Genetics of Human Cardiovascular Disease

Sekar Kathiresan[1,2,3,*] and Deepak Srivastava[4,5,*]
[1]Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Gladstone Institute of Cardiovascular Disease, San Francisco, CA, 94158, USA
[5]Departments of Pediatrics and Biochemistry & Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA
*Correspondence: skathiresan@partners.org (S.K.), dsrivastava@gladstone.ucsf.edu (D.S.)
DOI 10.1016/j.cell.2012.03.001

Cardiovascular disease encompasses a range of conditions extending from myocardial infarction to congenital heart disease, most of which are heritable. Enormous effort has been invested in understanding the genes and specific DNA sequence variants that are responsible for this heritability. Here, we review the lessons learned for monogenic and common, complex forms of cardiovascular disease. We also discuss key challenges that remain for gene discovery and for moving from genomic localization to mechanistic insights, with an emphasis on the impact of next-generation sequencing and the use of pluripotent human cells to understand the mechanism by which genetic variation contributes to disease.

Introduction

Cardiovascular disease (CVD) is a leading health problem, affecting more than 80,000,000 individuals in the United States alone. CVD encompasses a broad range of disorders, including diseases of the vasculature, diseases of the myocardium, diseases of the heart's electrical circuit, and congenital heart disease (Roger et al., 2012). For nearly all of these disorders, inherited DNA sequence variants play a role in conferring risk for disease. For example, in the general population, a history of premature atherosclerotic CVD in a parent confers ∼3.0-fold increase in CVD risk to offspring (Lloyd-Jones et al., 2004). The precise magnitude of the role of inheritance, however, varies by disease and by other factors such as age of disease onset and subtype of disease.

Over the past century, a key goal of biomedical research has been to correlate genotype with phenotype, i.e., to identify the specific genes and DNA sequence variants responsible for trait variation in humans. What is the principal reason to pursue this goal? Naturally occurring genetic variation has the unique potential to reveal causal biologic mechanisms in *humans*. This is particularly important, as some diseases like myocardial infarction (MI) are poorly modeled in nonhuman species.

In this Review, we consider the approaches used to discover genes for human CVD and the lessons learned from the study of Mendelian and of common, complex forms of CVD, and we take a look forward at research driven by next generation techniques, including sequencing and modeling human genetic disease in cells.

Approaches to Discover Genes for CVD

To discover genes for CVD and its risk factors in humans, two major approaches—linkage analysis and genetic association—have been utilized. The choice of approach has depended on the pattern of segregation, whether consistent with the ratios described by Mendel or more complex. Some forms of CVD exhibit a simple pattern of inheritance suggestive of a single causal gene that confers a large effect on phenotype. For many of these Mendelian forms of CVD, direct DNA sequencing and/or linkage analysis have successfully yielded the causal gene and mutation. In 1985, Lehrman and colleagues directly sequenced the low-density lipoprotein receptor (*LDLR*) gene in a patient with homozygous familial hypercholesterolemia and uncovered a 5 kilobase deletion that eliminated several exons, representing the first demonstration of a mutation for Mendelian CVD (Lehrman et al., 1985). In 1989, linkage analysis was used to localize the chromosomal position of a causal gene for hypertrophic cardiomyopathy, and in the subsequent year, mutations in the beta cardiac myosin heavy chain were discovered as causal for the phenotype (Geisterfer-Lowrance et al., 1990; Jarcho et al., 1989). Other prominent examples in the CVD field include long QT syndrome, severe hypercholesterolemia, a Mendelian family with early coronary artery disease, Mendelian forms of hypertension, Marfan's syndrome, and several forms of congenital heart disease, including septal defects and valve defects (Abifadel et al., 2003; Basson et al., 1997; Berge et al., 2000; Curran et al., 1995; Dietz et al., 1991; Garcia et al., 2001; Garg et al., 2003, 2005; Lifton et al., 2001; Mani et al., 2007; Schott et al., 1998; Soria et al., 1989; Tartaglia et al., 2001).

However, most CVD traits, such as MI or concentrations of plasma LDL cholesterol, show complex inheritance, suggestive of an interplay between multiple genes and nongenetic factors. Mapping gene loci associated with complex traits requires substantial levels of information and analysis, but since 2007, approaches to accomplish this goal have matured, and genetic mapping for complex traits in humans has become a reality. The intellectual foundations that enabled a systematic genome-wide screen of common variants (termed genome-wide association study [GWAS]) and results from this approach

were recently reviewed (Altshuler et al., 2008; O'Donnell and Nabel, 2011). The tools and methods included catalogs of polymorphisms, techniques to genotype these DNA sequence variants, and the analytical framework to distinguish true association signal from false positives. The initial focus has been on utilizing common DNA sequence variants (variants with allele frequency > 1:20) as a discovery tool, largely because it was practical to do so; recent advances in DNA sequencing and genotyping technology will allow interrogation of less frequent variants and will be considered below. The National Human Genome Research Institute hosts a catalog of published GWAS results, and as of January 14, 2012, the catalog includes 1,143 publications and 5,585 single-nucleotide polymorphisms (SNPs) with association evidence at $p < 10^{-5}$ (Hindorff et al., 2009).

What have we learned from these gene discovery efforts? Below, we discuss lessons that have emerged from genotype-to-phenotype correlation studies for Mendelian diseases, followed by lessons from studies of common, complex diseases.

Lessons Learned from the Study of Mendelian Forms of CVD

Rare Variants Lead to Broadly Relevant Insights

Because Mendelian diseases are rare in the population, there was initial skepticism about whether the genes and mechanisms that cause these diseases would inform our understanding of common forms of CVD. Linkage studies involve identifying and recruiting families with unique, often severe phenotypes, isolating a chromosomal segment that tracks with disease status in the family, and then pinpointing the causal gene and mutation in the linked segment. For a range of conditions, the genes identified by these linkage studies have transformed our understanding of CVD. Selected examples of Mendelian diseases, the responsible genes, and the gleaned biological and clinical insights are detailed in Table 1. Of particular significance is monogenic severe hypercholesterolemia, for which the six responsible genes have led to fundamental new biological concepts and have supported the development of new therapies.

Genotype-Phenotype Correlation Can Be Complex Even in Monogenic Disorders

Although there are cases in which single gene mutations lead to straightforward genotype-phenotype associations, other more complex relationships exist as well. This complexity can arise from three distinct genetic phenomena: pleiotropy, penetrance, and expressivity. Sometimes, mutations in a single gene can influence multiple phenotypic traits (i.e., pleiotropy). In 1995, Wang, Keating, and colleagues identified the α subunit of the type V voltage-gated sodium channel (SCN5A) as the cause of inherited long QT syndrome type 3 (Wang et al., 1995). Since then, mutations in the same gene have been demonstrated to cause Brugada syndrome (right precordial ST-segment elevation and increased risk for ventricular arrhythmias), cardiac conduction system disease, and dilated cardiomyopathy (Chen et al., 1998; McNair et al., 2004; Schott et al., 1999). This range of disease phenotypes may reflect the underlying functionality of the channel.

Among carriers of a Mendelian mutation in a given family, some may exhibit the condition and others may not. Penetrance refers to the proportion of individuals with a given genotype who exhibit the phenotype associated with the genotype. In many Mendelian cardiovascular conditions inherited in an autosomal-dominant manner, there is evidence for incomplete penetrance. For example, Hobbs and colleagues reported that, in a pedigree with familial hypercholesterolemia due to a point mutation in LDLR, only 12 out of 18 heterozygotes had high LDL cholesterol (>95th percentile), whereas some of the remaining 6 had LDL cholesterol as low as 28th percentile for the population (Hobbs et al., 1989). The lack of a high-cholesterol phenotype given the same genotype may be due to modifier genes or environmental influences.

Individuals with the same Mendelian genotype can also show different degrees of the same phenotype. Expressivity is the degree to which trait expression differs among individuals. Marfan's syndrome is a multisystem Mendelian disorder that can include a range of signs and symptoms involving the skeletal system (pectus excavatum, increased arm span to height ratio, craniofacial alterations), ocular system (eye lens dislocation, flat cornea), and cardiovascular system (aortic aneurysm, dissection of the ascending aorta, mitral valve prolapse), among others (Cañadas et al., 2010). Dietz and colleagues identified mutations in the FBN1 gene encoding the extracellular matrix protein fibrillin 1 as responsible for Marfan's syndrome (Dietz et al., 1991). When a specific mutation in the fibrillin 1 (FBN1) gene causes Marfan's syndrome in a family, carriers of the same mutation can display variable clinical manifestations (Faivre et al., 2007).

Pleiotropy, penetrance, expressivity, and nongenetic factors conspire to ensure that, even in a single gene disorder, genotype does not "equal" a specific phenotype. This complexity has several consequences. First, gene discovery is more difficult, as genotype may not segregate perfectly with phenotype, thereby reducing the power of linkage. Second, there is intense interest in identifying modifiers—genetic or environmental—that may modulate the relationship between genotype and phenotype. Finally, because of this complexity, in many Mendelian diseases, it has been difficult to develop genotype-specific prognostic or treatment recommendations.

The Long Road from Genotype to Mechanism to Treatment

For Marfan's syndrome, the path from the discovery of FBN1 as the causal gene to a breakthrough in the molecular understanding of the disease has spanned more than two decades. Historically, Marfan's syndrome had been viewed as a structural disease due to a defect in elastic fibers (Lindsay and Dietz, 2011). The identification of mutations in an extracellular matrix protein seemed to confirm this view. However, more recent studies suggest that microfibrils normally bind the large latent complex of the cytokine transforming growth factor β (TGF-β) and that failure of this event to occur results in increased TGF-β activation and signaling. Now, investigators are exploring the hypothesis that blocking TGF-β signaling will ameliorate the growth of aortic aneurysms in Marfan's syndrome. For further examples of therapeutic approaches derived from the study of Mendelian disorders, we refer the reader to a recent review on this topic (Dietz, 2010).

Table 1. Selected Examples of Mendelian Diseases and Insights Derived from the Study of Causal Genes

Mendelian Condition	Causal Genes	Key Biological and Clinical Insights	References
Severe hypercholesterolemia	*LDLR*, *APOB*, *ABCG5*, *ABCG8*, *ARH*, *PCSK9*	(1) Receptor-mediated endocytosis; (2) receptor recycling; (3) feedback regulation of receptors; (4) molecular mechanism of intestinal cholesterol absorption and biliary cholesterol excretion; (5) high LDL cholesterol is sufficient to cause MI.	(Abifadel et al., 2003; Brown and Goldstein, 1986; Garcia et al., 2001; Lehrman et al., 1985)
Familial hypobetalipoproteinemia	*APOB*, *PCSK9*, *ANGPTL3*	(1) Lifelong low-LDL cholesterol (from loss-of-PCSK9 function) is sufficient to protect from MI despite other coronary risk factors.	(Cohen et al., 2006; Musunuru et al., 2010a; Soria et al., 1989)
Mendelian forms of low and high blood pressure	*SLC12A3*, *SLC12A1*, *KCNJ1*, *CLCNKB*, *NR3C2*, *SCNN1A*, *SCNN1B*, *SCNN1G*; *CYP11B2*, *CYP11B1*, *HSD11B2*, *NR3C2*, *SCNN1B*, *SCNN1G*, *WNK1*, *WNK4*, *KLHL3*, *CUL3*	(1) Genes converge on a final common pathway of altering net renal sodium handling and balance; (2) identification of new targets for the treatment of blood pressure.	(Boyden et al., 2012; Chang et al., 1996; Geller et al., 2000; Geller et al., 1998; Hansson et al., 1995; Lifton et al., 1992a, 1992b, 2001; Mune et al., 1995; Shimkets et al., 1994; Simon et al., 1996a, 1996b, 1996c, 1997; Wilson et al., 2001)
Hypertrophic cardiomyopathy	*MYH7*, *TNNT2*, *TPM1*, *TNNl3*, *MYL2*, *MYBPC3*, *ACTC*, *MYL3*	(1) Mutations have expanded knowledge of the molecular mechanisms of heart muscle contraction; (2) mutations may cause increased TGF-β signaling in the myocyte with subsequent effects on neighboring fibroblasts, leading to fibrosis and scarring.	(Bonne et al., 1995; Carrier et al., 1993; Geisterfer-Lowrance et al., 1990; Kimura et al., 1997; Olson et al., 2000; Poetter et al., 1996; Seidman and Seidman, 2001; Thierfelder et al., 1994; Watkins et al., 1995)
Marfan's syndrome	*FBN1*	(1) Aneurysm formation is likely due to perturbations in cytokine signaling cascades and the smooth muscle contractile apparatus rather than defects in the extracellular matrix; (2) unexpected role for TGF-β pathway in disease.	(Dietz et al., 1991; Lindsay and Dietz, 2011)
Atrial or ventricular septal defects	*NKX2-5*, *GATA-4*, *TBX5*	(1) These transcription factors, originally discovered in flies and mice, are critical for proper heart development in humans and function in a common complex.	(Basson et al., 1997; Garg et al., 2003; Schott et al., 1998)
Bicuspid aortic valve, Calcific aortic valve disease	*NOTCH1*	(1) NOTCH1 functions to repress a default osteoblast fate of the valve mesenchymal cells; (2) NOTCH1 mutations likely result in a derepression of this fate choice and subsequent differentiation of valve cells into an osteoblast-like phenotype.	(Garg et al., 2005)

Lessons Learned from the Study of Common, Complex Forms of CVD
Gene Variants across the Spectrum of Allele Frequency Contribute to Most Complex, Common Diseases and Quantitative Traits

Variants associated with common, complex traits range in frequency from common (>1:20 frequency) to low-frequency (1:1,000 to 1:20) to very rare (<1:1,000). In other words, genetic heterogeneity from variants across the frequency spectrum may be the rule. Consider the example of plasma triglycerides, a phenotype that marks triglyceride-rich lipoproteins, including very low-density lipoprotein particles, chylomicrons, and remnant products of their metabolism. Roughly 50% of the inter-individual variability in plasma triglycerides is estimated to be

on the basis of DNA sequence variants. Johansen, Hegele, and colleagues studied individuals from the extremes of the plasma triglyceride distribution (438 individuals with high triglycerides [mean triglycerides = 14.2 mmol/l] and 327 individuals with low triglycerides [mean triglycerides = 1.2 mmol/l]) (Johansen et al., 2010) using both GWAS and resequencing of selected genes. In the GWAS, common variants at seven loci were associated with plasma triglycerides, and in the resequencing study, there was an excess of rare, nonsynonymous variants across four genes in individuals with high triglycerides when compared with those with low triglycerides. A comprehensive logistic regression model including clinical variables and both common and rare genetic variants explained 42% of total variation in hypertriglyceridemia diagnosis: clinical variables explained
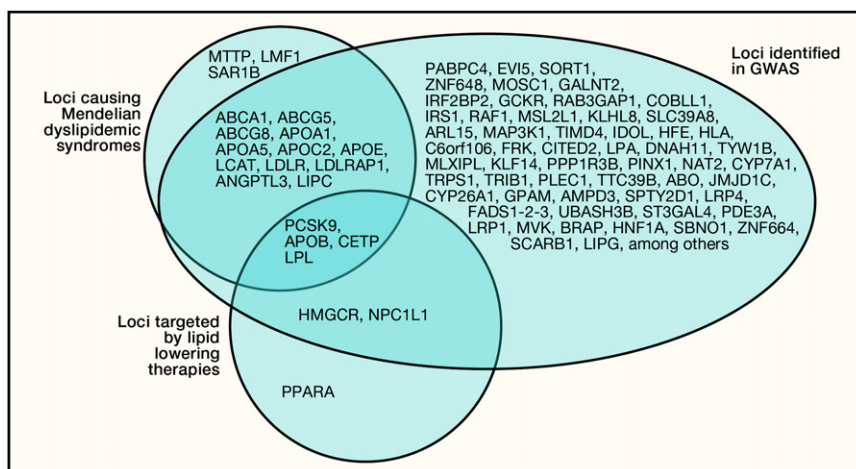
Figure 1. Overlap of Genetic Loci Causing Mendelian Dyslipidemic Syndromes, Those Targeted by Lipid-Lowering Therapies, and Those Identified in GWAS

Each GWAS locus is named according to a plausible biologic candidate in the locus or in the gene nearest to the lead single-nucleotide polymorphism (SNP). Note that, for most GWAS loci, the causal gene is not yet proven. Of 19 genes previously implicated in Mendelian lipid disorders, 16 of the genes underlying these Mendelian disorders lie within 100 kilobases of one of the lead SNPs mapped by GWAS, including nine that lie within 10 kilobases of the nearest lead SNP (Teslovich et al., 2010).

20%, common genetic variants in seven loci explained 21%, and rare genetic variants in four loci explained 1%. The genetic architecture for triglycerides in the population appears to be that of a mosaic comprised of large-effect variants rare in frequency, small-effect variants common in frequency, and environmental influences.

More generally, the concept of a mosaic model is supported by the fact that, for many cardiovascular traits and diseases, there is strong overlap between the genes mapped using GWAS and those identified earlier through Mendelian families. Nineteen genes have been identified as monogenic causes of extremely low or high levels of LDL cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides; loci harboring 16 of these genes were also mapped using GWAS (Figure 1) (Teslovich et al., 2010). Rare mutations in *FBN1* cause the thoracic aortic aneurysms and dissections seen in Marfan's syndrome, whereas common SNPs in the introns of *FBN1* are the top association result in a GWAS for spontaneous, nonsyndromic thoracic aortic aneurysm and dissection (Lemaire et al., 2011). Rare mutations in *SCN5A*, *KCNQ1*, *KCNH2*, *KCNE1*, and *KCNJ2* cause monogenic long QT syndrome, whereas common SNPs in these five genes are associated with QT interval measured on electrocardiograms in the population (Newton-Cheh et al., 2009).

New Biological Insights from GWAS Genes

Plasma lipids, platelets, and sickle cell disease represent three fields in which there has been progress toward new biology based on GWAS. GWAS for plasma LDL cholesterol, HDL cholesterol, and triglycerides have evaluated > 100,000 participants and have mapped 95 distinct loci associated with at least one of these traits at a stringent statistical threshold (p < 5 × 10^{-8}) (Kathiresan et al., 2008, 2009b; Pollin et al., 2008; Teslovich et al., 2010; Willer et al., 2008). Approximately one-third of the loci harbored genes previously appreciated to play a role in lipoprotein metabolism, including five targets of lipid-modifying therapies: *HMGCR* (statins), *NPC1L1* (ezetimibe), *APOB* (mipomersen), *CETP* (anacetrapib, dalcetrapib, and evacetrapib), and *PCSK9* (therapies in development by several pharmaceutical companies) (Figure 1).

Of note, the proportion of overall phenotypic variance explained by a genetic variant may have little correlation with the ultimate therapeutic or biological value of the gene mapped by the variant. Phenotypic variance explained by a variant is a function of two key parameters: allele frequency and effect size. For Mendelian diseases, the causal variants typically confer large effects but explain a small proportion of trait variance due to their rare frequencies. Variants from GWAS are common but explain a small proportion of trait variance due to modest effects. Nevertheless, variants that explain a small proportion of phenotypic variance may provide substantial biological or therapeutic insights. This has been highlighted for Mendelian genes in Table 1. Two examples for common SNPs include variants in the introns of *HMGCR* and *NPC1L1* (Teslovich et al., 2010). These SNPs confer a small effect on plasma LDL cholesterol at 3 mg/dl and 2 mg/dl, respectively; however, targeting of these genes with statins or ezetimibe, respectively, has a much more dramatic effect on LDL cholesterol. And to date, there have been no rare, large-effect Mendelian mutations described in *HMGCR*, presumably because such mutations are highly deleterious and not tolerated. Thus, some disease genes may only be discoverable through common, small-effect variants.

Approximately two-thirds of the 95 loci discovered for plasma lipid traits harbored genes not previously appreciated to play a role in the biology of lipoproteins. Several genes at the "novel" loci have now been manipulated in the mouse, and this manipulation led to plasma lipid changes analogous to that suggested by the human genetics. An example is the 8q24 locus containing the tribbles homolog 1 (*TRIB1*) gene (Burkhardt et al., 2010). DNA sequence variants downstream of the *TRIB1* gene were initially associated with plasma lipids, with the minor allele having association with lower plasma triglycerides, lower plasma LDL cholesterol, and higher plasma HDL cholesterol. Given this pattern of plasma lipid changes, minor allele carriers would be expected to have lower risk for coronary heart disease, and several groups have confirmed this expectation (Varbo et al., 2011). Targeted deletion of *Trib1* in mice led to elevated levels of plasma triglycerides and cholesterol, whereas liver-specific overexpression of *Trib1* reduced levels of plasma triglycerides and cholesterol. *SORT1*, *GALNT2*, *PPP1R3B*, and *TTC39B* are four other genes implicated in lipoprotein regulation by GWAS

and are confirmed in a similar manner using mouse models (Musunuru et al., 2010b; Teslovich et al., 2010).

Analogous to plasma lipids, well-powered GWAS and experimental follow-up in cells and model organisms have identified new regulators of thrombopoiesis and/or erythropoiesis. A GWAS study for blood platelet count and volume mapped 58 loci (Gieger et al., 2011). For genes at 11 mapped loci, gene silencing in either the zebrafish or *Drosophila* led to alterations in thrombopoiesis and/or erythropoiesis. The precise mechanisms by which these genes alter blood cell formation remain to be defined through additional studies; however, we now have a plethora of new genes for experimental follow-up, all with a foundation of relevance to human phenotypes.

The identification of BCL11A as a transcriptional regulator of fetal hemoglobin synthesis and as a potential therapeutic target for sickle cell disease represents more biology learned from GWAS. Sickle cell disease results from the substitution of a valine for glutamic acid in the β-globin chain of adult hemoglobin. The mutated hemoglobin (HbS) undergoes conformational change and polymerization upon deoxygenation, leading to hemolysis, red blood cell deformation, and pathology due to microvascular occlusion. Hereditary persistence of fetal hemoglobin (HbF) decreases the severity of sickle cell disease, and the level of HbF in adults is inherited as a quantitative trait. As a result, induction of HbF in adults has been a long-standing goal of therapies for sickle cell disease. In 2007, GWAS for the HbF phenotype revealed a single strong locus, with sequence variants in the intron of a transcription factor, B cell CLL/lymphoma 11A (*BCL11A*) (Menzel et al., 2007). A series of elegant studies led by Orkin and colleagues has now established that BCL11A represses the transcription of HbF expression in erythroid cells (Sankaran et al., 2008, 2009; Xu et al., 2010) and demonstrated that inactivation of BCL11A in a mouse model of sickle cell disease leads to pancellular induction of HbF and corrected the hematologic and pathologic defects of the disease (Xu et al., 2011).

Despite these successes, for the vast majority of mapped loci, it has been a challenge to move from genomic localization to biologic mechanism. Why? First, as discussed earlier in the context of discovery of Mendelian disease genes, inferring new biology from human genetics takes time, and only about 5 years have elapsed since the initial GWAS publications. Second, gene mapping and experimental follow-up require unique skill sets and expertise. In the examples highlighted above, collaborations between experimentalists and researchers focused on human genetics have been crucial to making progress, yet skepticism regarding the validity and value of discoveries made from GWAS may have dampened the enthusiasm for such collaborations (McClellan and King, 2010).

Third, genetic mapping by association gives us gene regions and not necessarily specific causal variants or causal genes. For each SNP and locus mapped by GWAS, we can generally conclude that, within ~100,000 bases of the locus, there exists a causal gene. Ideally, at each discovered locus, we need to understand: (1) the causal variant, (2) the causal gene, (3) the mechanism by which the variant affects the gene, and (4) the mechanism by which the gene affects the phenotype. At the 1p13 locus for LDL cholesterol and MI, fine-mapping,

resequencing, and manipulation of positional candidate genes in cell culture and model organisms have addressed these key questions (Musunuru et al., 2010b). However, at most loci, the answers remain a mystery.

Several features of genetic mapping using common variants have contributed to the difficulties. Local correlation (linkage disequilibrium) and weak effect size of common variants have made it difficult to identify causal variants. Mapped SNP variants have weak or modest effect (odds ratios of 1.05 to 1.40 for a dichotomous trait and < 1% of variance explained for continuous traits), and these weak effects mean that, in order to statistically distinguish between two correlated variants, the sample sizes need to be inordinately large. Nearly all of the mapped SNP variants are noncoding, and our ability to interpret the noncoding portion of the genome remains limited. Mechanisms by which a noncoding variant may affect a nearby gene include, among others, affecting a transcription factor or microRNA-binding site and regulating the local chromatin state.

In only a few instances has there been direct demonstration that a specific noncoding GWAS SNP affects the relevant gene through one of these mechanisms (Musunuru et al., 2010b). The example of chromosome 9p21 and risk for MI exemplifies the challenge. In 2007, several independent GWASs identified SNPs on chromosome 9p21 as associated with MI or coronary artery disease, with ~50% of the population carrying a risk allele and each copy of the risk allele conferring ~29% increase in risk for MI/coronary artery disease (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). The most strongly associated SNPs were noncoding and > 100 kilobases downstream of the nearest protein-coding genes *CDKN2B* (cyclin-dependent kinase inhibitor 2B encoding the protein $p15^{INK4b}$) and *CDKN2A* (cyclin-dependent kinase inhibitor 2A encoding the proteins $p16^{INK4a}$ and $p14^{ARF}$). Resequencing and fine-mapping studies in the gene region have identified a set of SNPs with indistinguishable statistical evidence, but not one causal variant (Shea et al., 2011). By what mechanism does the risk allele alter the relevant gene at the locus, and what is the gene that is responsible for atherosclerosis susceptibility? Answers to these two questions have been particularly elusive. Several mechanisms by which the noncoding variant might impact phenotype have been explored, including alteration of a noncoding RNA in the region (e.g., antisense noncoding RNA in the INK4 locus [*ANRIL*]) and disruption of the binding of a transcription factor, as observed with STAT1 (Harismendy et al., 2011; Holdt and Teupser, 2012). No definitive answers have emerged. Finally, a noncoding gene region homologous to the human-associated interval on 9p21 was deleted in mice, and this targeted deletion led to near absence of *CDKN2A* and *CDKN2B* expression in several tissues, increased rate of tumors, and increased proliferation of aortic smooth muscle cells (Visel et al., 2010). However, these mice did not develop increased atherosclerosis on a diet rich in fat and cholesterol. For the 9p21 locus, the path from genomic localization to functional insights has not been straightforward, likely due to our limited ability at present to interpret the noncoding portions of the human genome. However, as recently reviewed (Raychaudhuri, 2011), integration of GWAS findings with expression quantitative trait loci (eQTL) (eQTLs are genetic variants that correlate with

the transcript level of a gene) and genome-wide maps of chromatin state dynamics may improve our ability to interpret GWAS findings and understand how noncoding variation regulates genes (Ernst et al., 2011; Schadt et al., 2008).

Gene Variants May Distinguish Causal from Noncausal Biomarkers

Hypotheses concerning etiologic agents for complex diseases have often initially come from observational epidemiology. In 1961, in a paper entitled ''Factors of Risk in the Development of Coronary Heart Disease,'' Dr. William Kannel and colleagues at the Framingham Heart Study established an association of plasma total cholesterol with future risk for coronary heart disease (Kannel et al., 1961). Since then, hundreds of soluble biomarkers have similarly been associated with risk for coronary artery disease. How many of these biomarkers directly cause coronary artery disease, and how many simply reflect other causal processes? And why is this question important? Both causal and noncausal biomarkers may be helpful in terms of predicting risk for future disease. However, only a causal biomarker may be appropriate as a target of therapy. The ultimate proof of causality in humans is a randomized controlled trial testing whether a treatment that alters the biomarker will affect disease risk. However, as clinical trials are expensive and time-consuming, it would be helpful to have evidence in humans prior to engaging in a clinical trial.

In a technique termed Mendelian randomization, DNA variants are used to address the question of whether an epidemiological association between a risk factor and disease reflects a causal influence of the former on the latter (Davey Smith and Ebrahim, 2003; Gray and Wheatley, 1991; Katan, 1986). In principle, if a DNA variant is known to directly affect an intermediate phenotype (e.g., a variant in the promoter of a gene encoding a biomarker, affecting its expression) and the intermediate phenotype truly contributes to the disease, then the DNA variant should be associated with the disease to the extent predicted by: (1) the size of the effect of the variant on the phenotype and (2) the size of the effect of the phenotype on the disease (Musunuru and Kathiresan, 2010). If in an adequately powered sample, the predicted association between the variant and disease was not observed, it would argue against a purely causal role for the intermediate phenotype in the pathogenesis of the disease. The study design is akin to a prospective randomized clinical trial in that the randomization for each individual occurs at the moment of conception—genotypes of DNA variants are randomly ''assigned'' to gametes during meiosis, a process that should be impervious to the typical confounders observed in observational epidemiological studies. For example, a parent's disease status or socioeconomic status should not affect which of the parent's two alleles at a given SNP is passed to a child, with each allele having an equal (50%) chance of being transmitted via the gamete to the zygote. Thus, Mendelian randomization should be unaffected by confounding or reverse causation. Mendelian randomization has potential shortcomings, including: (1) the technique is only as reliable as the robustness of the estimates of the effect sizes of the variant on the phenotype and of the phenotype on disease and (2) it assumes that the DNA variant does not influence the disease by means other than the intermediate phenotype being studied (pleiotropy),

which may not be true. Nevertheless, Mendelian randomization has the potential to be as informative as a traditional randomized clinical trial.

Several Mendelian randomization studies have confirmed a causal relationship between LDL cholesterol and coronary artery disease. Nonsense variants in the *PCSK9* gene that significantly reduce plasma LDL cholesterol concentrations were observed to be associated with reduced incidence of coronary artery disease in an African American cohort (Cohen et al., 2006). Similarly, in European Americans, a common missense variant in *PCSK9* associated with lower LDL cholesterol levels was also found to be associated with lower risk of MI. These observations suggested that lower LDL cholesterol is sufficient to provide protection against coronary artery disease. Similar to LDL cholesterol, several recent genetic studies have confirmed prior observations that plasma lipoprotein(a) causally relates to coronary artery disease (Clarke et al., 2009; Kamstrup et al., 2009).

Unlike the results with plasma LDL cholesterol concentrations and plasma lipoprotein(a), three recent, large Mendelian randomization studies of C-reactive protein gene (*CRP*) variants that affect plasma CRP concentrations, performed in thousands of individuals, did not show an association between these variants and ischemic vascular disease or coronary artery disease (Elliott et al., 2009; Wensley et al., 2011; Zacho et al., 2008). It is unlikely that these studies were confounded or affected by pleiotropy, as the tested variants were within the *CRP* gene itself rather than being in other loci that might secondarily affect CRP levels. Although these studies cannot definitively rule out a causal role of CRP in cardiovascular disease, they strongly suggest that high CRP levels are indirectly rather than directly related to coronary artery disease.

A parallel line of genetic evidence also casts doubt on the notion that inflammatory biomarkers such as CRP are critical mediators of MI and coronary artery disease. Of 33 loci most highly associated with MI and coronary artery disease (Table 2), nine are related to plasma LDL cholesterol or lipoprotein(a), arguing for a strong causal relationship between LDL (or a modified LDL particle such as lipoprotein(a)) and disease (Schunkert et al., 2011). None of the other 24 loci are clearly related to inflammation. Overall, with the recent explosion in our ability to measure both soluble biomarkers (including metabolites and proteins) and genetic variation, Mendelian randomization will likely be an increasingly utilized approach to distinguish causal biomarkers from noncausal ones.

Next Generation of Approaches
Massively Parallel Sequencing as a Discovery Tool

As described above, considerable progress has been made in correlating genotype to phenotype for both Mendelian and common, complex diseases over the past several decades. Over the next decade, we are in a unique position to ''finish the job.'' For Mendelian diseases, the goal is nothing short of attempting to solve all Mendelian diseases not yet mapped. For common, complex diseases, the goal is to extend the range of genetic variation evaluated from common (>1:20 frequency) to low-frequency (1:1,000–1:20 frequency) and very rare (<1:1,000 frequency).

Table 2. Genetic Loci Mapped by GWAS for Myocardial Infarction or Coronary Artery Disease

Unique Locus	Chr	SNP	Risk Allele Frequency	Odds Ratio (95% CI) per Risk Allele	Genes of Interest within or Near Associated Interval	Association of SNP or Proxy with Other Cardiovascular Phenotypes	References
Also associated with established risk factor—plasma LDL cholesterol							
1	1p13	rs599839	78%	1.11 (1.08–1.15)	CELSR2, PSRC1, SORT1	eQTL for SORT1, CELSR2, and PSRC1 transcript levels in liver	(Samani et al., 2007; Schunkert et al., 2011)
2	19p13	rs6511720	89%	1.18 (1.11–1.25)	LDLR		(IBC 50K CAD Consortium, 2011; Kathiresan et al., 2009a)
3	1p32	rs11206510	82%	1.08 (1.05–1.11)	PCSK9		(Kathiresan et al., 2009a)
4	6q25	rs3798220	2%	1.51 (1.33–1.70)	LPA	lipoprotein(a)	(Clarke et al., 2009)
4	6q26	rs10455872	7%	1.68 (1.43–1.98)	LPA	lipoprotein(a)	(Clarke et al., 2009)
5	9q34	rs579459	21%	1.10 (1.07–1.13)	ABO	venous thromboembolism, ACE enzyme activity, plasma E-selectin level, plasma vWF level, among others	(Reilly et al., 2011; Schunkert et al., 2011)
6	11q23	rs964184	13%	1.13 (1.10–1.16)	ZNF259, APOA5, APOA1	triglycerides, HDL cholesterol	(Schunkert et al., 2011)
7	8q24	rs17321515	53%	1.06 (1.03–1.10)	TRIB1	triglycerides, HDL cholesterol	(IBC 50K CAD Consortium, 2011; Varbo et al., 2011)
8	2p21	rs4299376	32%	1.07 (1.04–1.11)	ABCG5, ABCG8	serum phytosterols	(IBC 50K CAD Consortium, 2011; Teupser et al., 2010)
9	19q13	rs2075650	14%	1.14 (1.09–1.19)	APOE		(IBC 50K CAD Consortium, 2011)
Also associated with established risk factor—blood pressure							
10	12q24	rs3184504	44%	1.07 (1.04–1.10)	SH2B3	LDL cholesterol, platelet count, plasma eosinophil count, among others	(Gudbjartsson et al., 2009; Schunkert et al., 2011)
11	10q24	rs12413409	89%	1.12 (1.08–1.16)	CYP17A1, CNNM2, NT5C2	intracranial aneurysm	(Schunkert et al., 2011; Yasuno et al., 2010)
Unknown mechanism by which variant confers risk for myocardial infarction or coronary artery disease							
12	9p21	rs4977574	46%	1.29 (1.23–1.36)	CDKN2A, CDKN2B, ANRIL	coronary artery calcification, intracranial aneurysm, abdominal aortic aneurysm, among others	(Wellcome Trust Case Control Consortium, 2007; Helgadottir et al., 2007; McPherson et al., 2007)
13	21q22	rs9982601	15%	1.18 (1.12–1.24)	SLC5A3, MRPS6, KCNE2	eQTL for MRPS6 transcript level in blood	(Kathiresan et al., 2009a)
14	1q41	rs17465637	74%	1.14 (1.09–1.20)	MIA3		(Samani et al., 2007)
15	10q11	rs1746048	87%	1.09 (1.07–1.13)	CXCL12	plasma CXCL12 level	(Samani et al., 2007)
16	6p24	rs12526453	67%	1.10 (1.06–1.13)	PHACTR1	coronary artery calcification	(Kathiresan et al., 2009a)
17	2q33	rs6725887	15%	1.14 (1.09–1.19)	WDR12, NBEAL1	eQTL for NBEAL1 transcript level in aortic media	(Kathiresan et al., 2009a)
18	3q22	rs9818870	18%	1.12 (1.07–1.16)	MRAS	eQTL for MRAS transcript level in aortic media	(Erdmann et al., 2009)
19	1p32	rs17114036	91%	1.17 (1.13–1.22)	PPAP2B		(Schunkert et al., 2011)
20	6p21	rs17609940	75%	1.07 (1.05–1.10)	ANKS1A		(Schunkert et al., 2011)
21	6q23	rs12190287	62%	1.08 (1.06–1.10)	TCF21	eQTL for TCF21 transcript level in liver, fat	(Schunkert et al., 2011)
22	7q32	rs11556924	62%	1.09 (1.07–1.12)	ZC3HC1		(Schunkert et al., 2011)
23	13q34	rs4773144	44%	1.07 (1.05–1.09)	COL4A1, COL4A2		(Schunkert et al., 2011)
24	14q32	rs2895811	43%	1.07 (1.05–1.10)	HHIPL1		(Schunkert et al., 2011)

Table 2. *Continued*

Unique Locus	Chr	SNP	Risk Allele Frequency	Odds Ratio (95% CI) per Risk Allele	Genes of Interest within or Near Associated Interval	Association of SNP or Proxy with Other Cardiovascular Phenotypes	References
25	15q25	rs3825807	57%	1.08 (1.06–1.10)	*ADAMTS7*		(Reilly et al., 2011; Schunkert et al., 2011)
26	17p13	rs216172	37%	1.07 (1.05–1.09)	*SMG6*		(Schunkert et al., 2011)
27	17p11	rs12936587	56%	1.07 (1.05–1.09)	*RASD1*, *PEMT*, *RAI1*	eQTL for RASD1 and PEMT transcript levels in monocytes	(Schunkert et al., 2011)
28	17q21	rs46522	53%	1.06 (1.04–1.08)	*UBE2Z*	eQTL for UBE2Z transcript level in blood	(Schunkert et al., 2011)
29	7q22	rs10953541	80%	1.08 (1.05–1.11)	*BCAP29*, *DUS4L*		(C4D, 2011)
30	10p11	rs2505083	38%	1.07 (1.04–1.09)	*KIAA1462*		(C4D, 2011)
31	11q22	rs974819	32%	1.07 (1.04–1.09)	*PDGFD*	eQTL for PDGFD transcript level in aortic media	(C4D, 2011)
32	10q23	rs1412444	42%	1.09 (1.07–1.12)	*LIPA*	eQTL for LIPA transcript level in monocytes	(C4D, 2011; Wild et al., 2011)
33	6p24	rs6903956	7%	1.65 (1.44–1.90)	*c6orf105*		(Wang et al., 2011)

Dramatic advances in sequencing and genotyping make such ambitious goals feasible. Next-generation sequencing platforms have markedly decreased the cost of DNA sequencing when compared with Sanger sequencing. Hybridization approaches have enabled selection of the portion of the genome that is protein coding (roughly 1% of the 3.2 billion bases), the so-called "exome." Sequencing just the exome (rather than the entire genome) is well justified in the search for genetic causes of rare inherited disorders (Ng et al., 2010). This is because the majority of alleles that are responsible for Mendelian disorders disrupt protein-coding sequence, and a large fraction of rare missense mutations in the human genome are predicted to be deleterious (Kryukov et al., 2007; Stenson et al., 2009).

Next-Generation Sequencing and Mendelian Disease

Over the past three years, whole-exome and whole-genome sequencing have been successfully utilized to identify new genes for several Mendelian forms of CVD. Most studies have sequenced the exomes of one or a few individuals affected with the disorder. Variants seen in these individuals are typically compared with those from reference individuals (unaffected individuals who are family members or unrelated). Variants that are shared by affected individuals and are not present in the unaffected population are considered causal candidates. With this strategy, investigators identified *ANGPTL3* mutations as a cause of familial combined hypolipidemia, *BAG* mutations for dilated cardiomyopathy, *NT5E* mutations for arterial calcification, *KCNJ5* mutations for hereditary hypertension due to aldosterone-producing adenomas, and *KLHL3* or *CUL3* mutations for hypertension and hyperkalemia (Boyden et al., 2012; Choi et al., 2011; Musunuru et al., 2010a; Norton et al., 2011; St Hilaire et al., 2011).

The discovery of *KLHL3* as a cause of pseudohypoaldosteronism type II, a Mendelian form of hypertension, is an elegant example of the value of exome sequencing (Boyden et al., 2012). The gene was identified by directly sequencing the exomes of 11 unrelated index cases with pseudohypoaldosteronism type II and contrasting with exome sequences from 699 population-based, unrelated controls without hypertension. Of 22 case chromosomes, 5 harbored a rare, protein-altering mutation in *KLHL3*, whereas only 2 of 1,398 control chromosomes did so, leading to a low probability that this observation was due to chance (p = 1×10^{-8}). Such successes have raised expectations for the exome sequencing approach.

However, the true yield of exome sequencing in solving Mendelian disorders is difficult to know at present, as negative results have not been routinely reported (Bamshad et al., 2011). It can be difficult to arrive at a single causal mutation after exome sequencing due to the following reasons. First, the causal variant may not be protein coding. Second, the causal variant may be protein coding, but the relevant gene may not be successfully captured and sequenced. Approximately 5%–10% of all exons may be poorly sequenced due to genomic features such as high GC content. Third, if the causal mutation is not fully penetrant, it will be present in individuals who are both phenotypically affected and unaffected and thus may be "filtered" out. Fourth, the segregation of a phenotype in a family may be due to nongenetic factors rather than a single gene of large effect. This problem is particularly true for CVD conditions that are common in the population like MI and atrial fibrillation. If one observes a multigeneration family in which multiple individuals are affected with MI, is this pattern due to a new Mendelian gene or poor lifestyle habits shared by the family? Fifth, in contrast to single-nucleotide substitutions, methods for calling small insertion-deletions and copy number changes from short-read sequence data are in need of improvement. If the causal mutation is a one or two base-pair insertion or deletion, it might be overlooked. Finally, exome sequencing and filtering are more efficient at reducing the number of candidate causal mutations for recessive rather than dominant disorders. After exome sequencing and filtering for a dominantly inherited
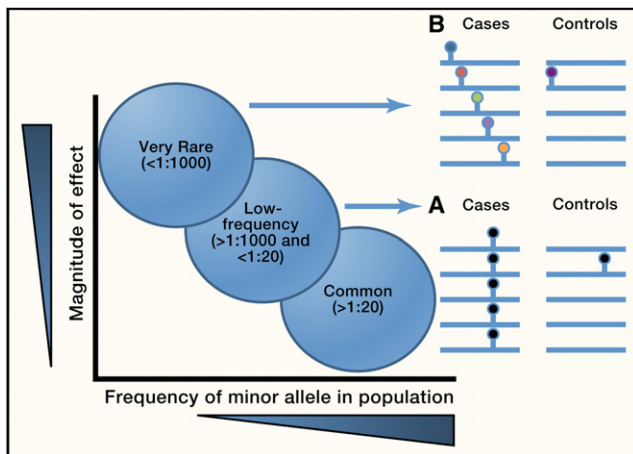
Figure 2. Association Signals Arising from Rare Variants
(A and B) Each line represents DNA sequence from a different individual, and each circle represents a DNA sequence variant. Each color denotes a DNA sequence variant allele at a different nucleotide site.
(A) "Low-frequency" variants can be defined as variants in frequency between 1:1000 and 1:20. Each variant can be assayed typically by genotyping and tested for association individually using single marker tests of association.
(B) "Very Rare" variants can be defined as variants with frequency less than 1:1,000, and often the variant may be seen only in a single person (i.e., a singleton). Variants in this frequency range need to be grouped together and tested in aggregate. Association signals from a burden of such very rare variants are typically best validated using resequencing of the same genomic interval in additional individuals. Figure adapted from Manolio et al. (2009).

condition, the number of mutations compatible with causing disease can be large.

Next-Generation Sequencing and Common, Complex Disease

What about exome or whole-genome sequencing to identify rare variants that confer a large effect on common, complex traits and diseases? The optimal study design for complex traits will depend on the frequency of the genetic variant(s) that are the source of the association signal. Though the term "rare" is used in the literature to refer to variants less than 1:20 frequency, we find it useful to distinguish between two types of "rare" variants, those that range in frequency from 1:1,000 to 1:20 and are denoted as "low-frequency" and those that are < 1:1,000 frequency and are denoted as "very rare" (Figure 2). This categorical distinction is important because the approach to discover and replicate the two signals can differ.

For low-frequency variants, it is now possible to catalog nearly all coding variants segregating in a population at a frequency of 1:1,000 or greater, directly genotype these variants, and test for association with disease (Figure 2A). For example, data from 12,000 human exome sequences have been mined, and all coding sites where the alternate allele is seen at least three times (twice for nonsense and splice) and in two separate studies have been cataloged. Investigators have designed a custom genotyping array—the "Exome Chip" —to directly genotype these ~300,000 variants at low cost. This genotyping array will allow for comprehensive testing of low-frequency coding variants for association with cardiovascular traits and diseases.

In contrast, a signal emerging from a burden of very rare variants can best be discovered by sequencing and can often only be replicated by sequencing (Figure 2B). The promise and challenges of this approach are evident in the candidate gene sequencing studies completed over the past few years. In seminal work, Cohen, Hobbs, and colleagues tested the hypothesis that rare, coding variation in candidate genes contributes to plasma HDL cholesterol variation in the population (Cohen et al., 2004). Three genes that cause Mendelian forms of low HDL cholesterol (*ABCA1*, *APOA1*, and *LCAT*) were sequenced in 128 individuals ascertained from the lowest 5th percentile (mean, 31 mg/dl) and in 128 individuals from the highest 95th percentile of the HDL cholesterol distribution (mean, 91 mg/dl) in a population-based cohort study. The investigators aggregated rare variants that met three criteria: (1) present within a gene, (2) changed amino acid, and (3) exclusively present in either the low or high group. Twenty such rare alleles were found in the low group, and two such rare alleles were found in the high group (Fisher's exact, p = 0.00006). This observation was replicated in an independent study, and additionally, cells from mutation carriers had reduced cholesterol efflux rates, a key function of ABCA1. This study established that rare nonsynonymous mutations in *ABCA1* with large phenotypic effects contribute to low HDL cholesterol in the population.

The challenge now is to utilize sequencing to enable de novo discovery of new genes that contribute to complex traits. It has become practical to extend the experiment described above from the sequencing of three genes to ~20,000 genes. Exomes can be rapidly sequenced, and single-nucleotide substitutions can be accurately called; however, will such sequencing for complex traits readily yield new discoveries? The key barriers at present are statistical. DNA sequence variants that are very rare (e.g., a mutation seen in a single person or a singleton) cannot be tested individually for association with phenotype. Instead, they need to be aggregated with similar rare variants to be tested collectively for association with phenotype (Figure 2B).

This requirement immediately brings forward several questions. Do we only aggregate variants within a single gene, or should it be extended to collections of genes in a pathway? Do we only group variants of a certain annotation—nonsynonymous, nonsense, synonymous, splice site, etc.? Should we impose a frequency threshold to define "rare" (i.e., only aggregate variants that are less than 1% frequency), and if so, what should this frequency threshold be? The answers to many of these questions are in development. We refer the reader to several recent publications that address these issues (Li and Leal, 2008; Madsen and Browning, 2009; Neale et al., 2011; Price et al., 2010).

Another key question is statistical power. What sample size will be required to identify a new disease gene based on a burden of rare coding mutations? An early computer simulation study conducted by Kryukov, Sunyaev, and colleagues suggested that 10,000 individuals (5,000 drawn from the lowest 5th percentile and 5,000 above the 95th percentile) will be needed to discover another gene with a mutation burden similar to *ABCA1* (Kryukov et al., 2009). The principal reason for this large sample size is the stringent statistical threshold that needs to be
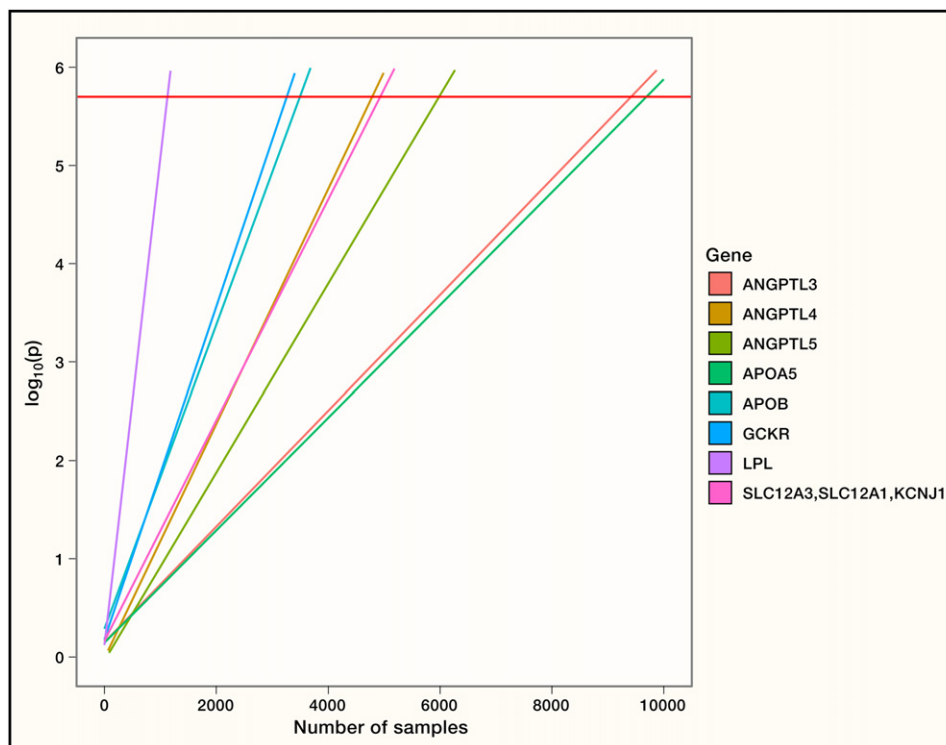
Figure 3. Estimate of Sample Sizes Required for Gene Discovery in Exome Sequencing Studies of Complex Traits
Extrapolations were performed using gene resequencing counts from published candidate gene studies. Fisher's exact test was performed for gene burden testing on mutation carrier status within each study. The red horizontal line indicates the genome-wide p value threshold (p = 2.5 × 10^{-6}) for gene burden of rare variants with minor allele frequencies < 1%. The extrapolations show that thousands of samples are needed to exceed statistical significance at a genome-wide level. Rare variant counts for *ANGPTL3*, *ANGPTL4*, and *ANGPTL5* for plasma triglycerides were obtained from Romeo and colleagues (Romeo et al., 2009); *APOA5*, *APOB*, *GCKR*, and *LPL* for hypertriglyceridemia from Johansen and colleagues (Johansen et al., 2010); *SLC12A3/SLC12A1/KCNJ1* from Ji and colleagues (Ji et al., 2008).

imposed when testing 20,000 genes. A first approximation of the appropriate statistical threshold would be Bonferroni correction for all of the genes sequenced (p = 2.5 × 10^{-6}). Our extrapolation of effect sizes and frequencies from published studies shows (Figure 3) that thousands of individuals are required to reach acceptable statistical power.

Statistical power would improve dramatically if we were able to distinguish rare mutations that affected protein function from those that did not. The bulk of discovered rare mutations are missense. Of all missense, it is estimated that ~1/5 are as deleterious as nonsense mutations, ~1/2 are mildly deleterious, and ~1/4 do not affect protein function (Kryukov et al., 2007). If one restricted mutation counts to only functional mutations, noise would be removed (MacArthur et al., 2012). At present, functional consequence of variants can be predicted based on comparative sequence analysis and protein structure analysis. Several software tools are available, and the accuracy of these methods remains a subject of debate (Sunyaev et al., 2001).

An alternative method is experimental evaluation of whether a given mutation is functional. Such experimental evaluation is possible for genes with known function and where that function can be readily evaluated with moderate- or high-throughput approaches. However, such evaluation of mutations assumes that the parameter evaluated is an appropriate surrogate for the mechanism by which the gene impacts the human phenotype. For complex phenotypes such as MI that arise after decades of pathology, it is often difficult to know whether this is the case. Nevertheless, distinguishing functional versus neutral missense mutations will be a key to improving statistical power.

Modeling Human Genetic Disease in Reprogrammed Cells
As an increasing number of genetic variants become associated with human disease, it will be essential to develop effective and accurate model systems to understand the mechanism of disease. This step is a prerequisite to translate genetic findings into new targets for therapy. Animal models have traditionally been used to mimic human mutations through gene knockin approaches or study of null mutations (Bruneau et al., 2001; Lindsay et al., 2001). Though this approach has been of value and has led to many discoveries, it is more common that mouse models fail to effectively recapitulate the human phenotype. This situation is even true for Mendelian traits, wherein heterozygous deletion of human disease genes is often well tolerated in mice, whereas homozygous deletions can have catastrophic consequences, failing to model the human condition. For many human disease genes, this has been a bottleneck, limiting the value of gene-hunting approaches described above.
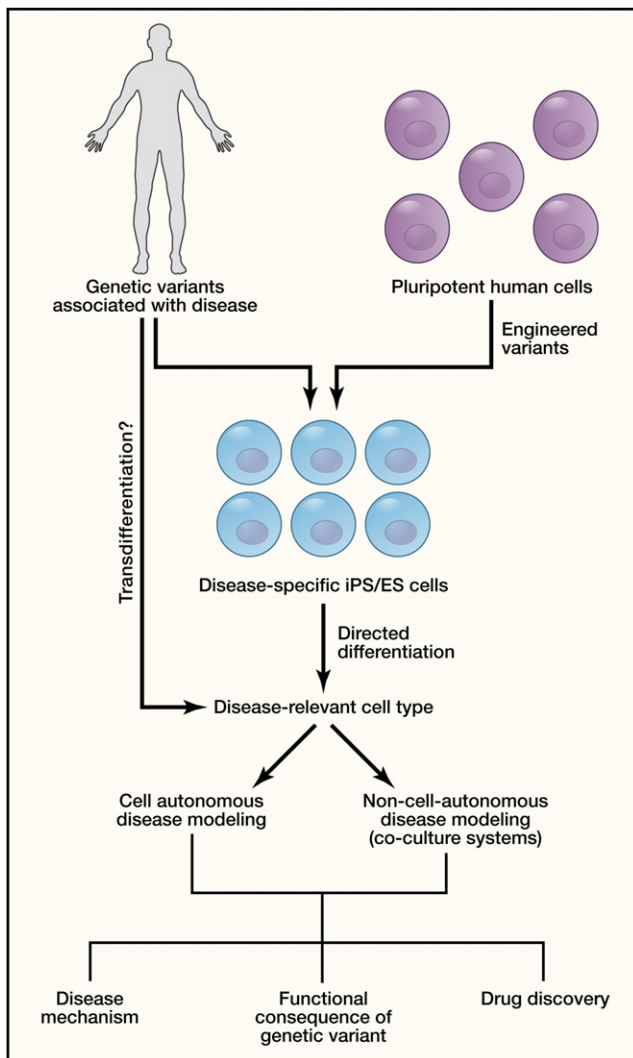
Figure 4. Modeling Human Genetic Disease in Reprogrammed Cells

Reprogrammed cells derived from skin fibroblasts or blood cells of individuals with genetic variants associated with disease can be used for disease modeling and drug discovery and to determine the impact of sequence variants on cellular biology. Reprogramming to induced pluripotent stem (iPS) cells and subsequent directed differentiation to relevant cell types can yield large numbers of human disease-related cell types for investigation. Specific genetic variants can also be engineered into human pluripotent stem cells using zinc-finger nucleases or TALENs to generate iPS or embryonic stem (ES) cells harboring disease-associated variants. Future approaches involving transdifferentiation may allow for a more direct approach to generate relevant cell types for study of CVD. In some cases, diseases will be cell autonomous, but other cases may require coculture of two or more cell types to recapitulate the disease process.

The recent revolution in cellular reprogramming technology provides a potential solution to the need for disease modeling (Figure 4). The striking observation by Yamanaka and colleagues that human adult somatic cells could be reprogrammed into a pluripotent state by expressing four transcription factors not only revolutionized the field of stem cell biology, but also has major consequences for the study of human genetics (Takahashi et al., 2007). Investigators can now reprogram skin or blood cells

from patients with defined genetic variants/mutations into induced pluripotent stem (iPS) cells and can then differentiate the iPS cells into the specific cell type affected by disease. This approach provides an unprecedented opportunity to investigate the consequences of human genetic variation on cellular phenotypes that may contribute to disease. There have been several reports of partial modeling of CVD using human iPS cells in the last 2 years, although success to date has been limited to diseases with discrete cell-autonomous cellular defects. We will consider a few examples below and highlight the advantages and limitations of this approach.

Autosomal-dominant long QT syndrome (LQTS) that predisposes to cardiac arrhythmias and sudden death involves single-gene mutations of sodium or potassium channels that provide a discrete and quantitative phenotype in iPS-derived cardiomyocytes. In LQTS, repolarization of cardiomyocytes is delayed, resulting in a prolonged action potential duration. Several groups have reported successful generation of iPS-derived cardiomyocytes from patients with LQTS, and an elongation of the action potential duration has been reported for a subset of genetically induced LQTS (Itzhaki et al., 2011; Moretti et al., 2010; Yazawa et al., 2011). LQTS can also be drug induced in selected individuals, and pluripotent stem cell-derived cardiomyocytes appear to be an effective model system to detect prolongation of action potentials due to cardiotoxic drugs (Braam et al., 2010). It is exciting to consider that identification of common genetic variants that predispose to drug-induced LQTS might be modeled in iPS-derived cardiomyocytes, thereby allowing a system in which to screen for this toxic side effect prior to in vivo use of new drugs.

Despite these successes, certain types of LQTS have been more difficult to model due to the incomplete maturation of iPS-derived cardiomyocytes and appropriate expression of ion channels in a consistent manner. It is possible that recent direct reprogramming strategies involving transdifferentiation, which do not involve a progenitor intermediate, will allow modeling of diseases in which a mature cell type is required (Ieda et al., 2010). The inability to expand mature, transdifferentiated cells is currently a rate-limiting factor for this approach, but future improvements in efficiency of reprogramming may overcome this hurdle.

Another example of iPS-based modeling recently came from the study of Noonan and LEOPARD syndromes, autosomal-dominant conditions that involve pulmonary valve stenosis and hypertrophic cardiomyopathy in a subset of patients (Gelb and Tartaglia, 2011). The genetic cause of Noonan and LEOPARD syndromes involves activating mutations in members of the Ras pathway. Correspondingly, iPS-derived cardiomyocytes from patients with cardiomyopathy associated with LEOPARD syndrome exhibit higher levels of Ras signaling that appear to contribute to excessive growth of cardiac cells (Carvajal-Vergara et al., 2010). Though such findings successfully model the pathway known to be affected, the ultimate goal will be to identify new therapeutics that can disrupt abnormal pathways and restore normal physiology. This has remained elusive thus far, in part because most iPS-based disease models lead to broad variability in phenotype, whereas high-throughput screening assays typically require a consistent phenotype with

a narrow range of variability. As protocols for generating iPS cells improve and directed differentiation methods become more consistent, ''disease in a dish'' models promise to alter the landscape of drug discovery (Figure 4).

Besides modeling the Mendelian syndromes described above, another potential use of iPS technology will be to annotate the function of SNPs identified by GWAS or rare mutations discovered by sequencing. iPS cells are currently being generated from large populations with defined genetic variants and CVD in the hope that certain aspects of the phenotype could be modeled in iPS-derived cells. It may prove difficult to identify consistent phenotypic differences associated with genetic variants of small effect; however, it is possible that the genetic background present in affected individuals will be sufficient to reveal a cellular defect.

The contribution of individual sequence variants in the context of complex genetic background issues can be interrogated through the use of recently developed DNA correction approaches in human pluripotent stem cells. The use of zinc-finger nucleases and transcription activator-like effector nuclease (TALEN) technology to specifically introduce or remove a sequence variant in stem cells will reveal the contribution of individual variants to cellular phenotypes (Yusa et al., 2011). Addition of environmental stresses to the cellular system may be necessary to provoke phenotypes and will be an important tool in disease modeling in iPS cells. Such studies are currently underway, and we look forward to the challenges and opportunities offered by this new approach to modeling human genetic variants.

Conclusions

The tools to decode human genetic variants on a large population scale have finally arrived and should begin to reveal the genetic contribution to disease at an exponential pace. Rigorous analysis of the functional consequences of sequence variants associated with disease will be critical, and new approaches leveraging stem cell technologies may facilitate study of human variants in relevant human cell types. Ultimately, the convergence of human genetics with functional biology will reveal new therapeutic targets for cardiovascular disease and will stimulate a new generation of drug discovery efforts.

REFERENCES

Abifadel, M., Varret, M., Rabès, J.P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., et al. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat. Genet. 34, 154–156.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science 322, 881–888.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. 12, 745–755.

Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T.A., Leblanc-Straceski, J., et al. (1997). Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. Nat. Genet. 15, 30–35.

Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J., Kwiterovich, P., Shan, B., Barnes, R., and Hobbs, H.H. (2000). Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science 290, 1771–1775.

Bonne, G., Carrier, L., Bercovici, J., Cruaud, C., Richard, P., Hainque, B., Gautel, M., Labeit, S., James, M., Beckmann, J., et al. (1995). Cardiac myosin binding protein-C gene splice acceptor site mutation is associated with familial hypertrophic cardiomyopathy. Nat. Genet. 11, 438–440.

Boyden, L.M., Choi, M., Choate, K.A., Nelson-Williams, C.J., Farhi, A., Toka, H.R., Tikhonova, I.R., Bjornson, R., Mane, S.M., Colussi, G., et al. (2012). Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. Nature 482, 98–102.

Braam, S.R., Tertoolen, L., van de Stolpe, A., Meyer, T., Passier, R., and Mummery, C.L. (2010). Prediction of drug-induced cardiotoxicity using human embryonic stem cell-derived cardiomyocytes. Stem Cell Res. (Amst.) 4, 107–116.

Brown, M.S., and Goldstein, J.L. (1986). A receptor-mediated pathway for cholesterol homeostasis. Science 232, 34–47.

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E., and Seidman, J.G. (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell 106, 709–721.

Burkhardt, R., Toh, S.A., Lagor, W.R., Birkeland, A., Levin, M., Li, X., Robblee, M., Fedorov, V.D., Yamamoto, M., Satoh, T., et al. (2010). Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. J. Clin. Invest. 120, 4410–4414.

C4D (Coronary Artery Disease Genetic Consortium). (2011). A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat. Genet. 43, 339–344.

Cañadas, V., Vilacosta, I., Bruna, I., and Fuster, V. (2010). Marfan syndrome. Part 1: pathophysiology and diagnosis. Nat. Rev. Cardiol. 7, 256–265.

Carrier, L., Hengstenberg, C., Beckmann, J.S., Guicheney, P., Dufour, C., Bercovici, J., Dausse, E., Berebbi-Bertrand, I., Wisnewsky, C., Pulvenis, D., et al. (1993). Mapping of a novel gene for familial hypertrophic cardiomyopathy to chromosome 11. Nat. Genet. 4, 311–313.

Carvajal-Vergara, X., Sevilla, A., D'Souza, S.L., Ang, Y.S., Schaniel, C., Lee, D.F., Yang, L., Kaplan, A.D., Adler, E.D., Rozov, R., et al. (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. Nature 465, 808–812.

Chang, S.S., Grunder, S., Hanukoglu, A., Rösler, A., Mathew, P.M., Hanukoglu, I., Schild, L., Lu, Y., Shimkets, R.A., Nelson-Williams, C., et al. (1996). Mutations in subunits of the epithelial sodium channel cause salt wasting with hyperkalaemic acidosis, pseudohypoaldosteronism type 1. Nat. Genet. 12, 248–253.

Chen, Q., Kirsch, G.E., Zhang, D., Brugada, R., Brugada, J., Brugada, P., Potenza, D., Moya, A., Borggrefe, M., Breithardt, G., et al. (1998). Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. Nature 392, 293–296.

Choi, M., Scholl, U.I., Yue, P., Björklund, P., Zhao, B., Nelson-Williams, C., Ji, W., Cho, Y., Patel, A., Men, C.J., et al. (2011). K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. Science 331, 768–772.

Clarke, R., Peden, J.F., Hopewell, J.C., Kyriakou, T., Goel, A., Heath, S.C., Parish, S., Barlera, S., Franzosi, M.G., Rust, S., et al; PROCARDIS Consortium.

(2009). Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N. Engl. J. Med. *361*, 2518–2528.

Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science *305*, 869–872.

Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. *354*, 1264–1272.

Curran, M.E., Splawski, I., Timothy, K.W., Vincent, G.M., Green, E.D., and Keating, M.T. (1995). A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. Cell *80*, 795–803.

Davey Smith, G., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. *32*, 1–22.

Dietz, H.C. (2010). New therapeutic approaches to mendelian disorders. N. Engl. J. Med. *363*, 852–863.

Dietz, H.C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M., et al. (1991). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature *352*, 337–339.

Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., Bennett, D., et al. (2009). Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. JAMA *302*, 37–48.

Erdmann, J., Grosshennig, A., Braund, P.S., König, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Trégouët, D.A., et al; Italian Atherosclerosis, Thrombosis, and Vascular Biology Working Group; Myocardial Infarction Genetics Consortium; Wellcome Trust Case Control Consortium; Cardiogenics Consortium. (2009). New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat. Genet. *41*, 280–282.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Faivre, L., Collod-Beroud, G., Loeys, B.L., Child, A., Binquet, C., Gautier, E., Callewaert, B., Arbustini, E., Mayer, K., Arslan-Kirchner, M., et al. (2007). Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. Am. J. Hum. Genet. *81*, 454–466.

Garcia, C.K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., Calandra, S., Bertolini, S., Cossu, F., Grishin, N., et al. (2001). Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. Science *292*, 1394–1398.

Garg, V., Kathiriya, I.S., Barnes, R., Schluterman, M.K., King, I.N., Butler, C.A., Rothrock, C.R., Eapen, R.S., Hirayama-Yamada, K., Joo, K., et al. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature *424*, 443–447.

Garg, V., Muth, A.N., Ransom, J.F., Schluterman, M.K., Barnes, R., King, I.N., Grossfeld, P.D., and Srivastava, D. (2005). Mutations in NOTCH1 cause aortic valve disease. Nature *437*, 270–274.

Geisterfer-Lowrance, A.A., Kass, S., Tanigawa, G., Vosberg, H.P., McKenna, W., Seidman, C.E., and Seidman, J.G. (1990). A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. Cell *62*, 999–1006.

Gelb, B.D., and Tartaglia, M. (2011). RAS signaling pathway mutations and hypertrophic cardiomyopathy: getting into and out of the thick of it. J. Clin. Invest. *121*, 844–847.

Geller, D.S., Rodriguez-Soriano, J., Vallo Boado, A., Schifter, S., Bayer, M., Chang, S.S., and Lifton, R.P. (1998). Mutations in the mineralocorticoid receptor gene cause autosomal dominant pseudohypoaldosteronism type I. Nat. Genet. *19*, 279–281.

Geller, D.S., Farhi, A., Pinkerton, N., Fradley, M., Moritz, M., Spitzer, A., Meinke, G., Tsai, F.T., Sigler, P.B., and Lifton, R.P. (2000). Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. Science *289*, 119–123.

Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y., et al. (2011). New gene functions in megakaryopoiesis and platelet formation. Nature *480*, 201–208.

Gray, R., and Wheatley, K. (1991). How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplant. *7*, 9–12.

Gudbjartsson, D.F., Bjornsdottir, U.S., Halapi, E., Helgadottir, A., Sulem, P., Jonsdottir, G.M., Thorleifsson, G., Helgadottir, H., Steinthorsdottir, V., Stefansson, H., et al. (2009). Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat. Genet. *41*, 342–347.

Hansson, J.H., Nelson-Williams, C., Suzuki, H., Schild, L., Shimkets, R., Lu, Y., Canessa, C., Iwasaki, T., Rossier, B., and Lifton, R.P. (1995). Hypertension caused by a truncated epithelial sodium channel gamma subunit: genetic heterogeneity of Liddle syndrome. Nat. Genet. *11*, 76–82.

Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.D., Topol, E.J., Rosenfeld, M.G., and Frazer, K.A. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature *470*, 264–268.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A., et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science *316*, 1491–1493.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA *106*, 9362–9367.

Hobbs, H.H., Leitersdorf, E., Leffert, C.C., Cryer, D.R., Brown, M.S., and Goldstein, J.L. (1989). Evidence for a dominant gene that suppresses hypercholesterolemia in a family with defective low density lipoprotein receptors. J. Clin. Invest. *84*, 656–664.

Holdt, L.M., and Teupser, D. (2012). Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. Arterioscler. Thromb. Vasc. Biol. *32*, 196–206.

IBC 50K CAD Consortium. (2011). Large-scale gene-centric analysis identifies novel variants for coronary artery disease. PLoS Genet. *7*, e1002260.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375–386.

Itzhaki, I., Maizels, L., Huber, I., Zwi-Dantsis, L., Caspi, O., Winterstern, A., Feldman, O., Gepstein, A., Arbel, G., Hammerman, H., et al. (2011). Modelling the long QT syndrome with induced pluripotent stem cells. Nature *471*, 225–229.

Jarcho, J.A., McKenna, W., Pare, J.A., Solomon, S.D., Holcombe, R.F., Dickie, S., Levi, T., Donis-Keller, H., Seidman, J.G., and Seidman, C.E. (1989). Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1. N. Engl. J. Med. *321*, 1372–1378.

Ji, W., Foo, J.N., O'Roak, B.J., Zhao, H., Larson, M.G., Simon, D.B., Newton-Cheh, C., State, M.W., Levy, D., and Lifton, R.P. (2008). Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat. Genet. *40*, 592–599.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., et al. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat. Genet. *42*, 684–687.

Kamstrup, P.R., Tybjaerg-Hansen, A., Steffensen, R., and Nordestgaard, B.G. (2009). Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. JAMA *301*, 2331–2339.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes, J., III. (1961). Factors of risk in the development of coronary heart disease—six

year follow-up experience. The Framingham Study. Ann. Intern. Med. *55*, 33–50.

Katan, M.B. (1986). Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet *1*, 507–508.

Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N.P., Rieder, M.J., Cooper, G.M., Roos, C., Voight, B.F., Havulinna, A.S., et al. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat. Genet. *40*, 189–197.

Kathiresan, S., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H., et al; Myocardial Infarction Genetics Consortium; Wellcome Trust Case Control Consortium. (2009a). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat. Genet. *41*, 334–341.

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. (2009b). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat. Genet. *41*, 56–65.

Kimura, A., Harada, H., Park, J.E., Nishi, H., Satoh, M., Takahashi, M., Hiroi, S., Sasaoka, T., Ohbuchi, N., Nakamura, T., et al. (1997). Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. Nat. Genet. *16*, 379–382.

Kryukov, G.V., Pennacchio, L.A., and Sunyaev, S.R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am. J. Hum. Genet. *80*, 727–739.

Kryukov, G.V., Shpunt, A., Stamatoyannopoulos, J.A., and Sunyaev, S.R. (2009). Power of deep, all-exon resequencing for discovery of human trait genes. Proc. Natl. Acad. Sci. USA *106*, 3871–3876.

Lehrman, M.A., Schneider, W.J., Südhof, T.C., Brown, M.S., Goldstein, J.L., and Russell, D.W. (1985). Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science *227*, 140–146.

Lemaire, S.A., McDonald, M.L., Guo, D.C., Russell, L., Miller, C.C., III, Johnson, R.J., Bekheirnia, M.R., Franco, L.M., Nguyen, M., Pyeritz, R.E., et al. (2011). Genome-wide association study identifies a susceptibility locus for thoracic aortic aneurysms and aortic dissections spanning FBN1 at 15q21.1. Nat. Genet. *43*, 996–1000.

Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am. J. Hum. Genet. *83*, 311–321.

Lifton, R.P., Dluhy, R.G., Powers, M., Rich, G.M., Cook, S., Ulick, S., and Lalouel, J.M. (1992a). A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. Nature *355*, 262–265.

Lifton, R.P., Dluhy, R.G., Powers, M., Rich, G.M., Gutkin, M., Fallo, F., Gill, J.R., Jr., Feld, L., Ganguly, A., Laidlaw, J.C., et al. (1992b). Hereditary hypertension caused by chimaeric gene duplications and ectopic expression of aldosterone synthase. Nat. Genet. *2*, 66–74.

Lifton, R.P., Gharavi, A.G., and Geller, D.S. (2001). Molecular mechanisms of human hypertension. Cell *104*, 545–556.

Lindsay, E.A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H.F., Scambler, P.J., et al. (2001). Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. Nature *410*, 97–101.

Lindsay, M.E., and Dietz, H.C. (2011). Lessons on the pathogenesis of aneurysm from heritable conditions. Nature *473*, 308–316.

Lloyd-Jones, D.M., Nam, B.H., D'Agostino, R.B., Sr., Levy, D., Murabito, J.M., Wang, T.J., Wilson, P.W., and O'Donnell, C.J. (2004). Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA *291*, 2204–2211.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.

Madsen, B.E., and Browning, S.R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. *5*, e1000384.

Mani, A., Radhakrishnan, J., Wang, H., Mani, M.A., Nelson-Williams, C., Carew, K.S., Mane, S., Najmabadi, H., Wu, D., and Lifton, R.P. (2007). LRP6 mutation in a family with early coronary disease and metabolic risk factors. Science *315*, 1278–1282.

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.

McClellan, J., and King, M.C. (2010). Genetic heterogeneity in human disease. Cell *141*, 210–217.

McNair, W.P., Ku, L., Taylor, M.R., Fain, P.R., Dao, D., Wolfel, E., and Mestroni, L.; Familial Cardiomyopathy Registry Research Group. (2004). SCN5A mutation associated with dilated cardiomyopathy, conduction disorder, and arrhythmia. Circulation *110*, 2163–2167.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R., et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. Science *316*, 1488–1491.

Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., et al. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat. Genet. *39*, 1197–1199.

Moretti, A., Bellin, M., Welling, A., Jung, C.B., Lam, J.T., Bott-Flügel, L., Dorn, T., Goedel, A., Höhnke, C., Hofmann, F., et al. (2010). Patient-specific induced pluripotent stem-cell models for long-QT syndrome. N. Engl. J. Med. *363*, 1397–1409.

Mune, T., Rogerson, F.M., Nikkilä, H., Agarwal, A.K., and White, P.C. (1995). Human hypertension caused by mutations in the kidney isozyme of 11 beta-hydroxysteroid dehydrogenase. Nat. Genet. *10*, 394–399.

Musunuru, K., and Kathiresan, S. (2010). Genetics of coronary artery disease. Annu. Rev. Genomics Hum. Genet. *11*, 91–108.

Musunuru, K., Pirruccello, J.P., Do, R., Peloso, G.M., Guiducci, C., Sougnez, C., Garimella, K.V., Fisher, S., Abreu, J., Barry, A.J., et al. (2010a). Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N. Engl. J. Med. *363*, 2220–2227.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., et al. (2010b). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466*, 714–719.

Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S.M., Roeder, K., and Daly, M.J. (2011). Testing for an unusual distribution of rare variants. PLoS Genet. *7*, e1001322.

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.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. *42*, 30–35.

Norton, N., Li, D., Rieder, M.J., Siegfried, J.D., Rampersaud, E., Züchner, S., Mangos, S., Gonzalez-Quintana, J., Wang, L., McGee, S., et al. (2011). Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. Am. J. Hum. Genet. *88*, 273–282.

O'Donnell, C.J., and Nabel, E.G. (2011). Genomics of cardiovascular disease. N. Engl. J. Med. *365*, 2098–2109.

Olson, T.M., Doan, T.P., Kishimoto, N.Y., Whitby, F.G., Ackerman, M.J., and Fananapazir, L. (2000). Inherited and de novo mutations in the cardiac actin gene cause hypertrophic cardiomyopathy. J. Mol. Cell. Cardiol. *32*, 1687–1694.

Poetter, K., Jiang, H., Hassanzadeh, S., Master, S.R., Chang, A., Dalakas, M.C., Rayment, I., Sellers, J.R., Fananapazir, L., and Epstein, N.D. (1996). Mutations in either the essential or regulatory light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. Nat. Genet. *13*, 63–69.

Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A., et al. (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. Science *322*, 1702–1705.

Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., and Sunyaev, S.R. (2010). Pooled association tests for rare variants in exon-resequencing studies. Am. J. Hum. Genet. *86*, 832–838.

Raychaudhuri, S. (2011). Mapping rare and common causal alleles for complex human diseases. Cell *147*, 57–69.

Reilly, M.P., Li, M., He, J., Ferguson, J.F., Stylianou, I.M., Mehta, N.N., Burnett, M.S., Devaney, J.M., Knouff, C.W., Thompson, J.R., et al; Myocardial Infarction Genetics Consortium; Wellcome Trust Case Control Consortium. (2011). Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet *377*, 383–392.

Roger, V.L., Go, A.S., Lloyd-Jones, D.M., Benjamin, E.J., Berry, J.D., Borden, W.B., Bravata, D.M., Dai, S., Ford, E.S., Fox, C.S., et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. (2012). Executive summary: heart disease and stroke statistics—2012 update: a report from the American Heart Association. Circulation *125*, 188–197.

Romeo, S., Yin, W., Kozlitina, J., Pennacchio, L.A., Boerwinkle, E., Hobbs, H.H., and Cohen, J.C. (2009). Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. J. Clin. Invest. *119*, 70–79.

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E., et al; WTCCC and the Cardiogenics Consortium. (2007). Genomewide association analysis of coronary artery disease. N. Engl. J. Med. *357*, 443–453.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839–1842.

Sankaran, V.G., Xu, J., Ragoczy, T., Ippolito, G.C., Walkley, C.R., Maika, S.D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M.A., et al. (2009). Developmental and species-divergent globin switching are driven by BCL11A. Nature *460*, 1093–1097.

Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the genetic architecture of gene expression in human liver. PLoS Biol. *6*, e107.

Schott, J.J., Benson, D.W., Basson, C.T., Pease, W., Silberbach, G.M., Moak, J.P., Maron, B.J., Seidman, C.E., and Seidman, J.G. (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science *281*, 108–111.

Schott, J.J., Alshinawi, C., Kyndt, F., Probst, V., Hoorntje, T.M., Hulsbeek, M., Wilde, A.A., Escande, D., Mannens, M.M., and Le Marec, H. (1999). Cardiac conduction defects associate with mutations in SCN5A. Nat. Genet. *23*, 20–21.

Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., et al; Cardiogenics; CARDIoGRAM Consortium. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat. Genet. *43*, 333–338.

Seidman, J.G., and Seidman, C. (2001). The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. Cell *104*, 557–567.

Shea, J., Agarwala, V., Philippakis, A.A., Maguire, J., Banks, E., Depristo, M., Thomson, B., Guiducci, C., Onofrio, R.C., Kathiresan, S., et al; Myocardial Infarction Genetics Consortium. (2011). Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. Nat. Genet. *43*, 801–805.

Shimkets, R.A., Warnock, D.G., Bositis, C.M., Nelson-Williams, C., Hansson, J.H., Schambelan, M., Gill, J.R., Jr., Ulick, S., Milora, R.V., Findling, J.W., et al. (1994). Liddle's syndrome: heritable human hypertension caused by mutations in the beta subunit of the epithelial sodium channel. Cell *79*, 407–414.

Simon, D.B., Karet, F.E., Hamdan, J.M., DiPietro, A., Sanjad, S.A., and Lifton, R.P. (1996a). Bartter's syndrome, hypokalaemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2. Nat. Genet. *13*, 183–188.

Simon, D.B., Karet, F.E., Rodriguez-Soriano, J., Hamdan, J.H., DiPietro, A., Trachtman, H., Sanjad, S.A., and Lifton, R.P. (1996b). Genetic heterogeneity of Bartter's syndrome revealed by mutations in the K+ channel, ROMK. Nat. Genet. *14*, 152–156.

Simon, D.B., Nelson-Williams, C., Bia, M.J., Ellison, D., Karet, F.E., Molina, A.M., Vaara, I., Iwata, F., Cushner, H.M., Koolen, M., et al. (1996c). Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter. Nat. Genet. *12*, 24–30.

Simon, D.B., Bindra, R.S., Mansfield, T.A., Nelson-Williams, C., Mendonca, E., Stone, R., Schurman, S., Nayir, A., Alpay, H., Bakkaloglu, A., et al. (1997). Mutations in the chloride channel gene, CLCNKB, cause Bartter's syndrome type III. Nat. Genet. *17*, 171–178.

Soria, L.F., Ludwig, E.H., Clarke, H.R., Vega, G.L., Grundy, S.M., and McCarthy, B.J. (1989). Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. Proc. Natl. Acad. Sci. USA *86*, 587–591.

St Hilaire, C., Ziegler, S.G., Markello, T.C., Brusco, A., Groden, C., Gill, F., Carlson-Donohoe, H., Lederman, R.J., Chen, M.Y., Yang, D., et al. (2011). NT5E mutations and arterial calcifications. N. Engl. J. Med. *364*, 432–442.

Stenson, P.D., Mort, M., Ball, E.V., Howells, K., Phillips, A.D., Thomas, N.S., and Cooper, D.N. (2009). The Human Gene Mutation Database: 2008 update. Genome. Med. *1*, 13.

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., III, Kondrashov, A.S., and Bork, P. (2001). Prediction of deleterious human alleles. Hum. Mol. Genet. *10*, 591–597.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Tartaglia, M., Mehler, E.L., Goldberg, R., Zampino, G., Brunner, H.G., Kremer, H., van der Burgt, I., Crosby, A.H., Ion, A., Jeffery, S., et al. (2001). Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat. Genet. *29*, 465–468.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

Teupser, D., Baber, R., Ceglarek, U., Scholz, M., Illig, T., Gieger, C., Holdt, L.M., Leichtle, A., Greiser, K.H., Huster, D., et al. (2010). Genetic regulation of serum phytosterol levels and risk of coronary artery disease. Circ. Cardiovasc. Genet. *3*, 331–339.

Thierfelder, L., Watkins, H., MacRae, C., Lamas, R., McKenna, W., Vosberg, H.P., Seidman, J.G., and Seidman, C.E. (1994). Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. Cell *77*, 701–712.

Varbo, A., Benn, M., Tybjærg-Hansen, A., Grande, P., and Nordestgaard, B.G. (2011). TRIB1 and GCKR polymorphisms, lipid levels, and risk of ischemic heart disease in the general population. Arterioscler. Thromb. Vasc. Biol. *31*, 451–457.

Visel, A., Zhu, Y., May, D., Afzal, V., Gong, E., Attanasio, C., Blow, M.J., Cohen, J.C., Rubin, E.M., and Pennacchio, L.A. (2010). Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature *464*, 409–412.

Wang, Q., Shen, J., Splawski, I., Atkinson, D., Li, Z., Robinson, J.L., Moss, A.J., Towbin, J.A., and Keating, M.T. (1995). SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. Cell *80*, 805–811.

Wang, F., Xu, C.Q., He, Q., Cai, J.P., Li, X.C., Wang, D., Xiong, X., Liao, Y.H., Zeng, Q.T., Yang, Y.Z., et al. (2011). Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. Nat. Genet. *43*, 345–349.

Watkins, H., Conner, D., Thierfelder, L., Jarcho, J.A., MacRae, C., McKenna, W.J., Maron, B.J., Seidman, J.G., and Seidman, C.E. (1995). Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. Nat. Genet. *11*, 434–437.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

Wensley, F., Gao, P., Burgess, S., Kaptoge, S., Di Angelantonio, E., Shah, T., Engert, J.C., Clarke, R., Davey-Smith, G., Nordestgaard, B.G., et al; C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC). (2011). Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. BMJ *342*, d548.

Wild, P.S., Zeller, T., Schillert, A., Szymczak, S., Sinning, C.R., Deiseroth, A., Schnabel, R.B., Lubos, E., Keller, T., Eleftheriadis, M.S., et al. (2011). A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. Circ. Cardiovasc. Genet. *4*, 403–412.

Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M., et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat. Genet. *40*, 161–169.

Wilson, F.H., Disse-Nicodème, S., Choate, K.A., Ishikawa, K., Nelson-Williams, C., Desitter, I., Gunel, M., Milford, D.V., Lipkin, G.W., Achard, J.M., et al. (2001). Human hypertension caused by mutations in WNK kinases. Science *293*, 1107–1112.

Xu, J., Sankaran, V.G., Ni, M., Menne, T.F., Puram, R.V., Kim, W., and Orkin, S.H. (2010). Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. *24*, 783–798.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W., and Orkin, S.H. (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993–996.

Yasuno, K., Bilguvar, K., Bijlenga, P., Low, S.K., Krischek, B., Auburger, G., Simon, M., Krex, D., Arlier, Z., Nayak, N., et al. (2010). Genome-wide association study of intracranial aneurysm identifies three new risk loci. Nat. Genet. *42*, 420–425.

Yazawa, M., Hsueh, B., Jia, X., Pasca, A.M., Bernstein, J.A., Hallmayer, J., and Dolmetsch, R.E. (2011). Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. Nature *471*, 230–234.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordóñez, A., Hannan, N.R., Rouhani, F.J., et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391–394.

Zacho, J., Tybjaerg-Hansen, A., Jensen, J.S., Grande, P., Sillesen, H., and Nordestgaard, B.G. (2008). Genetically elevated C-reactive protein and ischemic vascular disease. N. Engl. J. Med. *359*, 1897–1908.