* SNP rs2238018 significantly modifies the RR interval prolonging (heart rate lowering) effect of diltiazem.

INTRODUCTION

Calcium antagonists are a chemically diverse group of drugs (i.e. phenylalkylamines such as verapamil, benzothiazepines such as diltiazem and dihydropyridines such as nifedipine) that share the property of blocking the inward calcium ion flow through L-type calcium channels. They exert this effect by binding to the α1c subunit, which is the pore forming subunit of the L-type calcium channel[1]. Clinically, only verapamil and diltiazem depress sinus node function and atrio-ventricular conduction. Although all calcium antagonists give vasodilatation through L-type calcium channel blockade in vascular smooth muscle cells, only dihydropyridines are associated with rebound tachycardia as the result of sympathetic activation in response to dihydropyridine induced vasodilatation which antagonizes their negative chronotropic effects[1]. Diltiazem and verapamil are prescribed for a variety of indications, including angina pectoris and supraventricular tachyarrhythmias, since the negative chronotropic and inotropic effects results in a reduction of oxygen consumption and rate control, which is beneficial in these conditions. Therefore, the strength by which the drugs exert these negative chronotropic effects might be relevant for drug therapy efficacy.

Recently, we performed a genome-wide association study on RR interval (inverse heart rate) that led to the identification of several genetic loci. In additional analyses, sub-borderline single nucleotide polymorphisms (SNPs) contributed to an increase in explained variance of the population variation of heart rate when these variants were included in a polygenic allele count that was tested in an independent sample. One of the SNPs included in the polygenic model that led to the maximal explained variance was rs2238018[2] located in *CACNA1C*, the gene encoding the α1c subunit of the L-type calcium channel[3].

In summary, diltiazem and verapamil modulate heart rate, act on the L-type calcium channel and common variant rs2238018 within the gene encoding the α1c subunit of the L-type calcium channel is associated with heart rate. Therefore, we conducted a population based study to investigate whether the effect of diltiazem and verapamil on RR interval is modified by rs2238018. Additionally, since the L-type calcium channel in the myocardium is also under hormonal regulation by dihydrotestosterone in vitro[4,5] and basal heart rate levels differ between men and women and vary with testosterone levels among men[6] we tested whether the effects were sex dependent.

METHODS

Setting and design

The study we report on was primarily carried out in the Rotterdam Study. The Rotterdam Study is an ongoing prospective population-based cohort study of chronic diseases in Caucasian elderly, which started in 1990. The Medical Ethics Committee of the Erasmus University approved the study. All inhabitants of Ommoord, a Rotterdam suburb in the Netherlands, aged 55 years and over (n=10,278) were invited to participate. Of them, 78% (n=7,983) gave their written informed consent for participation. Baseline examinations took place from March 1990 through

July 1993. Follow up examinations are carried out periodically with completion of the fourth examination cycle in 2004. This cohort is designated RS-I. In addition, in 2000 a second cohort (first extended cohort, RS-II) was enrolled. All inhabitants of Ommoord aged 55 years and above at that time and not yet participating in RS-I were invited (n=4,504). Of them, 3,011 (67%) entered the study with consent. In 2005 the second follow up examination of the RS-II cohorts was completed. In 2006 a further extension of the cohort (second extended cohort, RS-III) was initiated in which 3,932 subjects, 65% out of 6,057 invited, living in the Ommoord district and aged 45 years and over were included. The RS-III cohort completed the first examination round in 2008. Overall, there were 24,903 ECGs available in 12,961 individuals.

In addition to the periodical examinations, all participants were continuously monitored for major morbidity and mortality through linkage with general practitioner and municipality records. Detailed information on design, objectives and methods of the Rotterdam Study was described elsewhere[7-9].

Study population

All members of the Rotterdam Study cohorts who had at least one ECG and successful genotyping were eligible for inclusion in the study population. This resulted in 19,897 ECGs available in 9,685 participants. ECGs were excluded for presence of left or right bundle branch block, absence of P-waves or a RR interval <600 milliseconds (>100 beats per minute) or >1500 milliseconds (<40 beats per minute). Consequently, 18,063 ECGs from 8,967 participants were included.

RR interval measurement

All electrocardiograms were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the RR intervals were made using the Modular ECG Analysis System (MEANS)[10]. The MEANS program locates the QRS complexes and determines a stable reference point in each complex. The QRS detector of MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the QRS complexes among the other parts of the signal. RR intervals are taken as the intervals between the reference points in adjacent QRS complexes. The median RR interval was computed, after exclusion of RR intervals that immediately precede and follow any premature ventricular complex.

Drug exposure

For all participants exposure to medication is continuously monitored since January 1, 1991, through computerized pharmacy records of the pharmacies in the Ommoord district. The pharmacy data include the Anatomical Therapeutical Chemical-code[11], the dispensing date, the total amount of drug units per prescription, the prescribed daily number of units, and product name of the drugs. This provides us with information on start and duration of use of all prescribed medication.

ECGs were classified as exposed to the drug of interest when the date of the ECG was covered by an exposure period of the participant. The exposure period starts at the prescription filling date and the duration of this period is defined as the number of units issued per prescription

divided by the prescribed number of daily units. The exposure of interest was defined as use of mainly cardiac acting L-type calcium channel antagonists (ATC-code C08D) at the time of ECG measurement. Drug exposure to each of the two subgroups, the phenylalkylamine derivatives (ATC-code C08DA) and the benzothiazepine derivatives (ATC-code C08DB), was assessed specifically. In practice only one drug within each of these groups is used, namely verapamil (ATC-code C08DA01) and diltiazem (ATC-code C08DB01).

Covariates

Hypertension, myocardial infarction, heart failure as well as body mass index as well as beta blocker (ATC-code C07), digoxin (ATC-code C01AA), sotalol (ATC-code C07AA07) and class-I/III anti-arrhythmic medication (ATC-code C01B) use were included as covariates. Blood pressure and BMI measures were obtained at the same day as the ECG recordings during a visit to the research center. Hypertension was defined as a systolic blood pressure ≥140 mmHG and/or diastolic blood pressure ≥90mmHG and/or use of antihypertensive medication treatment, corresponding to the lower limit for grade 1 hypertension according to the World Health Organization guidelines[12]. Myocardial infarction and heart failure ascertainment have been described previously[13-15]. The presence of these covariates was assessed at each ECG index date.

Genotyping

For the presented study we extracted the genotype for rs2238018 based on our previous report[2]. This SNP is included on the used genome-wide genotyping arrays in the Rotterdam Study cohorts and passed all quality checks. For detailed information on genotyping procedures we refer to previous reports[2,16]. No other genotypes were extracted and analyzed.

Statistical analysis

The association between calcium channel blocker exposure and RR interval duration was studied by use of linear regression for repeated measurements as implemented in the PROC MIXED function (SAS, version 9.2, SAS Institute Inc., Cary, NC, USA) to account for correlation between RR interval measures in repeated ECG recording within individuals. First, the association between the drug exposure of interest and RR interval was assessed. Subsequently, this association was studied in genotype strata to assess effect modification which was formally tested by inclusion of interaction terms. Both stratified as well as interaction analysis were performed primarily using additive genetic models. In a second analysis a dominant genetic model was tested, but this was considered a secondary outcome. To minimize confounding by indication we additionally assessed the effect of variant allele carriage within strata of drug exposure. For diltiazem, the analysis were repeated within sex strata to assess if the interaction was sex dependent. Due to low numbers, we did not perform sex stratified analysis for verapamil. Analyses were at least adjusted for age at the moment of the ECG recording, sex and the cohort of origin. In more extensive models, hypertension, myocardial infarction, heart failure as well as body mass index and concomitant drug use were included as time dependent covariates.

RESULTS

Study participants

Out of the total of 19,897 ECGs (from 9,685 participants with genotype data), 18,063 ECGs (7,601 ECGs from male subjects, 42.1%) from 8,967 participants (3,799 men, 42.4%) were included after applying the exclusion criteria. The baseline characteristics (characteristics at the first ECG included in the analysis) are presented in Table 1. There were 98 ECGs recorded while a subject was taking verapamil and 292 ECGs were recorded during use of diltiazem. These 98 and 292 ECGs belonged to 74 and 223 individuals, respectively. The observed T-allele frequency of rs2238018 was 19.6%.

Table 1 *Baseline characteristics of the study populations.*

	RS-I	RS-II	RS-III	Total
Number of subjects	4997	1958	2012	8967
Sex, male	2045 (40.9%)	877 (44.8%)	877 (43.6%)	3799 (42.4%)
Mean age at first ECG (SD)	68.8 (8.5)	64.9 (7.5)	56.0 (5.6)	65.1 (9.3)
Mean heart rate at first ECG (SD), bpm	69.8 (11.2)	69.0 (10.6)	69.2 (10.1)	69.5 (10.8)
Mean RR interval at first ECG (SD), ms	882.2 (143.4)	890.7 (138.3)	885.1 (130.8)	884.7 (139.6)
Number of ECGs	12952	3099	2012	18063
Number of exposed ECGs*				
verapamil	86	9	3	98
diltiazem	237	46	9	292

RS-I: Rotterdam Study-I cohort, RS-II: Rotterdam Study-II cohort; RS-III: Rotterdam Study-III cohort; SD: standard deviation; bpm: beats per minute; ms: milliseconds; ECG: electrocardiogram;
*no simultaneous exposure to verapamil or diltiazem was observed

Heart rate

The mean RR interval at the baseline ECG was significantly ($P=5.7\times10^{-40}$) lower in females (868.0 milliseconds) than in males (907.3 milliseconds). Use of verapamil was associated with a 24.4 milliseconds longer RR interval than in non-users but this association was borderline non-significant ($P=0.06$). In contrast, diltiazem use was strongly associated with RR interval [+35.4 milliseconds (95%CI: 21.1;49.6), $P=1.1\times10^{-6}$] after adjustment for cohort sex, age, heart failure, myocardial infarction, hypertension, body mass index, beta-blocker, digoxin and anti-arrhythmic use. (Table 2)

Drug specific interactions with *CACNA1C* genotype on heart rate

We stratified for rs2238018 genotype to asses the effect of verapamil or diltiazem use on RR interval per genotype. For both verapamil and diltiazem a larger effect was observed within the strata of heterozygous and homozygous variant carriers (Table 3). Testing the interaction between rs2238018 genotype and use of verapamil or diltiazem showed a significant interaction

Table 2 *Verapamil, Diltiazem and heart rate*

	Unexposed ECG's	RR interval prolongation	Exposed ECG's	RR interval prolongation*	*P*-value
Verapamil	17965	Reference	98	+24.4 (-1.1-50.0)	$6.1x10^{-2}$
Diltiazem	17771	Reference	292	+35.4 (21.1-49.6)	$1.1x10^{-6}$

*adjusted for sex, age and cohort, heart failure, myocardial infarction, hypertension, BMI, beta-blocker use, digoxin use, sotalol or class-I/-III anti-arrhythmic use

Table 3 *Verapamil, Diltiazem, rs2238018 and heart rate*

	rs2238018 genotype stratum	Unexposed ECG's	RR interval prolongation	Exposed ECG's	RR interval prolongation*	*P*-value
Verapamil	CC	11596	Reference	54	+9.3 (-24.6-43.2)	0.59
	CT	5730	Reference	41	+42.8 (2.6-83.0)	$3.7x10^{-2}$
	TT	639	Reference	3	+94.2 (-66.7-255.1)	0.25
	Interaction					0.18
	CT+TT	6369	Reference	44	+45.1 (6.0-84.2)	$2.4x10^{-2}$
	Interaction					0.20
Diltiazem	CC	11445	Reference	205	+21.0 (4.2-37.7)	$1.4x10^{-2}$
	CT	5696	Reference	75	+74.4 (47.5-107.3)	$5.2x10^{-7}$
	TT	630	Reference	12	+77.2 (4.1-150.3)	$3.9x10^{-2}$
	Interaction					$3.1x10^{-3}$
	CT+TT	6326	Reference	87	+73.5 (46.7-100.3)	$8.0x10^{-8}$
	Interaction					$1.7x10^{-3}$

* adjusted for ,sex, age and cohort, heart failure, myocardial infarction, hypertension, BMI, beta-blocker use, digoxin use, sotalol or class-I/-III anti-arrhythmic use

between rs2238018 variant allele carriage and use of diltiazem with an additive genetic model ($P_{interaction}$ =3.1×10^{-3}). To minimize confounding by indication when comparing use and non-use within drug strata we additionally addressed the effect of the T-allele within drug strata. Within non-users of verapamil each additional T-allele was associated with a 3.7 millisecond longer RR interval (95%CI: -0.8;8.3, P=0.11). Within users this effect was stronger but also non-significant [+23.4 milliseconds per additional T-allele (95%CI: -53.4;100.2), P=0.55]. For diltiazem, the effects in non-users and users were +3.3 milliseconds (95%CI: -1.3;7.9), P=0.16 and +39.6 milliseconds (95%CI: 8.4;70.8), P=0.01 per additional T-allele, respectively. These results are consistent with the genotype stratified analyses.

Sex dependency

Use of diltiazem had an RR interval prolonging effect within both males [+32.6 milliseconds (95%CI: 10.7; 54.6), P=3.5×10^{-3}] and females [+39.1 milliseconds (95%CI: 20.4; 57.9), P=4.4×10^{-5}]. The effect within rs2238018 variant carriers was stronger in women. Within female carriers of at least one variant allele of rs2338018, use of diltiazem was associated with a 101.5-millisecond

longer RR interval. This corresponds to approximately 7-8 beats per minute. Sensitivity analysis were performed to assess the impact of low numbers of exposed ECGs in the homozygous carriers on the additive genetic model by applying a dominant genetic model and excluding homozygous variant allele carriers from the analysis. This showed consistent results (Table 4).

Table 4 *Diltiazem, sexe, rs2238018 and heart rate*

	rs2238018 genotype stratum	Unexposed ECG's	RR interval prolongation	Exposed ECG's	RR interval prolongation*	*P*-value
Male	CC	4805	Reference	97	26.1 (0.1-52.1)	$4.9x10^{-2}$
	CT	2363	Reference	41	56.6 (13.7-99.5)	$9.8x10^{-3}$
	TT	291	Reference	4	-5.7 (-139.3-127.8)	0.93
	Interaction					0.65
	CT+TT	2654	Reference	45	49.8 (8.9-90.8)	$1.7x10^{-2}$
	Interaction					0.45
Female	CC	6640	Reference	108	16.5 (-5.5-38.4)	0.14
	CT	3333	Reference	34	98.7 (58.8-138.6)	$1.4x10^{-6}$
	TT	339	Reference	8	137.4 (42.3-232.5)	$4.9x10^{-3}$
	Interaction					$1.5x10^{-4}$
	CT+TT	3672	Reference	42	101.5 (65.5-137.5)	$3.7x10^{-8}$
	Interaction					$1.0x10^{-4}$

* adjusted for sex, age and cohort, heart failure, myocardial infarction, hypertension, BMI, beta-blocker use, digoxin use, sotalol or class-I/-III anti-arrhythmic use

Formal testing of this higher order interaction by including all factorial interaction terms showed a significant interaction for diltiazem * sexe * rs2238018 (P=0.04 and P=0.03 under additive and dominant genetic models, respectively, adjusted for cohort, age, heart failure, myocardial infarction, hypertension, body mass index, beta-blocker use, digoxin use and sotalol or class-I/-III anti-arrhythmic use). Given the low numbers we refrained from sex stratified analysis for verapamil.

DISCUSSION

In this prospective population based cohort study we showed that the RR interval prolonging (heart rate lowering) effect of diltiazem is potentiated by a common variant (rs2238018) within the *CACNA1C* gene encoding the alpha subunit of the L-type calcium channel. The genotype-stratified results suggested a similar effect for verapamil but the interaction term was non-significant. Additionally, we showed that the potentiating effect of genotype on of RR interval prolongation by diltiazem was strongest in women. No other studies reported so far on the interaction between diltiazem or verapamil and genetic variation within *CACNA1C* on heart rate. Previous pharmacogenetic studies on calcium antagonists and *CACNA1C* are difficult to compare. Beitelshees *et al.* studied the combined clinical endpoint of death, nonfatal myocardial in-

farction or nonfatal stroke[17] using a case-control approach nested within a genetic sub study of a randomized trial on atenolol versus verapamil. They reported odds ratio for the combined outcome by comparing treatment strategies within genotype strata, which is difficult to interpret in other settings. Furthermore, the variant they report as relevant (rs1051375) is monomorphic in the HapMap European reference population, so linkage disequilibrium between rs2238018 studied by us and this variant could not be assessed. Other pharmacogenetic studies focused on dihydropyridine calcium antagonists, used outcome definitions based on blood pressure response and all included a different set of SNPs within *CACNA1C*[18,19].

Any inferences on the underlying mechanisms by which rs2238018 and diltiazem interact on RR interval are speculative. Underlying mechanisms could be direct effects on the affinity of the drug for binding the L-type calcium channel or indirect through decreased L-type calcium ion channel density. In addition, the vascular effects of verapamil and diltiazem could be of relevance. If the SNP were associated with a lower drug affinity in smooth muscle myocytes only, this would result in less vasodilatation and perhaps a decrease in sympathetic activity. This would counteract the negative chronotropic effects and might result in a net increase of negative chronotropy. The observation that the interacting effect is mainly present in women might be related to potential differences in ion channel density between men and women due to hormonal differences[4-6].

One of the strengths of our study is the availability of a large number of digitally stored and analyzed ECGs with repeated measures for a large proportion of the individuals which allows for more precision due to repeated measures. Additionally, the use of digital recordings all analyzed with the validated MEANS system[10,20,21] probably resulted in low reader variability. Furthermore, detailed information on drug dispensing records allowed for detailed ascertainment of drug exposure. Due to the prospective nature of the Rotterdam Study, with collection of the data independent of disease outcomes or (digital) ECG measures, information bias is unlikely. ECGs were excluded only if they showed left or right bundle branch block, or a RR interval <600 milliseconds (>100 beats per minute) or >1500 milliseconds (<40 beats per minute) to minimize measurement error and impact of non-physiological heart rate states. We additionally excluded ECGs with an absence of P-waves to minimize confounding by indication since this can be the result of atrial fibrillation or flutter and this is related with both the outcome (RR interval) and exposure (verapamil, diltiazem) studied. In theory residual confounding by indication could exist if the genotype (carriage of the variant T-allele) is associated with a protective effect on e.g. angina pectoris due to coronary disease and this on itself is associated with shorter RR interval (higher heart rate) and use of the drug. However, no association of genetic variation within *CACNA1C* is know to be associated with coronary heart disease so this is unlikely to be the case. Confounding was also minimized by adjustment for known factors which influence heart rate. Inclusion of the covariables hardly changed any of the effect estimates indicating that confounding was minimal. A limitation of the current study is the low number of verapamil users. Although a trend of incremental longer RR interval for exposure to verapamil was observed with additional variant alleles, we could not provide statistical significance for this interaction. Unfortunately, testosterone levels were only available in a small subset of the popu-

lation[6], so that we were not able to study the effect on the interacting effect of rs2238018 and calcium antagonists. The presence of such an interaction would be informative on the potential underlying mechanism of the sex depencency. Finally, in the current study we limited the analysis to non-dihydropyridine calcium antagonists only based on the hypothesis that the negative chronotropic effects of non-dihydropyridine calcium antagonists were potentiated by *CACNA1C* genotype. What remains unanswered is a possible interacting effect of dihydropyridine calcium antagonists and *CACNA1C* genotype. Although these drugs do not have a clinical effect on heart rate in general, *CACNA1C* genotype might be an effect modifier resulting in negative chronotropic effects in a subgroup of individuals due to a shift in balance of cardiac and vascular effects.

In conclusion, we demonstrated that *CACNA1C* SNP rs2238018 significantly modifies the RR interval prolonging (heart rate lowering) effect of diltiazem. Additionally, we showed that this potentiating effect is mainly observed in women. Since the clinical relevance of these drugs for angina pectoris and supraventricular arrhythmias is based on the negative chronotropic and inotropic effects, this gene-drug interaction may be of clinical relevance for drug efficacy.

REFERENCES

1. Abernethy, D.R. and Schwartz, J.B. (1999) Calcium-antagonist drugs. *N Engl J Med*, 341, 1447-1457.
2. Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P.I., Muller, M., Morrison, A.C., Smith, A.V., Isaacs, A., Sanna, S., Dorr, M. *et al.* (2010) Genome-wide association analysis identifies multiple loci related with resting heart rate. *Hum Mol Genet*.
3. Soldatov, N.M. (1994) Genomic structure of human L-type Ca2+ channel. *Genomics*, 22, 77-87.
4. Er, F., Gassanov, N., Brandt, M.C., Madershahian, N. and Hoppe, U.C. (2009) Impact of dihydrotestosterone on L-type calcium channels in human ventricular cardiomyocytes. *Endocrine research*, 34, 59-67.
5. Er, F., Michels, G., Brandt, M.C., Khan, I., Haase, H., Eicks, M., Lindner, M. and Hoppe, U.C. (2007) Impact of testosterone on cardiac L-type calcium channels and Ca2+ sparks: acute actions antagonize chronic effects. *Cell calcium*, 41, 467-477.
6. van Noord, C., Dorr, M., Sturkenboom, M.C., Straus, S.M., Reffelmann, T., Felix, S.B., Hofman, A., Kors, J.A., Haring, R., de Jong, F.H. *et al.* (2010) The association of serum testosterone levels and ventricular repolarization. *Eur J Epidemiol*, 25, 21-28.
7. Hofman, A., Breteler, M.M., van Duijn, C.M., Janssen, H.L., Krestin, G.P., Kuipers, E.J., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. *et al.* (2009) The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol*, 24, 553-572.
8. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22, 819-829.
9. Hofman, A., Grobbee, D.E., de Jong, P.T. and van den Ouweland, F.A. (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*, 7, 403-422.
10. van Bemmel, J.H., Kors, J.A. and van Herpen, G. (1990) Methodology of the modular ECG analysis system MEANS. *Methods Inf Med*, 29, 346-353.
11. Fleischmann, B.K., Duan, Y., Fan, Y., Schoneberg, T., Ehlich, A., Lenka, N., Viatchenko-Karpinski, S., Pott, L., Hescheler, J. and Fakler, B. (2004) Differential subunit composition of the G protein-activated inward-rectifier potassium channel during cardiac development. *J Clin Invest*, 114, 994-1001.
12. Kjeldsen, S.E., Erdine, S., Farsang, C., Sleight, P. and Mancia, G. (2002) 1999 WHO/ISH Hypertension Guidelines--highlights & ESH update. *Journal of hypertension*, 20, 153-155.
13. Bleumink, G.S., Knetsch, A.M., Sturkenboom, M.C., Straus, S.M., Hofman, A., Deckers, J.W., Witteman, J.C. and Stricker, B.H. (2004) Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure The Rotterdam Study. *Eur Heart J*, 25, 1614-1619.

14. Bots, M.L., Hoes, A.W., Koudstaal, P.J., Hofman, A. and Grobbee, D.E. (1997) Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study. *Circulation*, 96, 1432-1437.
15. Mosterd, A., Hoes, A.W., de Bruyne, M.C., Deckers, J.W., Linker, D.T., Hofman, A. and Grobbee, D.E. (1999) Prevalence of heart failure and left ventricular dysfunction in the general population; The Rotterdam Study. *Eur Heart J*, 20, 447-455.
16. Lango Allen, H. and Estrada, K. and Lettre, G. and Berndt, S.I. and Weedon, M.N. and Rivadeneira, F. and Willer, C.J. and Jackson, A.U. and Vedantam, S. and Raychaudhuri, S. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467, 832-838.
17. Beitelshees, A.L., Navare, H., Wang, D., Gong, Y., Wessel, J., Moss, J.I., Langaee, T.Y., Cooper-DeHoff, R.M., Sadee, W., Pepine, C.J. *et al.* (2009) *CACNA1C* gene polymorphisms, cardiovascular disease outcomes, and treatment response. *Circ Cardiovasc Genet*, 2, 362-370.
18. Bremer, T., Man, A., Kask, K. and Diamond, C. (2006) *CACNA1C* polymorphisms are associated with the efficacy of calcium channel blockers in the treatment of hypertension. *Pharmacogenomics*, 7, 271-279.
19. Kamide, K., Yang, J., Matayoshi, T., Takiuchi, S., Horio, T., Yoshii, M., Miwa, Y., Yasuda, H., Yoshihara, F., Nakamura, S. *et al.* (2009) Genetic polymorphisms of L-type calcium channel alpha1C and alpha1D subunit genes are associated with sensitivity to the antihypertensive effects of L-type dihydropyridine calcium-channel blockers. *Circ J*, 73, 732-740.
20. de Bruyne, M.C., Kors, J.A., Hoes, A.W., Kruijssen, D.A., Deckers, J.W., Grosfeld, M., van Herpen, G., Grobbee, D.E. and van Bemmel, J.H. (1997) Diagnostic interpretation of electrocardiograms in population-based research: computer program research physicians, or cardiologists? *J Clin Epidemiol*, 50, 947-952.
21. Willems, J.L., Abreu-Lima, C., Arnaud, P., van Bemmel, J.H., Brohet, C., Degani, R., Denis, B., Gehring, J., Graham, I., van Herpen, G. *et al.* (1991) The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*, 325, 1767-1773.

PART 6

General discussion

6 General discussion

INTRODUCTION

Sudden cardiac death (SCD) has a genetic risk component and identifying the genetic pathways contributing to SCD can lead to knowledge of novel mechanistic pathways underlying ventricular arrhythmogenesis. Possibly, it may also support clinicians in future risk stratification of patients. Inherent to the operational character of its definition the etiology of SCD is heterogeneous. SCD is defined as a natural death due to cardiac causes, heralded by abrupt loss of consciousness, within one hour after the onset of acute symptoms or an unwitnessed, unexpected death of someone seen in a stable medical condition less than 24 hours previously with no evidence of a non-cardiac cause[1]. Important reasons to study the quantitative traits that are associated with an increased risk of SCD are its etiological heterogeneity and the assumption that common genetic complex disorders such as SCD have a normally distributed genetic liability conferred by Mendelian inherited genes and thus the trait would be normally distributed as a quantitative trait[2, 3].

The main objectives of this thesis were to identify the genetic loci associated with SCD and quantitative ECG derived intermediate traits and to develop methodology to enhance our ability to do so. This discussion will place the main findings in perspective and some generic methodological considerations relevant for genome-wide association studies will be discussed in more depth, both in general and more specifically for pharmacogenetics as one of the promises of human genetics and of relevance for SCD as a consequence of drug induced arrhythmia[4-7].

MAIN FINDINGS

QT interval duration, Sudden Cardiac Death and *NOS1AP*

The QT interval is an ECG based measure for cardiac repolarization and prolongation is associated with SCD risk. Using genome-wide analysis, a common variant in the nitric oxide synthase 1 adaptor protein (*NOS1AP*) gene was identified[8] and consistently associated with QT interval variation across independent replication studies[9-11]. The genome-wide association study that identified single nucleotide polymorphism rs10494366 associated with QT interval used one of the earlier Affymetrix genotyping platforms (including ~100,000 variants) with limited coverage of the genome variation. Since the identified variant might not be the causative variant, we studied the *NOS1AP* region using dense genotype information to localize the region associated with the QT interval with more precision. We identified common variant rs12134842 in the

NOS1AP region as the strongest genetic predictor of QT interval (chapter 2.1). Each additional minor allele of rs12143842 was associated with a 4.4 millisecond longer QT interval duration, compared to 3.5 milliseconds for each additional minor allele of rs10494366. The rs12143842 variant was moderately linked to rs10494366 (r^2= 0.46) and conditional analysis showed that the associations of both variants with QT interval were statistically independent of each other. This led to the conclusion that possibly multiple genetic variants at *NOS1AP* modulate the QT interval, which was later confirmed by us and others[12-15]. It should be kept in mind that a non-identified causative variant might exist that is correlated to both rs12143842 and rs10494366 and drives both observed associations, even though they appear to be independent. We did perform haplotype analysis to see if incorporating genotype and linkage disequilibrium information improved the association signal since it was shown that multimarker haplotype analysis could improve the power of genetic association studies if an unmeasured marker is underlying the association observed[16,17]. Haplotype analysis did not show more significant results compared with rs12143842 alone. Till the causative variants are identified the genetic architecture of the well-established *NOS1AP*-QT interval association remains unclear. There are increasing efforts to identify the causative variants underlying genome-wide association study findings. It has been shown by simulation studies that through linkage rare causative variants can create significant associations for common variants, including multiple associations that appear independent[18]. The simulations additionally showed that these so called synthetic associations could be located far away from the causative rare variant[18]. This requires careful interpretation of the results in genome-wide association analyses in follow up studies like the Rotterdam Study, and it is important to consider the possibility of such potential pitfalls when designing studies to identify the causative variant[19]. Fortunately, there is compelling evidence that these synthetic associations do not underlie many of the common variant associations[20].

After the presence of multiple independent association signals at the *NOS1AP* locus was recognized, the first study soon appeared assessing the association of multiple *NOS1AP* variants with SCD[12]. In chapter 2.2 the association between these *NOS1AP* genetic variants and SCD in the Rotterdam Study is described. No strong associations were observed between the genetic variants studied and SCD. Limiting the analysis to witnessed SCD strengthened all results strengthened and showed a significant risk increase for carriage of each additional rs12143842 T-allele. In this study we also meta-analysed the results of the initial report and our report. Irrespective of the case definition used the association between rs16847548 (a proxy for rs12143482) and SCD strengthened upon combining the individual study results. Strengthening of the association of a second marker (rs12567209), independent of rs16847548, was conditional on the case definition used and was only observed when limiting to witnessed SCD in the Rotterdam Study. We interpreted these results as replication of the original report since it met most criteria for adequate replication[21]. For instance, it was conducted in an independent sample with a similar phenotype observed in a comparable population based sample. In addition, the same SNPs or good proxies were studied using the same statistical methods and models and the joint analysis led to a smaller *P*-value than seen in the initial report. However, we observed a dependency of the significance in the Rotterdam Study and joint analysis on the case-definition used. Within

the Rotterdam Study SCD is defined according to the commonly used operational definition that includes both witnessed as well as unwitnessed cases. We expected that this may introduce misclassification of the outcome. For instance, if some cases were in fact strokes with a rapidly fatal course this might give a dilution of effects (see section on *Methodological considerations* for a more general discussion on misclassification). The actual consequence of restricting the case definition to witnessed SCD only is dependent on the ratio of increasing power due to increased precision and loss of power due to lower numbers by exclusion of correctly classified cases. Therefore the effect of stricter case definitions will vary for studies that have different case ascertainment procedures and different case mixes.

Genome-wide association studies of ECG derived traits: discovery of novel quantitative trait loci

With the increasing availability of genome-wide genotype information several studies joined forces and formed consortia to study complex traits and diseases. In 2008 the QTGEN consortium formed, including participants from three cohort studies, i.e. the Framingham Heart Study, Cardiovascular Health Study and the Rotterdam Study, including 13,685 subjects. The aim of this consortium was to identify genetic loci that influence QT interval duration in the general population (*chapter 3.1*). As time moved on more studies generated genome-wide genotype data and logistic and methodological difficulties in performing consortium based genome-wide association studies were overcome. This led to the initiation of larger initiatives like the QT-International Genetics Consortium (QT-IGC), which is ongoing at the moment, but also large consortia focussing on other ECG based measures such as the RR interval (*chapter 3.2*) and QRS duration (*chapter 3.3*). All these efforts were successful in identifying novel loci associated with these traits. However they explain only 5-6% of the variation observed in QT and QRS duration and even less of the variation in RR interval (<1%), while for all these traits it is estimated that the heritability is 30-40%. In general, both variants within loci known to be of relevance and at genetic loci not previously recognized were identified. In the study on QT for example, five out of ten observed loci included ion channel genes that were known to be involved in myocardial repolarization, namely *NOS1AP, KCNQ1, KCNH2, KCNE1* and *SCNA5*. These ion channel coding genes harbouring QT associations are known from Mendelian long-QT syndromes (Table 1).

Table 1 *Overview of loci associated with QT interval in the QTGEN study within genes known for their involvement in cardiac repolarization.*

Gene	Protein	Mendelian disorder	Known associations in the general population	Novel associations in general population
NOS1AP	Nitric oxide synthtase 1 activaor protein	–	Yes[8,22]	Yes
KCNQ1	Alpha subunit slow delayed rectifier potassium channel	LQT1, SQT2[23]	Yes[24]	Yes
KCNH2	Alpha subunit rapid delayed rectifier potassium channel	LQT2[23]	Yes[25]	Yes
KCNE1	Beta subunit potassium channel (assembles with *KCNQ1*)	LQT5[23]	Yes[26]	No
SCN5A	Alpha subunit sodium channel	LQT3[27]	No	Yes

LQT= long-QT syndrome, SQT= short-QT syndrome

In addition to these known loci, additional associations were observed at novel loci not previously recognised as relevant, including 16q21 near *NDRG4, SETD6, CNOT1, SLC38A7* and *GINS3*. It is difficult to identify the gene truly impacting cardiac repolarization solely on the observed association signal. Existing literature might aid in interpreting these results, but will not always give a conclusive answer. At the moment of conducting the study, *NDRG4* and *GINS3* were considered promising candidates at the 16q21 locus for further study based on available expression data and results from functional studies in zebrafish[28,29]. At that moment *CNOT1* was not considered amongst the more likely candidates at this region. Only later *CNOT1* was added to the list of potential relevant genes at this locus based on the identification of the CCR4-Not complex as a central regulator of adult heart function in Drosophila[30]. It remains to be elucidated through which genes at the associated genomic regions genetic variation influences QT interval. In addition, as discussed above for chapter 2.1, rare variants can cause significant associations for common variants and therefore this possibility should be considered in interpreting the results for both the association signals in both the genes known to be involved in Mendelian QT interval disorders as well as the novel loci[18]. The observation that several of the associations map to genes known to cause Mendelian forms of repolarization disease argues against synthetic effects caused by rare variants located far away, since the results do point at well established functional key elements.

Another example of the difficulties in interpreting the results of genome-wide association studies can be found in our RRGEN study (*chapter 3.2*). Several of the identified loci, e.g *GJA1* and *PLN*, are known to harbor rare mutations that cause congenital structural heart disease[31-34]. This might implicate that common variation can result in variation in cardiac structure in the general population and so influence heart rate or that the common variant associates with heart rate due to linkage with rare variants with large effects on cardiac structure that impact heart rate. However, *PLN* and *GJA1* are reported to have multiple other effects on the heart. *GJA1* encodes the major gap junction responsible for impulse propagation in the heart[35] and is an important regulator of gene expression during human cardiogenesis[36]. *PLN* encodes phospholamban, a well known protein that modulates heart rate in the sinus node[37,38] and was previously associated with QT interval in our QTGEN study[13] and left ventricular diastolic dimensions[39]. This suggests that alternative pathways exist through which these loci might influence heart rate. All in all, even for loci of known cardiac relevance no inference could be made on the mode of action on heart rate without further study. In the report on QRS duration we therefore included functional studies on one of the identified loci in order to create a better understanding of one of the findings. The most significant region included multiple significant associations that spanned multiple genes encoding sodium channels, both known and previously unrecognized to play a role in cardiac conduction. Using a mouse model our collaborators showed that the sodium channel previously unrecognized to be of relevance for cardiac conduction (*SCN10A*) was actually expressed in the heart, preferentially in the ventricular conduction system, and selective blockade of the encoded channel resulted in a cardiac conduction delay.

After identifying a set of loci associated with the trait of interest a genetic risk score is often generated and tested for association with a clinically more relevant outcome based on the continuous trait studied. In the Rotterdam Study, it has been shown that a prolonged heart rate corrected QT interval (QTc) over 450 milliseconds in men and over 470 milliseconds in women was associated with an 2.5 fold increased risk of SCD. The 20% of the population with the highest genetic risk score (carrying the most alleles identified as QT prolonging in the QTGEN study, weighted by their effect) had on average 12 milliseconds longer QT intervals and a 3 fold increased risk of a prolonged QTc as defined above when compared to the 20% of the population carrying the least alleles with QT prolonging effect. A 12 millisecond longer QT is in excess of the effect observed for QT prolonging drugs which were withdrawn from the market due to increased arrhythmia risk[4]. This supports the clinical relevance of the genetic variants identified, although this should be placed in perspective. It is not said that all arrhythmia and SCD risk associated with certain drugs and the long-QT syndromes is caused by prolongation of the QT interval. Alternative pathways through which drugs and genetic mutations confer SCD susceptibility might exist. This is supported by the observation that not all drugs that prolong the QT interval increase the risk of SCD risk and not all drugs that are associated with SCD do prolong the QT interval. Therefore, not all variation at the identified loci might increase the risk of SCD or the risk increase might not be proportional to the effect on QT interval. Well-sized and adequately phenotyped samples will be required to answer these questions. Additionally, the method used in the QTGEN study to asses the effect of the genetic risk score on the risk of QT prolongation might suffer from overestimation of the true effect due to the fact that the people with longer QT intervals were included in the genome-wide association analysis. A sort of similar approach was applied for QRS duration with the advantage that the subjects with the clinical outcome (ventricular conduction defect defined as QRS duration of more than 120 milliseconds) were independent of the subjects included in the association analysis identifying the loci. Here it was shown that each additional copy of an allele that was associated with longer QRS duration increased the risk of a ventricular conduction defect by 8%.

Combining genome-wide association study results of correlated traits

As the genome-wide association studies on multiple individual traits within the same or partly overlapping collection of samples ignores the additional information (increase in phenotype accuracy), that is present when traits are correlated, we sought a method to efficiently combine the results of the univariate genome-wide association studies in order to increase the power to identify truly associated loci (*chapter 4*). The main advantage would be increased efficiency as the analyses are relatively simple and easy because already available GWA analyses are combined. Although we have used simulations to describe the behaviour of the proposed method, its additional value has to be proven in practice by applying it to real data with adequate replication of new loci identified by means of this novel approach. We intend to introduce this method on the short term to assess its value in practice. In addition, extension of the applicability of this method to more than two traits would be desirable.

Common variation and effect modification

Based on the observed association between genetic variation within *CACNA1C* and RR interval[40] as well as the fact that drugs that block the L-type calcium channel do this by binding the protein encoded by *CACNA1C*[41], we hypothesized that the effects of the phenylalkylamine (verapamil) and benzothiazepine (diltiazem) calcium antaganonists on RR interval, and thus heart rate, are modified by the *CACNA1C* genotype. Although another group of calcium antagonists is also commonly prescribed (i.e. dihydropyridines) these were not included since they do not lead to slowing of the heart rate.

We observed a significant interaction between *CACNA1C* genotype and diltiazem. Although verapamil showed a similar trend no significant associations were observed, perhaps due to lower numbers of exposed RR interval measurements. Additionally, we performed a sex stratified analysis based on the observation of a relation between testosterone levels and heart rate[42], large differences in heart rate between men and women in general and the rate dependent effects of calcium antagonists (larger effect at higher heart rates)[41]. We showed that the interacting effect between diltiazem and *CACNA1C* genotype was sex dependent, with the largest effect in women. These observations raise a number of questions which we are not yet able to answer. First, what is the underlying mechanism of the drug-gene interaction. Although there seem to be plausible explanations, they are all speculative. In short, alterations in the functional properties and affinity for calcium channel blockers could be relevant as well as changes in ion channel density and thus number of occupied and open channels after drug administration. The second question we are unable to answer with our observational study is on the underlying mechanism of the observed effect modification by sex. Based on the rationale that led to doing the analysis, several explanations are plausible. Differences in hormone levels between men and women, especially testosterone, might result in differences in gene expression and ion channel densities[43,44]. In a situation of higher ion channel expression in men the ion channel reserve in women could be lower compared to men. When two factors impact this ion channel density the effects could be potentiated and result in stronger interaction in women. To complicate things further, the rate dependent effects could be of relevance.

We investigated the possibility to study the effect of testosterone on the interaction following on a previous study on the effects of testosterone on heart rate[42]. Unfortunately, the numbers were very low due to the limited number of testosterone level measurements and such an analysis was not feasible.

METHODOLOGICAL CONSIDERATIONS

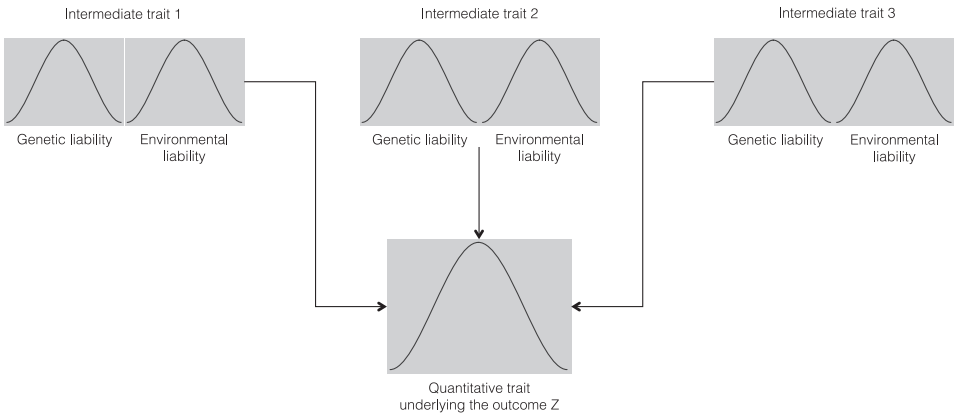
The methodological considerations relevant to the presented studies have been discussed in the individual chapters. In this paragraph a more detailed discussion is presented on the framework of genome-wide association studies in general and pharmacogenomic genome-wide association studies in particular. More specifically, the impact of outcome and exposure definition is discussed since it might affect the statistical power and therefore contribute to the missing heritability[45]. The missing heritability refers to the difference between the heritability

estimates and variance explained by the genetic variants identified in GWA studies[46,47]. Here we discuss the basic principles of the effects of phenotypic heterogeneity in the framework of genome-wide association studies on common complex disorders.

Discussion on the general framework in genome-wide association studies

The general principle underlying genome-wide association studies is the common variant – common disease assumption[48]. This assumption is based on a polygenic quantitatively distributed disease liability in the population that results from a combination of Mendelian inherited genes. Common disorders are the extremes of quantitative traits resulting from the quantitative genetic liability[3]. This genetic liability is accompanied by an environmental liability that in aggregate result in the population's risks distribution. The quantitative trait underlying a disease will be the net result of multiple intermediate traits. (Figure 1)

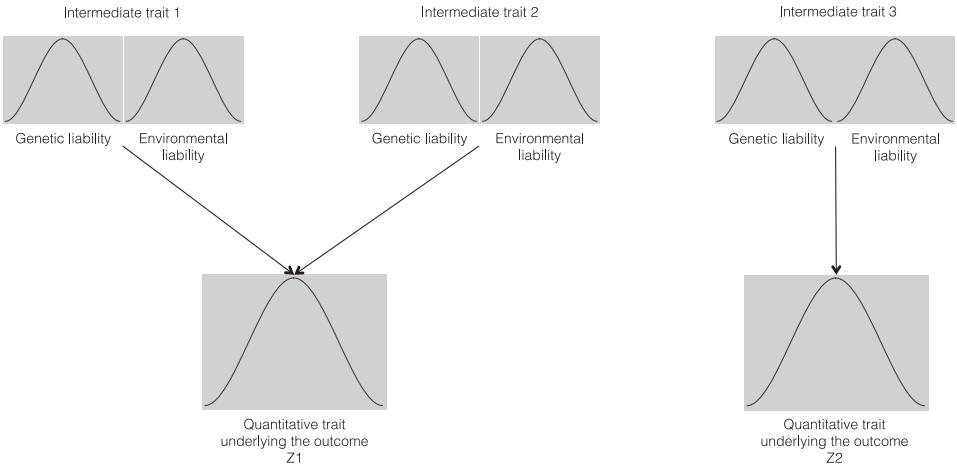
Figure 1 *If disease outcome Z is seen as the extreme of a quantitative distributed trait, other quantitative traits can underlie this distribution, each with their own genetic and environmental quantitatively distributed liability that in aggregate result in the disease liability distribution.*



If the clinical disease outcome studied (Z) consists of different patho-physiological entities with different underlying intermediate traits (Figure 2) the power to identify associated genetic variation will be affected. For the variants that are not associated with Z1 or Z2 such misclassification of the outcome would not lead to spurious associations since the genotype distributions between cases and controls will not be altered. In the situation where variants do associate with Z1 but not Z2 (or vice versa), using a phenotype definition that includes both Z1 and Z2 will bring the genotype distribution closer to unity. This means that the frequency of the truly causative variant within those classified as diseased will be closer to the total population mean and the genotype distribution in the control sample. The measure of effect will be smaller and more difficult to detect (i.e. loss of statistical power). This situation can occur when lack of information or knowledge leads to imprecise definitions of the disease or when case ascertainment pro-

cedures can not further specify the phenotype (e.g. sudden cardiac death in population based studies where the underling cause can be unknown and heterogeneous) resulting in a case mix actually consisting of separate disease entities. A nice example of a strict case definition in the field of SCD genetics can be found in the paper of Bezzina *et al.*[49] where the outcome of interest is ventricular fibrillation in the presence of acute ischemia. As described in this discussion for chapter 2.2, approaches in population based studies could include defining different versions of the phenotype by limiting the case definition to a subset that is more likely to be homogeneous, like only witnessed SCD or restricting to events occurring at younger age groups. However, this does not mean that case mix heterogeneity is always something to avoid. In the situation where different disease entities share intermediate traits and/or genetic risks, such a case mixture could be more powerful in identifying shared risk factors. In the example of SCD this might occur if a shared vulnerability to arrhythmia exists in ischemic heart disease, heart failure and (drug induced-) long QT syndrome. Studying the appropriate intermediate traits will limit the problem of outcome misclassification. However, lack of knowledge might limit the identification of relevant intermediate traits since they should not be limited to symptoms or diagnostic criteria of a disease entity only.

Figure 2 *If a clinical disease entity Z actually consists of two distinct phenotypes (Z1 and Z2) they should be studied separately since combining them into one outcome is inappropriate and results in misclassification with a loss of power to identify the genetic liability as a result. Studying the appropriate intermediate trait might overcome this problem.*



It is important to keep these principles in mind when defining the aim of a genome-wide association study and interpreting the results. Continuous re-evaluation of disease definitions based on developing (patho-) physiological knowledge and novel insights into genetic liability will lead to better outcome definitions and thus more power. Perhaps, such a step-wise approach will help in identifying the so far unexplained heritability of common disorders.

Pharmacogenetics

In contrast to GWA studies that have been successful in identifying many loci for complex disorders, pharmacogenetic studies could be described 'disappointing'. In pharmacogenetics some successes are reported for adverse events (e.g. simvastatin induced myopathy[50] and flucloxacillin induced liver injury[51]) and efficacy (e.g interferon-α treatment in hepatitis C infection[52,53]), where common variants with large effects were identified using relatively small numbers of cases. In the following section several methodological considerations are discussed that are relevant to conduct and interpretation of genome-wide association studies in pharmacogenetics with a particular focus on definitions of drug exposure, phenotype of interest and use of intermediate traits.

We have to realize that drugs undergo selection during their development and their effects are often studied on quantitative intermediate traits. This gives us a clue on the genetic effects we can expect. In general, a drug that has serious adverse events due to genetic variation will not be marketed if the genetic variation underlying this reaction is common and the effects are large since the adverse event rate will be unacceptably high. The same holds if a very large group is non-responsive to the intended effect. Furthermore, drugs are often studied on intermediate end-points, like HDL and LDL cholesterol levels changes for statins and platelet aggregation inhibition parameters for anti-thrombotic agents. Variation within these intermediate traits does not per se translate directly into variation in the occurrence of the relevant clinical endpoints (prevention of myocardial infarction, stroke, death). Since the complete range of actions (and thus relevant intermediate traits) of drugs are often not known it can not be taken for granted that genetic variation underlying variation of the intermediate traits usually studied also explain variation in the clinical endpoints. This is in line with the difficulty of identifying the relevant intermediate traits for complex disorders due to lack of knowledge of underlying mechanisms.

In principle the same framework can be applied as described above with the addition of a drug exposure and redefining the intermediate traits and outcome definition (Figure 3), which are now based on the drug exposure. In other words, the phenotypes are redefined according to the context. If we take the example of drug induced arrhythmia, a major problem in drug development and safety, the dichotomous outcome Z is defined as the occurrence of fatal arrhythmia due to exposure to drug Y. However, the dichotomous outcome Z has an underlying quantitative trait and liability distribution that is the result of intermediate traits 1, 2 and 3. An example of an intermediate trait can be the effect of drugs on cardiac repolarization duration (QT interval). These intermediate traits are quantitatively distributed, each with a genetic and environmental quantitatively distributed liability. The introduction of drug exposure complicates matters in the sense that especially in the early phase of marketing the available numbers of exposed subjects are often low. The definition of the exposure is an additional factor that can influence the power to detect genetic variation associated with the (intermediate) trait of interest. If not a specific drug but a whole class of drugs is studied this can impact the results if some drugs within that class do not influence the trait studied (i.e. no consistent class effect). The same argument holds if the drug exposure is defined based on a common association with outcome Z but drugs act through different intermediate traits. In this case the analysis will

Figure 3 *Introducing drug exposure. If outcome Z is defined as the occurrence due to exposure to Y, the intermediate traits can be defined as the effect of Y on a certain biological process or product, each with their own genetic and environmental quantitatively distributed liability, that in aggregate result in the underlying liability distribution of the dichotomous outcome.*

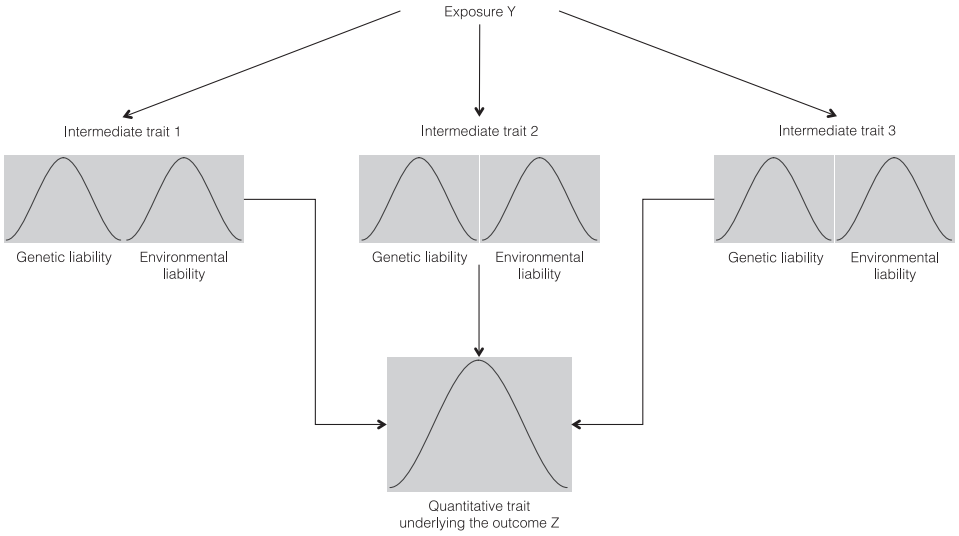
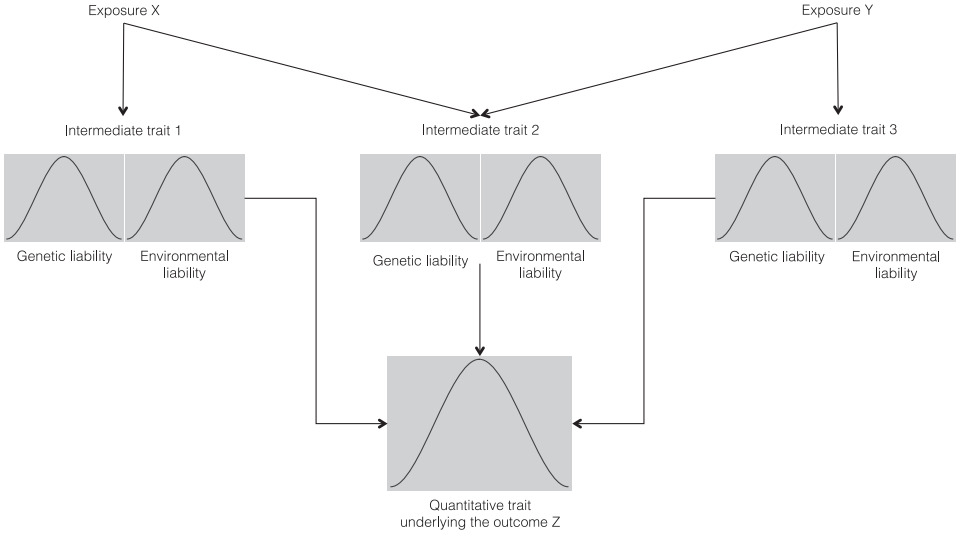


Figure 4 *Two types of exposure with an effect on outcome Z through shared and different mechanisms (intermediates).*

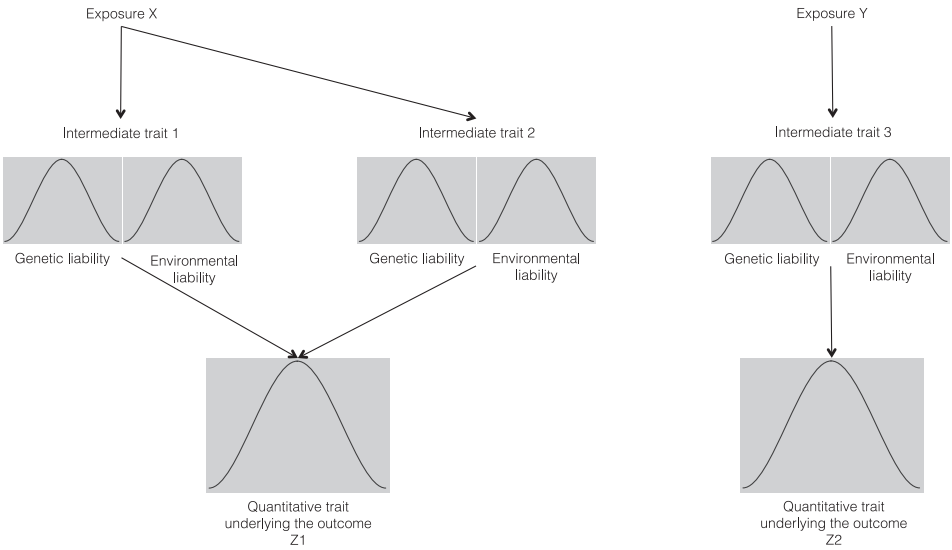


suffer from dilution toward the null hypothesis of no association and a reduction in power. In figure 4 this is relevant when not the quantitative trait underlying Z is studied but any of the intermediate traits 1 or 3, and exposure X and Y are combined (e.g. anti-psychotics as exposure X and tri-cyclic antidepressants as exposure Y) as the exposure of interest since they both increase the risk of outcome Z (e.g. fatal arrhythmia). If intermediate trait 2 is studied, the impact of the exposure definition is dependent on the underlying genetic liability. If the genetic liability is shared for exposure X and Y combining the exposure groups will increase the power to detect the genetic liability underlying intermediate trait 2 (e.g. drug induced QT prolongation).

Finally, in the situation as depicted in Figure 5 with outcome Z being a clinical composite of biologically two different outcomes Z1 and Z2 as a consequence of exposure to X and Y, respectively, insufficient specificity of the exposure definition results in a loss of power, regardless of the trait studied (intermediate, Z1 or Z2). If on the other hand the exposure is adequately defined, but Z1 and Z2 are combined, again this will result in misclassification of the outcome and dilution towards the null of no association.

In general and especially in pharmacogenetics it is not only sample size that matters. An equal amount of effort invested in sample collection should be invested on refinement of case and exposure definitions and thorough phenotyping (including drug and environmental exposures) of study subjects.

Figure 5 *Outcome Z being biologically two distinct phenotypes as a consequence of exposure to X and Y respectively each with their own intermediates.*



FUTURE DIRECTIONS

In this thesis we showed that genome-wide association studies could identify genetic variation influencing SCD risk and ECG derived intermediate traits. The identified loci leave much of the heritability unexplained. A lot of attention is focused on rare variants with larger effects although the contribution to the explained heritability remains unclear at his moment. Efforts like the 1000 Genomes project and individual studies deep-sequencing large numbers of individuals will create more insight into the role of rare variants in the near future. As brought forward in the main part of this discussion phenotype heterogeneity might be a relevant issue in discovering novel loci influencing human disease and drug response. Not only is this relevant to discovery but perhaps even more important for follow up studies that try to identify the causal variants. Therefore we should not only invest in improving (technical) genetic methods and knowledge, but also in generating databases with extensive phenotype data on large numbers of subjects. Identifying relevant intermediate traits can lead to better understanding and redefinition of diseases, ultimately resulting in more power to detect genetic variation underlying the genetic liability. An approach that may identify novel intermediate traits could be a 'phenome-wide' association analysis. For example, this could include linking the human metabolome[54-56], just one of the deep phenotyping tools[57], to disease phenotypes identifying relevant intermediate traits at the level of small molecule metabolites found in the human body. Subsequently, the identified traits could be studied in relation to human genetic variation[58,59]. Such an approach would be applicable to both complex diseases as well as complex pharmacogenetic phenotypes. Better human phenotyping would lead to better understanding of the recent findings and increase the statistical power for future research[57]. The relevance for clinical medicine of such higher resolution phenotyping is only present if decision and or treatment consequences are linked to further differentiating diagnosis, which is were pharmacogenetics comes into play.

CONCLUDING REMARK

The underlying genetic liability for SCD is far from resolved and has mainly been successful for several of its intermediate traits. It remains to be seen what part of the identified loci contribute to SCD risk. Nevertheless, important insights into normal physiology are being generated by genome-wide association studies. We only touched upon the tip of the iceberg and need to look for the larger part under water for which not only progression of human molecular genetics is relevant but also more precise phenotyping.

REFERENCES

1. Priori, S.G., Aliot, E., Blomstrom-Lundqvist, C., Bossaert, L., Breithardt, G., Brugada, P., Camm, A.J., Cappato, R., Cobbe, S.M., Di Mario, C. *et al.* (2001) Task Force on Sudden Cardiac Death of the European Society of Cardiology. *Eur Heart J*, 22, 1374-1450.
2. Fisher, R.A. (1918) The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinb.*, 52, 399-433.
3. Plomin, R., Haworth, C.M. and Davis, O.S. (2009) Common disorders are quantitative traits. *Nat Rev Genet*, 10, 872-878.
4. Roden, D.M. (2004) Drug-induced prolongation of the QT interval. *N Engl J Med*, 350, 1013-1022.
5. Roden, D.M. (2008) Cellular basis of drug-induced torsades de pointes. *Br J Pharmacol*, 154, 1502-1507.
6. Roden, D.M. and Viswanathan, P.C. (2005) Genetics of acquired long QT syndrome. *J Clin Invest*, 115, 2025-2032.
7. Roepke, T.K. and Abbott, G.W. (2006) Pharmacogenetics and cardiac ion channels. *Vascular pharmacology*, 44, 90-106.
8. Arking, D.E., Pfeufer, A., Post, W., Kao, W.H., Newton-Cheh, C., Ikeda, M., West, K., Kashuk, C., Akyol, M., Perz, S. *et al.* (2006) A common genetic variant in the *NOS1* regulator *NOS1AP* modulates cardiac repolarization. *Nat Genet*, 38, 644-651.
9. Lehtinen, A.B., Newton-Cheh, C., Ziegler, J.T., Langefeld, C.D., Freedman, B.I., Daniel, K.R., Herrington, D.M. and Bowden, D.W. (2008) Association of *NOS1AP* genetic variants with QT interval duration in families from the Diabetes Heart Study. *Diabetes*, 57, 1108-1114.
10. Post, W., Shen, H., Damcott, C., Arking, D.E., Kao, W.H., Sack, P.A., Ryan, K.A., Chakravarti, A., Mitchell, B.D. and Shuldiner, A.R. (2007) Associations between genetic variants in the *NOS1AP* (CAPON) gene and cardiac repolarization in the old order Amish. *Hum Hered*, 64, 214-219.
11. Tobin, M.D., Kahonen, M., Braund, P., Nieminen, T., Hajat, C., Tomaszewski, M., Viik, J., Lehtinen, R., Ng, G.A., Macfarlane, P.W. *et al.* (2008) Gender and effects of a common genetic variant in the *NOS1* regulator *NOS1AP* on cardiac repolarization in 3761 individuals from two independent populations. *Int J Epidemiol*, 37, 1132-1141.
12. Kao, W.H., Arking, D.E., Post, W., Rea, T.D., Sotoodehnia, N., Prineas, R.J., Bishe, B., Doan, B.Q., Boerwinkle, E., Psaty, B.M. *et al.* (2009) Genetic Variations in Nitric Oxide Synthase 1 Adaptor Protein Are Associated With Sudden Cardiac Death in US White Community-Based Populations. *Circulation*.
13. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M., de Bakker, P.I., Yin, X., Estrada, K., Bis, J.C., Marciante, K., Rivadeneira, F., Noseworthy, P.A. *et al.* (2009) Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet*, 41, 399-406.
14. Pfeufer, A., Sanna, S., Arking, D.E., Muller, M., Gateva, V., Fuchsberger, C., Ehret, G.B., Orru, M., Pattaro, C., Kottgen, A. *et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet*, 41, 407-414.
15. Arking, D.E., Khera, A., Xing, C., Kao, W.H., Post, W., Boerwinkle, E. and Chakravarti, A. (2009) Multiple independent genetic factors at *NOS1AP* modulate the QT interval in a multi-ethnic population. *PloS one*, 4, e4333.
16. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat Genet*, 37, 1217-1223.
17. Pe'er, I., de Bakker, P.I., Maller, J., Yelensky, R., Altshuler, D. and Daly, M.J. (2006) Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet*, 38, 663-667.
18. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. and Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS biology*, 8, e1000294.
19. Wang, K., Dickson, S.P., Stolle, C.A., Krantz, I.D., Goldstein, D.B. and Hakonarson, H. Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet*, 86, 730-742.
20. Orozco, G., Barrett, J.C. and Zeggini, E. (2010) Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet*, 19, R137-144.
21. Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E. *et al.* (2007) Replicating genotype-phenotype associations. *Nature*, 447, 655-660.
22. Eijgelsheim, M., Aarnoudse, A.L., Rivadeneira, F., Kors, J.A., Witteman, J.C., Hofman, A., van Duijn, C.M., Uitterlinden, A.G. and Stricker, B.H. (2009) Identification of a common variant at the *NOS1AP* locus strongly associated to QT-interval duration. *Hum Mol Genet*, 18, 347-357.

23. Splawski, I., Shen, J., Timothy, K.W., Lehmann, M.H., Priori, S., Robinson, J.L., Moss, A.J., Schwartz, P.J., Towbin, J.A., Vincent, G.M. *et al.* (2000) Spectrum of mutations in long-QT syndrome genes. *KVLQT1, HERG, SCN5A, KCNE1,* and *KCNE2*. *Circulation*, 102, 1178-1185.
24. Pfeufer, A., Jalilzadeh, S., Perz, S., Mueller, J.C., Hinterseer, M., Illig, T., Akyol, M., Huth, C., Schopfer-Wendels, A., Kuch, B. *et al.* (2005) Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. *Circ Res*, 96, 693-701.
25. Newton-Cheh, C., Guo, C.Y., Larson, M.G., Musone, S.L., Surti, A., Camargo, A.L., Drake, J.A., Benjamin, E.J., Levy, D., D'Agostino, R.B., Sr. *et al.* (2007) Common genetic variation in *KCNH2* is associated with QT interval duration: the Framingham Heart Study. *Circulation*, 116, 1128-1136.
26. Friedlander, Y., Vatta, M., Sotoodehnia, N., Sinnreich, R., Li, H., Manor, O., Towbin, J.A., Siscovick, D.S. and Kark, J.D. (2005) Possible association of the human *KCNE1* (minK) gene and QT interval in healthy subjects: evidence from association and linkage analyses in Israeli families. *Ann Hum Genet*, 69, 645-656.
27. Splawski, I., Timothy, K.W., Tateyama, M., Clancy, C.E., Malhotra, A., Beggs, A.H., Cappuccio, F.P., Sagnella, G.A., Kass, R.S. and Keating, M.T. (2002) Variant of *SCN5A* sodium channel implicated in risk of cardiac arrhythmia. *Science*, 297, 1333-1336.
28. Milan, D.J., Kim, A.M., Winterfield, J.R., Jones, I.L., Pfeufer, A., Sanna, S., Arking, D.E., Amsterdam, A.H., Sabeh, K.M., Mably, J.D. *et al.* (2009) Drug-sensitized zebrafish screen identifies multiple genes, including *GINS3*, as regulators of myocardial repolarization. *Circulation*, 120, 553-559.
29. Qu, X., Jia, H., Garrity, D.M., Tompkins, K., Batts, L., Appel, B., Zhong, T.P. and Baldwin, H.S. (2008) ndrg4 is required for normal myocyte proliferation during early cardiac development in zebrafish. *Dev Biol.*
30. Neely, G.G., Kuba, K., Cammarato, A., Isobe, K., Amann, S., Zhang, L., Murata, M., Elmen, L., Gupta, V., Arora, S. *et al.* (2010) A global in vivo Drosophila RNAi screen identifies *NOT3* as a conserved regulator of heart function. *Cell*, 141, 142-153.
31. Dasgupta, C., Martinez, A.M., Zuppan, C.W., Shah, M.M., Bailey, L.L. and Fletcher, W.H. (2001) Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). *Mutat Res*, 479, 173-186.
32. Carniel, E., Taylor, M.R., Sinagra, G., Di Lenarda, A., Ku, L., Fain, P.R., Boucek, M.M., Cavanaugh, J., Miocic, S., Slavov, D. *et al.* (2005) Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation*, 112, 54-59.
33. Ching, Y.H., Ghosh, T.K., Cross, S.J., Packham, E.A., Honeyman, L., Loughna, S., Robinson, T.E., Dearlove, A.M., Ribas, G., Bonser, A.J. *et al.* (2005) Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet*, 37, 423-428.
34. Niimura, H., Patton, K.K., McKenna, W.J., Soults, J., Maron, B.J., Seidman, J.G. and Seidman, C.E. (2002) Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation*, 105, 446-451.
35. Rohr, S. (2004) Role of gap junctions in the propagation of the cardiac action potential. *Cardiovasc Res*, 62, 309-322.
36. Moore, J.C., Tsang, S.Y., Rushing, S.N., Lin, D., Tse, H.F., Chan, C.W. and Li, R.A. (2008) Functional consequences of overexpressing the gap junction Cx43 in the cardiogenic potential of pluripotent human embryonic stem cells. *Biochem Biophys Res Commun*, 377, 46-51.
37. Vinogradova, T.M., Lyashkov, A.E., Zhu, W., Ruknudin, A.M., Sirenko, S., Yang, D., Deo, S., Barlow, M., Johnson, S., Caffrey, J.L. *et al.* (2006) High basal protein kinase A-dependent phosphorylation drives rhythmic internal Ca2+ store oscillations and spontaneous beating of cardiac pacemaker cells. *Circ Res*, 98, 505-514.
38. Vinogradova, T.M., Sirenko, S., Lyashkov, A.E., Younes, A., Li, Y., Zhu, W., Yang, D., Ruknudin, A.M., Spurgeon, H. and Lakatta, E.G. (2008) Constitutive phosphodiesterase activity restricts spontaneous beating rate of cardiac pacemaker cells by suppressing local Ca2+ releases. *Circ Res*, 102, 761-769.
39. Vasan, R.S., Glazer, N.L., Felix, J.F., Lieb, W., Wild, P.S., Felix, S.B., Watzinger, N., Larson, M.G., Smith, N.L., Dehghan, A. *et al.* (2009) Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *Jama*, 302, 168-178.
40. Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P.I., Muller, M., Morrison, A.C., Smith, A.V., Isaacs, A., Sanna, S., Dorr, M. *et al.* (2010) Genome-wide association analysis identifies multiple loci related with resting heart rate. *Hum Mol Genet.*
41. Abernethy, D.R. and Schwartz, J.B. (1999) Calcium-antagonist drugs. *N Engl J Med*, 341, 1447-1457.
42. van Noord, C., Dorr, M., Sturkenboom, M.C., Straus, S.M., Reffelmann, T., Felix, S.B., Hofman, A., Kors, J.A., Haring, R., de Jong, F.H. *et al.* (2010) The association of serum testosterone levels and ventricular repolarization. *Eur J Epidemiol*, 25, 21-28.

43. Er, F., Gassanov, N., Brandt, M.C., Madershahian, N. and Hoppe, U.C. (2009) Impact of dihydrotestosterone on L-type calcium channels in human ventricular cardiomyocytes. *Endocrine research*, 34, 59-67.
44. Er, F., Michels, G., Brandt, M.C., Khan, I., Haase, H., Eicks, M., Lindner, M. and Hoppe, U.C. (2007) Impact of testosterone on cardiac L-type calcium channels and Ca2+ sparks: acute actions antagonize chronic effects. *Cell calcium*, 41, 467-477.
45. van der Sluis, S., Verhage, M., Posthuma, D. and Dolan, C.V. (2010) Phenotypic Complexity, Measurement Bias, and Poor Phenotypic Resolution Contribute to the Missing Heritability Problem in Genetic Association Studies. *PLoS One*, 5, e13929.
46. Maher, B. (2008) Personal genomes: The case of the missing heritability. *Nature*, 456, 18-21.
47. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
48. Reich, D.E. and Lander, E.S. (2001) On the allelic spectrum of human disease. *Trends Genet*, 17, 502-510.
49. Bezzina, C.R., Pazoki, R., Bardai, A., Marsman, R.F., de Jong, J.S., Blom, M.T., Scicluna, B.P., Jukema, J.W., Bindraban, N.R., Lichtner, P. *et al.* (2010) Genome-wide association study identifies a susceptibility locus at 21q21 for ventricular fibrillation in acute myocardial infarction. *Nat Genet*, 42, 688-691.
50. Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M. and Collins, R. (2008) SLCO1B1 variants and statin-induced myopathy – a genomewide study. *N Engl J Med*, 359, 789-799.
51. Daly, A.K., Donaldson, P.T., Bhatnagar, P., Shen, Y., Pe'er, I., Floratos, A., Daly, M.J., Goldstein, D.B., John, S., Nelson, M.R. *et al.* (2009) HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genet*, 41, 816-819.
52. Suppiah, V., Moldovan, M., Ahlenstiel, G., Berg, T., Weltman, M., Abate, M.L., Bassendine, M., Spengler, U., Dore, G.J., Powell, E. *et al.* (2009) IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet*, 41, 1100-1104.
53. Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., Sakamoto, N., Nakagawa, M., Korenaga, M., Hino, K., Hige, S. *et al.* (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet*, 41, 1105-1109.
54. Forsythe, I.J. and Wishart, D.S. (2009) Exploring human metabolites using the human metabolome database. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al,* Chapter 14, Unit 14 18.
55. Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res*, 37, D603-610.
56. Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S. *et al.* (2007) HMDB: the Human Metabolome Database. *Nucleic Acids Res*, 35, D521-526.
57. Tracy, R.P. (2008) 'Deep phenotyping': characterizing populations in the era of genomics and systems biology. *Current opinion in lipidology*, 19, 151-157.
58. Gieger, C., Geistlinger, L., Altmaier, E., Hrabe de Angelis, M., Kronenberg, F., Meitinger, T., Mewes, H.W., Wichmann, H.E., Weinberger, K.M., Adamski, J. *et al.* (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*, 4, e1000282.
59. Illig, T., Gieger, C., Zhai, G., Romisch-Margl, W., Wang-Sattler, R., Prehn, C., Altmaier, E., Kastenmuller, G., Kato, B.S., Mewes, H.W. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*, 42, 137-141.

Summary

Sudden cardiac death (SCD) has a genetic risk component and identifying the genetic pathways contributing to SCD risk can lead to knowledge of novel mechanistic pathways underlying ventricular arrhythmogenesis and aid in risk stratification. Inherent to its definition, which is operational of character, the underlying etiology of SCD is heterogeneous. SCD is defined as a natural death due to cardiac causes, heralded by abrupt loss of consciousness, within one hour after the onset of acute symptoms or an unwitnessed, unexpected death of someone seen in a stable medical condition less than 24 hours previously with no evidence of a non-cardiac cause. Important reasons to study the quantitative traits that underlie the common disorder of interest are this heterogeneity, together with the assumption that common genetic complex disorders, such as SCD, have a normally distributed genetic liability conferred by Mendelian inherited genes and thus the trait would be normally distributed as a quantitative trait. For SCD, several of such endophenotypical risk factors measured on electrocardiographic recordings (ECG) are available, namely: 1) the QT interval, 2) the QRS duration and 3) the RR interval.

In addition to identification of genetic loci contributing to SCD risk, the results from genome-wide association studies on risk factors for SCD might generate novel candidate genes for further study with respect to gene-drug interactions.

The main objectives of this thesis were: 1) refining the *NOS1AP* – QT interval association signal and assessing the relation with SCD, 2) to identify, by means of genome wide association studies, the genetic variants underlying the genetically determined variation within quantitative endophenotypes of SCD, 3) to develop methodology to enhance our ability to do so and, finally, 4) to perform hypothesis driven gene-drug interaction studies. Studies described in this thesis were conducted within the Rotterdam Study, a population based cohort study, often in collaboration with other studies in larger consortia.

In *part 2*, the association between genetic variation at the *NOS1AP* locus, QT-interval and SCD is studied within the Rotterdam Study. In *chapter 2.1* fine-mapping of the association at the *NOS1AP* region and QT interval is presented. We identified common variant rs12134842 as the strongest genetic predictor of QT interval at the *NOS1AP* region. We were the first to show that multiple genetic variants at *NOS1AP* independently associate with the QT interval, which was later confirmed in other studies. In *chapter 2.2* the association between *NOS1AP* genetic variants and SCD in the Rotterdam Study is described. We provided additional evidence that rs12567209 and rs16847548 are associated with SCD as was previously reported in an independent sample.

The strength of the association was dependent on the definition of SCD that was used. We performed secondary analysis that used the stricter outcome definition of witnessed SCD only. This resulted in stronger associations despite a reduction of fifty percent in the number of cases (from 208 to 109).

In *part 3* we describe the results of large-scale consortium based genome-wide association studies on intermediate traits and risk factors for SCD, such as QT interval (*chapter 3.1*), RR interval (*chapter 3.2*) and QRS duration (*chapter 3.3*). These studies were performed in order to identify novel loci influencing the trait variability in the general population. These analyses are not based on prior hypotheses and asses the association of approximately 2.5 million genetic variants with these traits. For all three traits multiple novel loci were identified and for many of these identified loci the mechanisms underlying the associations remain to be elucidated. In a collaborative effort we showed that *SCN10A*, the locus at which we observed the most significant association with QRS duration, is expressed in the specialized His-Purkinje conduction system and selective blockade of this channel resulted in QRS duration prolongation (*chapter 3.3*). In *chapter 3.2* we showed that loci not reaching genome-wide significance are informative since inclusion of these loci increased the explained variance for RR interval. Increasing the sample size would likely lead to additional identification of relevant loci.

The observation that some of the identified loci in *part 3* were overlapping, together with the fact that some of these traits are correlated, resulted in the development of an efficient method that takes benefit from additional power that can be derived from studying correlated traits in order to identify additional pleiotropic loci. In *part 4* we describe a method using univariate genome-wide association results from two traits that, via several simple steps, can be combined to derive significance statistics taking both directionality and trait correlation into account. Using simulations we describe the test statistic behaviour under different conditions and show that the results correlate highly with standard multivariate methods. We are aiming to introduce this method on the short term to assess its value in practice.

In *part 5* we applied the candidate gene approach after we formulated the hypothesis that genetic variant rs2238018 in *CACNA1C* (*chapter 3.2*), encoding the L-type calcium channel α1C-subunit targeted by calcium channel blockers, interacts with verapamil and dilitiazem on RR interval. We observed a significant interaction between use of diltiazem and carriage of the variant allele on RR interval prolongation (heart rate lowering). This effect was strongest in women. For verapamil a similar trend was observed but did not reach statistical significance, possibly due to low numbers.

Part 6 places the main findings into perspective and discusses some relevant methodological considerations in more depth. The difficulties in interpreting genome-wide association results are discussed as well as the relevance of phenotype definitions for successful genetic (pharmaco-)epidemiological research. The discussion ends with some recommendations, including investing not only in human molecular genetics but also human phenotyping.

Samenvatting

Het risico op plotse hartdood is deels genetisch bepaald en het identificeren van de componenten die bijdragen aan dit risico kan nieuwe inzichten geven in de mechanismen die ten grondslag liggen aan het ontstaan van ventriculaire ritmestoornissen. De onderliggende oorzaken van plotse hartdood zijn zeer divers van aard, inherent aan de operationele definitie. Plotse hartdood wordt gedefinieerd als een natuurlijke dood met cardiale oorzaak, voorafgegaan door een plotseling bewustzijnsverlies, binnen één uur na de eerste symptomen, of, een niet waargenomen en onverwacht overlijden van een persoon die tot 24 uur voorafgaand nog in een stabiele medische conditie gezien is en zonder aanwijzingen voor een niet cardiale oorzaak. Deze heterogeniteit, samen met de veronderstelling dat genetisch complexe aandoeningen zoals plotse hartdood in de algemene populatie een normaal verdeelde genetische gevoeligheid hebben die resulteert in een kwantitatieve normaal verdeelde onderliggende eigenschap, zijn belangrijke redenen om de kwantitatieve risicofactoren te bestuderen van complexe en vaak voorkomende aandoeningen. Voor plotse hartdood zijn enkele van deze risicofactoren te meten op het elektrocardiogram, namelijk: 1) het QT interval, 2) het RR interval en 3) de QRS duur. Naast het identificeren van loci die een rol spelen in plotse hartdood kunnen de resultaten van genoomwijde associatie studies nieuwe kandidaten voor studies naar gen-medicatie interactie voortbrengen.

De doelstelling van dit proefschrift waren: 1) het verfijnen van het associatie signaal tussen *NOS1AP* en QT interval en het bestuderen van de associatie tussen *NOS1AP* en plotse hartdood, 2) door middel van genoomwijde associatiestudies de genetische varianten onderliggend aan de genetische bepaalde variatie tussen individuen van de kwantitatieve risicofactoren van plotse hartdood te identificeren, 3) nieuwe methoden te ontwikkelen om ons vermogen dit te doen te vergroten, en 4) het verrichten van hypothese gedreven gen-medicatie interactiestudies. De beschreven studies zijn verricht binnen de Erasmus Rotterdam Gezondheids Onderzoek (ERGO) studie (internationaal bekend als de Rotterdam Study), vaak in nauwe samenwerking met andere studies.

In *deel 2* wordt de associatie tussen genetische variatie in de *NOS1AP* regio, QT interval en plotse hartdood bestudeerd in de ERGO studie. In *hoofdstuk 2.1* wordt de verfijning van het associatie signaal tussen *NOS1AP* en QT interval beschreven. De veelvoorkomende variant rs12143842 werd geïdentificeerd als de sterkste genetische voorspeller van QT interval in de *NOS1AP* regio. Met deze studie waren wij ook de eersten die aantoonden dat meerdere varianten onafhankelijk van elkaar het QT interval beïnvloeden, wat later bevestigd is door anderen. In *hoofdstuk 2.2*

worden de resultaten beschreven van de studie naar de associatie tussen *NOS1AP* varianten en plotse hartdood binnen de ERGO studie. In deze studie hebben wij additioneel bewijs geleverd voor een associatie van de varianten rs12567209 en rs16847548 met plotse hartdood, zoals eerder was beschreven in een onafhankelijke studie populatie. De sterkte van de associatie in de ERGO studie was afhankelijk van de striktheid die gehanteerd werd in de definitie van plotse hartdood. Secundaire analyses waarbij de definitie beperkt werd tot alleen waargenomen plotse hartdood resulteerden in sterkere associaties ondanks een vermindering in het aantal cases van bijna vijftig procent (van 208 naar 109).

In *deel 3* staan de resultaten beschreven van grote genoomwijde associatiestudies centraal die verkregen zijn door in grote consortia kwantitatieve risicofactoren, als QT interval (*hoofdstuk 3.1*), RR interval (*hoofdstuk 3.2*) en QRS duur (*hoofdstuk 3.3*), van plotse hartdood te bestuderen. Deze studies hebben tot doel nieuwe loci te ontdekken die bijdragen aan de genetisch bepaalde variabiliteit van de bovengenoemde eigenschappen in de algemene bevolking. Aan deze analyses liggen geen hypotheses ten grondslag en worden 2.5 miljoen genetische varianten bestudeerd op hun associatie met deze risicofactoren. Voor alle drie bestudeerde risicofactoren zijn er meerdere nieuwe loci geïdentificeerd, waarbij voor het merendeel de mechanismen die de associatie onderliggen nog onbekend zijn. In consortiumverband hebben we voor *SCN10A*, de meest sterke associatie met QRS duur, kunnen aantonen dat het natriumkanaal, waarvoor dit gen codeert, aanwezig is in het gespecialiseerde His-Purkinje geleidingssysteem en dat selectieve blokkade van dit kanaal leidt tot QRS duur verlenging (*hoofdstuk 3.3*). In *hoofdstuk 3.2* beschrijven we dat een genotype score met inclusie van loci die niet voldeden aan de genoomwijde significantie eisen de verklaarde variantie kan doen toenemen. Vergroting van het aantal deelnemers in de studie kan leiden tot de identificatie van extra geassocieerde loci.

In *deel 3* hebben we gezien dat er een gedeeltelijke overlap bestaat tussen de geïdentificeerde loci voor QT interval, RR interval en QRS duur. Samen met het feit dat de bestudeerde eigenschappen gedeeltelijk gecorreleerd zijn heeft dit ons er mede toe bewogen een methode te ontwikkelen die op efficiënte wijze pleiotrope loci kan identificeren en die daarnaast profiteert van de extra informatie die uit de correlatie kan worden afgeleid. In *deel 4* beschrijven we een methode welke univariate genoomwijde resultaten via enkele simpele stappen combineert, daarbij rekening houdend met de correlatie en richting van de effecten. Met simulaties hebben we het gedrag van de test statistiekdistributie bestudeerd in verschillende situaties en hebben aangetoond dat de resultaten van onze methode sterk correleren met standaard multivariate methoden. Het doel is om deze methode op korte termijn te implementeren en de waarde ervan in de praktijk te bepalen.

In *deel 5* worden de resultaten van een kandidaat gen-analyse gepresenteerd. Deze analyse was gebaseerd op de hypothese dat de genetische variant rs2238018 in het *CACNA1C* gen (*hoofdstuk 3.2*), coderend voor de α1c-subunit van de L-type calcium kanaal waarop calcium antagonisten hun blokkerende werking uitoefenen, de effecten van diltiazem en verapamil op RR interval beïnvloeden. De resultaten lieten een significante interactie zien tussen het gebruik van diltiazem en dragerschap van een variant allele op RR intervalverlenging (hartfrequentie-

verlaging). Dit effect was het sterkst bij vrouwen. Voor verapamil werden gelijkende effecten gezien, echter geen statistische significantie werd bereikt.

In *deel 6* worden de belangrijkste bevindingen in perspectief geplaatst en relevante methodologische overwegingen in meer detail besproken. Onder meer de moeilijkheden in het interpreteren van genoomwijde associatieresultaten en de relevantie van fenotype definitie voor genetisch (farmaco-)epidemiologisch onderzoek worden besproken. De discussie eindigt met enkele aanbevelingen, waaronder de noodzaak om te investeren in betere fenotypering naast genotypering.

Dankwoord

Het meest gelezen deel van het proefschrift is het dankwoord. Dit is terecht, want een proefschrift kan alleen tot stand komen met de hulp van anderen.

Als eerste wil ik mijn promotoren Prof.dr. B.H.Ch. Stricker en Prof.dr. A.G. Uitterlinden bedanken. Beste Bruno, allereerst ben ik je dankbaar voor mijn eerste ervaring met wetenschappelijk onderzoek. In 2005, als geneeskunde student, heeft dit geleid tot het aangaan van het promotietraject dat nu ten einde is gekomen. Je manier van werken en de vrijheid die daarmee samenging heb ik als zeer prettig ervaren. Het is een enorme stimulans geweest om verantwoordelijkheid en initiatieven te nemen in een vroeg stadium van mijn wetenschappelijke carrière. De ervaringen die daarbij horen zijn enorm vormend en waardevol geweest. Daarnaast is het altijd een plezier om je anekdotes en visie op de maatschappij aan te horen. Ik hoop dat we onze samenwerking door kunnen zetten.

Beste André, ook bij jou staan eigen verantwoordelijk en ruimte voor initiatieven van jonge onderzoekers centraal. In combinatie met de infrastructuur voor genetisch onderzoek die mede dankzij jou inspanningen tot stand is gekomen heeft dit geleid tot enorme kansen voor mij en vele anderen. Hartelijk dank hiervoor.

Prof.dr. C.M. van Duijn, Prof.dr. J.L.C.M. van Saase en Prof.dr. J.W. Jukema wil hartelijk danken voor hun bereidheid zitting te nemen in de kleine commissie en het beoordelen van het manuscript.

Graag wil ik ook Prof.dr. A. Hofman, hoofd van de afdeling Epidemiologie, bedanken. Beste Bert, onze eerste kennismaking was in 2002, aan het einde van mijn eerste studiejaar geneeskunde. Op dat moment kwam ik op sollicitatie gesprek voor het Master of Science programma Klinische Epidemiologie voor medisch studenten. Ruim 8 jaar later schrijf ik dit dankwoord, in een proefschrift gebruikmakend van de Erasmus Rotterdam Gezondheids Onderzoek (ERGO) gegevens. Dank voor de geboden mogelijkheden!

A special word of appreciation is directed to dr. C. Newton-Cheh. Dear Chris, I just started my PhD project and already I was part of the international QTGEN collaboration initiated by you. This has been an amazing educational experience. In addition you hosted my short-term stay at two scientific inspiring institutions: Massachusetts General Hospital and the Broad Institute. Here I was forced to move beyond applying state-of-the art methods towards trying to develop them. In this light I also would like to thank dr. P.I.W. de Bakker. Dear Paul, together with Chris

we started thinking about multi-trait GWAS methods using aggregate level data. We have had very interesting discussions that helped in creating what is now chapter 4.1. This small chapter gives the thesis just that bit extra.

Since most of the co-authors are international collaborators I address all co-authors in English to adequately convey my sincere word of appreciation and express the pleasure I had working with them. Thank you all for the enormous support and hard work. Our fruitful collaborations have been made possible by the formation of the CHARGE-consortium. I therefore would like to thank the senior investigators of all participating cohort studies that made the CHARGE-consortium the perfect scientific environment for junior researchers. I especially would like to thank dr. C.J. O'Donnell and Prof.dr. B.M. Psaty, with whom I had the pleasure to work with in a more direct way. Dear Chris and Bruce, it was and still is a pleasure to work with you on a variety of CHARGE-projects.

Daarnaast wil ik collectief de deelnemers, in binnen- en buitenland, aan alle studies die bijgedragen hebben aan het werk beschreven in dit proefschrift danken voor hun deelname. Bijzondere dank gaat uit naar de deelnemers van ERGO, evenals de betrokken huisartsen en apothekers. Zonder hun was dit werk niet mogelijk geweest.

De afdeling kan niet functioneren zonder alle onderzoeksmedewerkers op het onderzoekscentrum en de faculteit, de fuppers, de computer ondersteuning en het secretariaat Epidemiologie. Dank voor jullie ondersteuning. Ook collega's zijn cruciaal voor het functioneren van een afdeling. En daarvan zijn er nogal wat op de afdeling Epidemiologie en Medische Informatica. Ik wil bij dezen alle (ex-)collega's bedanken, want jullie zijn met teveel om op te noemen… ik zou vast iemand vergeten! (En dat ligt aan mij, niet aan jullie.)

Mijn paranimfen, Wouter en Daan, bedankt dat jullie aan mijn zijde willen staan op deze dag. Wouter, mijn grote broer, zo anders, zo belangrijk… ik ben blij dat je naast me staat.

Daan, hoe laat is het? Jouw moment moet nog komen. Daar kan je je klok op gelijk zetten! In ieder geval hebben we het gezellig gehad op de kamer (en daarbuiten). Je weet wat ze zeggen: 'Gezelligheid kent geen tijd.'

Mijn schoonfamilie, altijd belangstellend en betrokken.

Lieve Wouter, Cindy, Sil en Amy. Tijdens het werk aan dit boekje waren mijn gedachten vaak bij jullie.

Pa en Ma, niets was jullie te veel om ons alle mogelijkheden te geven. Dankzij jullie ben ik waar ik ben, mijn dank hiervoor is niet in woorden uit te drukken.

Lieve Marieke, 'it's always better when we're together, we're better together!'

Bibliography

MANUSCRIPTS BASED ON THE STUDIES DESCRIBED IN THIS THESIS

Chapter 2.1
Identification of a common variant at the NOS1AP locus strongly associated to QT interval duration.
Eijgelsheim M, Aarnoudse AL, Rivadeneira F, Kors JA, Witteman JC, Hofman A, van Duijn CM, Uitterlinden AG, Stricker BH.
Hum Mol Genet. 2009 Jan 15;18(2):347-357. Epub 2008 Oct 16.

Chapter 2.2
Genetic variation in NOS1AP is associated with sudden cardiac death: evidence from the Rotterdam Study.
Eijgelsheim M, Newton-Cheh C, Aarnoudse AL, van Noord C, Witteman JC, Hofman A, Uitterlinden AG, Stricker BH.
Hum Mol Genet. 2009 Nov 1;18(21):4213-4218. Epub 2009 Jul 30.

Chapter 3.1
Common variants at ten loci influence QT interval duration in the QTGEN Study.
Newton-Cheh C*, Eijgelsheim M*, Rice KM*, de Bakker PI*, Yin X, Estrada K, Bis JC, Marciante K, Rivadeneira F, Noseworthy PA, Sotoodehnia N, Smith NL, Rotter JI, Kors JA, Witteman JC, Hofman A, Heckbert SR, O'Donnell CJ, Uitterlinden AG, Psaty BM, Lumley T#, Larson MG#, Stricker BH#.
Nat Genet. 2009 Apr;41(4):399-406. Epub 2009 Mar 22.

Chapter 3.2
Genome-wide association analysis identifies multiple loci related to resting heart rate.
Eijgelsheim M*, Newton-Cheh C*, Sotoodehnia N* ,de Bakker PI*, Müller M#, Morrison AC#, Smith AV#, Isaacs A#, Sanna S#, Dörr M#, Navarro P#, Fuchsberger C#, Nolte IM#, de Geus EJ#, Estrada K, Hwang SJ, Bis JC, Rückert IM, Alonso A, Launer LJ, Hottenga JJ, Rivadeneira F, Noseworthy PA, Rice KM, Perz S, Arking DE, Spector TD, Kors JA, Aulchenko YS, Tarasov KV, Homuth G, Wild SH, Marroni F, Gieger C, Licht CM, Prineas RJ, Hofman A, Rotter JI, Hicks AA, Ernst F, Najjar SS, Wright AF, Peters A, Fox ER, Oostra BA, Kroemer HK, Couper D, Völzke H, Campbell H, Meitinger T, Uda M, Witteman JC, Psaty BM, Wichmann HE, Harris TB, Kääb S, Siscovick DS, Jamshidi Y, Uitterlinden AG, Folsom AR, Larson MG, Wilson JF, Penninx BW, Snieder H, Pramstaller PP, van Duijn CM, Lakatta EG, Felix SB, Gudnason V, Pfeufer A, Heckbert SR, Stricker BH, Boerwinkle E, O'Donnell CJ.
Hum Mol Genet. 2010 Oct 1; 19(19): 3885-3894. Epub 2010 July 16.

Chapter 3.3
Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction.
Sotoodehnia N*, Isaacs A*, de Bakker PI*, Dörr M*, Newton-Cheh C*, Nolte IM*, van der Harst P*, Müller M*, Eijgelsheim M*, Alonso A*, Hicks AA*, Padmanabhan S*, Hayward C*, Smith AV*, Polasek O*, Giovannone S*, Fu J*, Magnani JW, Marciante KD, Pfeufer A, Gharib SA, Teumer A, Li M, Bis JC, Rivadeneira F, Aspelund T, Köttgen A, Johnson T, Rice K, Sie MP, Wang YA, Klopp N, Fuchsberger C, Wild SH, Leach IM, Estrada K, Völker U, Wright AF, Asselbergs FW, Qu J, Chakravarti A, Sinner MF, Kors JA, Petersmann A, Harris TB, Soliman EZ, Munroe PB, Psaty BM, Oostra BA, Cupples

* and # denote equal author contributions

LA, Perz S, de Boer RA, Uitterlinden AG, Völzke H, Spector TD, Liu FY, Boerwinkle E, Dominiczak AF, Rotter JI, van Herpen G, Levy D, Wichmann HE, van Gilst WH, Witteman JC, Kroemer HK, Kao WH, Heckbert SR, Meitinger T, Hofman A, Campbell H, Folsom AR, van Veldhuisen DJ, Schwienbacher C, O'Donnell CJ, Volpato CB, Caulfield MJ, Connell JM, Launer L, Lu X, Franke L, Fehrmann RS, Te Meerman G, Groen HJ, Weersma RK, van den Berg LH, Wijmenga C, Ophoff RA, Navis G, Rudan I#, Snieder H#, Wilson JF#, Pramstaller PP#, Siscovick DS#, Wang TJ#, Gudnason V#, van Duijn CM#, Felix SB#, Fishman GI#, Jamshidi Y#, Ch Stricker BH#, Samani NJ#, Kääb S#, Arking DE#.
Nat Genet. 2010 Dec;42(12): 1068-1076. Epub 2010 Nov 14.

Chapter 4
Meta-analysis of genome-wide association studies of correlated traits.
Mark Eijgelsheim, Karol Estrada, Christopher Newton-Cheh, André G Uitterlinden, Bruno HCh Stricker, Paul IW de Bakker
Manuscript in preparation.

Chapter 5
Common variation in the CACNA1C gene modulates the effect of diltiazem on heart rate.
Mark Eijgelsheim, Eline M. Rodenburg, Jan A. Kors, Jacqueline C. M. Witteman, Albert Hofman, André G. Uitterlinden, Bruno H. Ch. Stricker
Manuscript in preparation.

OTHER PUBLICATIONS

2010

The effect of short-term NSAID use on echocardiographic parameters in the elderly: a population-based cohort study.
van den Hondel KE, Eijgelsheim M, Ruiter R, Witteman JCM, Hofman A, Stricker BH.
Heart. 2010 Nov 20 [Epub ahead of print]

Dependency of phenprocoumon dosage on polymorphisms in the VKORC1, CYP2C9 and CYP4F2 genes.
Teichert M, Eijgelsheim M, Uitterlinden AG, Buhre PN, Hofman A, de Smet OA, Visser LE, Stricker BH.
Pharmacogenet Genomics. 2010 Nov 6 [Epub ahead of print]

Drug- and non-drug-associated QT interval prolongation.
van Noord C, Eijgelsheim M, Stricker BH.
Br J Clin Pharmacol. 2010 Jul;70:16-23.

High-ceiling diuretics are associated with an increased risk of basal cell carcinoma in a population-based follow-up study.
Ruiter R, Visser LE, Eijgelsheim M, Rodenburg EM, Hofman A, Coebergh JW, Nijsten T, Stricker BH.
Eur J Cancer. 2010 Sep;46(13):2467-2472. Epub 2010 Jun 3.

Single nucleotide polymorphisms in genes that are associated with a modified response to statin therapy: the Rotterdam Study.
de Keyser CE, Eijgelsheim M, Hofman A, Sijbrands EJ, Maitland-van der Zee AH, van Duijn CM, Uitterlinden AG, Witteman JC, Ch Stricker BH.
Pharmacogenomics J. 2010 Mar 2. [Epub ahead of print]

Adrenergic receptor polymorphisms and platelet reactivity after treatment with dual antiplatelet therapy with aspirin and clopidogrel in acute coronary syndrome.
Cuisset T, Hamilos M, Delrue M, Frère C, Verhamme K, Bartunek J, Saut N, Bonnet JL, Eijgelsheim M, Wijns W, Alessi MC, Barbato E.
Thromb Haemost. 2010 Mar 31;103(4):774-779. Epub 2010 Feb 2.

2009

Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function.
Hancock DB*, Eijgelsheim M*, Wilk JB*, Gharib SA*, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, Schabath MB, Couper DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JI, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T#, Stricker BH#, O'Connor GT#, London SJ#.
Nat Genet. 2010 Jan;42(1):45-52. Epub 2009 Dec 13

Hedgehog-interacting protein is a COPD susceptibility gene: the Rotterdam Study.
Van Durme YM*, Eijgelsheim M*, Joos GF, Hofman A, Uitterlinden AG, Brusselle GG, Stricker BH.
Eur Respir J. 2010 Jul;36(1):89-95. Epub 2009 Dec 8.

Common genetic variation near the phospholamban gene is associated with cardiac repolarisation: meta-analysis of three genome-wide association studies.
Nolte IM, Wallace C, Newhouse SJ, Waggott D, Fu J, Soranzo N, Gwilliam R, Deloukas P, Savelieva I, Zheng D, Dalageorgou C, Farrall M, Samani NJ, Connell J, Brown M, Dominiczak A, Lathrop M, Zeggini E, Wain LV; Wellcome Trust Case Control Consortium; DCCT/EDIC Research Group, Newton-Cheh C, Eijgelsheim M, Rice K, de Bakker PI; QTGEN consortium, Pfeufer A, Sanna S, Arking DE; QTSCD consortium, Asselbergs FW, Spector TD, Carter ND, Jeffery S, Tobin M, Caulfield M, Snieder H, Paterson AD, Munroe PB, Jamshidi Y.
PLoS One. 2009 Jul 9;4(7):e6138

No modification of the beneficial effect of NSAIDs on colorectal cancer by CYP2C9 genotype.
Siemes C, Eijgelsheim M, Dieleman JP, van Schaik RH, Uitterlinden AG, van Duijn CM, Hofman A, Coebergh JW, Stricker BH, Visser LE.
Neth J Med. 2009 Apr;67(4):134-141.

A genome-wide association study of acenocoumarol maintenance dosage.
Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, De Smet PA, van Gelder T, Visser LE, Stricker BH.
Hum Mol Genet. 2009 Oct 1;18(19):3758-3768. Epub 2009 Jul 4.

Calcium channel blockers, NOS1AP, and heart-rate-corrected QT prolongation.
van Noord C, Aarnoudse AJ, Eijgelsheim M, Sturkenboom MC, Straus SM, Hofman A, Kors JA, Newton-Cheh C, Witteman JC, Stricker BH.
Pharmacogenet Genomics. 2009 Apr;19(4):260-266.

2008

Protective effect of a GRK5 polymorphism on heart failure and its interaction with beta-adrenergic receptor antagonists.
Eijgelsheim M, Visser LE, Uitterlinden AG, Stricker BH.
Pharmacogenomics. 2008 Oct;9(10):1551-1555.

* and # denote equal author contributions

PhD portfolio

Name: Mark Eijgelsheim
Erasmus MC Department: Epidemiology
PhD Period: 2007–2011
Supervisors: Prof.dr. B.H.Ch. Stricker, Prof.dr. A.G. Uitterlinden

1. PhD training

Research skills

2002–2005 Master of Science in Clinical Epidemiology, Netherlands Institute for Health Sciences, Rotterdam, The Netherlands

Oral Presentations

2010 'Genoombrede associatie studies in intermediaire traits van cardiovasculaire ziekte en de bijdrage hiervan aan de farmacogenetica', Nederlands Netwerk Onderzoek Farmacogenetica, Utrecht, The Netherlands
2010 'Multivariate analysis of ECG-traits', CHARGE-meeting Houston, TX, USA
2009 'Genome-wide Association Analysis of 25,330 Individuals Identifies Multiple Loci Associated With Resting Heart Rate", American Heart Association Scientific Sessions, Orlando, FL, USA
2009 'Genome-wide Association Analysis of resting heart rate', CHARGE-meeting Rotterdam, The Netherlands
2009 'Common variants at 10 loci influence myocardial repolarization: The QTGEN Consortium', 49th American Heart Association Epidemiology and Prevention Annual conference, Tampa, FL, USA
2008 'QTGEN: Meta analysis of GWAS of QT interval in 13,109 individuals', CHARGE-meeting Seattle, WA, USA
2008 'ACE inhibitors, common variants in the Renin-Angiotensin System and risk of Type 2 Diabetes Mellitus', 24th International Conference on Pharmacoepidemiology and Therapeutic Risk, Copenhagen, Denmark

Poster Presentations

2008 'Identification of a common variant at the *NOS1AP* strongly associated to QT interval', 7e Wetenschapsdag Nederlandse Hartstichting, Amsterdam, The Netherlands

International Conferences

2011 CHARGE-meeting Boston, MA, USA
2010 Dutch-German Cardiovascular Meeting, Rotterdam, The Netherlands
2010 CHARGE-meeting Houston, TX, USA
2009 CHARGE-meeting Washington DC, USA
2009 American Heart Association Scientific Sessions, Orlando, FL, USA
2009 CHARGE-meeting Rotterdam, The Netherlands
2009 49th American Heart Association Epidemiology and Prevention Annual conference, Tampa, FL, USA
2008 CHARGE-meeting Seattle, WA, USA
2008 24th International Conference on Pharmacao-epidemiology and Therapeutic Risk, Copenhagen, Denmark

Seminars/Workshops/Courses

2010 Colloquium 'Pharmacogenetics of cardiovascular drugs: Implications for a safer and more efficiënt drug therapy', Koninklijke Nederlandse Akademie van Wetenschappen, Amsterdam, The Netherlands
2010 Nederlands Netwerk Onderzoek Farmacogenetica, Utrecht, The Netherlands

2009	Arrhytmia Research Methodology, COEUR, Rotterdam, The Netherlands
2009	Primer on Medical Population Genetics, Broad Institute, Cambridge, MA, USA
2009	Broad Retreat, Broad Institute, Cambridge, MA, USA
2008	A Workshop for Clinical Investigators on the Genetics of Complex Disorders, Broad Institute, Cambridge, MA, USA
2008	Advanced Epidemiology, Netherlands Institute for Health Sciences, Lunteren, The Netherlands
2008	Adanced topics in Pharmacoepidemiological methods, special skills workshops, 24th International Conference on Pharmacao-epidemiology and Therapeutic Risk, Copenhagen, Denmark
2007	SNPs and Human Disease, MolMed, Rotterdam, The Netherlands
2007 – 2011	Research Seminars, Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

2. Teaching activities

Supervising practicals

2008 – 2011	Data-analysis in Pharmaco-epidemiology, NIHES, Rotterdam, The Netherlands
2008 – 2009	Principles of Research in Medicine, NIHES, Rotterdam, The Netherlands
2007 – 2010	Pharmacoepidemiology, 4th year medical students, Erasmus MC, Rotterdam, The Netherlands
2007	Statistics, 4th year medical students, Erasmus MC, Rotterdam, The Netherlands

Supervising Master of Science students

2009 – 2010	K. van den Hondel, Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
2008 – 2009	C.E. de Keyser, Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

About the author

Mark Eijgelsheim was born on September 29th, 1981 in Rotterdam, the Netherlands. In 1999 he graduated from 'osg. Hugo de Groot' (athenaeum-beta) and started with studying Health, Policy and Management at the Erasmus University Rotterdam. In 2001 he was admitted by the Erasmus University Rotterdam, via admittance exams, to medical school and subsequently started his medical training at this institution. During his second year in medical school he started with a Master of Science programme in Clinical Epidemiology at the Netherlands Institute for Health Sciences for which he attended the Harvard University Summer School and, under the supervision of Prof.dr. B.H.Ch. Stricker, worked at the Department of Epidemiology of the Erasmus MC in Rotterdam (head: Prof.dr. A. Hofman). In 2005 he received his 'doctorandus' degree (cum laude) in Medicine as well as his 'Master of Science' degree in Clinical Epidemiology. After completing his medical training and receiving his medical degree (cum laude) in 2007 he started the work described in this thesis under the supervision of Prof.dr. B.H.Ch. Stricker and Prof.dr. A.G. Uitterlinden at the department of Epidemiology, Erasmus MC (head: Prof.dr. A. Hofman). In 2009 he was a short term visiting scientist at the Broad Institute in Cambridge, Massachusetts, USA and the Center for Human Genetic Research of the Massachusetts General Hospital, Boston, Massachusetts, USA under supervision of dr. C. Newton-Cheh, MPH. In March 2011 he will start his residency in Internal Medicine at the IJsselland Hospital in Capelle a/d IJssel (head: dr. H.E. van der Wiel) as part of his specialty training at the Erasmus MC (head: Prof. dr. J.L.C.M van Saase).