Genomics in Sudden Cardiac Death

Dan E. Arking, Sumeet S. Chugh, Aravinda Chakravarti, Peter M. Spooner

Abstract—Sudden cardiac death (SCD) remains a public health problem of major magnitude. Contrary to earlier expectations, and despite decreased overall cardiac mortality, SCD rates appear to be rising in concert with escalating global prevalence of coronary disease and heart failure, the two major conditions predisposing to SCD. With the exception of the implantable defibrillator, there are few effective approaches to SCD prevention and even fewer clues concerning patient phenotypes predisposed to life-threatening arrhythmias. Clinical variables such as ejection fraction predict mortality but are not sensitive enough to identify many high SCD risk patients. The predictive power of autonomic dysregulation and markers such as lipid levels, hypertension, diabetes, and smoking is quite low in subclinical heart disease, the population in which the majority of SCDs occur. This review addresses advances in genomic science applicable to the SCD public health problem in both rare and common forms of heart disease. These include novel bioinformatic approaches to both identify candidate genes/pathways and identify previously unknown functional genetic elements, as well as methods to comprehensively screen these elements. We also discuss the possibility of applying high-density genome-wide SNP analyses to examine genetic contributions to arrhythmia susceptibility in community-based, case-control studies of common forms of SCD. The development of novel strategies to identify contributors to susceptibility in common cardiac phenotypes is most likely to lead to new and relevant therapeutic targets for SCD. (Circ Res. 2004;94:712-723.)

Key Words: genomics ■ sudden cardiac death ■ arrhythmias ■ risk stratification

Deaths due to cardiovascular disease remain the largest contributor to premature mortality in most developed societies. Of the ~2 400 000 US deaths in 1999, ~720 000 (30%) were directly attributed to cardiac diseases. Of this number, the US Centers for Disease Control and Prevention estimated that ~462 000, or 64% of the subtotal, were “sudden cardiac deaths” (SCDs), using their definition of SCD as including all deaths “...due to cardiac disease that...”

Original received December 15, 2003; revision received February 2, 2004; accepted February 11, 2004.
From the McKusick-Nathans Institute of Genetic Medicine (D.E.A., A.C.), Johns Hopkins University School of Medicine, Baltimore, Md; Division of Cardiology (S.S.C.), Oregon Health and Science University, Portland, Ore; and D.W. Reynolds Clinical Cardiovascular Center (P.M.S.), Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, Md.
Aravinda Chakravarti is a paid member of the Scientific Advisory Board of Affymetrix. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.
This manuscript was sent to Harry A. Fozzard, Consulting Editor, for review by expert referees, editorial decision, and final disposition.
Correspondence to Dan E. Arking, Johns Hopkins University School of Medicine, 733 N Broadway, Room 580, Baltimore, MD 21205. E-mail arking@jhmi.edu
© 2004 American Heart Association, Inc.
Circulation Research is available at http://www.circresaha.org DOI: 10.1161/01.RES.0000123861.16082.95
occurred out of hospital (≈341 000) or in an emergency department, or one in which the decedent was reported ‘dead on arrival’ ” (page 123).1 Contributing comorbidities and actual precipitating causes of death are, unfortunately, not reflected in these data. And, although recorded as “sudden,” the majority of deaths tabulated in this statistic are the ultimate result of complex pathologies that develop progressively over a period of years, encompassing many different clinical and etiological phenotypes able to influence terminal arrhythmogenesis through multiple mechanisms. The concept of “sudden arrhythmic death” (SAD), implying the “sudden” occurrence of fibrillation or a potentially lethal ventricular tachy- or bradyarrhythmia as a cause of death, would be a more informative phenotype, but the terminal arrhythmia is often not documented in the community. One estimate of the potential magnitude by which ventricular arrhythmias underlie the SCD problem was provided recently by a study of deaths occurring within the community at large. Using electrocardiographic data from first responders alone, Cobb et al2 estimated that ≈185 000 SADs occur annually in the US, results similar to the estimate of ≈250 000 provided by Myerburg et al using a somewhat different approach. Given the prevalence of cardiovascular diseases potentially associated with lethal ventricular arrhythmia (≈13 000 000 affected US individuals),4 approximately 5% of the middle-aged US population has a presently indeterminable, but significant predisposition to SCD.

Potential of New Approaches

The burgeoning public health problem of SCD is all too familiar to the clinical and research communities, yet progress in developing effective approaches to prevent SAD has been difficult to achieve. Despite over two decades of federal and industrial support for the evaluation of several hundred compounds there are few antiarrhythmic drugs that reduce SAD incidence, even in patients with high-risk pathologies. Even for these agents, efficacy is observed in less than one case in five.5 6 Alternative therapies, chiefly implantable cardiac defibrillators, are expensive (≈$75 000 per patient for the device, implantation, and follow-up) requiring medical resources not readily available to many segments of the population. Although optimization of implantable cardioverter-defibrillator (ICD) use continues to evolve,7 these devices may not deliver therapies in as many as half of recipients. In the context of individual risk reduction, however, ICDs remain without doubt, an effective therapeutic modality for SAD, especially in the initial years after implantation.

The discovery of meaningful biological markers indicating elevated risk of SCD in specific individuals with differing cardiac disease etiologies could thus add much to present therapeutic approaches. Prior reliance on markers such as elevated cholesterol subtraction levels, hypertension, or markers of inflammation, although useful in guiding community public health education efforts, have proven insufficiently predictive when applied to specific individuals. Individual clinical evaluation, such as indices of autonomic tone, hemodynamics, or arrhythmia inducibility in the electrophysiology laboratory, cannot be feasibly applied to the asymptomatic population, where SCD prevalence is in fact the highest.8

The ideal SCD risk indicators would be inexpensive and testable in large numbers of individuals that currently have clinically silent susceptibility to risk. The repeated observation that lethal ventricular arrhythmias may often be the first and only manifestation of disease, argues strongly for improved individual stratification of SCD risk rather than today’s approach to nonspecific prevention in large populations. The potential of applying new approaches, such as evaluation of genomic information or proteomic expression, is one long recognized as important, but difficult to achieve in practice. As noted by multiple investigators,9 it is toward identifying individuals with unidentified, subclinical, inherited, environmental, or acquired disease risks who have not yet appeared on the established disease screen, that the greatest progress is needed. The problem is thus of individual diagnostics, with genomic approaches emerging as one of the few means by which this can be achieved.10 11

Genetic Factors in SCD

Evidence that assessment of inherited sequence variation in relevant genes offers a potential means of SCD risk identification comes from three lines of investigation.

Rare Disease Paradigm

The best known evidence supporting this idea is the extensive series of studies showing that inherited mutations in coding sequences in at least seven cardiac sarcolemmal, Na+, K+, and Ca2+ ion channel subunit genes (ie, KVLQT1 [KCNQ1], HERG [KCNH2], SCN5A, minK [KCNJ1], RYR2, MiRP1 [KCNJ2], and Kir2.1 [KCNJ2]),1 result in increased propensity to SCD. Electrophysiological dysfunction resulting in delayed myocardial cell depolarization and repolarization caused by sequence variation in the proteins encoded by these genes, as originally discovered by Keating, Schwartz, Moss, Priori, and others during the 1990s, is now known to underlie a whole family of related proarrhythmic conditions, exemplified by the long-QT and Brugada syndromes.12 13 More recent evidence suggests that mutations in these same genes that cause enhanced cell repolarization may result in converse disorders, such as “short-QT syndrome” also believed to enhance SCD risk (Arthur M. Wilde, personal communication, 2003).14 15 Ion channel gene sequence variations have now been identified as being arrhythmogenic in more than a half dozen rare inherited conditions, including Anderson’s disease (Kir2.1 [KCNJ2]),16 Brugada syndrome (SCN5A),17 one form of catecholaminergic-induced ventricular tachycardias (RYR2),18 and one form of arrhythmogenic cardiomyopathy (RYR2).19 The fact that multiple disease elements are likely to contribute to some of these conditions has become increasingly apparent as the rich and diverse nature of genotype-phenotype associations in these conditions has been revealed. One such set of possibilities, clearly important but beyond the scope of the present discussion, are those susceptibilities that arise from genetic variations in sarcomeric proteins, such as β-myosin heavy chain (MyHC), cardiac troponin T (cTnT), and myosin binding protein-C (MyBP-C), which underlie SCDs that occur in patients with
inherited hypertrophic cardiomyopathies. 20 Another important example, and again, one outside the bounds of this discussion, are those genetic changes that impact on early patterning events during embryogenesis, and subsequently cause disturbances in cardiac electrical function from development through maturation. Chien and collaborators, for example, were among the first to report that genetic alterations impacting early transcription factor expression may lead to enhanced arrhythmia susceptibility. 21 Similar alterations in developmental factors were also recently implicated in rare familial vascular defects which appear to result in enhanced susceptibility to myocardial infarction. 22

**Secondary Gene Variation and Environmental Interactions**

The second line of implicating evidence is the discovery of interactions between several of the previously mentioned channel proteins, notably the HERG K channel, and a number of “proarrhythmic” drugs and environmental agents. These include familiar antibiotics (eg, erythromycin), antihistamines (eg, terfenadine), antipsychotics, gastrointestinal (eg, cisapride) and paradoxically, antiarrhythmics such as sotalol and quinidine. 23 Parallel work has shown that polymorphisms in hepatic P450 clearance pathways for these compounds may increase risk of ventricular “torsades-de-pointe” arrhythmias by prolonged exposure to higher circulating drug levels. 24 Natural inhibitors of the P450 metabolizing pathways involved, including many present in certain foods (eg, licorice and grapefruit extracts), may also result in a similar threat by the same causative mechanisms. Enhanced SCD risk from these sources has been loosely referred to as acquired or aLQTS. The extent to which susceptibility of this nature contribute to enhancement of SCD risk in the overall population is believed to be low, but remains of serious concern to both the FDA and the pharmaceutical enterprise. Investigations as to whether individuals who demonstrate enhanced clinical susceptibility to episodes of LQTS may have identifiable sequence variation in the major ion channel proteins have been negative to date. 25

**Family History**

The third line of evidence suggesting a broad influence of genetic factors on SCD susceptibility in common forms of cardiac disease comes from two large retrospective epidemiological case-control studies conducted in broad community-based populations. The first, conducted in the United States by Friedlander et al, 26 analyzed associations with “primary cardiac events” in a cohort of men and women attended by first responders in King County, Washington (235 cases, 374 controls). The second done in Paris by Jouven et al 27 analyzed deaths in a cohort of 7746 asymptomatic middle-aged males, using retrospective autopsy and clinical data analyses to ascribe cardiac deaths to either SCD or MI. Although direct evidence for this may have been limited, both studies categorized deaths into “sudden arrhythmic events” versus those more likely precipitated by acute myocardial infarction. Multifactorial statistical analyses indicated that the occurrence of SCD in a parent results in a 1.6- to 1.8-fold increase in SCD susceptibility despite controlling for conventional risk factors indicative of coronary disease (eg, cholesterol subfractions, blood pressure, obesity, tobacco use, etc). In a very limited number of cases in the Parisian study, where there was a history of both maternal and paternal SCD events (n = 19), the relative risk in offspring was ~9 (P = 0.01), indicating an additive genetic model. Elevated incidence of SCD in the Paris study segregated independently of elevated familial incidence of myocardial infarction, with a number of conventional markers showing preferential influence on either mode of death. What is not known, of course, is whether the SCDs in these studies resulted from an influence on electrophysiological pathways, as believed to occur with the rare inherited arrhythmias, or by novel, presently unknown mechanisms. As noted previously, such influences could be both compound and complex, and related to rare or common sequence variation in multiple elements of pathways that affect either substrate electrical susceptibilities or precipitating influences (eg, sympathetic neural activation) (Figure 1). Pathways of both arrhythmia initiation and arrhythmia propagation, as well as largely unknown mechanisms of innate protection, could all be potentially involved individually or cooperatively in a multiple-gene pattern of pathology.

**Limitations of the Rare Disease Paradigm**

The idea that ion channel sequence variations that alter cardiac de- or repolarizing currents in patients with rare inherited syndromes, like LQTS, may also contribute to
enhanced SCD susceptibility seen with more common forms of cardiac disease represents a probable, but as yet unproven premise. One available piece of information supporting a more widespread translation of the rare disease paradigm comes from recent work by Splawski and colleagues. In this study, a single nucleotide sequence variant in the SCN5A Na+ channel gene found in African Americans was associated with a small enhancement of arrhythmia risk likely only important in the context of other, potentially proarrhythmic influences, eg, aLQTS-inducing drugs. The aberrant allele, a single base replacement, was estimated present in up to 4.6 million African Americans: a level of prevalence far beyond all previously established SCD susceptibility alleles combined and was not identified in other ethnic populations sampled. The specific mutation was electrophysiologically meaningful in cellular studies and may in fact be a precursor of future single or multigene discoveries in other populations. Although the extent to which this type of influence might affect incidence of arrhythmic SCDs in the broad cardiac disease population was not studied, this work is important because it represents the first successful translation of this paradigm to a wider population.

To identify the forms of genetic variation (Figure 1) that contribute most substantially to SCD susceptibility in common cardiac diseases, prospective community-based case-control studies will be needed. Although the prospect is an exciting one, it should be reiterated that firm evidence connecting specific forms of genetic variation to more common forms of SCD is not yet available. Also, the concept of multiple, small, individually insignificant genetic contributions combining to enhance overall susceptibility to SCD remains unproven.

**Applying Genomics to SCD**

**Linkage Studies and Their Limitations**

The classic paradigm for identifying disease genes has emerged largely from experience limited to Mendelian diseases—rare single gene disorders with high penetrance and individually rare mutations. Under this scenario, linkage analysis is an extremely powerful tool for disease gene identification, or positional cloning. This method relies on meiotic recombination and the screening of DNA markers in families segregating the disease phenotype. Markers closest to the disease gene will show the strongest correlation with the disease, as meiotic recombination events occurring between a marker and the disease locus will decrease this correlation. By analyzing the patterns of meiotic recombination, one can narrow the region harboring the disease gene to as little as 100 kb to several 1000 kb (depending on the study size). Although a great deal of additional effort is often required to ultimately identify the disease gene, linkage analysis can essentially reduce the pool of candidate genes from ~30,000 to ~30, a much more manageable number. Similar strategies have been applied to complex disease, but success has been limited. In complex phenotypes, disease penetrance, that is phenotypic expression, can be variable and problems arise when subjects are misclassified, ie, a subject is categorized as unaffected despite carrying the disease allele.

Furthermore, even when analysis is limited to obviously affected individuals (eg, affected sibpair analysis), nonallelic heterogeneity, a situation in which mutations at different loci cause the same phenotype, confounds these studies. Indeed, Risch has presented an analysis indicating that in diseases with multiple susceptibility loci, each with moderate effect (genotype relative risk <2.0) and intermediate allele frequencies (0.05 to 0.50), linkage analysis is unlikely to be successful, even with unrealistically large family datasets (>3000 sibpairs). Linkage studies are further compromised by the need to collect families, which can often be difficult in the case of relatively late onset diseases such as SCD. SCD has an additional complication, in that the low survival rates (<10%) often preclude the collection of DNA from multiple affected family members. Thus, studies which do not require family data, such as association studies, which are performed by comparing unrelated affected individuals to appropriate control populations, have engendered a great deal of interest, albeit with their own limitations.

**Population Selection for Case-Control Studies**

For the genomic analysis of complex phenotypes, defining the affected individual in a consistent manner is another critical, although historically difficult prerequisite. Whereas diverse cardiac disease conditions predispose to lethal arrhythmias through multiple mechanisms involving influences on both the myocardial electrical substrate, as well as initiating triggers, population studies have implied that ~80% of patients that suffer SCD have some form of CAD. As a result of this association, most studies on genetic influences in SCD have equated risk factors for SCD with those for CAD—at best a remote and mechanistically distal oversimplification. Despite several decades of investigation, however, discovering more proximal determinants of SCD susceptibility has remained elusive, especially for individuals with the more prevalent forms of coronary disease. There is obviously much potential overlap between molecular mechanisms of coronary ischemia and genesis of lethal arrhythmia in coronary artery disease, as suggested by Figure 1. This discussion will be limited primarily to influences and technologies useful in detecting ventricular arrhythmia susceptibility in primary forms of cardiac disease in adults.

Two key features of study design may provide improved, mechanistically oriented approaches to this problem. The first would be to approach SCD as a unified, yet complex disease phenotype requiring more accurate phenotyping than has been common practice. Whereas multiple disease etiologies predispose to SCD, the single unifying factor common to all individuals at risk is the initiation of a lethal arrhythmia as the terminal event. Any search for common genetic determinants of SCD should thus also provide detailed information regarding common clinical variables as well as confounders. For example, histological evidence of chronic myocardial scar or myofiber disarray could be more directly linked to SAD pathophysiology, but alterations in Ca2+ handling that occur in heart failure, inflammatory or cytokine influences associated with bacterial or viral myocarditis, or more diffuse alterations such as alterations in autonomic tone or cardiac energetics, may or may not be contributory (Figure 1).
Combined with more detailed electrocardiographic phenotyping, such an approach might ideally begin to encompass as many relevant mechanistic clinical factors as could reasonably be believed to be associated with SCD. It would be important to continue to include markers suggestive of more general forms of cardiovascular deterioration, such as deterioration in hemodynamic performance (eg, ejection fraction), which may be related to yet underappreciated proarrhythmic molecular alterations such as activation of stretch receptors. Further, use of established clinical markers substantiated in many well-defined populations over prolonged periods of time may eventually be mechanistically informative as well. The transition to more accurately defined disease phenotypes would obviously also facilitate comparative genotypic evaluations, providing a logical platform for identifying improved determinants of SAD susceptibility.

Because the majority of unpredicted SCD cases occur in the general population,34 the second and equally important component of a renewed approach to the genetic dissection of SCD would be to enhance more rigorous evaluations of its occurrence in a population-based manner. Community-based investigations of SCD have largely focused on “primary” cardiac arrest events or unexpected events ascertained from first responder agencies, and these may be only partially representative of the etiologies that contribute to many SCDs of clinical importance, for example in the heart failure population. A related issue is that the very definition of SCD varies greatly in the broad studies, which might shed light on this issue, and thus, combining such data may or may not be indicated.35 This is a central problem in the SCD research area, and one that seems unlikely to be resolved readily, as discussed recently by Myerburg,36 without more accurately defined prospective studies. More uniform definitions of different SAD phenotypes, including as much information as possible on cardiac disease context, would likely greatly improve power to identify common mechanisms.

Another means of achieving this goal is the initiation of case-control studies in the general population. SCD cases can be compared with controls that have not suffered a SCD event, but are matched for cardiac disease conditions. These provide both mechanistic and statistical advantages, yet such principles have only recently been applied. Ore-SUDS (The Oregon Sudden Unexplained Death Study), undertaken by the US Centers for Disease Control and Prevention, is a prospective, population-based investigation of SCD. Since February 1, 2002, all SCDs occurring among 660 486 residents of Multnomah County, Oregon, have been identified on an annual basis. In addition to providing a prospective annual incidence of SCD in the general population, all patients are comprehensively phenotyped and contribute to a growing genome bank of SCD.37 There are few such repositories of DNA from phenotypically well-defined SCD populations. It is possible that genetic substudies could be undertaken in large publicly funded cohort studies such as the Cardiovascular Health Study (CHS),38 Framingham Heart Study (FHS), Atherosclerosis Risk in Communities (ARIC),39 and others where there is excellent clinical as well as outcome data. However, a careful prospective determination of the SCD phenotype would remain an essential prerequisite.

Association Study Design

Once appropriate populations are available, study design assumes paramount importance, and must be guided by our understanding of both genetic characteristics of disease (eg, heredity, prevalence, etc) and genomic structure. SCD is a common, complex phenotype, and therefore, unlike a Mendelian disease, is likely due to relatively high-frequency allelic variation in one or more genes. This idea is known as the “common disease, common variant” hypothesis.40,41 The association between a specific sequence variant and a phenotype can arise in one of two ways. First, many independent and recurrent mutations can directly cause disease, as in achondroplasia.42 In this scenario, although the mutation introduced is the same, recurrent mutations occur in different individuals with different genetic backgrounds. Thus, there is association between the mutation and the phenotype but no association with neighboring polymorphisms. If this is the case for SCD, then association screens will fail unless the actual causal variant is directly screened. More commonly, a mutation may have arisen once or a few times in evolution, such as for sickle cell anemia,43 and the cystic fibrosis ΔF508 mutation,44 and each time occurred on a distinct genetic background, or haplotype. This second scenario leads to a specific haplotypic association such that sequence variants flanking the disease-causing mutation are nonrandomly associated with it, a situation known as linkage disequilibrium (LD). These associations are robust, and can even be recognized in cases with several distinct mutations at the same locus, known as allelic heterogeneity. Using the most common form of variation in the DNA sequence, single-nucleotide polymorphisms (SNPs), one could test for association between SNPs and SCD under the assumption of LD between the causal mutation and flanking SNPs. With a SNP occurring roughly every 1000 bp,45 and useful LD extending approximately 10 000 bp,46 there are sufficient numbers of SNPs available to undertake the mapping of a complex phenotype like SCD. Until recently, SNP association studies have been limited to candidate gene analysis due to both cost and the lack of genotyping technologies compatible with genome-wide association studies. For genetic dissection of SCD, as for most complex phenotypes, both of these approaches will be necessary.

Candidate Genes Previously Screened for SCD Susceptibility

Given limitations in clinical phenotyping, as well as experimental genotyping capacities, alternative strategies for case-control candidate gene selection and methodological approaches for their analysis have advanced relatively slowly. Existing investigations have largely involved selection of one or two candidate genes with presumed biological relevance (eg, a clotting protein or myocardial channel subunit) followed by screening with a limited number of known SNPs. Several published case-control studies of associations between community-encountered forms of SCD and specific candidate genes reported over the recent past illustrate some of the issues that arise in these types of explorations, including the use of varying definitions of SCD. Mikkelsson and colleagues47 investigated the prevalence of the A2 allele
of the Pr\textsuperscript{Al/A2} polymorphism of the gene for glycoprotein (GP) IIIa among 666 white, middle-aged Finnish men who died unexpectedly of multiple causes. The A2 allele was associated with acute fatal coronary thrombosis, and A2 carriers under the age of 50 were at an increased risk of SCD compared with death due to violent causes or other diseases (odds ratio [OR]=2.5; 95% confidence interval [CI]=1.2 to 5.3). This suggests a polymorphism that is either a possible predictor of SCD secondary to coronary thrombosis in early middle age, or is in LD with a variant located close to it. A subsequent study in the same population examined the \(\alpha_\text{rg}\)-adrenoreceptor insertion/deletion polymorphism, previously associated with increased risk for acute myocardial infarction.\textsuperscript{38} The deletion/deletion genotype was associated with an increased risk for SCD in logistic regression analysis with an OR versus the insertion/deletion genotype=2.0, (95% CI=1.1 to 3.7), whereas the OR versus insertion-insertion genotype was 2.1 (95% CI=1.1 to 4.0). When stratified by age <55 years, an even greater effect was seen (OR=4.5; 95% CI=2.1 to 10). Another study by Anvari et al\textsuperscript{49} examined the SCD impact of the plasminogen activator inhibitor type 1 (PAI-I) 4G/5G polymorphism. The 4G polymorphism was selected based on association with both PAI-I plasma levels and an apparent increase in coronary ischemia.\textsuperscript{50} They genotyped 97 CAD-positive survivors of SCD who received implantable defibrillators and had known associated CAD as well as 113 controls with CAD and a negative history of ventricular arrhythmia. An additive genetic risk was associated with the 4G allele: 4G/5G OR=1.9; 95% CI=1.2 to 2.9, versus 4G/4G OR=3.6; 95% CI=1.4 to 8.9. They also demonstrated that PAI-I plasma levels were higher in SCD subjects than controls, even when stratified by genotype. Together the data suggested a role for PAI-I in SCD, but there remain, however, serious caveats concerning appropriate matching between cases and controls. The 4G/5G polymorphism is in Hardy-Weinberg equilibrium in the cases, but not in the controls (\(P<0.01\)), indicating a potential for population stratification. Furthermore, the polymorphism is not in Hardy-Weinberg equilibrium in the combined dataset (\(P<0.05\)). In a separate study, Reiner et al\textsuperscript{51} found no association for the prevalence of two well-characterized prothrombotic mutations, factor V Leiden and prothrombin G20210A, in a population-based study of 145 cases and 592 controls. However, criteria for the selection of cases were different from the Anvari et al\textsuperscript{49} study, such that cases/controls with a history of clinically recognized CAD were excluded. Each of these studies illustrate some of the problems in approaching the SCD susceptibility problem using candidate gene approaches, which have been common in many earlier investigations. Thus an important question to address is how can we improve on these approaches?

**Identifying New SCD Candidate Genes**

Candidate genes for SCD could, as suggested by Figure 1, involve many different sites of variation within the diverse pathways that result in a life-threatening arrhythmia. Potential sources of variation will clearly depend on etiological context for individual cases and for different SCD phenotypes. For example, variations affecting the thrombotic cascade would obviously be much more relevant in cases of familial infarction as opposed to individuals who suffered a SCD event as a result of structural remodeling as might occur in arrhythmogenic ventricular cardiomyopathy (ARVC). Especially important might be those related to final common pathways of arrhythmogenesis common to multiple phenotypes. Thus, an obvious focus would be defects in electrogenic protein coding genes implicated in the rare inherited syndromes, such as LQTS, including the ion channel genes (see section The Rare Disease Paradigm). Cytoskeleton proteins, such as ankyrin (recently associated with LQT4\textsuperscript{52}), and mutations in genes coding for structural proteins, such as those linked to SCDs in familial hypertrophic cardiomyopathies,\textsuperscript{53-60} and in ARVC and catecholaminergic polymorphic ventricular tachycardia (CPVT) are also of interest.\textsuperscript{18} Obviously, many other genes are involved in modulating cardiac excitability and conduction in different disease states and phenotypes, eg, matrix metalloproteinases (MMPs) with an important role in fibrosis, as well as processes such as inflammation, cell-to-cell communication, and electrical and structural remodeling. Instead of systematically sampling various potential pathways of high biological risk, selection of SCD candidate genes in the past has often been a haphazard process, not straying far from those types of molecules implicated in rare disease studies. In most cases, it has been the focus on a new clinical phenotype that has led to new gene discovery. With advances in genotyping technology, comprehensively screening hundreds of genes is now feasible.

As a result, a more pressing problem is the lack of computational tools to systematically and comprehensively identify candidate genes. A casual search of the literature can identify numerous genes but will likely miss many critical genes that could play a role. Additionally, the function of the vast majority of genes is not currently known, resulting in an unacceptable level of a priori exclusion of candidate genes. One approach to address this issue is to incorporate known protein-protein interactions into the selection and prioritizing of candidate genes. Working together with Dr Akhilesh Pandey, who established the Human Protein Reference Database\textsuperscript{61} (http://www.hprd.org), we have generated a protein-protein interactions schematic for one example of a protein whose dysfunction is believed to enhance arrhythmogenesis, RYR2, a candidate gene implicated in ARVD and CPVT (Figure 2). Despite the inherent complexity of protein-protein interaction data, it is noteworthy that such an analysis readily identifies another candidate gene with a known connection to SCD, FKBP12.6,\textsuperscript{62} involved in excitation-contraction coupling. Algorithms for fully mining this new approach to identify additional candidate pathways are still being developed, but could prove useful once the more obvious candidates are exhausted. A more comprehensive analysis would involve identifying all genes with known involvement in SCD and those with biological function that make them SCD susceptibility candidates (eg, RYR1, RYR3, FKBP12.6, FKBP). Genes with unknown biological function or no obvious connection to potential SCD susceptibility pathways (eg, PP2A, GRB2) would then be ranked based on the number of interactions with the known/putative SCD susceptibility genes initially identified. Note that this analysis would
not be limited to “hub” proteins, especially when information from numerous SCD candidate genes is generated. Indeed, the notion of investigating candidate genes to identify susceptibility factors in complex phenotypes is a vast oversimplification. Instead, the ability to investigate candidate pathways, which can be identified through protein interaction data, in which one can explore biochemical interactions, will likely play a much more crucial role in identifying underlying genetic susceptibility to SCD.

Candidate Gene Analysis
With current genotyping costs decreasing rapidly, the ability to screen thousands of SNPs in thousands of samples should soon be feasible—a result which will markedly change strategic approaches in genome analysis for many diseases. It may also change the focus of efforts on SCD susceptibilities in previously unexplored ways. In most previous studies, for example, the focus has been on evaluation of SNPs in protein coding regions, largely due to the strong functional influences that changes in amino acid primary structure were found to have on SCD susceptibility in the rare inherited syndromes. Results of screens for many Mendelian disease-causing alleles have also contributed to this bias, as they have often identified missense and nonsense mutations as the primary culprit, as for example with cystic fibrosis or sickle cell disease. For a complex phenotype, such as the various commonly encountered forms of SCD, it is plausible, if not likely, that sequence variants that affect the regulation of gene expression could also contribute to enhanced disease susceptibility. The identification of regulatory regions has been technically very difficult. However, recent studies involving cross-species comparisons using readily available computer software (see review) to identify conserved noncoding elements have successfully identified elements directly implicated in gene regulation. For the comprehensive screening of a gene, SNPs in or adjacent to these conserved noncoding regions should be included, in addition to those in known exonic, promoter, and enhancer regions.

Although it is readily apparent that a positive association can implicate a gene in disease susceptibility, the converse is not so readily interpretable. Because association studies depend on LD between the markers screened and the causal mutation (see section Association Study Design), the absence of association for a SNP only indicates that that particular SNP, and flanking SNPs in LD, are not associated with disease (assuming sufficient sample size to detect associa-

Figure 2. Protein-protein interaction data for RYR2 are displayed. Test protein is displayed in green, hub proteins (interact with 3 or more additional proteins) in red, and spoke proteins (single interactions) as yellow circles. This schematic illustrates how protein-protein interaction data generated from a test protein (eg, RYR2) can be used to identify additional candidate SCD susceptibility genes (eg, FKBP12.6).
tion). Thus, a systematic approach is required to comprehensively screen a gene. Without knowing the distance over which LD extends between SNPs, the optimal number of SNPs required to cover a genomic region cannot be known in advance. Several studies have shown that LD blocks average 10 kb, albeit with a great deal of fine-scale variation. Thus for an average gene spanning 100 kb, 10 to 15 SNPs, occurring roughly every 10 kb may be required, perhaps incorporating a bias toward those found in conserved noncoding regions, known promoters, and exons. After an initial pass with these SNPs, the LD between these SNPs can be examined using publicly available software (http://innateimmunity.net/IIPGA2/Bioinformatics/) to determine the need for additional SNPs, as well as the optimal placement of these SNPs. The HapMap project (http://hapmap.cshl.org/) has already provided data for 237,000 SNPs, and over the next 2 years, optimal SNP sets will be available for all regions of the genome, greatly facilitating the ability to comprehensively screen candidate genes. The UW-FHCRC Variation Discovery Resource is also providing this information for a large number of genes hypothesized to be involved in cardiac phenotypes (http://pga.gs.washington.edu/).

**Genome-Wide Association Studies**

Although it is possible that screening candidate genes for susceptibility to SCD may prove successful, there is no guarantee that the appropriate gene will be identified for screening, or that the culprit genetic variant(s) resides within a currently identified gene locus. In fact, the effect of regulatory mutations at sites distant from the known structural gene has long been suspected in complex noncoding regions, known promoters, and exons. After an initial pass with these SNPs, the LD between these SNPs can be examined using publicly available software (http://innateimmunity.net/IIPGA2/Bioinformatics/) to determine the need for additional SNPs, as well as the optimal placement of these SNPs. The HapMap project (http://hapmap.cshl.org/) has already provided data for 237,000 SNPs, and over the next 2 years, optimal SNP sets will be available for all regions of the genome, greatly facilitating the ability to comprehensively screen candidate genes. The UW-FHCRC Variation Discovery Resource is also providing this information for a large number of genes hypothesized to be involved in cardiac phenotypes (http://pga.gs.washington.edu/).

**Assessment of Genomic Structure**

The approaches for scanning genetic variation described above make specific assumptions regarding the nature of the molecular defect(s) underlying SCD. Indeed, much of human genetics is geared to the finding of a genetic susceptibility mutation as a sequence alteration in a gene. However, it is not difficult to imagine that structural changes in the human genome (insertions, deletions, and inversions) may influence a complex phenotype like SCD. Indeed, a structural change can be a potent mutation in a complex disease, because it can alter the effects at multiple genes and be dosage-dependent across individuals. Interestingly, although the role of multiple structural changes is expected in a complex disorder such as cancer, these are seldom investigated in other phenotypes.

Current approaches to examine genomic structure are limited in their ability to detect relatively small changes (<1
Cautions and Concerns

Population Stratification
In pursuing analysis of genetic variation in the types of case-control SCD susceptibility studies discussed in this review, there are surely many potential pitfalls in deciding which identified sequence variants have real disease relevance and which do not. Some of these are related to well known or predictable statistical or methodological issues commonly encountered in association analyses. Others stem from problems in establishing a sufficiently unique or powerful phenotype or adequately assessing potentially important environmental or family information. Whenever analyzing case-control studies, one must always be aware that case and control populations may not be appropriately matched. When samples are drawn from a genetically heterogenous population, false associations due to population stratification can occur under a specific set of conditions: differences in allele frequency and differences in disease prevalence among the populations sampled (see review80). The extent to which population stratification is actually a confounding factor in association studies is subject to a great deal of debate,61 but with the decreased costs of genotyping, there is no reason not to control for potential population stratification. Several methods are available to both test and adjust for population structure. Although a comprehensive comparison of these methods is beyond the scope of this review, a detailed discussion has been provided by Devlin and colleagues.82 Briefly, the methods fall into two categories, both of which rely on genotyping markers unlinked to the candidate disease locus. “Genomic control”83 uses the data from unlinked markers to adjust the critical value for statistical significance. In contrast, “structured association”85 uses latent class analysis to estimate the number of homogenous subpopulations and assigns a probability for each subject’s membership of a subpopulation. Thus, analysis is performed within subpopulations, in which there is no confounding due to population substructure. Both these methods are widely applicable, and specific study design will dictate which technique is more appropriate.

Validation of Association
With numerous association studies being undertaken, and multiple tests within each study, reproducibility in an independent sample is essential to establishing the validity of a gene-disease association. This is especially true because the “common disease, common variant” model of SCD would predict that the causal variant may not be readily identifiable such as (ie, is unlikely to be a nonsense or obviously disruptive missense mutation, which would be strongly selected against through evolution). Indeed, a number of studies for genes involved in complex disease have found strong, reproducible association between a locus and disease phenotype, but despite extensive investigation, have been unable to identify a causal variant.86-88 Although replication is important, association to a specific locus does not conclusively demonstrate that the local gene causes disease, because it is still possible that variants in nearby regulatory elements, which may influence distant genes, actually cause the phenotype. One powerful method to address this issue was used by Carrasquillo et al87 in the examination of genes involved in Hirschsprung disease (HSCR), a common congenital intestinal obstruction. Using a genome-wide association study, they were able to show strong association between transmission of the G protein-coupled receptor EDNRB Trp276Cys hypomorphic mutation and an HSCR-susceptibility haplotype of RET, a receptor tyrosine kinase. Even though a causal variant could not be identified in RET, conclusive evidence implicating RET in Hirschsprung disease was generated by recapitulating the EDNRB/RET interaction in a mouse model. In the absence of recapitulation in a model organism, which can be costly and time-consuming, direct biochemical functional analyses of genetic variants associated with disease can provide compelling evidence linking a specific gene to disease, and should be undertaken. This may be especially true for SCD, for which transgenic mouse models may not be entirely applicable given the vast electrophysiological differences between mice and men.89

Concluding Remarks

New genomic approaches, based on high-density mapping of marker SNPs and assessment of disease susceptibility to lethal arrhythmia. As costs decrease, gene-scanning technologies improve, and new candidate disease loci are identified, the usefulness of these approaches should expand over the next several years. In the progression from genotype to complex phenotype, there probably exist multiplex, dynamic, interlinked systems and to elucidate SCD etiology, the identification of gene defects is likely to be necessary, and should translate directly into improved prediction of risk in both asymptomatic and clinically manifest populations at risk of SCD. With these advances, new clues to both prevention and treatment are likely to follow. Such information will also complement and extend prior discoveries concerning genetic variations that enhance SCD susceptibility in individuals with rare inherited arrhythmias. The problem of discerning which of these known and as yet to be revealed events are most important in understanding SCD in different cardiac disease populations, represents an enormous scientific challenge. However, the new conceptual and technological tools of genomic science coupled with the success of the human genome project are making the solution increasingly certain.

Acknowledgments
D.E.A. is supported by National Institute of Mental Health grant MH070172. S.S.C. is funded by US Centers for Disease Control and Prevention/Association of Teachers of Preventive Medicine grant
References


Genomics in Sudden Cardiac Death

Dan E. Arking, Sumeet S. Chugh, Aravinda Chakravarti and Peter M. Spooner

Circ Res. 2004;94:712-723
doi: 10.1161/01.RES.0000123861.16082.95
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2004 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/94/6/712

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/