Genetics: Implications for Prevention and Management of Coronary Artery Disease

**Brief Title:** Genetics and the Prevention and Management of CAD

Themistocles L. Assimes, MD, Ph.D, Robert Roberts, MD

From the a Department of Medicine, Stanford University School of Medicine, Stanford, California; International Society for Cardiovascular Translational Research, University of Arizona College of Medicine, Phoenix, Arizona.

**Disclosures:** Dr. Assimes has reported that he has no relationships relevant to the contents of this paper to disclose. Dr. Roberts is a Scientific Advisor to Cumberland Pharmaceuticals.

**Reprint requests and correspondence:**
Dr. Themistocles L. Assimes, MD
Assistant Professor of Medicine
Stanford University School of Medicine
Stanford, California 94305
E-mail: tassimes@stanford.edu
Abstract
An exciting new era has dawned for the prevention and management of CAD utilizing genetic risk variants. The recent identification of over 60 susceptibility loci for coronary artery disease (CAD) confirm not only the importance of established risk factors, but also the existence of many novel causal pathways that are expected to improve our understanding of the genetic basis of CAD and facilitate the development of new therapeutic agents over time. Concurrently, Mendelian randomization studies have provided intriguing insights on the causal relationship between CAD-related traits, and highlight the potential benefits of long-term modifications of risk factors. Lastly, genetic risk scores of CAD may serve not only as prognostic, but also as predictive markers and carry the potential to considerably improve the delivery of established prevention strategies. This review will summarize the evolution and discovery of genetic risk variants for CAD and their current and future clinical applications.

Key Words: genetic variation, Mendelian randomization, risk scores, sequencing

Abbreviations and Acronyms
CAD = coronary artery disease
GRS = genetic risk score
GWAS = genome-wide association study
HDL = high-density lipoprotein
LDL = low-density lipoprotein
MI = myocardial infarction
MR = Mendelian randomization
SNP = single-nucleotide polymorphism
WES = whole-exome sequencing
WGS = whole-genome sequencing
Introduction

Coronary artery disease (CAD), with its clinical sequelae of angina, myocardial infarction (MI), heart failure, and sudden death, has long been the number 1 cause of death in the developed world (1). Recently, it became the number 1 killer in the developing world, and so is now a global, sustained pandemic (1). This increased incidence is due to several factors, including increasing lifespans, changes in lifestyle, adoption of Western diets, and many others (1). Despite the global increased prevalence of CAD, its incidence in the developed world decreased substantially after the implementation of effective primary and secondary prevention strategies in the latter part of the 20th century (2,3). These strategies include the promotion of a healthy lifestyle anchored on optimal nutrition, smoking cessation, and increased physical activity, as well as the development of effective pharmacological therapies, such as platelet and 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) reductase inhibitors. Collectively, these strategies are able to reduce the incidence of CAD by almost 50% in high-risk populations (2,3).

Improved application of established primary prevention strategies undoubtedly has the potential to further reduce the incidence of CAD (4,5). However, many individuals have substantial residual risk, remain reluctant to take medications on the basis of assessments of risk with current risk calculators, and/or are unable to institute and maintain changes in lifestyle necessary to substantially reduce risk (6-9). In this context, recent impressive progress in genomic medicine, guided by astonishing breakthroughs in laboratory technology and computing power, provides us with a golden opportunity to dissect the genetic basis of CAD. This knowledge, in turn, is being leveraged to improve our ability to identify subjects at high risk and to develop novel therapies. The Central Illustration summarizes progress to date, as well as the topics covered in this review. The pharmacogenetics of drugs used to treat or prevent CAD,
including oral antithrombotic agents and lipid-lowering drugs is of substantial interest, but is reviewed in detail elsewhere (10,11).

**Origin of Human DNA Variation Responsible for Predisposition to Disease**

The human genome possesses 3.2 billion base pairs, of which approximately 99.5% are identical between any 2 individuals (12,13). The DNA sequence defining the features unique to each human is contained in approximately 0.5% of the sequence (~15 million bases). Most of these 15 million base differences consist of large segments of DNA that are rearranged, or repetitive sequences that vary in the number of copies per genome (copy number variation [CNV]) (13). However, the most common type of mutation contributing to the genome’s uniqueness is a substitution of one base (either adenine, guanine, thymine, and cytosine) for another, commonly referred to as a single-nucleotide polymorphism (SNP) (12,13).

SNPs are due to DNA copying errors during replication of the 2 strands of DNA. DNA turns over every few days, and the base pairs are replaced appropriately and accurately, with only 1 error per billion base pairs created (mutation rate of $1.4 \times 10^{-8}$) (14,15). This error rate is enough to induce a significant number of mutations, which may be transmitted to the next generation if they occur in the germ line (egg or sperm) (16). SNPs are distributed throughout the genome, and their effect on the phenotype may be neutral, benign, or markedly influential. The number of SNPs is fairly constant, at about 3 million per human genome, and these variations are responsible for a sizable fraction of the variation in physiological traits, including predisposition to disease. However, SNPs are constantly evolving with each new generation of humans. Given a mutation rate of $1.0 \times 10^{-8}$, it is estimated that 40 to 60 new nonembryonic lethal mutations are introduced per generation. With the world’s population of 7 billion, over 300 billion new mutations exist in the current generation alone (17).
New mutations that are randomly assigned to each generation are initially rare, and their frequency in subsequent genomes will evolve according to Darwinian principles. SNPs that provide an advantage to or have no effect on survival and fitness may increase in frequency, whereas those that are detrimental will remain rare or disappear (18). In this context, genetic determinants of most common chronic diseases, including CAD, have not been under substantial selective pressure, given that these conditions have not been primary causes of morbidity and mortality until only recently (18). Even as CAD emerged as a leading cause of death in the last century, it generally still does not cause death before the end of the reproductive phase of life in humans.

Family Studies Suggest That Common Presentations of CAD Are Moderately Heritable

Several epidemiology studies involving unrelated individuals provided the first clues that non-Mendelian common presentations of CAD in middle to late adulthood were heritable (19). These familial aggregation studies estimated the contribution of genetic factors by comparing rates of CAD among subjects with a positive family history of CAD to rates among subjects with no such history. Overall, these studies have documented ~2.5- to 4-fold higher rates of CAD among individuals with a family history compared with individuals with no family history when adjusting only for age and sex (19-21). When further controlling for traditional risk factors, the excess risk of CAD is reduced to about 1.5- to 2.5-fold. This is consistent with the notion that at least part of this excess risk is a consequence of the familial aggregation of traditional risk factors (19-21). Familial aggregation studies also provide compelling evidence that a stronger family history or an earlier age of onset of disease in a family member further increases the risk for close relatives (19-21).
A weakness of familial aggregation studies is rooted in the possibility that a positive family history may serve as a proxy of exposure to common familial environmental risk factors (e.g., pollution and nongenetic influences, such as smoking, diet, and physical activity) in addition to genetic factors (22). Such confounding may result in an overestimate of the contribution of genetic factors. To address this concern, several twin studies have been conducted, which offer the advantage of nullifying the effect of a common familial environment by comparing concordance rates of disease among monozygotic twins to those in dizygotic twins. Difference in concordance rates are then transformed to unbiased estimates of heritability, assuming the extent to which common familial environmental risk factors are shared between either set of twins is equal. Arguably, the most informative twin studies conducted to date for CAD have involved registries in Denmark and Sweden, which have estimated the heritability of CAD to be in the range of 40% to 60% (23-25). Consistent with the familial aggregation studies, the twin studies have also observed higher heritability when disease occurs at a younger age, and a partial attenuation of estimates when they are adjusted for traditional risk factors (23-25).

Linkage Studies Fail to Identify Genetic Determinants of Common Presentations of CAD

In the 1980s, the identification and mapping of several hundred highly polymorphic DNA markers that are equally spaced throughout the genome (referred to as microsatellites) enabled scientists to identify a region of the genome harboring highly penetrant mutations responsible for rare monogenic disorders through linkage analysis. Such work was possible even before the sequencing of the entire human genome was complete. In such studies, microsatellites are typed in families with at least 1 affected member, and analyses are then performed to identify which markers within the family members segregate best with disease status. Subsequent DNA
sequencing around the region of linkage is then used to refine the signal and to ultimately identify the causal variant.

Linkage analysis has proven to be a very effective means of identifying highly penetrant variants responsible for disease in large multigenerational families that include many affected individuals with a Mendelian or Mendelian-like mode of inheritance. For example, this approach was used to identify mutations in the beta-myosin heavy chain at the 14q1 chromosomal locus as a cause of familial hypertrophic cardiomyopathy (26-28) in the early 1990s. However, such families are not common and/or not easy to find for CAD. For many of these families identified to date, the culprit mutation has proved to be a highly damaging mutation within a gene that strongly influences circulating low-density lipoprotein (LDL) cholesterol levels (e.g., familial hypercholesterolemia (FH) genes, including \textit{LDLR}, \textit{APOB}, \textit{PCSK9}, and, less commonly, \textit{LDLRAP1}, \textit{ABCG5}, and \textit{ABCG8})(29). In more recent studies, the culprit mutation has been found in genes predisposing to severe forms of the metabolic syndrome (e.g., \textit{LRP6}, \textit{DYRK1B})(30,31). The initial observation that supported the highly multigenic nature of CAD for a majority of subjects presenting with CAD was the failure of linkage studies performed on families with a more typical low penetrance of disease to detect any susceptibility loci (29,32).

The limited power of linkage studies to detect variants with low to modest effects for multigene disorders became glaringly clear by the mid 1990s (32,33). Researchers also recognized that association studies focused on directly comparing the frequency of variants among affected and unaffected subjects were much better powered than linkage studies to identify polygenic determinants of common disorders, such as CAD (32). All that was needed to institute such an approach was a complete catalog of human genetic variants.

\textbf{The 21st Century Ushers in Genetics of Multigene Disorders}
Completion of The Human Genome Project and Failure of Early Candidate Gene Association Studies

The initial sequencing of the human genome in the late 1990s and early 2000s resulted in the discovery of approximately 1.5 million SNPs (34-36). Scientists jumped on this resource to perform small-scale association studies limited to variants within or near genes previously suspected of being involved in the pathogenesis of disease (37). Unfortunately, this biased candidate gene approach was unsuccessful, as many positive initial reports of association could not be replicated (37). The reason for the failure of candidate genes studies are many, but are rooted in the biased and premature approach of examining only a subset of SNPs within a subset of genes in search of what ultimately proved to be very modest effects, using relatively small sample sizes (37). As a result, many published studies were woefully underpowered and carried a high rate of false positive findings (38,39).

Genome-Wide Association Studies

Whole-genome sequencing of more humans by various groups led to the cataloging of several additional millions of SNPs in the first half of the 2000s. Between 2002 and 2007, the International HapMap Consortium subsequently carefully documented the allele frequencies of approximately 3.1 million of these SNPs in 270 subjects, representing 4 races/ethnic groups (40,41). This rich resource provided the DNA markers necessary to span the entire genome, and set the stage for initial genome-wide association studies (GWAS) using newly developed, cost-efficient array technology, which allowed for the simultaneous accurate genotyping of up to 1 million SNPs (42,43).

A GWAS represents an unbiased approach to detect genetic variants associated with clinical outcomes (43,44). In essence, it extends the candidate gene association study approach
not only to variants within all genes, but also to variants between genes that may be regulating the expression of genes nearby (Figure 1). Similar to candidate gene association studies, GWAS compares the frequency of each genetic variant among subjects with a disease of interest to the frequency among controls. Variants with frequencies that statistically differ between the 2 groups are assumed to either be pathogenic or to be highly correlated with the pathogenic variant(s) in the region (43,44).

The overwhelming failure of candidate gene association studies motivated the scientific community of genetic epidemiologists, statistical geneticists, and journal editors to adopt strict criteria for establishing novel findings from GWAS (43,45). Testing 1 million SNPs simultaneously can be problematic with respect to traditional thresholds of significance, as about 50,000 of the tests would be expected to have a p < 0.05 by chance. Thus, to account for the multitude of tests, a Bonferroni correction to the p value threshold of significance was embraced at $5 \times 10^{-8}$, a threshold commonly referred to as the *genome-wide significance threshold* (33).

Ideally, for a given SNP, both a discovery and a replication study would demonstrate similar magnitude of effect in the same direction, with a joint or combined p value leading to a smaller p value than that seen in the initial report (45). Additional criteria suggested for establishing a positive replication in GWAS include examining the same or a very similar population and phenotype in the replication study, using the same genetic model in both the discovery and the replication datasets, and providing a strong rationale for the SNPs selected for replication from the initial study (45).

**Discovery of the First Locus Predisposing to Common Presentations of CAD**

In 2007, three groups independently reported the first region of the genome to be linked to CAD utilizing the GWAS approach (46-48). All 3 groups made the discovery in European
populations (Table 1). The susceptibility locus is located on the short arm (p) of chromosome 9 at band 2.1 and thus is commonly referred to as 9p21. The locus includes approximately 60 strongly correlated SNPs over ~53,000 base pairs, and is approximately 100,000 to 150,000 base pairs upstream of the genes encoding 2 cyclin-dependent kinase (CDK) inhibitors and known tumor suppressors, CDKN2B and CDKN2A (46-48). The locus also overlaps the last section of a long noncoding RNA, which is transcribed antisense to CDKN2B (CDKN2BAS or ANRIL) (46-48) (Figure 2). The lead variants at 9p21 are common, with a minor allele frequency of about 50% in Europeans, resulting in about 75% of individuals of European ancestry carrying at least 1 allele that raises risk. The initial increase in risk for CAD observed was about 25% for 1 copy and about 50% for 2 copies, with somewhat higher risks per allele observed among those with early-onset CAD (46,47). The risk mediated by 9p21 appeared from the outset to be completely independent of all known risk factors, implying that other unknown factors are contributing substantially to the pathogenesis of CAD (46,47). The findings at 9p21 were quickly replicated in many other cohorts of Europeans and East Asians (49,50). Somewhat surprisingly, robust replication in African Americans remains elusive to this day, although some signals in the region have been detected (51-53). Replication studies have also demonstrated that the true excess risk per risk allele is likely closer to 20% in Europeans (49,54).

The exact mechanism of action linking genetic variation at 9p21 to the risk of CAD remains unclear to this day, although several sets of observations at the population level and in the laboratory have helped to narrow down the possibilities (55). First, the exact same variants have been linked to complications of atherosclerosis outside of the heart, including carotid plaque, ischemic stroke, and peripheral arterial disease, suggesting that the locus predisposes to atherosclerosis in all vascular beds (55). Secondly, the exact same variants have been linked to
both abdominal aortic and intracranial arterial aneurysms, suggesting that the cells that are affected by these variants reside in the vessel wall of the artery and play some role in maintaining its integrity (55). Thirdly, a growing body of evidence suggests that the SNPs in the high-risk region disrupt or create transcription factor binding sites that alter the expression levels, or the relative abundance of different transcripts of the noncoding RNA, ANRIL, which in turn affects the expression levels of CDKN2B and/or 2A. The protein products of these 2 genes, p15\textsuperscript{INK4a} and p16\textsuperscript{INK4a}, then alter the function of resident macrophages and/or vascular smooth muscle cells, facilitating the formation of atherosclerotic plaque (55). Lastly, animal model studies to date suggest that these effects could involve increased proliferation and reduced apoptosis of resident macrophages and/or vascular smooth muscle cells (55).

**Further Discovery of Susceptibility Loci for CAD through the Formation of Multiple International Consortia and Meta-Analysis of GWAS**

Eleven other loci for CAD were identified within a very short interval of the discovery of 9p21 by several groups, including the Welcome Trust Case-Control Consortium (WTCCC), the Cardiogenics Consortium, and the Myocardial Infarction Genetics Consortium (MIGEN) (56-59). The result of these initial GWAS studies confirmed that common susceptibility variants for CAD carried minimal incremental risk, and would require very large sample sizes, in the tens of thousands to the hundreds of thousands, to be uncovered. Larger national and international consortia were formed to tackle this challenge, including the Coronary Artery Disease Genome Wide Replication and Meta-analysis (CARDIoGRAM) and the Coronary Artery Disease (C4D) Genetics consortia, which in 2011 confirmed almost all of the previously reported loci and uncovered 17 new loci through meta-analysis (60,61). Through further collaboration between these 2 consortia, and the testing of a larger fraction of SNPs with a minor allele frequency >1%
using advanced imputation algorithms, CARDIoGRAMplusC4D reported an additional 15 novel loci in 2013 and 8 loci in 2015, after the examination of >60,000 cases and >120,000 controls (54,62). These studies brought the total loci having reached genome-wide significance in largely European and, to a lesser extent, South Asian populations to 58 (54,62).

East Asians represent the next most studied race/ethnic group. A total of 12 loci have convincingly reached genome-wide significance for CAD in studies involving either Han Chinese, Korean, or Japanese subjects, with at least 6 of these loci overlapping with loci uncovered in Europeans (63-67). Thus, both ethnic-specific as well as cross-ethnic loci have been uncovered to date. The overlap in susceptibility loci across race/ethnic groups is expected to increase as larger studies are conducted in non-European populations, given that the genetic basis of most common diseases is believed to be identical across race/ethnic groups (68-70).

The decreasing cost of genotyping over time also allowed the IBC 50k CAD Consortium to revisit the candidate gene approach using a customized chip that comprehensively surveyed common genetic variation at ~2,100 candidate genes, representing a diverse spectrum of cardiovascular diseases (71). Well-powered studies using this array firmly established associations between CAD and genetic variants in \textit{LPA}, the gene encoding lipoprotein(a), and confirmed several early GWAS discoveries for CAD (9p21, \textit{COL4A1/COL4A2}, \textit{ZC3HC1}, and \textit{CYP17A1}) (72,73). However, these studies also confirmed a lack of association for a vast majority of variants within candidate genes related to CAD (72,73).

\textbf{Many Susceptibility Loci for Conventional Risk Factors for CAD Have Also Been Identified through GWAS}

The genetics of several known risk factors for CAD including body mass index, blood lipid levels, blood pressure (BP), diabetes, insulin resistance, and smoking are also being
explored in parallel to the CAD outcome. In total, several hundred susceptibility loci have been identified for these traits with many correlating with more than 1 trait (74-78). The genetic risk variants involved with these conventional risk factors may have indirect or direct effects on the risk of CAD. Indeed, as highlighted above for 9p21 locus, the pattern of association between a susceptibility SNPs and multiple traits may provide important clues on the mechanism of association (79).

**Features of Genetic Risk Variants for CAD**

*Table 1* lists regions of the genome where at least 1 genetic risk variant has reached genome-wide significance for association with CAD to date since 2007. To minimize inflation of effects from winner’s curse, the table is restricted to the results of the most significant SNP in the region, as determined by the latest GWAS meta-analysis conducted by CARDIoGRAMplusC4D (54).

*Table 1* highlights several features of these variants. First, the variants are common, have small effects, and are generally not located within protein-coding regions of the genome. These findings support the common variant-common disease hypothesis put forth many years ago by leading geneticists, which predicted that common disease-causing polymorphisms found in all major populations of the world are predominantly responsible for complex polygenic diseases through subtle effects on regulatory regions of the genome (80). The findings also highlight the lack of power and the accompanying high risk of false positive associations inherent in the candidate gene studies that preceded GWAS, which typically involved no more than a few hundred cases (37-39). Secondly, only about one-third of the variants appear to be influencing the risk of CAD through effects on traditional risk factors, namely, lipids and BP (62). This estimate was derived from cross-referencing the GWAS signals for CAD with those observed for
risk factors (62). Thus, the mechanisms whereby most of these risk variants mediate risk remain unknown.

The small effect sizes observed with common variants should not be interpreted as pointing to genes or pathways with little clinical relevance. Although such effect sizes make risk prediction more challenging, they may still involve genes or pathways that have the potential to dramatically affect risk of disease through more extreme modulation of gene function or expression. For example, the initial GWAS signals for CAD include common variants with modest effects in or near all 3 of the FH genes (LDLR, APOB, and PCSK9). Collectively, these loci point to the cholesterol synthesis pathway within the hepatocyte, a pathway that has already been very effectively targeted by statins and proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors (81,82). Thus, a better understanding of the causal genes and pathways responsible for GWAS findings to date for CAD should not only provide insight into the pathogenesis of coronary atherosclerosis, but also facilitate the development of effective novel therapeutic agents.

**Genetic Risk Variants for CAD Appear to Primarily Promote The Formation of Coronary Atherosclerosis**

In the pursuit of genetic risk for coronary atherosclerosis, CAD and MI have been treated as the same phenotype. This approach is justified on the basis that MI, with rare exceptions, occurs secondary to significant coronary atherosclerosis (83,84). The pathophysiology of coronary atherosclerosis involves many factors, including endothelial dysfunction, monocyte adhesion, cholesterol deposition, and an inflammatory response with release of cytokines (83,84). The genetic predisposition leading to a MI would be through different targets and
mechanisms than that of atherosclerosis, including mechanisms associated with plaque rupture and thrombosis (83,84).

A major limitation in determining whether a susceptibility locus for CAD predisposes specifically to coronary atherosclerosis versus MI is our inability to accurately quantify the burden of coronary atherosclerosis, as well as the number of prior symptomatic and silent plaque ruptures among individuals with coronary atherosclerosis (85-87). One way to partially circumvent this difficulty is to compare subjects with prior MIs to subjects with no prior MIs among subjects with documented clinically significant CAD. Such a design was leveraged to identify the ABO locus as an MI-predisposing locus, with the A and B risk variants shown to increase the risk of MI in the setting of CAD by ~20% (88). The protein coded by the ABO locus, α-1-3-N-acetylgalactoseaminytransferase, is known to transfer a carbohydrate onto von Willebrand Factor (vWF), prolonging its half-life and leading to levels in the plasma that are 25% higher in individuals with A, B, or AB blood groups compared with those with blood group O (89). The blood group O gene also codes for the same transferase protein, but was mutated a long time ago, resulting in loss of any biochemical activity. Individuals with blood group O do not have increased plasma levels of vWF and have a lower risk for MI. Thus, the biological plausibility of this association is reasonable, given that plasma vWF levels have been repeatedly shown to predict the occurrence of adverse cardiac events, including death, among patients with preexisting vascular disease (90). However, the ABO locus has also been found to associate with LDL levels, as well as with CAD outcomes in conventional case-control studies, making it difficult to conclude with certainty whether the locus predisposes to MI independent of its ability to promote atherosclerosis (61,88).
Examination of correlation patterns between established variants for CAD and the burden of atherosclerosis on coronary angiography may provide some additional insight on mechanisms of association. Multiple studies have shown that genetic variants with CAD susceptibility loci correlate with burden of disease on angiography (91-93). However, such correlations could still be driven by a genetic predisposition to subclinical extra- or intra-plaque ruptures and/or erosions of plaques, which have been previously shown to be responsible for the nonuniform growth of many moderate and high-grade lesions used to quantify an individual’s burden of disease on angiography (87,94).

The study of the earliest lesions of coronary atherosclerosis not yet prone to plaque rupture or erosion represents another means to help distinguish between the 2 potential mechanisms of association. Salfati et al. recently took advantage of the PDAY (Pathobiological Determinants of Atherosclerosis in Youth) resource to document associations between a genetic risk score (GRS) of high-risk alleles for clinical CAD and the degree of subclinical coronary atherosclerotic lesions documented through autopsy in young adults dying between 15 and 35 years of age from noncardiovascular causes (95). They found that the magnitude of association between the GRS and early lesions documented in PDAY was comparable to the association observed with coronary artery calcification in older subjects, and that these associations persisted, even when risk factor SNPs were removed from the GRS (95). Overall, these findings support the hypothesis that a majority of newly identified GWAS loci for CAD are atherosclerosis predisposing loci, rather than loci that predispose to plaque rupture or thrombosis (95). However, the PDAY resource is too small to allow for definitive conclusions regarding associations on a SNP by SNP basis. Thus, more studies using this approach are necessary before definitive conclusions can be made for each CAD susceptibility locus.
Determining Causality of Risk Factors for CAD through Mendelian Randomization

A pervasive challenge in observational epidemiology is determining whether a link between a risk factor and an outcome is causal (96). Although a noncausal risk factor may still serve as a useful predictor, an intervention developed specifically to modify the degree of exposure to such a risk factor is not likely to succeed in improving outcomes (96).

The most reliable way to assess a causal association between a risk factor and an outcome has been through randomized controlled trials (RCTs) of interventions that modify risk factor levels (96). In some instances, such interventions have demonstrated the limitations of statistical adjustment of measured confounders in observational studies. Notable examples include a plethora of studies on vitamin levels and cardiovascular health, whose benefits in the observational setting could never be replicated through corresponding RCTs (97).

The advent of large-scale genotyping of common variants over the last 15 years has facilitated the development of Mendelian randomization (MR) studies (Figure 3) as an alternate method of assessing causality between a risk factor and an outcome in an observational setting (96). MR studies leverage the well-established method of instrumental variable (IV) analysis, used successfully in econometrics for many years to deal with issues of confounding, reverse causality, and regression dilution (96). In an analogous fashion, MR studies use 1 or more genetic variants associated with the risk factor of interest as the instrument to achieve the same goal in medicine (96) (Figure 3). A key prerequisite for successful MR studies is the identification of a subset of genetic variants that have been randomly allocated at conception, found to robustly influence the exposure levels of a risk factor, and are not linked to the outcome in any way other than through effects on the risk factor (96) (Figure 3).
MR studies have provided some intriguing insights on the causality of a variety of cardiometabolic risk factors and, in some cases, have shattered previously widely held beliefs (98). For example, MR studies to date suggest that homocysteine, fibrinogen, secretory phospholipase IIA (sPLA2-IIA), lipoprotein-associated PLA2 (Lp-PLA2), high-sensitivity C-reactive protein, and high-density lipoprotein (HDL) are not causally associated with CAD (99). These observations are supported by the failure of RCTs of folic acid supplementation, sPLA2-IIA and Lp-PLA2 inhibitors, and, most recently, cholesterol ester transfer protein (CETP) inhibitors (100-104). The failure of the CETP inhibitors has been particularly enigmatic, given the profound elevations of HDL achieved with these drugs.

Other MR studies have supported a causal role for elevated body mass index, elevated BP, diabetes, LDL, lipoprotein(a), and triglycerides (99,105). These findings are supported by prior RCTs of weight loss and of a variety of antihypertensive drugs, statins, and ezetimide (99,105). However, not all interventions influencing risk factors deemed causal through IV analysis have proven successful, indicating that MR studies cannot be trusted entirely to predict the success of all interventions influencing one’s degree of exposure to causal risk factors. Notable failures in this respect include recent trials of niacin performed on a background of full-dose statin therapy, which did not result in a reduction of CAD events, despite marked lowering of triglyceride levels (100). Similarly, fibrates and omega-3 fatty acid supplements have not been convincingly shown to reduce the risk of CAD, despite impressive reductions in triglyceride levels (100,106). Lastly, evacetrapib, a CETP inhibitor was recently reported to have no effect on outcomes, despite lowering LDL by ~37%, in addition to raising HDL (107). These failures are likely a consequence of an incomplete understanding of the various physiological pathways that control biomarker levels, and highlight a need to identify and target the subset of pathways that
also influence the risk of CAD. Discrepancies between MR studies and results of drug trials could be resolved in the future, as we gain a better understanding of the mechanisms behind each genetic association related to risk factors of CAD. For example, MR studies restricted to the subset of variants associated with triglyceride levels, as well as effects on the level of expression of genes influenced by niacin or fibrates, would be expected to be negative.

Another insight derived through MR studies on risk factors for CAD has been a better appreciation of the potential benefit of long-term pharmacological therapy. Causal estimates derived from MR studies for quantitative traits, such as lipids, reflect the lifetime effects of a risk factor, as compared to the typical 3- to 5-year intervention achieved through a clinical trial (108). For example, MR studies suggest that the beneficial effects of LDL and BP lowering over a lifetime are substantially greater than has been observed in short term observational studies, and RCTs of statin and antihypertensive drug therapies (98,109-111). In a comprehensive MR study involving 9 LDL-related SNPs in 6 genomic regions, as well as a meta-analysis of multiple studies involving >300,000 individuals and thousands of events, Ference et al. documented that individuals inheriting LDL lowering experience a marked reduction in cardiac morbidity and mortality over their lifetime (111). They estimated that a decrease in 1 unit of LDL (38.7 mg/dl) since birth is associated with 54.5% reduction in cardiac events, which is 3-fold greater than that observed in clinical trials involving statin therapies given over shorter periods of time later in life (111). In a more recent study, individuals with LDL cholesterol >190 mg/dl on a single measure, who were carrying damaging mutations within FH genes (LDLR, APOB, and PCSK9), were found to be at 2 to 3 times higher risk of clinical CAD than individuals who were not carrying such mutations within every stratum of LDL (112). Furthermore, FH mutation carriers were found to have a higher cumulative exposure to LDL cholesterol than noncarriers in an analysis of
participants with repeated measures over many years (112). These results reinforce the potential of early primary prevention of CAD, especially among those at very high lifetime risk at a young age, such as individuals suffering from FH or other polygenic causes of very elevated LDL levels.

An increasing appreciation of the strengths and weaknesses of MR studies has evolved over the last few years. First, the consequences of an underpowered MR study, either because of the use of a weak instrument or due to an inadequate number of events, is now very well understood (96). Over the next few years, mega-cohort studies involving biobanks and genetic data, combined with stronger instruments as more genetic discoveries are made, are expected to optimize the power of IV analyses and minimize the probability of either missing a causal association or falsely declaring its presence. Secondly, important new methods have been recently developed to identify potentially problematic instruments, which may result in bias and false conclusions about the causality of a risk factor (113-115). Thus, MR studies are expected to continue to serve as an important means of assessing the causal nature of a risk factor and estimating the probability of success of a novel therapeutic agent in improving outcomes of CAD.

**CAD Genetic Risk Scores May Improve Our Ability to Identify Individuals at High Risk Who Are Most Likely to Respond to Therapy**

A major anticipated clinical application for genetic risk variants predisposing to CAD is their utilization to improve our ability to risk-stratify individuals (116,117). Although current clinical risk prediction scores for CAD perform relatively well compared with scores for other chronic diseases, substantial room for improvement exists, as a large pool of individuals with
incident events carry either only 1 modifiable risk factor or only borderline risk factors (118-120).

The dilemma in selecting individuals for primary prevention can be illustrated in the case of 47-year-old white premenopausal woman with total plasma cholesterol of 320 mg/dl, an HDL of 50 mg/dl, and an LDL of 175 mg/dl, untreated BP of 120/80 mm Hg, and no other risk factors. According to the American College of Cardiology (ACC)/American Heart Association (AHA) pooled cohort risk calculator, her 10-year risk is only 2%, and treatment with statins to lower her LDL is not indicated unless she develops other risk factors or has a cardiac event (121). This individual is likely to have appreciable subclinical CAD, which may not be identified through a coronary artery calcification scan. Decreasing her plasma LDL would be expected to delay or prevent development of clinical CAD. Improving our ability to better predict which subset of women with similar risk factor profiles will develop an event through the addition of a novel biomarker, such as one’s genetic susceptibility to CAD, would be expected to improve outcomes through more efficient application of established primary prevention therapies.

The most practical way to currently integrate genetics into risk prediction models, such as the Framingham Risk Score or the ACC/AHA pooled cohorts calculator, is through the calculation of a genetic risk score (GRS) for individuals (122). A GRS is a single variable that summarizes one’s exposure to variants that increase risk for CAD (123). A GRS is typically calculated by summing the product of the number of high-risk variants inherited by each individual for each susceptibility variant and the log of the odds ratio previously determined in a GWAS for the same variant (122). Figure 4 provides an example of how a GRS can be integrated into the Framingham Risk Score to update an individual’s 10-year risk.
The use of GRS in clinical practice has been slow to materialize for several reasons, including the high cost of genotyping, the more modest effects of genetic variants on the risk of CAD than originally anticipated, and the challenge of improving a clinical risk score, such as the Framingham or ACC/AHA risk score, that already performs quite well (117,122,124). Consequently, it has been difficult to demonstrate substantial improvements to standard model performance metrics, such as the C-statistic, with the addition of a GRS, even though ample evidence now exists that a GRS of known CAD loci predicts incident CAD events independent of all other traditional risk factors (122,125-130). The challenge of demonstrating the clinical utility of a novel biomarker through observational studies alone has been further amplified by concerns of the power and reliability of some of the most commonly used model performance metrics for novel biomarkers, including the C-statistic, as well as the net reclassification index (131-133).

Family history would intuitively serve as a substitute for genetic risk. However, individual variants, as well as GRS of CAD, have been shown to predict clinical complications of CAD independent of family history (129). Furthermore, the magnitude of the association between family history and the risk of incident CAD is not obviously reduced when a GRS is introduced into a multivariate prediction model that includes family history (129). These observations support the notion that family history reflects exposure to common familial environmental risk factors of CAD to a larger degree than previously suspected, and the proportion of the heritability of CAD reflected by current GRS is not high enough to noticeably erode the predictive power of family history.

A recent report highlights the great potential for a GRS of CAD to improve outcomes, not only through primary prevention, but also possibly secondary prevention (126). In this report, a
GRS involving 27 variants previously proven to be associated with risk for CAD was constructed after genotyping DNA biobanked at baseline from participants in 1 community-based cohort study (the Malmo Diet and Cancer Study), 2 primary prevention trials of statins (JUPITER and ASCOT), and 2 secondary prevention trials assessing the efficacy of statin therapy (CARE and PROVE IT-TIMI 22) (126). Among the 48,421 individuals and 3,477 events included in this study, investigators showed that the GRS not only predicted incident CAD events, but also recurrent CAD events, independent of all traditional risk factors, as well as family history (126). Furthermore, a significant gradient of increasing relative risk reductions (p = 0.0277), summarized in Figure 5, as well as absolute risk reductions (p = 0.0101), was documented across the low, intermediate, and high genetic risk categories among the 4 randomized trials (126). The absolute risk reductions estimated a roughly 3-fold decrease in the number needed to treat to prevent 1 CAD event in the primary prevention trials (126). Specifically, in the primary prevention trials, the number needed to treat to prevent 1 such event in 10 years was 66 in people at low genetic risk, 42 in those at intermediate genetic risk, and 25 in those at high genetic risk in JUPITER, and 57, 47, and 20, respectively, in ASCOT. Thus, a GRS has the potential to serve not only as a prognostic marker, but also as a marker that can predict response to the single most important primary and secondary therapy already available (134). Such a combination of prognostic and predictive power for a single biomarker is uncommon, and has generally been observed only for certain markers of cancer (135). This combination offers the potential to substantially improve the efficiency of both primary and secondary established primary prevention strategies.

These findings, if successfully replicated, have important implications for the prevention and clinical management of CAD. Utilizing GRS, one could target large populations at high risk,
such as premenopausal women with high genetic risk and optimal response to statin therapy, for primary prevention. Furthermore, the incremental predictive ability of the GRS is expected to improve over time, although the pace and magnitude of improvement remains unclear (117). However, a substantial improvement in the predictive ability of a GRS may eventually justify the early application of primary prevention strategies, such as statins, to asymptomatic individuals who are at very high genetic risk, despite not having any other apparent risk factor. Such an approach is already indirectly supported by recent Mendelian randomization studies for LDL and CAD risk (111). Of course, such approaches would have to be supported by carefully conducted randomized clinical trials demonstrating improvement in outcomes or acceptable intermediate traits with the use of a GRS. Thus, the testing of GRS in prospective randomized clinical trials holds great promise for reducing the incidence of CAD.

**Whole-Exome and Whole-Genome Sequencing Studies of CAD**

GWAS have been exceptionally successful in identifying susceptibility loci harboring common variants with modest effects on the risk of CAD. These common variants are likely responsible for the majority of the heritability of CAD (136-140). However, less common variants with larger effect sizes exist both within loci already uncovered through GWAS, as well as in loci not yet linked to CAD. As many of these rare causal variants have not been previously observed, the only way to reliably identify them is through the direct sequencing of the DNA of subjects with disease.

Next-generation sequencing platforms have been developed and continuously enhanced over the last decade (141). They are now mature enough to allow for large-scale genome sequencing projects. Two types of genome sequencing are possible: genome sequencing limited to the exons of all protein coding genes, referred to as whole-exome sequencing (WES), and
genome sequencing of the entire genome, referred to as whole-genome sequencing (WGS). WES became economically feasible to conduct in large numbers approximately 6 years ago (142). In 2010, the NHLBI funded several groups to conduct WES association studies through the Exome Sequencing Project (ESP), a collaborative endeavor that included several sample sets with individuals suffering from clinical CAD (143). A subset of the rare variants uncovered through the NHLBI’s WES project and other similar projects across the world were then incorporated into custom exome genotyping chips to facilitate the large-scale and cost-efficient replication of the findings in the discovery WES studies (144). This subset was restricted to variants predicted to most likely have severe consequences on protein function, such as damaging missense variants and variants affecting intron-exon splice sites. Overall, the discoveries to date using this strategy have been limited, primarily because effect sizes of less common and rare variants have been observed to be not much higher than the effect sizes observed for more common variants (145). Consequently, the study of very large sample sizes involving tens of thousands of individuals has been necessary to reliably identify novel associations involving rare variants uncovered through WES (144). Furthermore, it has been extremely challenging to examine rare variants individually (144). Instead, inference in WES studies have generally relied on comparing the aggregate number of rare variants within a gene observed among cases to that observed among controls (145).

Discoveries to date using WES and exome chip genotyping related to CAD have almost exclusively pointed to previously established genes involved in lipoprotein metabolism. These genes include genes influencing plasma LDL levels (PSCK9, LDLR, and NPC1L1) as well as genes influencing plasma triglyceride levels (APOA5, APOC3, LPL, and ANGPTL4)(146-149). A low-frequency variant in SVEP1 was also found to be associated with CAD whose mechanism
of action most likely involves effects on BP (149). Despite the limited novelty of these findings with respect to mechanisms of disease, they include PCSK9 and NPC1L1, 2 genes with FDA approved drugs on the market that inhibit the activity of the gene’s protein and lead to both a reduction of plasma LDL levels as well as the risk of CAD (105,150,151). These findings combined with the results of MR studies of triglyceride levels and CAD have reinvigorated the search for, and the development of, novel therapeutic agents that reduce the risk of CAD by interfering with the metabolism of triglyceride rich lipoproteins (152-155). In other smaller studies, WES has recently been used to suggest that rare nonsynonymous variants in SPTBN5, NID2, and ADAMTSL4 may protect against CAD and to implicate a 2 base insertion in RECQL5 as a cause of early onset MI in a Chinese Han family(156,157).

Discoveries from WES are expected to expand as larger sample sizes become available. However, WES will undoubtedly be supplanted by WGS as the differential in price between the 2 approaches dwindles over time. WGS offers the additional benefit of documenting the presence of less common and rare variants in noncoding regulatory regions of an individual’s genome whose degree of existence until recently was substantially underappreciated (158). Empirical evidence already exists that the effects sizes for such variants are not likely to be substantially larger than the effect sizes observed for causal variants within exons or common noncoding causal variants discovered through GWAS (139) (Figure 6). Thus, the ability to link such variants will also be heavily dependent on the profiling of very large sample sizes. The discovery of loci not previously implicated through GWAS or WES may be particularly challenging (Figure 6).

The first large-scale NIH funded projects to explore the WGS approach in the context of CAD and its risk factors have only recently been initiated, and include the NHLBI’s WGS
project within the Trans-Omics for Precision Medicine (TOPMed) (159) program, as well as the NHGRI’s new Genome Sequencing Program (GSP), which includes 4 Centers for Common Disease Genomics (CCDG) (160). TOPMed is currently completing WGS in >60,000 samples from 28 studies, of which at least 14 relate to coronary atherosclerosis or its risk factors. The initial disease focus of the CCDG is early onset myocardial infarction, hemorrhagic stroke, and autism.

**Future Directions**

We have undoubtedly entered a golden era in the genetics of CAD early in the 21st century, highlighted by a rapid pace of discovery that has been facilitated by technological advances, which have not only made it feasible to perform high-throughput genetic profiling in a cost-efficient manner, but also allowed research groups around the world to share data and collaborate. These collaborations have resulted in the identification of over 60 susceptibility loci, which not only highlight the importance of well-established mechanisms of disease, such as cholesterol metabolism, but also reveal the presence of many novel pathways related to coronary atherosclerosis. The same large-scale population genetic studies are also providing us with opportunities to better understand the relationship between established or emerging risk factors and CAD through Mendelian randomization and to improve our ability to identify individuals at high risk of CAD through the incorporation of genetic risk into clinical risk scores.

Many challenging tasks remain to build on this remarkable progress and maximize the clinical utility of the knowledge gained. These tasks include the identification of the remaining susceptibility loci for CAD, proving the clinical utility of genetic data in the prevention of CAD, and acquiring a solid appreciation of the cellular and/or extracellular mechanisms responsible for genetic associations observed at the population level. Such knowledge will serve as the
foundation for the development of novel therapeutic agents. Extremely large sample sizes are needed for additional discoveries, given the distribution of effect sizes observed to date for both common and rare variants, as well as the estimated proportion of the heritability of CAD explained by these variants to date. In the coming years, this need should be fulfilled by megabiobanks involving at least a half-million participants, including, but not limited to, the UK Biobank, the China Kadoorie Biobank, the Million Veteran Program, and the soon-to-be-established NIH Precision Medicine Initiative cohort (161-164). Such biobanks will likely also be leveraged to gain a better understanding of the clinical utility of genetic risk scores, and to conduct additional, well-powered MR studies to complement studies published to date. Perhaps the most challenging of these tasks will be to understand the cellular mechanisms behind these associations. This task requires the careful interpretation of genetic association signals in combination with ongoing initiatives, such as the ENCODE (ENCyclopedia Of DNA Elements), GTEx (Genotype-Tissue Expression), and Roadmap Epigenomics projects, which are carefully documenting the complex and abundant regulatory regions of the genome across all human cell types (158). This task also requires compelling evidence to support the causality of variants within susceptibility loci through in vitro and in vivo targeted genome editing experiments (158). These new resources and techniques have already provided important mechanistic insights for several novel susceptibility loci for CAD, including those regions harboring the genes CDKN2B, SORT1, TCF21, ADAMTS7, SMAD3, and other loci (165-172). Although this understanding has not yet been translated to therapeutic agents targeting these loci, the rapid path of translation for other susceptibility loci, such as PCSCK9, provides some reassurance that such developments are forthcoming and are likely to provide us with groundbreaking opportunities to further reduce, and possibility eliminate, the threat of CAD in the 21st century.
References


48. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.
49. Schunkert H, Götz A, Braund P, et al. Repeated replication and a prospective meta-


155. Kathiresan S. Developing medicines that mimic the natural successes of the human genome: lessons from NPC1L1, HMGCR, PCSK9, APOC3, and CETP. J Am Coll Cardiol 2015;65:1562-6.


Figure Legends

Central Illustration: Genetics of CAD: Prevention and Management. A summary of recent progress and future directions in identifying the genetic determinants of CAD as well as the application of this knowledge to the management and prevention of this disease. The figure summarizes topics covered by this review. CAD = coronary artery disease; PCSK9 = proprotein convertase subtilisin/kexin type 9; SNP = single-nucleotide polymorphism.

Figure 1. Stages of a GWAS. In the first stage, SNPs across the genome are genotyped, almost exclusively on chip-based products generated by 1 of 2 companies, Illumina (San Diego, California) or Affymetrix (Santa Clara, California). The genotyping content of these products differs, but can be combined through imputation algorithms that allow for the reliable calling of thousands of un-typed SNPs based on the pattern of variation observed in SNPs genotyped in the vicinity. Second, the SNP data are subjected to quality control and cleaning procedures to allow for the removal of samples from ethnically distant subjects and adjustment for any systematic differences between or within cohorts. Third, each SNP that survives quality control and cleaning is then tested for association with a disease or trait. Shown is a Manhattan plot, which is typically used in GWAS and plots the negative log of the p value against chromosomal position. A high false positive rate is expected given the number of statistical tests. Therefore, genome-wide statistical significance is often set at p < 5.0 × 10^{-8} or less at this stage of the analysis. The additive model is the most frequently tested model, which assumes that the presence of 1 risk allele confers an intermediate risk between having no allele and having 2 alleles. Fourth, SNPs or loci are selected for replication in an independent sample set, ideally of the same or larger size than the sample analyzed in the genome-wide association. The selection of loci may be based on statistical significance alone or a combination of statistical significance and biologic plausibility.
Fifth, replication experiments lead to any combination of 3 results: selected loci show clear and unequivocal association with disease, show no association signal, or show an association with disease that is not of sufficient magnitude to pass a predetermined statistical threshold. Sixth, genotyping is performed in independent replication cohorts to determine whether an association with a disease is real. Seventh, data mining at unequivocally associated loci reveals transcripts in and around a locus, in addition to the mapping of all known genetic variation within the region. Further fine mapping of the locus is performed by a combination of deep-resequencing methods to discover new variants and the genotyping of untyped variants to determine which are most significantly associated with disease. The examination of association patterns in the region in other race/ethnic groups may help narrow the high risk region or pinpoint a causal variant. Further analysis of the region is performed to determine the most critical variants, the pathologically relevant gene, and the likely biological effect. SNP = single-nucleotide polymorphism. Reproduced from N Engl J Med, Hardy J, Singleton A, Genomewide association studies and human disease, 2009;360:1759-1768. Copyright © 2009 Massachusetts Medical Society (44).

**Figure 2: Genes and Genetic Variants Around the CAD Core Risk Region on 9p21.** The figure illustrates the blocks of correlated variants in Europeans (CEU). Numbers on the scale bar represent the chromosome 9 location (kb = kilobases). Shading indicates correlation (or linkage disequilibrium [LD]), ranging from D’ = 1 in red (strong correlation) to D’ = 0 in white (no correlation). Genes are shown above the bar in blue, with arrows indicating the direction of transcription. The core risk region overlaps the last section of CDKN2BAS, a long noncoding RNA, which is transcribed antisense to CDKN2B. Reproduced from Curr Atheroscler Rep, Genetic Mechanisms Mediating Atherosclerosis Susceptibility at the Chromosome 9p21 Locus,
Figure 3: MR study Provides Evidence of Causality Between a Risk Factor and an Outcome. (A) Similarities between RCTs and MR studies. Both designs eliminate confounding through randomization (B) The Mendelian randomization model. In order to infer a causal role of a risk factor, X, in a disease, Y, 3 assumptions must be held. First, the genetic instrument, G, must associate with X. Secondly, G must be independent of confounders, U, of the X-Y relationship. Thirdly, G must be associated with Y only through its effect on X. If these 3 relationships can be demonstrated, the risk factor, X, can be said to be causally related to the disease. MR = Mendelian randomization; RCT = randomized controlled trial. Reproduced from Curr Cardiol Rep, Genetic Risk Factors and Mendelian Randomization in Cardiovascular Disease (2015) 17: 33, Swerdlow, D.I., Hingorani, A.D. & Humphries, S.E., © Springer Science+Business Media New York 2015, with permission of Springer (174).

Figure 4: Incorporating a GRS Into Standard Clinical Risk Assessment for CAD. The figure shows how a GRS of CAD variants can be integrated with a clinical risk score to update an individual’s 10-year risk. The increase in the estimated 10-year risk illustrated here could result in a recommendation to initiate statins. CAD = coronary artery disease; GRS = genetic risk score. Reproduced from Front. Genet., Simple, standardized incorporation of genetic risk into non-genetic risk prediction tools for complex traits: coronary heart disease as an example, 2015 Jul 7; 7: 231, open access (122).

Figure 5: A GRS of CAD Variants Identifies Optimal Responders to Established Prevention Therapy. The figure shows hazard ratios for CAD with statin therapy across GRS categories. The boxes indicate the point estimates, and the size of each box represents the weight
of a trial’s data within that subgroup. The horizontal lines are the 95% CIs. The diamonds provide summary data. In PROVE IT-TIMI 22, the control group is moderate-intensity statin therapy (pravastatin 40 mg) and the statin group is high-intensity statin therapy (atorvastatin 80 mg). Subjects in the high genetic risk score category show the largest reductions in relative risk, whereas subjects at low risk have a nonsignificant reduction in relative risk. CI = confidence interval. Other abbreviations as in Figure 4. Reproduced from The Lancet, 385(9984), Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials, p. 2264-71, Copyright (2015), with permission from Elsevier (126).

**Figure 6: Genetic Variants by Risk-Allele Frequency and Strength of Genetic Effect.**

Genome-wide association studies to date for CAD and other complex traits have explored common variants and established many associations with modest to low effect size on risk (brown). Whole-exome sequencing and whole-genome sequencing studies seek to further identify all low-frequency and rare variants within genes or across the entire genome, respectively, that cannot be genotyped or imputed. The effect sizes of less common and rare variants are on average greater than more common variants, although the initial exome sequencing studies for CAD suggest these differences in effect sizes are small, requiring very large sample sizes for discovery. CAD = coronary artery disease; GWA = genome-wide association. Reproduced from Nature, Teri A. Manolio, Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff et al., Finding the missing heritability of complex diseases, 2009. 41(1): p. 5-6, with permission (140).
<table>
<thead>
<tr>
<th>Lead SNP</th>
<th>Chromosome</th>
<th>Nearest Genes</th>
<th>Frequency of Allele Raising Risk</th>
<th>OR (95% CI)</th>
<th>P Value</th>
<th>Potential Mechanism Of Action</th>
<th>Year Locus First Reported to Reach Genome-Wide Significance</th>
<th>Consortium and/or Author and Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11206510</td>
<td>1</td>
<td>PCSK9</td>
<td>0.85</td>
<td>1.08 (1.05-1.11)</td>
<td>2.340 × 10⁻⁸</td>
<td>LDL levels</td>
<td>2009</td>
<td>MIGen (59)</td>
</tr>
<tr>
<td>rs7528419</td>
<td>1</td>
<td>SORT1</td>
<td>0.79</td>
<td>1.12 (1.10-1.15)</td>
<td>1.970 × 10⁻²³</td>
<td>LDL levels</td>
<td>2007</td>
<td>WTCCC and Cardiogenics (56)</td>
</tr>
<tr>
<td>rs515135</td>
<td>2</td>
<td>APOB</td>
<td>0.79</td>
<td>1.07 (1.04-1.10)</td>
<td>3.090 × 10⁻⁸</td>
<td>LDL levels</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs6544713</td>
<td>2</td>
<td>ABCG5-ABCG8</td>
<td>0.32</td>
<td>1.05 (1.03-1.07)</td>
<td>8.880 × 10⁻⁷</td>
<td>LDL levels</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs56289821</td>
<td>19</td>
<td>LDLR</td>
<td>0.90</td>
<td>1.14 (1.11-1.18)</td>
<td>4.440 × 10⁻¹⁵</td>
<td>LDL levels</td>
<td>2009</td>
<td>MIGen (59)</td>
</tr>
<tr>
<td>rs4420638</td>
<td>19</td>
<td>APOE-</td>
<td>0.17</td>
<td>1.10 (1.07-1.13)</td>
<td>7.070 × 10⁻¹¹</td>
<td>LDL levels</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs</td>
<td>SNV</td>
<td>Gene</td>
<td>Haplo</td>
<td>Effect</td>
<td>S.E.</td>
<td>p-value</td>
<td>Outcome</td>
<td>Year</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>------------</td>
<td>-------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>--------------</td>
<td>------</td>
</tr>
<tr>
<td>rs3184504</td>
<td>12</td>
<td>APOC1</td>
<td>0.42</td>
<td>1.07 (1.04-1.09)</td>
<td>1.030 × 10^{-9}</td>
<td>LDL levels, BP</td>
<td>2009</td>
<td>deCODE (58)</td>
</tr>
<tr>
<td>rs55730499</td>
<td>6</td>
<td>SH2B3</td>
<td>0.06</td>
<td>1.37 (1.31-1.44)</td>
<td>5.390 × 10^{-9}</td>
<td>Lp(a) levels</td>
<td>2009</td>
<td>WTCCC and Cardiogenics (174)</td>
</tr>
<tr>
<td>rs264</td>
<td>8</td>
<td>LPL</td>
<td>0.85</td>
<td>1.06 (1.03-1.09)</td>
<td>1.060 × 10^{-3}</td>
<td>TRIG levels</td>
<td>2013</td>
<td>CARDIoGRAM+</td>
</tr>
<tr>
<td>rs2954029</td>
<td>8</td>
<td>TRIB1</td>
<td>0.55</td>
<td>1.04 (1.03-1.06)</td>
<td>2.610 × 10^{-6}</td>
<td>TRIG levels</td>
<td>2013</td>
<td>CARDIoGRAM+</td>
</tr>
<tr>
<td>rs964184</td>
<td>11</td>
<td>ZNF259 - APOA5 - APOA1</td>
<td>0.18</td>
<td>1.05 (1.03-1.08)</td>
<td>5.600 × 10^{-3}</td>
<td>TRIG levels</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs17609940</td>
<td>6</td>
<td>ANKS1A</td>
<td>0.82</td>
<td>1.03 (1.00-1.05)</td>
<td>3.000 × 10^{-2}</td>
<td>HDL levels, height</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs3918226</td>
<td>7</td>
<td>NOS3</td>
<td>0.06</td>
<td>1.14 (1.09-1.19)</td>
<td>1.690 × 10^{-9}</td>
<td>BP</td>
<td>2015</td>
<td>1GP CARDIoGRAM+</td>
</tr>
<tr>
<td>rs number</td>
<td>Loci</td>
<td>Gene</td>
<td>MAF</td>
<td>Effect Allele</td>
<td>Odds Ratio (95% CI)</td>
<td>p value</td>
<td>Trait</td>
<td>Year</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>-----</td>
<td>---------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>------</td>
</tr>
<tr>
<td>rs2681472</td>
<td></td>
<td>ATP2B1</td>
<td>0.20</td>
<td>1.08 (1.05-1.10)</td>
<td>6.170 \times 10^{-11}</td>
<td>BP</td>
<td>2012</td>
<td>Lu et al. (63)</td>
</tr>
<tr>
<td>rs17514846</td>
<td></td>
<td>FURIN-FES</td>
<td>0.44</td>
<td>1.05 (1.03-1.07)</td>
<td>3.100 \times 10^{-7}</td>
<td>BP</td>
<td>2013</td>
<td>CARDIoGRAM+</td>
</tr>
<tr>
<td>rs72689147</td>
<td></td>
<td>GUCY1A3</td>
<td>0.82</td>
<td>1.07 (1.05-1.10)</td>
<td>6.070 \times 10^{-9}</td>
<td>BP, cell growth/differentiation/apoptosis</td>
<td>2013</td>
<td>CARDIoGRAM+</td>
</tr>
<tr>
<td>rs11830157</td>
<td></td>
<td>KSR2</td>
<td>0.36</td>
<td>1.12 (1.08-1.16)</td>
<td>2.120 \times 10^{-9}</td>
<td>BMI</td>
<td>2015</td>
<td>1GP</td>
</tr>
<tr>
<td>rs663129</td>
<td></td>
<td>PMAIP1-MC4R</td>
<td>0.26</td>
<td>1.06 (1.04-1.08)</td>
<td>3.200E-08</td>
<td>BMI</td>
<td>2015</td>
<td>1GP</td>
</tr>
<tr>
<td>rs4252185</td>
<td></td>
<td>PLG</td>
<td>0.06</td>
<td>1.34 (1.28-1.41)</td>
<td>1.640 \times 10^{-52}</td>
<td>coagulation</td>
<td>2013</td>
<td>CARDIoGRAM+</td>
</tr>
<tr>
<td>rs2519093</td>
<td></td>
<td>ABO</td>
<td>0.19</td>
<td>1.08 (1.06-1.11)</td>
<td>1.190 \times 10^{-11}</td>
<td>coagulation, LDL levels</td>
<td>2011</td>
<td>CARDIoGRAM+</td>
</tr>
<tr>
<td>rs9349379</td>
<td></td>
<td>PHACT</td>
<td>0.43</td>
<td>1.14 (1.12-1.16)</td>
<td>1.810 \times 10^{-42}</td>
<td>Arterial vessel wall</td>
<td>2009</td>
<td>MIGen (59)</td>
</tr>
<tr>
<td>rs</td>
<td>R1</td>
<td>Gene</td>
<td>Minor Allele</td>
<td>Minor Allele Frequency</td>
<td>Major Allele Frequency</td>
<td>Effect Size</td>
<td>95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------------</td>
</tr>
<tr>
<td>rs9319428</td>
<td>13</td>
<td>FLT1</td>
<td>0.31</td>
<td>1.04 (1.02-1.06)</td>
<td>7.130 × 10^{-5}</td>
<td>Arterial vessel wall endothelial cell</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs8042271</td>
<td>15</td>
<td>MGE8</td>
<td>0.90</td>
<td>1.10 (1.06-1.14)</td>
<td>3.680 × 10^{-8}</td>
<td>Arterial vessel wall endothelial cell</td>
<td>2015</td>
<td>1GP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-ABHD2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7212798</td>
<td>17</td>
<td>BCAS3</td>
<td>0.15</td>
<td>1.08 (1.05-1.11)</td>
<td>1.880 × 10^{-8}</td>
<td>Arterial vessel wall endothelial cell</td>
<td>2015</td>
<td>1GP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4593108</td>
<td>4</td>
<td>EDNRA</td>
<td>0.80</td>
<td>1.07 (1.05-1.10)</td>
<td>8.820 × 10^{-10}</td>
<td>Arterial vessel wall smooth muscle cell</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs17087335</td>
<td>4</td>
<td>REST-NOA1</td>
<td>0.21</td>
<td>1.06 (1.04-1.09)</td>
<td>4.590 × 10^{-8}</td>
<td>Arterial vessel wall smooth muscle cell</td>
<td>2015</td>
<td>1GP</td>
</tr>
<tr>
<td>rs12202017</td>
<td>6</td>
<td>TCF21</td>
<td>0.70</td>
<td>1.07 (1.05-1.09)</td>
<td>1.980 × 10^{-11}</td>
<td>Arterial vessel wall smooth muscle cell</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs2107595</td>
<td>7</td>
<td>HDAC9</td>
<td>0.20</td>
<td>1.08 (1.05-1.10)</td>
<td>8.050 × 10^{-11}</td>
<td>Arterial vessel wall smooth muscle cell</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs</td>
<td>Pos</td>
<td>Gene</td>
<td>rs</td>
<td>Chromosome and Location</td>
<td>Minor Allele</td>
<td>Minor Allele Frequency</td>
<td>Minor Allele Effect</td>
<td>Common Disease Risk</td>
</tr>
<tr>
<td>-------------</td>
<td>-----</td>
<td>--------</td>
<td>----</td>
<td>-------------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>rs2891168</td>
<td>9</td>
<td>CDKN2</td>
<td></td>
<td>9p21</td>
<td>0.49</td>
<td>1.21 (1.19-1.24)</td>
<td>2.290 × 10^{-98}</td>
<td>Arterial vessel wall - smooth muscle cell</td>
</tr>
<tr>
<td>rs11191416</td>
<td>10</td>
<td>CYP17A</td>
<td></td>
<td>1-CNNM2-NT5C2</td>
<td>0.87</td>
<td>1.08 (1.05-1.11)</td>
<td>4.650 × 10^{-9}</td>
<td>Arterial vessel wall smooth muscle cell</td>
</tr>
<tr>
<td>rs4468572</td>
<td>15</td>
<td>ADAMTS7</td>
<td></td>
<td></td>
<td>0.59</td>
<td>1.08 (1.06-1.10)</td>
<td>4.440 × 10^{-16}</td>
<td>Arterial vessel wall smooth muscle cell</td>
</tr>
<tr>
<td>rs10840293</td>
<td>11</td>
<td>SWAP70</td>
<td></td>
<td></td>
<td>0.55</td>
<td>1.06 (1.04-1.08)</td>
<td>1.280 × 10^{-8}</td>
<td>Arterial vessel wall smooth muscle cell, inflammation/immune system/cell migration-adhesion</td>
</tr>
<tr>
<td>rs17678683</td>
<td>2</td>
<td>ZEB2-ACO740</td>
<td>93.1</td>
<td></td>
<td>0.09</td>
<td>1.10 (1.07-1.14)</td>
<td>3.000 × 10^{-9}</td>
<td>Cell growth/differentiation/apoptosis</td>
</tr>
<tr>
<td>rsID</td>
<td>SNP</td>
<td>Gene</td>
<td>Minor Allele</td>
<td>Minor Allele Frequency</td>
<td>Major Allele Frequency</td>
<td>p-Value</td>
<td>Year</td>
<td>Study/Database</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>-------</td>
<td>--------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>rs2128739</td>
<td>11</td>
<td>PDGFD</td>
<td>A</td>
<td>0.32</td>
<td>1.07 (1.05-1.09)</td>
<td>$7.050 \times 10^{-11}$</td>
<td>2011</td>
<td>C4D (60)</td>
</tr>
<tr>
<td>rs56062135</td>
<td>15</td>
<td>SMAD3</td>
<td>G</td>
<td>0.79</td>
<td>1.07 (1.05-1.10)</td>
<td>$4.520 \times 10^{-9}$</td>
<td>2015</td>
<td>1GP CARDIoGRAM+ C4D (54)</td>
</tr>
<tr>
<td>rs46522</td>
<td>17</td>
<td>UBE2Z</td>
<td>T</td>
<td>0.51</td>
<td>1.04 (1.02-1.06)</td>
<td>$1.840 \times 10^{-5}$</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs9970807</td>
<td>1</td>
<td>PPAP2B</td>
<td>C</td>
<td>0.92</td>
<td>1.13 (1.10-1.17)</td>
<td>$5.000 \times 10^{-14}$</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs2487928</td>
<td>10</td>
<td>KIAA14</td>
<td>G3</td>
<td>0.42</td>
<td>1.06 (1.04-1.08)</td>
<td>$4.410 \times 10^{-11}$</td>
<td>2011</td>
<td>C4D (60)</td>
</tr>
<tr>
<td>rs1870634</td>
<td>10</td>
<td>CXCL2</td>
<td>A</td>
<td>0.64</td>
<td>1.08 (1.06-1.10)</td>
<td>$5.550 \times 10^{-15}$</td>
<td>2007</td>
<td>WTCCC and Cardiogenics (56)</td>
</tr>
<tr>
<td>SNP</td>
<td>Chrom</td>
<td>Gene</td>
<td>Minor allele</td>
<td>Major allele</td>
<td>OR</td>
<td>95% CI</td>
<td>P-value</td>
<td>Functional role</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>rs1412444</td>
<td>10</td>
<td>LIPA</td>
<td>0.37</td>
<td>1.07 (1.05-1.09)</td>
<td>5.150 × 10^{-12}</td>
<td>Inflammation/immune system/cell migration-adhesion</td>
<td>2011</td>
<td>C4D (60)</td>
</tr>
<tr>
<td>rs6689306</td>
<td>1</td>
<td>IL6R</td>
<td>0.45</td>
<td>1.06 (1.04-1.08)</td>
<td>2.600 × 10^{-9}</td>
<td>Inflammation/immune system/cell migration-adhesion, cell growth/differentiation/apoptosis</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs67180937</td>
<td>1</td>
<td>MIA3</td>
<td>0.66</td>
<td>1.08 (1.06-1.11)</td>
<td>1.010 × 10^{-12}</td>
<td>Extracellular matrix</td>
<td>2007</td>
<td>WTCCC &amp; Cardiogenics (56)</td>
</tr>
<tr>
<td>rs11838776</td>
<td>13</td>
<td>COL4A1/A2</td>
<td>0.26</td>
<td>1.07 (1.05-1.09)</td>
<td>1.830 × 10^{-10}</td>
<td>Extracellular matrix</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs16986953</td>
<td>2</td>
<td>AK097927</td>
<td>0.10</td>
<td>1.09 (1.06-1.12)</td>
<td>1.450 × 10^{-8}</td>
<td>Other/unknown</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs7568458</td>
<td>2</td>
<td>VAMP5-VAMP8-GGCX</td>
<td>0.45</td>
<td>1.06 (1.04-1.08)</td>
<td>3.620 × 10^{-10}</td>
<td>Other/unknown</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs6725887</td>
<td>2</td>
<td>WDR12</td>
<td>0.11</td>
<td>1.14 (1.11-1.18)</td>
<td>9.510 × 10^{-18}</td>
<td>Other/unknown</td>
<td>2009</td>
<td>MIGen (59)</td>
</tr>
<tr>
<td>SNP</td>
<td>Chr</td>
<td>Gene/Org</td>
<td>Minor Allele</td>
<td>Minor Allele Frequency</td>
<td>Risk Allele</td>
<td>Odds Ratio (95% CI)</td>
<td>P Value</td>
<td>Association</td>
</tr>
<tr>
<td>-------------</td>
<td>-----</td>
<td>--------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>rs9818870</td>
<td>3</td>
<td>MRAS</td>
<td>0.14</td>
<td>1.07 (1.04-1.10)</td>
<td>2.210 × 10^-6</td>
<td>Other/unknown</td>
<td>2009</td>
<td>Cardiogenics (57)</td>
</tr>
<tr>
<td>rs273909</td>
<td>5</td>
<td>SLC22A4</td>
<td>0.12</td>
<td>1.06 (1.03-1.09)</td>
<td>1.240E-04</td>
<td>Other/unknown</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs6903956</td>
<td>6</td>
<td>ADTRP-C6orf10</td>
<td>0.35</td>
<td>1.00 (0.98-1.02)</td>
<td>9.600 × 10^-1</td>
<td>Other/unknown</td>
<td>2011</td>
<td>Wang et al. (176)</td>
</tr>
<tr>
<td>rs56336142</td>
<td>6</td>
<td>KCNK5</td>
<td>0.81</td>
<td>1.07 (1.04-1.09)</td>
<td>1.850 × 10^-8</td>
<td>Other/unknown</td>
<td>2013</td>
<td>CARDIoGRAM+ C4D (62)</td>
</tr>
<tr>
<td>rs10953541</td>
<td>7</td>
<td>7q22</td>
<td>0.78</td>
<td>1.05 (1.03-1.08)</td>
<td>1.020 × 10^-5</td>
<td>Other/unknown</td>
<td>2011</td>
<td>C4D (60)</td>
</tr>
<tr>
<td>rs11556924</td>
<td>7</td>
<td>ZC3HC1</td>
<td>0.69</td>
<td>1.08 (1.05-1.10)</td>
<td>5.340 × 10^-11</td>
<td>Other/unknown</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs10139550</td>
<td>14</td>
<td>HHIPL1</td>
<td>0.42</td>
<td>1.06 (1.04-1.08)</td>
<td>1.380 × 10^-8</td>
<td>Other/unknown</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs216172</td>
<td>17</td>
<td>SMG6</td>
<td>0.35</td>
<td>1.05 (1.03-1.07)</td>
<td>5.070 × 10^-7</td>
<td>Other/unknown</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rs12936587</td>
<td>17</td>
<td>RAI1-</td>
<td>0.61</td>
<td>1.03 (1.01-1.05)</td>
<td>8.240 × 10^-8</td>
<td>Other/unknown</td>
<td>2011</td>
<td>CARDIoGRAM (61)</td>
</tr>
<tr>
<td>rsID</td>
<td>Chr</td>
<td>Gene (Additional Information)</td>
<td>OR (95% CI)</td>
<td>p-value</td>
<td>Population</td>
<td>Year</td>
<td>Database</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>--------------------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>rs12976411</td>
<td>19</td>
<td>ZNF507 - LOC400684 (gene desert)</td>
<td>0.91 (1.38-1.67)</td>
<td>$1.180 \times 10^{-14}$</td>
<td>Other/unknown</td>
<td>2015</td>
<td>1GP CARDioGRAM+C4D (54)</td>
<td></td>
</tr>
<tr>
<td>rs28451064</td>
<td>21</td>
<td>KCNE2 (gene desert)</td>
<td>0.12 (1.10-1.17)</td>
<td>$1.330 \times 10^{-15}$</td>
<td>Other/unknown</td>
<td>2009</td>
<td>MGen (59)</td>
<td></td>
</tr>
<tr>
<td>rs180803</td>
<td>22</td>
<td>POM12 1L9P-ADORA2A 2A (gene desert)</td>
<td>0.97 (1.13-1.27)</td>
<td>$1.640 \times 10^{-10}$</td>
<td>Other/unknown</td>
<td>2015</td>
<td>1GP CARDioGRAM+C4D (54)</td>
<td></td>
</tr>
</tbody>
</table>

*Table shows susceptibility loci having reached genome-wide significance for CAD at any time, ORs, and p values for the lead SNP within each locus from the most recent CARDioGRAMplusC4D meta-analysis, involving largely subjects of European (77%) and South Asian (13%) descent, sorted by potential mechanism of action and chromosomal location. Adapted from supplementary Tables 2 and 4 in Nikpay et al. (54).

†The most significant SNP in the locus, as determined by the CARDIOGRAMplusC4D 1,000 genomes meta-analysis.
The nearest 1 to 3 genes in the locus where the signal has been found. Although these genes are the most likely to be affected, it is possible that another gene that is slightly further away, but still nearby, may be responsible for the association, which would change the potential mechanism of action. For many loci, the causal gene in the region remains unclear.

The p value for several of these loci that have reached genome-wide significance ($p < 5 \times 10^{-8}$) in prior studies is greater than the genome-wide significance in the latest meta-analysis by CARDIoGRAMplusC4D, most likely due to statistical fluctuation. All of the lead SNPs, except for rs17609940 (ANKS1A) and rs6903956 (ADTRP-C6orf105), show a strong trend towards genome-wide significance, and likely represent true CAD susceptibility loci. The rs17609940 (ANKS1A) region requires further validation. rs6903956 (ADTRP-C6orf105) is a locus that was first identified in East Asian in 2011, but never replicated in both additional East Asian and non-East Asian populations, despite adequate power to detect the original effect size.

Potential mechanisms of action are on the basis of a combination of cross-referencing these regions with findings in the region for GWAS of risk factors for CAD and other related phenotypes (e.g., other complications of atherosclerosis, arterial aneurysms, migraines, and so on) and/or what is already known about the function of the nearby genes.

These loci were identified using a recessive model of inheritance; all other loci were identified using an additive model of inheritance.

BMI = body mass index; BP = blood pressure; CAD = coronary artery disease; CI = confidence interval; HDL = high-density lipoprotein; LDL = low-density lipoprotein; Lp(a) = lipoprotein(a); OR = odds ratio; SNP = single-nucleotide polymorphism; TRIG = triglycerides.