Abstract: Observational epidemiological studies have associated plasma lipid concentrations with risk for coronary heart disease (CHD), but these studies cannot distinguish cause from mere correlation. Human genetic studies, when considered with the results of randomized controlled trials of medications, can potentially shed light on whether lipid biomarkers are causal for diseases. Genetic analyses and randomized trials suggest that low-density lipoprotein is causal for CHD, whereas high-density lipoprotein is not. Surprisingly, human genetic evidence suggests that lipoprotein(a) and triglyceride-rich lipoproteins causally contribute to CHD. Gene variants leading to higher levels of plasma apolipoprotein B-containing lipoproteins [low-density lipoprotein, triglyceride-rich lipoproteins, or lipoprotein(a)] consistently increase risk for CHD. For triglyceride-rich lipoproteins, the most compelling evidence revolves around lipoprotein lipase and its endogenous facilitator (APOA5 [apolipoprotein A-V]) and inhibitory proteins (APOC3 [apolipoprotein C-III], ANGPTL4 [angiopoietin like 4]). Combined, these genetic results anticipate that, beyond low-density lipoprotein, pharmacological lowering of triglyceriderich lipoproteins or lipoprotein(a) will reduce risk for CHD, but this remains to be proven through randomized controlled trials. (Circ Res. 2016;118:579-585. DOI: 10.1161/CIRCRESAHA.115.306398.)

Key Words: atherosclerosis ■ coronary disease ■ genetics ■ lipoproteins ■ triglycerides
Atherosclerotic vascular disease, particularly coronary heart disease (CHD) and its complication of myocardial infarction (MI), is the leading cause of death worldwide. Although many environmental factors influence the risk of CHD, genetics play an important role as well. A parental history of premature CHD is associated with a 2- to 3-fold increase in one’s personal CHD risk. Plasma lipoproteins—low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglyceride-rich lipoproteins (TRLs), and lipoprotein(a) (Lp(a))—all have been found to be associated with CHD risk in observational epidemiological studies.

Whereas LDL as assessed by its cholesterol content (LDL-C) and TRLs (as assessed by triglycerides) are directly associated with CHD, HDL as assessed by its cholesterol content (HDL-C) is inversely associated with CHD risk. These lipid traits are themselves partly genetically determined, with about one half of the interindividual variation in plasma lipid concentrations attributable to genetic variants.

Over the past few decades, a variety of human genetics approaches have identified genes that are involved in lipoprotein metabolism and, thus, have possible involvement in the pathogenesis of CHD. Traditional linkage analyses and candidate gene sequencing have pinpointed genes responsible for familial hypercholesterolemia as well as other inherited lipid disorders. These genes harbor rare DNA variants—usually deleterious loss-of-function mutations, but occasionally gain-of-function mutations—that so greatly affect protein function that they singlehandedly produce highly aberrant plasma lipid levels that can potentially result in CHD.

Common DNA variants with small or moderate effects on gene function also modulate plasma lipid levels, albeit in a polygenic fashion. Family-based studies are not well suited to identifying these variants; instead, population-based association studies are needed. The primary population-based study design used to date has been the common variant association study (GWAS), in which large cohorts of unrelated individuals are genotyped at millions of single nucleotide polymorphisms (SNPs) across the genome. For each SNP, individuals with one genotype at the SNP are compared with individuals with another genotype to assess whether there is a phenotypic difference. Such lead SNPs indicate a genomic locus harboring a gene or other functional element that influences the phenotype. With respect to plasma lipid levels, a GWAS with ≈100000 people identified 95 loci with LDL-C, HDL-C, triglycerides, or total cholesterol, and a subsequent association study with ≈190000 people increased the yield to 157 loci.

More recently, next-generation sequencing technologies have enabled the rare variant association study. As rare variants occur too infrequently to allow association tests of individual variants, rare variant association studies require aggregating rare variants into sets and comparing the aggregate frequency in cases versus controls or mean quantitative trait values in those who carry a set versus those who do not. Rare variant association studies have identified coding variants with large effects on gene function in large cohorts of unrelated individuals, effectively combining the advantages of GWASs and traditional family-based studies. Such rare variant association studies have been able to identify several novel genes associated with lipids.

As related below, these discoveries have yielded surprising insights into the causal roles of plasma lipids in the pathogenesis of CHD and into the appropriate use of lipid biomarkers to predict the clinical efficacy of lipid-modifying agents in the reduction of CHD risk.

### The Mendelian Randomization Study Design

Starting in the 1960s with the seminal finding in the Framingham Heart Study that plasma total cholesterol concentration was associated with future risk for CHD, hundreds of biomarkers have been reported to be associated with CHD risk in observational epidemiological studies. What is unclear is how many of these biomarkers are themselves causal in the pathobiology of CHD, and how many are simply proxies for other causal processes. Although any of these biomarkers is potentially useful for cardiovascular risk prediction, only the causal biomarkers represent potential therapeutic targets. The gold standard for proving that a biomarker is causal is a randomized controlled trial (RCT) that demonstrates that an intervention specifically targeting the biomarker reduces the risk of CHD. Such RCTs typically require following thousands or tens of thousands of individuals for several years, making them a time-consuming and costly proposition.

The principles of human genetics offer an alternative study design, called Mendelian randomization, that is, akin to an RCT that has already been performed by nature. DNA variants can be used as instruments to assess whether a biomarker that has been found to have an epidemiological association with risk for a disease is truly causal for the disease. If (1) a DNA variant is known to directly influence the biomarker level (eg, a noncoding variant in a promoter or enhancer that alters the expression of the gene that encodes the biomarker) or the activity of a protein that directly influences the biomarker level (eg, a coding variant that affects the function of an enzyme that metabolizes the biomarker) and (2) the biomarker is truly causal for a disease, then (3) the DNA variant should be associated with disease risk to an extent consistent with the size of the effect of the DNA variant on the biomarker level and, in turn, the size of the effect of the biomarker on the
Musunuru and Kathiresan  Surprises From Genetic Analyses of Lipids  581

disease process. Assuming that the Mendelian randomization study is adequately powered—which can require hundreds of thousands of individuals if the effect of the DNA variant on the biomarker is small—if the expected association between the DNA variant and disease risk is not apparent, it would argue that the biomarker is not causal for the disease.

The similarity between a Mendelian randomization study and an RCT arises from Mendel’s first law—the law of segregation—which dictates that each of a parent’s 2 alleles at the site of a DNA variant has a 50% chance of being passed to a given gamete and, thus, being transmitted to an offspring. In other words, there is random assignment of alleles to offspring. This assignment is unaffected by traditional confounders of observational epidemiology studies, eg, age, disease status, and socioeconomic status. (There is the possibility of confounding via epigenetic phenomena affecting allele transmission, but this is a largely theoretical concern.) The law of segregation renders Mendelian randomization studies fairly impervious to confounding or reverse causation, which is also a major advantage of RCTs.

There are 2 possible shortcomings to a Mendelian randomization study. First, the study will only be as reliable as the estimates of effect sizes of the DNA variant on the biomarker and the biomarker on disease risk, which are obtained from observational epidemiological studies. If the estimates are unreliable, so too may be the conclusions drawn from a Mendelian randomization study. Second, there is an assumption that the DNA variant only affects disease risk via the biomarker in question, and not through any other mechanisms. If there is pleiotropy, ie, the DNA variant affects multiple mechanisms, then an observed association between the DNA variant and disease risk may not be because of the biomarker. In this scenario, the biomarker may not itself be causal but instead act as a proxy for another mechanism influenced by the DNA variant.

These issues notwithstanding, Mendelian randomization is potentially a powerful tool to discriminate between causal and noncausal biomarkers for a disease. It is particular useful when considered in combination with RCTs in which pharmacological agents targeting the same biomarkers have been assessed for their effects on disease risk.

Plasma LDL Cholesterol as a Causal Biomarker

There is now ample evidence from human genetics that the plasma LDL-C concentration represents a causal risk factor for CHD. Initial studies of patients with familial hypercholesterolemia identified loss-of-function mutations in the LDLR gene to be linked to high plasma LDL-C levels and premature CHD, with disease manifesting as early as childhood. Biological plausibility emerged from the recognition that the protein product of LDLR, the LDL receptor, resides on the plasma membrane and is responsible for the uptake of LDL particles out of the bloodstream into the cell, within which the LDL particles are degraded. In subsequent studies, mutations in the APOB and PCSK9 genes were identified in familial hypercholesterolemia patients in whom LDLR mutations had been ruled out.\(^{15,16}\) APOB encodes apolipoprotein B (apoB), a key component of LDL particles that is the protein via which the LDL receptor binds to LDL particles and promotes their uptake into cells. Specific APOB mutations result in disruption of the interaction between the LDL receptor and its ligand, apoB, leading to an increased plasma LDL-C concentration and premature CHD. PCSK9 encodes a protein that acts as an antagonist to the LDL receptor by promoting its degradation. Gain-of-function mutations in PCSK9, thus, cause familial hypercholesterolemia and premature CHD by inhibiting the removal of LDL particles from the bloodstream. Conversely, loss-of-function mutations in PCSK9 result in increased levels of the LDL receptor on cells and reduction of the blood LDL-C concentration, translating into as much as an 88% reduction of CHD risk.\(^{17}\)

These studies of LDLR, APOB, and PCSK9 variants make a strong argument for a causal link between LDL-C and CHD. However, they all impinge on a single biological pathway, cellular LDL particle uptake that directly modulates the plasma LDL-C concentration in a specific way. What remains to be answered is whether all biological pathways that alter blood LDL-C concentration also alter CHD risk. A partial answer to this question was provided by an analysis of 10 lead SNPs in loci previously identified in GWAS to be primarily associated with LDL-C.\(^{18}\) These SNPs were assessed in a case–control study of \(\approx\)20000 individuals with MI and 50000 control individuals. Nine of the 10 SNPs were found to be associated not only with LDL-C but also with risk of MI in the concordant direction (ie, the same allele associated with both a decrease in LDL-C and a decrease in MI risk). Although the loci included the LDLR, APOB, and PCSK9 genes, they also included a variety of other genes that influence LDL-C through other mechanisms, such as APOE, HMGCR (which encodes the cholesterol synthesis enzyme targeted by the statin drugs), and LPA [which encodes apolipoprotein (a), a component of the LDL-like Lp(a) particle].\(^{19-20}\)

In the same study, a more formal Mendelian randomization analysis was performed for LDL-C in more than 50000 cases and controls, using a genetic score comprising 13 SNPs in loci primarily associated with LDL-C.\(^{18}\) Strikingly, whereas a 1-standard deviation (1-SD) increase in LDL-C (\(\approx35\) mg/dL increase) was expected to be associated with a 54% increase in MI risk using data from observational epidemiological studies, a 1-standard deviation increase in LDL-C because of genetic score was found to confer a 113% increase in risk (\(P=2\times10^{-10}\); Figure 1). Thus, the analysis suggested that the genetic contribution to the plasma LDL-C level has, if anything, an outsize effect on CHD risk, strongly arguing for a generalized causal relationship between LDL-C and CHD.

The most widely used LDL-C–lowering medications are the statins, which have been demonstrated in numerous RCTs in a broad variety of populations to reduce the risk of cardiovascular events.\(^{21}\) In isolation, these RCTs suggest but do not prove that LDL-C is causal for CHD, because statins have well-known pleiotropic effects—besides reducing plasma LDL-C levels, they also reduce plasma C-reactive protein levels, which are a marker for inflammation.\(^{22}\) In principle, the beneficial effects of statins could be because of anti-inflammatory effects rather than lipid modification. However, in light of the strong genetic evidence that LDL-C is causal for CHD, it is reasonable to interpret the RCTs as demonstrating that the reduction of LDL-C is a primary mechanism by which statins protect against CHD.
Besides the statins, one of the more commonly used lipid-modifying drugs is ezetimibe, which works to reduce plasma LDL-C levels at least in part by inhibiting the protein product of the NPC1L1 gene. NPC1L1 regulates the absorption of dietary and biliary cholesterol in the gastrointestinal tract. Because of negative results in an early RCT assessing the effects of ezetimibe on a nonclinical end point, carotid intima-media thickness, ezetimibe was felt by many commentators to be an unproven medication despite its unequivocal, safe reduction of plasma LDL-C levels by 15% to 20%. Recently, Mendelian randomization studies using variants in or near the NPC1L1 gene found that carriers of the variants not only had lower plasma LDL-C levels but also had decreased CHD risk. Indeed, the degree of risk reduction exceeded that, which would be predicted from observational epidemiological studies. However, carriers of the LIPG Asn396Ser variant were found to have a negligible change in MI risk (7% decrease; P=0.63; Figure 1).

The totality of the existing genetic data and RCT data strongly argue that LDL-C is causal for CHD risk, at least in a simplistic sense. Although the data cannot rule out that there are some biological mechanisms that lead to increased plasma LDL-C levels that also protect against CHD, it seems fair to conclude that not all interventions that raise HDL-C will reduce CHD risk. Further support for this conclusion is provided by RCT data, most notably with the cholesteryl ester transfer protein (CETP) inhibitors, which substantially raise plasma HDL-C levels. Three inhibitors of CETP—torcetrapib, dalcetrapib, and evacetrapib—all raised HDL-C substantially, and each failed to reduce risk for CHD in large-scale RCTs. All 3 studies—the Investigation of Lipid Level Management to Understand Its Impact in Atherosclerotic Events (ILLUMINATE) trial of torcetrapib, which increased HDL-C by 70%; the dal-OUTCOMES trial of dalcetrapib, which increased HDL-C by 30%; and the Assessment of Clinical Effects of Cholesteryl Ester Transfer Protein Inhibition With Evacetrapib in Patients at a High-Risk for Vascular Outcomes (ACCELERATE) trial of evacetrapib, which was projected to raise HDL-C by >90%—were all terminated prematurely because of lack of clinical efficacy.

Interestingly, CETP has pleiotropic effects on blood lipid levels, and SNPs in or near the CETP gene are associated with SNPs was found to have a strong relationship with MI risk. The same study performed a parallel Mendelian randomization study in >50,000 cases and controls using a genetic score comprising 14 GWAS SNPs primarily associated with plasma HDL-C levels. Whereas a 1-standard deviation increase in HDL-C (≈15 mg/dL) was expected to be associated with a 38% decrease in MI risk using data from observational epidemiological studies, a 1-standard deviation decrease in HDL-C because of genetic score conferred no significant change in MI risk (7% decrease; P=0.63; Figure 1).

The same study performed a more focused Mendelian randomization analysis on a coding SNP (Asn396Ser) in the LIPG gene, which encodes endothelial lipase, an enzyme that metabolizes HDL particles but has little effect on plasma LDL-C and triglycerides. To obtain adequate power for the analysis, the SNP was genotyped in >20,000 individuals with MI and 95,000 control individuals. Carriers of the LIPG Asn396Ser variant had increased plasma HDL-C levels, on average ≈5.5 mg/dL. This degree of increase in HDL-C was expected to be associated with a 13% decrease in MI risk using data from observational epidemiological studies. However, carriers of the LIPG Asn396Ser variant were found to have a negligible change in MI risk (1% decrease; P=0.85; 95% confidence interval ranging from an 11% increase to a 12% decrease), essentially ruling out an effect of LIPG on the pathogenesis of CHD.

In contrast to LDL-C, the collective genetic data suggest that HDL-C is not causal for CHD risk, at least in a simplistic sense. Although the data cannot rule out that there are some biological mechanisms that lead to increased plasma LDL-C levels that also protect against CHD, it seems fair to conclude that not all interventions that raise HDL-C will reduce CHD risk.
not only altered plasma HDL-C levels but also plasma LDL-C in the opposite direction, albeit to a lesser degree. Multiple genetic studies have found that CETP variants are associated with modest changes in CHD risk.\textsuperscript{18,34,35} Given the pleiotropy at work, it seems likely that the altered CHD risk is due primarily to the effect of CETP on LDL-C rather than HDL-C. This observation leaves open the possibility that the ongoing RCT of a fourth CETP inhibitor—anacetrapib——may find some degree of clinical utility, if the LDL-C–lowering effect is large enough to be beneficial and the HDL-C–raising effect is not harmful.

**Plasma Triglyceride-Rich Lipoproteins as Casual**

The epidemiological association of plasma triglyceride levels with CHD risk is not as strong as those of LDL-C and HDL-C.\textsuperscript{3} Nonetheless, genetic evidence is emerging thatTRLs as assessed by plasma triglycerides represent a causal risk factor for CHD. SNPs in at least 6 genes that modulate plasma triglyceride levels—apolipoprotein A-V (APOA5), apolipoprotein C-III (APOC3), angioptietin like 4 (ANGPTL4), LPL, APOA4, and TRIB1——have been persuasively linked to CHD.\textsuperscript{18,36–43} However, Mendelian randomization studies using SNPs associated with plasma triglyceride levels are difficult to interpret because of most of those SNPs having pleiotropic relationships with lipids. Of 185 lead SNPs for GWAS loci associated with lipids, 1 or more plasmid lipid traits, 94 are associated with triglycerides; of those, just 7 are only associated with triglycerides, with the other 87 also associated with LDL-C or HDL-C.\textsuperscript{44} This is perhaps not surprising, because triglycerides are carried by multiple classes of lipoprotein particles in the blood and, accordingly, the measured plasma triglyceride concentration reflects contributions from multiple physiological processes.

Given this pleiotropy, a statistical framework—termed multivariable Mendelian randomization—to separate the triglyceride-associated effects on CHD risk from the LDL-C– and HDL-C–associated effects on CHD risk was recently developed.\textsuperscript{44} Confirming previous observations, the isolated LDL-C genetic effect (ie, contribution of all LDL-C–associated SNPs, after adjustment for the HDL-C and triglyceride effects) confers significantly increased risk of CHD, whereas the isolated HDL-C genetic effect on CHD risk is negligible. In contrast to HDL-C, the isolated triglyceride effect increases the risk of CHD to a comparable degree as the isolated LDL-C effect (Figure 1). This finding suggests that the plasma triglyceride concentration captures risk causes for CHD that is independent of the plasma LDL-C concentration.

Exactly what risk factor or, potentially, risk factors are embodied in the plasma triglyceride concentration remain to be fully defined, although presumably they involve the metabolism of TRLs that carry triglycerides in the blood. The aforementioned APOA5, APOC3, ANGPTL4, and LPL genes all share the common characteristic that they encode either lipoprotein lipase or encode regulators of lipoprotein lipase, a key enzyme that hydrolyzes triglycerides in various lipoprotein particles. This suggests that lipoprotein lipase is central to a causal pathway for CHD. Another possible causal risk factor may lie in postprandial cholesterol metabolism, specifically the amount of cholesterol in remnant lipoproteins, with which plasma triglyceride levels are strongly correlated. Remnant lipoproteins seem to promote atherosclerosis in much the same way as LDL particles.\textsuperscript{45}

Of note, RCTs of triglyceride-lowering therapies have yielded ambiguous results, with a trial of gemfibrozil and a trial of fenofibrate resulting in reduced cardiovascular events\textsuperscript{46,47} (although the latter did not show a difference in the primary end point, coronary events) but another trial of fenofibrate and a trial of omega-3 fatty acids showing no such reduction.\textsuperscript{46,49} This may reflect that the specific mechanism by which plasma triglycerides are lowered, in combination with other factors, such as the degree of triglyceride reduction and characteristics of the study population, determines the extent of clinical benefit. Triglyceride-lowering therapies that do so by altering TRLs via the lipoprotein lipase pathway may prove to be particularly efficacious.

**Lipoprotein(a)- and Lipoprotein-Associated Phospholipase A\textsubscript{2} as Causal and Noncausal Biomarkers**

Lp(a) is an LDL-like particle that is covalently linked to a protein called apolipoprotein(a), expressed by the LPA gene. The plasma Lp(a) level is notable in that it varies up to 1000-fold among individuals, with the vast majority of the variation determined by genetic variation.\textsuperscript{50} With plasma Lp(a) being associated with CHD risk,\textsuperscript{4} a natural question has been whether Lp(a) is a causal risk factor for disease. Mendelian randomization studies using variants in or near the LPA gene have unequivocally demonstrated that genetically elevated Lp(a) results in increased risk of CHD.\textsuperscript{50,20} Thus, in principle, therapies that specifically reduce plasma Lp(a) concentrations should confer cardiovascular protection.

Lipoprotein-associated phospholipase A\textsubscript{2} (Lp-PLA\textsubscript{2}) is an enzyme that is encoded by the \textit{PLA2G7} gene and circulates in the plasma, primarily associated with LDL particles. Both Lp-PLA\textsubscript{2} mass and activity in the plasma are associated with CHD risk.\textsuperscript{51} These observations prompted 2 large RCTs with darapladib, an inhibitor of Lp-PLA\textsubscript{2}. Both the Stabilization of Atherosclerotic Plaque by Initiation of Darapladib Therapy (STABILITY) trial and The Stabilization Of pLaques using Darapladib-Thrombolysis In Myocardial Infarction 52 Trial (SOLID-TIMI 52) found that darapladib did not reduce the risk of CHD,\textsuperscript{52,53} calling into question whether Lp-PLA\textsubscript{2} is a causal risk factor for disease. A subsequently reported Mendelian randomization study using variants in \textit{PLA2G7} found no association with CHD risk.\textsuperscript{54} The consistency of the results of RCTs and Mendelian randomization studies for statins, ezetimibe, and darapladib and their target genes supports the notion that Mendelian randomization studies could potentially be used to prioritize RCTs for those agents most likely to result in the desired clinical outcome.

**Implication of Lipid and Nonlipid Causal Factors in Atherosclerosis**

The weight of the genetic evidence suggests that the plasma LDL-C, triglycerides, and Lp(a) concentrations reflect causal risk factors for CHD, whereas the plasma HDL-C concentration does not. This is contrary to the expectations one would have if going purely by observational epidemiological studies,
which generally find that HDL-C has the strongest association with CHD. The disparity highlights the need to distinguish between association and causation with respect to biomarkers of disease, with genetic approaches, such as Mendelian randomization studies proving to be helpful in this regard.

Another notable finding to emerge from human genetic studies is the overall importance of lipid causal factors in atherosclerosis. The largest GWAS reported to date for coronary artery disease identified a total of 55 loci associated with the clinical phenotype. Remarkably, 13 of the loci are clearly linked to plasma LDL-C, triglycerides, or Lp(a) (Figure 2). The common feature of LDL particles,TRLs, and Lp(a) is that they all carry apoB, reinforcing the notion that apoB-containing particles are generally atherogenic and that approaches that reduce any of these classes of apoB-containing particles may have therapeutic efficacy. Of course, any such approaches will need to be validated with RCTs before clinical use.

A much smaller subset of the 55 loci are linked to other recognized risk factors for CHD, including blood pressure and inflammation. The remaining loci seem to contribute to CHD by as of yet unrecognized pathogenetic mechanisms, suggesting that there are quite a few causal risk factors (and possibly therapeutic targets) remaining to be discovered.

Disclosures

None.

References


Surprises From Genetic Analyses of Lipid Risk Factors for Atherosclerosis
Kiran Musunuru and Sekar Kathiresan

Circ Res. 2016;118:579-585
doi: 10.1161/CIRCRESAHA.115.306398
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/118/4/579

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

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

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