



2018 George Lyman Duff Memorial Lecture

Genetics and Genomics of Coronary Artery Disease: A Decade of Progress

Ruth McPherson

ABSTRACT: Recent studies have led to a broader understanding of the genetic architecture of coronary artery disease and demonstrate that it largely derives from the cumulative effect of multiple common risk alleles individually of small effect size rather than rare variants with large effects on coronary artery disease risk. The tools applied include genome-wide association studies encompassing over 200 000 individuals complemented by bioinformatic approaches including imputation from whole-genome data sets, expression quantitative trait loci analyses, and interrogation of ENCODE (Encyclopedia of DNA Elements), Roadmap Epigenetic Project, and other data sets. Over 160 genome-wide significant loci associated with coronary artery disease risk have been identified using the genome-wide association studies approach, 90% of which are situated in intergenic regions. Here, I will describe, in part, our research over the last decade performed in collaboration with a series of bright trainees and an extensive number of groups and individuals around the world as it applies to our understanding of the genetic basis of this complex disease. These studies include computational approaches to better understand missing heritability and identify causal pathways, experimental approaches, and progress in understanding at the molecular level the function of the multiple risk loci identified and potential applications of these genomic data in clinical medicine and drug discovery.

VISUAL OVERVIEW: An online [visual overview](#) is available for this article.

Key Words: alleles ■ coronary artery disease ■ Genome Wide Association Study ■ genomics ■ humans

It is a great honor to be chosen as the first Canadian to present the George Lyman Duff Memorial Lecture established in 1956 by the Society for the Study of Arteriosclerosis. Dr Duff was one of Canada's most distinguished pathologists whose studies of experimental atherosclerosis in the rabbit provided a foundation for years of work that followed. Here, I will describe, in part, our research over the last decade performed in collaboration with a series of bright trainees and an extensive number of groups and individuals around the world as it applies to our understanding of the genetic basis of this complex disease.

GENETICS OF CORONARY ARTERY DISEASE

Rare coding variants, primarily in genes regulating LDL (low-density lipoprotein) cholesterol (LDL-C) metabolism, for example, *LDLR*, *PCSK9*, and *APOB*, contribute

significantly to premature coronary artery disease (CAD) risk and the importance of early diagnosis and treatment of familial hypercholesterolemia is well established. However, CAD is a complex phenotype, largely driven by the cumulative effect of multiple common variants affecting numerous biological processes, and the majority of individuals with premature CAD do not harbor a causative mutation in a single gene.

Recent progress in understanding the genetics of CAD and other complex disease has been driven by technological advances including high-throughput DNA microarray technology using chips containing up to a million DNA markers consisting of single-nucleotide polymorphisms (SNPs) that tag common variation (allele frequency $\geq 5\%$ of individuals) across the human genome. Of note, these are tag SNPs that point to a causative locus, but are rarely themselves functional variants. This approach makes use of linkage disequilibrium (LD) that is the nonrandom

Correspondence to: Ruth McPherson, MD, PhD, Atherogenomics Laboratory, Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, 40 Ruskin St, H4Z0Z3, Ottawa, Canada K1Y 4W7. Email rmcpherson@ottawaheart.ca

For Sources of Funding and Disclosures, see page 1935.

© 2019 The Authors. *Arteriosclerosis, Thrombosis, and Vascular Biology* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](#) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

Arterioscler Thromb Vasc Biol is available at www.ahajournals.org/journal/atvb

Nonstandard Abbreviations and Acronyms

AP-1	activator protein-1
CAD	coronary artery disease
CARDIoGRAM	Coronary Artery Disease Genome-Wide Replication and Meta-Analysis
CRISPR-Cas9	clustered regularly interspaced short palindromic repeats/clustered regularly interspaced short palindromic repeat-associated 9
CRP	C-reactive protein
eQTL	expression quantitative trait loci
GENIUS-CHD	Genetics of Subsequent Coronary Heart Disease
GWAS	genome-wide association studies
HDL	high-density lipoprotein
IL	interleukin
LD	linkage disequilibrium
LDL	low-density lipoprotein
LDL-C	low-density lipoprotein cholesterol
Lp-PLA2	lipoprotein-associated phospholipase A2
MIGEN	Myocardial Infarction Genetics Consortium
miRQTL	miRNA expression quantitative trait loci
MLM	mixed linear model
MR	Mendelian randomization
PRS	polygenic risk score
QTL	quantitative trait loci
SNP	single-nucleotide polymorphism
TGFβ	transforming growth factor-β
VSMC	vascular smooth muscle cell

coinheritance of genetic variants across the human genome. Comparison of the allele frequency of each SNP in cases and controls provides an agnostic approach with discovery potential, unlike the candidate gene studies of the previous century. Given that comparison of a million markers implies a P of $\leq 5 \times 10^{-8}$ for genome-wide association studies (GWAS) significance, success has been contingent on the allele frequency and effect size of a given variant and requires large meta-analyses involving thousands of carefully phenotyped CAD cases and controls and collaboration among many groups across the world.

Using this approach, concurrently with 2 other groups, we identified the first robust association with CAD in a large LD block containing multiple highly correlated SNPs at 9p21.3.¹⁻³ The early discovery of this risk locus was facilitated by its large effect size (>1.3) and high-risk allele frequency (≈ 0.48). Approximately 25% of Europeans carry 2 copies of the risk allele and have a 40%

Highlights

- Genetic risk for coronary artery disease is largely due to multiple common variants of small effect size, interacting in pathways and networks, many related to immune function, cell proliferation, interactions at the arterial wall.
- The majority of the over 160 genome-wide association studies-identified loci are in noncoding regions of the genome, implying local or long-distance effects on the expression of genes via a number of mechanisms or on their proximal regulators including microRNAs or long noncoding RNAs. Functional characterization of each of these loci remains challenging.
- The vast amount of available genome-wide association studies data can be used to create robust polygenic risk scores to identify otherwise healthy individuals at high future risk for coronary artery disease.

increased risk of CAD in general and a 2-fold risk of premature CAD. The estimated population attributable risk is 10% to 15% of CAD incidence in the United States, making it the largest known genomic contributor to direct and indirect healthcare costs.⁴ For our discovery sample, we applied an extreme phenotype approach comparing young, multivessel CAD patients to elderly, physically active, asymptomatic subjects. We demonstrated that 9p21.3 as a particularly strong locus for early disease as later confirmed by the CARDIoGRAM study (Coronary Artery Disease Genome-Wide Replication and Meta-Analysis) where the allele-specific odds ratio for CAD in subjects with onset before 50 years of age was 1.41 versus 1.24 for later onset CAD.⁵ The 9p21 locus associates with the extent and severity of atherosclerosis.^{6,7} It has also been consistently shown that the risk conferred by this locus is independent of known CAD risk factors and associates with other vascular phenotypes including carotid atherosclerosis,⁸ stroke,⁹ peripheral arterial disease,¹⁰ abdominal aortic, and intracranial aneurysms,¹¹ highlighting possible effects on vascular wall integrity. A recent study by the GENIUS-CHD (Genetics of Subsequent Coronary Heart Disease) consortium, albeit a survivors' cohort, reveals that this locus is not associated with recurrent events in patients with established CAD. It does, however, confer an increased risk for revascularization,¹² consistent with a primary role in atherosclerosis.

As of 2019, over 160 GWAS-significant loci for CAD have been identified, with close to twice this number when false discovery rate <0.05 variants are included.¹³⁻¹⁶ This success has been driven by collaboration among numerous groups across the world, important initiatives such as the UK Biobank and the ability to use whole genome data from the 1000 Genomes Project to impute lower frequency variants. The 1000 Genomes Project phase 1, v3 data set includes over 38 million variants, half of

which have a minor allele frequency (MAF) <0.005, as well as insertions and deletions (indels). These data were leveraged in a meta-analysis led by groups at Oxford, the Broad and the Ottawa Heart Institute, that included over 185 000 CAD cases and controls from 48 studies providing imputed data on 9.4 million SNPs including 2.7 million low-frequency SNPs (MAF, 0.005–0.05).¹⁶

Several of these loci contain multiple independent signals. With the exception of a few including 9p21.3 and *LPA*, the effect size is generally ≤1.08. Given that the heritability of CAD is ≈50% and conventional risk factors are moderately predictive of atherosclerotic burden,¹⁷ it is not surprising that the majority of novel loci do not relate to lipoprotein metabolism, hypertension, or glycemic traits. Indeed, multiple biological processes are highlighted (Figure 1).

Individually, several GWAS findings provide new understanding of processes related to atherosclerosis or its metabolic precedents. The poster child for this approach was the identification of a 1p13 locus encompassing *CELSR2/PSRC1/SORT1* for LDL-C and CAD that prompted a series of elegant functional studies leading to a previously unknown role for sortilin-1 in lipoprotein metabolism.¹⁸

The discovery of the 9p21.3 risk locus has been instrumental to our understanding of vascular smooth

muscle cell (VSMC) behavior in atherosclerosis. The risk haplotype consists of ≈60 SNPs in perfect LD, some of which overlap the terminal exons of a long noncoding RNA, *ANRIL/CDKN2B-AS1*. Both long (19 exons) and short (13 exons) transcripts of *ANRIL* are identified by RNA sequencing. In an earlier study, we demonstrated increased expression of short variants of *ANRIL* and increased expression of the long variant in carriers of the risk alleles. Relevant to atherosclerosis, genome-wide expression profiling demonstrated upregulation of gene sets modulating cellular proliferation in carriers of the risk allele.¹⁹ In a recent elegant study, Lo Sardo et al²⁰ generated multiple hiPSC (human induced pluripotent stem cells) clones from homozygous carriers of the risk (RR) and nonrisk (NN) alleles and then used TALENs (transcription activator like effector nucleases) to delete the 60-kb haplotype in each of the hiPSC clones. Whole-transcriptome analysis during VSMC development showed that deletion of the risk haplotype in RR clones resulted in transcriptional profiles similar to that of the NN clones. Thus the RR haplotype essentially confers a dedifferentiated state with more proliferative and less contractile VSMCs. The RR VSMCs exhibit globally altered transcriptional networks that encompass a number of GWAS-identified CAD genes converging on aberrant adhesion,

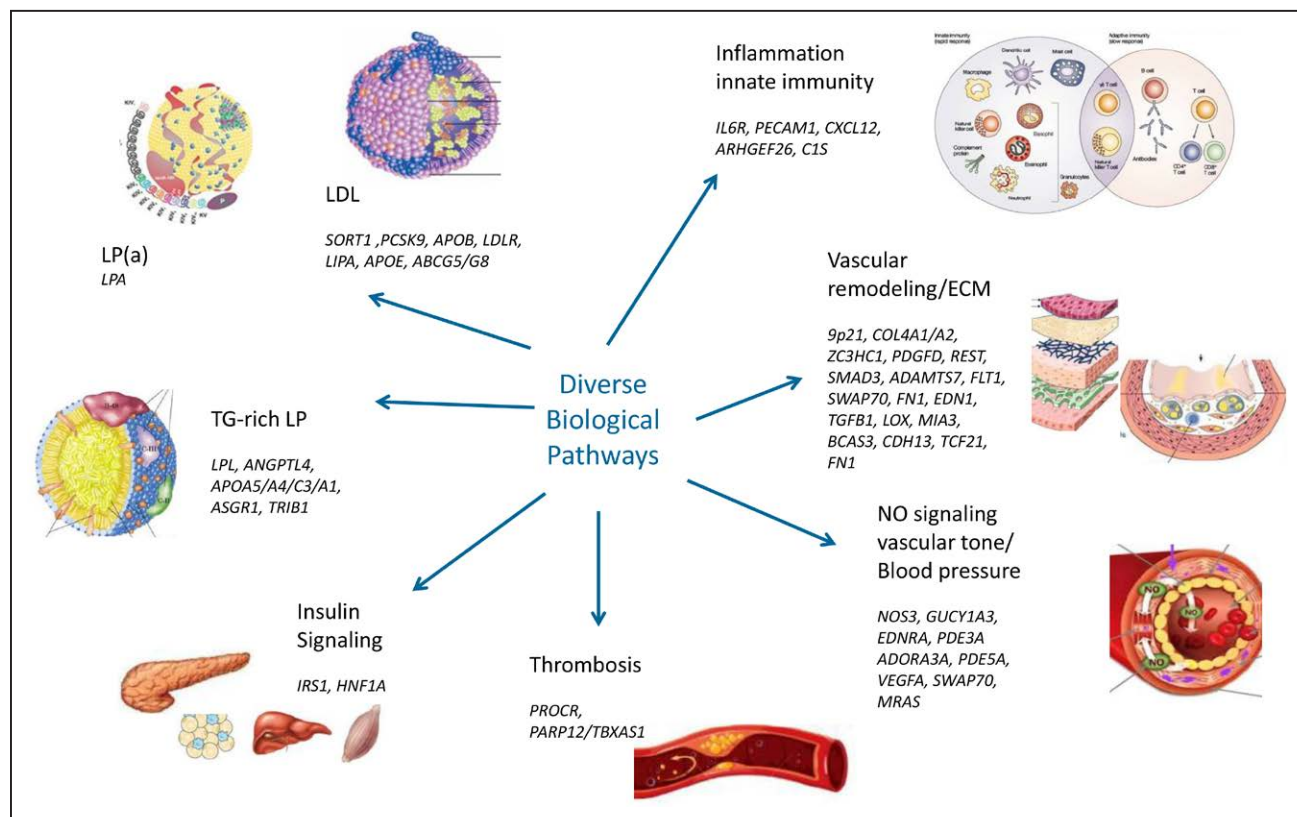


Figure 1. Genome-wide association studies (GWAS)-identified coronary artery disease (CAD) loci indicate a role for multiple biological processes including lipid metabolism, inflammation and innate immunity, thrombosis, vascular remodeling and extracellular matrix integrity, and NO signaling.

For many novel GWAS signals, the link between the associated gene(s) and processes underlying CAD risk remain unknown. ECM indicates extracellular matrix; LDL, low-density lipoprotein; LP(a), lipoprotein (a); and TG-rich LP, triglyceride-rich lipoproteins.

contraction, and proliferation pathways. The risk haplotype is associated with preferential expression of the shorter *ANRIL* isoform that appears to mediate, in part, these phenotypic and functional changes.²⁰

Identification of SNPs at 8q24 associated with plasma lipoproteins and hepatic fat downstream from the *TRIB1* gene uncovered an unknown role for tribbles-1 in hepatic lipid metabolism—an effect that may be mediated by a long noncoding RNA that we have named *TRIBAL*.²¹ Seven independent signals at the *COL4A1/4A2* locus have highlighted the role of the extracellular matrix in atherosclerosis.²²

INTEGRATIVE ANALYSES TO EXTEND GWAS FINDINGS

In recent studies, we have mined additional data sets to extend GWAS findings. Complex phenotypes including atherosclerosis and its clinical sequelae derive from multiple nonlinear and interactive processes. Thus, genetic associations with disease can only be understood by the integration of multiple layers of information and the knowledge that many genetic effects are isolated in time and place that is susceptible to epigenetic effects that take place during a particular phase of development or cell cycle and are specific to a particular cell type or tissue.²³

The simplest level of integration includes pairwise quantitative trait loci (QTL) analysis of mRNA (expression QTL [eQTL]), miRNA-eQTL (miRQTL), protein eQTL, etc. An eQTL effect, that is, an effect on mRNA expression on a nearby (*cis*-eQTL) or distant (*trans*-eQTL) gene, provides strong evidence for a functional effect of a given SNP or one in close LD. Whereas some eQTLs are shared among tissues, others are tissue specific and interrogation of eQTL databases, for example, genotype-tissue expression requires careful consideration of the cell type or tissue of interest.²⁴ Finally, as noted above, the direct contribution of any given locus may be temporally restricted; for instance, early transient activation followed by sustained epigenetic effects.

Gene-Environment Interaction

Identifying gene-environment interactions can further our understanding of the architecture and underlying biology of complex traits.²⁵ A difference in the effect size of a genetic variant in individuals differing in an environmental or clinical exposure suggests the presence of a gene-environment interaction, rather than simply additive effects of the 2 exposures.²⁶ Unfortunately, the ability to measure accurately environmental exposures is a limiting factor. However, there has been limited success. Saleheen et al²⁷ demonstrated that the CAD-protective effect of a GWAS that lowers expression of *ADAMTS7* is attenuated in smokers—an effect attributed to their higher vascular *ADAMTS7* expression. More recently, Hindy et al²⁸ showed

highly significant interaction of smoking with a polygenic risk score (PRS) for CAD with a marked exacerbation of risk in current and to a lesser extent, past smokers. We utilized a weighted PRS constructed from loci previously reported by the Global Lipids Genetics Consortium and demonstrated, in a large population of lean and obese individuals, a significant interaction between obesity status and the effect of risk alleles on each of triglycerides and HDL (high-density lipoprotein) cholesterol but not LDL-C. For plasma triglycerides, the effect size (β) of a weighted PRS_{TG} in the obese population was nearly double that of the lean population (0.48 versus 0.26).²⁹

Pathway Analysis

Clinically informative polymorphisms related to atherosclerosis occur in systems of closely interacting genes.³⁰ Thus, weakly associated variants that do not reach significance in GWAS may provide important information on its biological basis when such variants cluster within a common functional module or pathway.³¹ Thus, beyond single SNP interrogation of GWAS analysis, novel insights into the genetic architecture of CAD can be obtained by interpreting genetic findings in the context of biological processes and functional interactions among genes. We performed a 2-stage pathway-based gene set enrichment analysis of 16 GWAS data sets for CAD, comprised of over 25 000 CAD patients and >66 000 controls³¹ using the i-GSEA4GWAS tool³² and the Reactome pathway database.³³ A total of 32 Reactome pathways demonstrated convincing association with CAD (Figure 2). These resided in core biological processes and included pathways relevant to innate immunity, extracellular matrix integrity, axon guidance, PDGF (platelet-derived growth factor) signaling, NOTCH, and the TGF β (transforming growth factor- β)/SMAD receptor complex, confirming and adding to a substantial body of basic research.

To extend this analysis, we then interrogated molecular networks based on interactions among diverse molecules including genes, proteins, and metabolites that underlie the complex architecture of complex disease.³⁴ As described,³¹ network analysis of unique genes within the replicated pathways further confirmed known processes such as lipid metabolism but also revealed a number of interconnected functional and topologically interacting modules representing novel associations. These included the semaphorin-regulated axonal guidance pathway—a known modulator of cellular adhesion, migration, proliferation, differentiation, survival, and synaptic plasticity, processes involving highly conserved families of guidance molecules, including netrins, slits, semaphorins, and ephrins, and their cognate receptors.³⁵ Network centrality analysis further identified genes (eg, *NCAM1*, *FYN*, and *FURIN*) that may be central to the maintenance and functioning of several of the replicated pathways.³¹

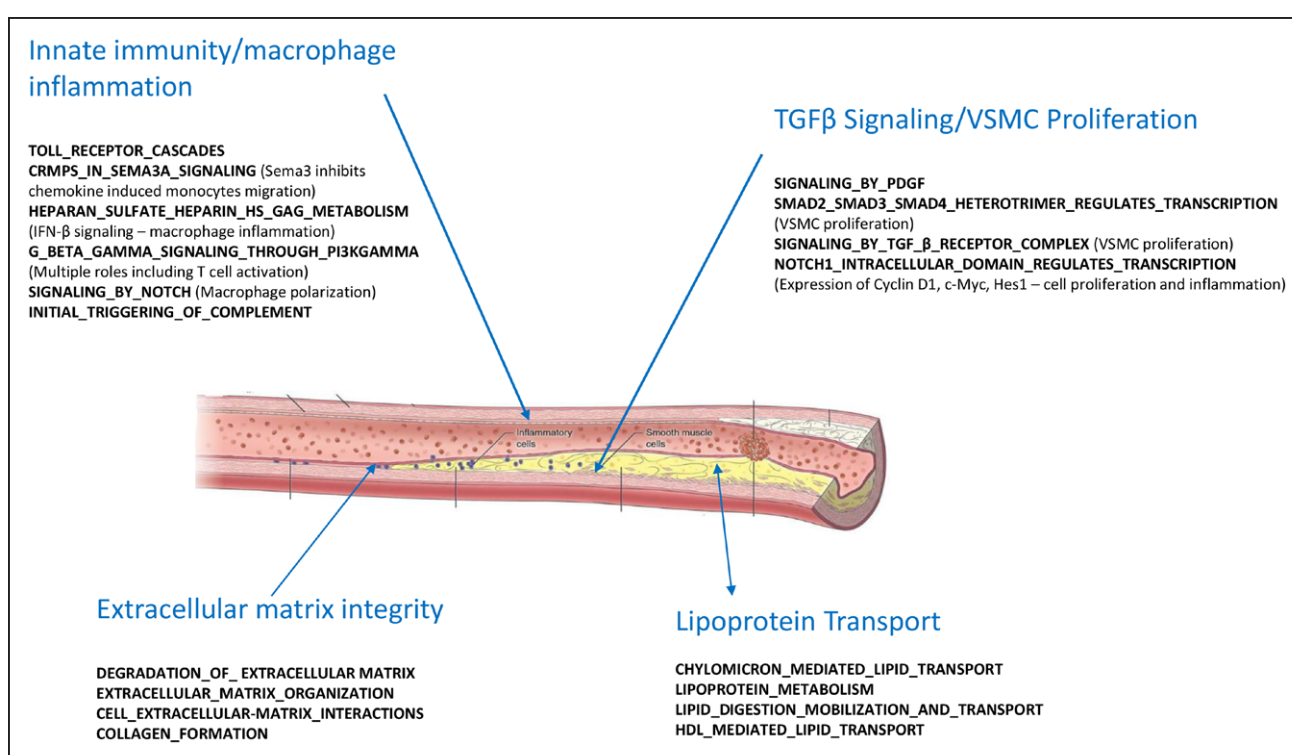


Figure 2. Major biological processes highlighted by pathway analysis of genome-wide association studies including 29 391 coronary artery disease cases and 66 819 controls.³¹

Significant Reactome pathways are indicated in bold. c-Myc indicates c-master regulator of cell cycle entry and proliferative metabolism; HDL, high-density lipoprotein; Hes-1, hairy and enhancer of split-1; IFN- β , interferon-beta; TGF β , transforming growth factor- β ; and VSMC, vascular smooth muscle cell.

Heritability of CAD

Decades of epidemiological research have identified multiple psychosocial and lifestyle variables associated with CAD risk. These risk factors are partially modifiable, account for approximately half of the burden of CAD risk in Western countries, and are used clinically in the application of imperfect risk scores to estimate individual risk. The other half of CAD susceptibility is heritable, that is due to genetic and possibly epigenetic factors. This is termed the broad-sense heritability (H^2) of CAD, and family-based and twin-pair studies estimate H^2 to be 40% to 60%. The analysis of the contribution of GWAS findings commonly focuses on the extent to which they explain narrow-sense heritability (h^2), that is fraction of risk estimated to be genetic in pathogenesis.³⁶

One of the challenges in understanding the genetics of CAD and other complex traits is the discrepancy between the fraction of CAD that is estimated to be heritable and that explained by GWAS-identified variants. The 58 GWAS-significant loci identified in our 2015 article explained only $\approx 11\%$ of h^2 for CAD. Beyond considering only GWAS-significant loci ($P \leq 5 \times 10^{-8}$), a study also leveraging UK Biobank data and using a definition of CAD that included angina, 304 independent variants associated at 5% false discovery rate were identified that together explain 21.2% of estimated CAD heritability (h^2).¹⁴ The phenomenon of missing heritability is attributed to the notion that much

of the heritability of a complex trait lies in low-frequency SNPs with extremely low effect size that are not detectable at current sample sizes. Beyond considering only GWAS or false discovery rate–significant associations, a mega-analysis comprising 1 745 180 genetic variants across the genome increased this number to 26.8%.³⁷

Computational Approaches to the Understanding of Missing Heritability

To address this issue, we and others have used mixed linear models (MLMs). Unlike the univariate GWAS approach, MLMs quantify the overall contribution of SNPs to variation of a trait without testing the SNPs individually or setting a significance threshold level. MLMs can decompose the phenotypic variance into genetic and environmental components and estimate the heritability of a trait or the genetic correlations between 2 groups. Because these methodologies require knowledge of the genetic relatedness of individuals, individual genome-wide genotypes are required rather than the summary data provided by many GWAS centers. MLMs also allow investigation of heritability enrichment in a subset of SNPs defined according to functional annotation, providing the opportunity to leverage annotation information and genome-wide genotype data to gain novel insights into the biology of complex disease.

Exploiting the 1000 Genomes imputed data, we used the MLM approach implemented in genome-wide complex trait analysis to further investigate the genetic basis of CAD.³⁸ We included only genetically unrelated cases and controls, genotyped at this site, and calculated the variance explained by genome-wide autosomal SNPs. The variance estimate derives from the average genome-wide similarity between all pairs of individuals determined using all SNPs. Heritability is estimated when case-case pairs and control-control pairs are on average more similar across the genome than case-control pairs. We then partitioned this heritability by sex, gene modules, and SNP annotation. Using over 3 million autosomal genome-wide SNPs ($MAF \geq 0.01$) and MLMs in a sample of genetically unrelated 4535 cases and 2977 controls, we reported that genome-wide SNPs explain 22% of liability to CAD (55% of narrow-sense heritability). Consistent with our earlier pathway analysis,³¹ we identified a number of CAD-associated modules related to immune responses, cell adhesion, and proliferation. Of note, genes involved in inflammation accounted for one-fifth of SNP-based heritability. As expected, given that 90% of identified CAD loci are in noncoding regions of the genome, heritability-enrichment analysis highlighted epigenetic sites associated with transcriptional activity.

Taken together, these recent studies interrogating pathways and networks derived from a large body of GWAS data^{31,38,39} confirm an important role for immune-mediated processes in CAD. At a single-gene level, there are robust data for a direct role of the interleukin (IL)-6 signaling pathway^{40,41} in atherosclerosis, consistent with clinical trial evidence for a protective effect of IL-6 inhibition using an IL-1 β monoclonal antibody, canakinumab.⁴²

Identification of GWAS Signal for microRNAs Regulating Cardiometabolic Phenotypes

GWAS signals for a complex trait commonly associate with expression of a protein-coding gene (eQTLs). We queried whether pleiotropic effects associated with several GWAS loci may be due to effects on expression of microRNAs. Leveraging miRNA sequencing data and the 1000 Genomes imputed genotypes, we explored this question in a sample of 710 unrelated subjects of European ancestry, using data from the Framingham and the Geuvadis studies to replicate our findings.⁴³ At least 1 GWAS-significant miRQTL was identified for 143 circulating miRNAs. Although the majority accounted for a small portion (<1%), a few explained 4% to 20% of variation in plasma miRNA levels. Identified *cis*-miRQTLs associated with their counterpart mature miRNAs ($P < 0.0001$) suggesting that they mainly regulate the expression of primary miRNAs unlike the *trans*-miRQTLs that appear to exert their effect through processes that affect the stability of mature miRNAs. We then used the identified miRQTLs to investigate the links between

circulating miRNAs and plasma biomarkers through Mendelian randomization (MR) analysis. We found miR-1908-5p plays an important role in regulating LDL-C levels and glycemic traits, whereas miR-10b-5p mediates the *trans*-regulatory effect of the *ABO* locus on several plasma proteins, LDL-C, and on the risk for CAD. We also demonstrated a causal relationship between plasma levels of miR-199a and LDL-C (Figure 3).⁴³ This analysis provides new insights relevant to lipid metabolism and is a focus of ongoing functional studies in our laboratory.

INTERROGATION OF RARE CODING VARIANTS CONTRIBUTING TO CAD

The large body of GWAS data strongly supports the conclusion that genetic susceptibility to CAD is largely derived from the effect of multiple common SNPs individually of small effect size.²³ GWAS is limited by the inability to reliably interrogate rare variants ($MAF < 0.005$), that is present in <1 in 200 individuals, including nonsynonymous mutations in genes of possible relevance to CAD. This approach can, however, be successful in genetic isolates. In a GWAS that included only 809 Amish participants, Shuldiner et al identified an 11q23 SNP, 823 kb away from the *APOA1/C3/A4/A5* gene cluster that associated with markedly reduced plasma TG and protection. Resequencing of the region identified the causal variant, a null mutation R19X in *APOC3*, extremely rare in the general population, but with an MAF of 0.05 in the Amish.⁴⁴

Whole-exome sequencing is a comprehensive approach to identify and classify coding variants linked to human disease and is of particular importance in Mendelian disease where the effect size is high and affected and unaffected family members are available for study.⁴⁵ Musunuru et al⁴⁶ performed whole-exome sequencing in a large kindred previously reported to have low levels of both triglycerides and LDL-C without evidence of atherosclerosis to uncover the protective effects of ANGPTL3 deficiency, now a novel and important drug target.^{47,48}

There has been a rapid increase in the number and scale of large human sequencing projects. These include DiscovEHR,⁴⁹ UK BioBank,⁵⁰ ExAC/gnomAD,⁵¹ and others. With reference to CAD per se, the MIGEN (Myocardial Infarction Genetics Consortium) Exome Sequencing Project reported on whole-exome sequencing efforts in $\approx 11\,000$ individuals. Of note, the identified coding variants were limited to genes relevant to lipoprotein metabolism including *LDLR*,⁵² *APOA5*,⁵² *APOC3*,⁵³ and *NPC1L1*⁵⁴ and for which there was a large body of preexisting information. In a more recent MIGEN study that included validation in the Geisinger Health System DiscovEHR cohort, rare damaging mutations in *LPL* were associated with CAD. Overall, these findings were

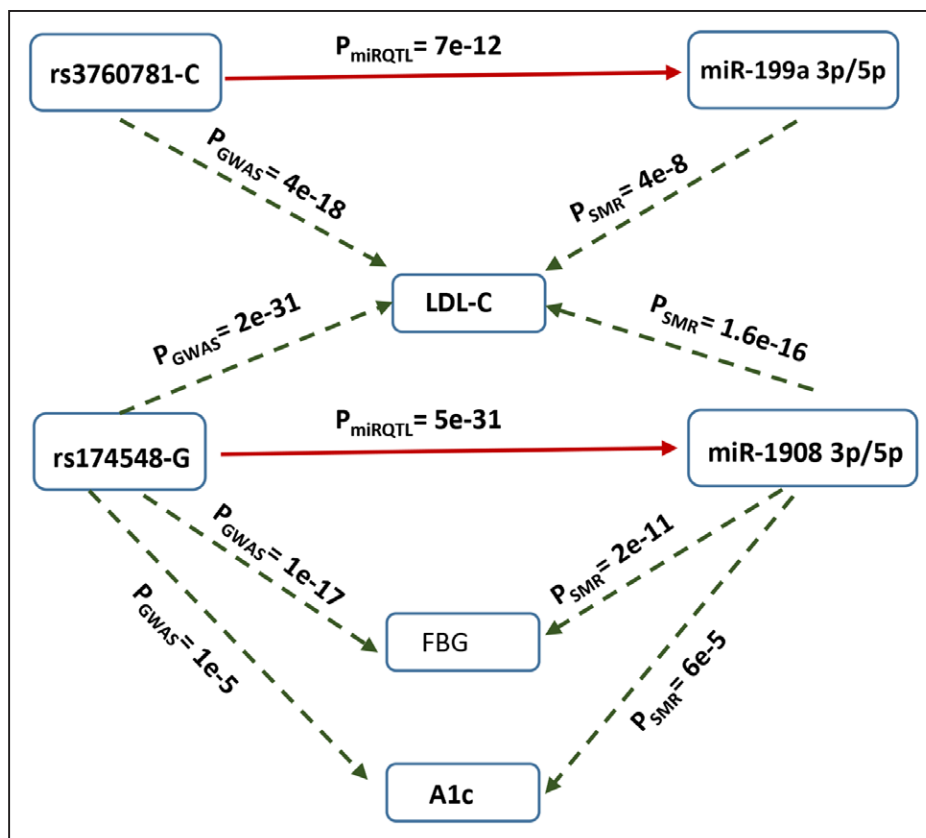


Figure 3. miRNA expression quantitative trait loci (eQTL) data were integrated with summary statistics from previous genome-wide association studies to identify single-nucleotide polymorphisms (SNPs) for which the effect on blood phenotypes is mediated by miRNAs.

An SNP name indicates its rsID-effect allele. At the *FADS* locus, miR-1908—an eQTL for rs174548—plays an important role in regulating LDL (low-density lipoprotein) cholesterol (LDL-C), fasting blood glucose (FBG), and A1c. At the *TMED* locus on chromosome 11, miR-199-3p/5p—an eQTL for rs3760781—regulates LDL-C levels. Positive associations are indicated by solid red lines, negative associations by dotted green lines. Summary-based Mendelian randomization differentiates between a pleiotropic effect and a linkage effect.⁴³

somewhat disappointing but did provide further credence to the relevance of triglyceride metabolism to atherosclerosis^{55,56} and the validity of NPC1L1 as a drug target.^{57,58} Given the perceived need for large sample sizes and subsequent replication, the sample included CAD cases with limited phenotypes such as acute MI without angiographic documentation of disease burden and controls that may have harbored a burden of nonobstructive atherosclerosis. The challenges inherent in the demonstration of cosegregation in family members, often with a multitude of other CAD risk factors, also limit the ability to confirm novel mutations linked to atherosclerosis.

SOMATIC MUTATIONS

In a novel extension to our understanding of the molecular pathogenesis of CAD, Jaiswal et al⁵⁹ have recently described an age-related disorder characterized by the acquisition of somatic mutations in hematopoietic cells. This phenomenon termed clonal hematopoiesis of indeterminate potential allows selective expansion of mutated clones in stem cells, lymphocytes, monocytes, and

granulocytes over time. Clonal hematopoiesis of indeterminate potential is detectable in $\approx 10\%$ of individuals over the age of 70 years and confers an increased risk for hematologic malignancies but also CAD. Increased CAD risk varies from 2-fold for individuals with a mean age of 60 years to 4-fold for those under the age of 50 years. Four of the most commonly mutated genes, *DNMT3A*, *TET2*, *ASXL1*, and *JAK*, showed individually significant associations with CAD. This study that included 8255 participants from 4 case-control cohorts adds credence to the role of inflammation in atherosclerosis. *DNMT3A*, *TET2*, and *JAK* have reported anti-inflammatory functions. Mice transplanted with *Tet2* heterozygous or homozygous deficient bone marrow exhibited increased expression of several CXC chemokines including *Cxcl1*, *Cxcl2*, *Cxcl3*, *Pf4*, and *Ppbb* and proinflammatory cytokine genes including *Il6* and *Il1b* and accelerated atherosclerosis. The *JAK2* V617F mutation was associated with a 12.1-fold CAD risk.⁵⁹ Relevant to this finding, a *JAK* inhibitor tofacitinib is approved for rheumatoid arthritis and is currently under investigation for other immune-mediated diseases. The overall contribution of clonal

hematopoiesis of indeterminate potential to age-related atherosclerosis remains to be fully understood but is likely to be significant.

FROM LOCUS TO FUNCTION

Although GWASs have identified >150 loci associated with CAD, it has been a much more tedious process to understand their function, the majority of which are in noncoding regions of the genome. Over 90% of GWAS-identified loci for CAD are situated in intergenic regions. Few are nonsynonymous coding variants with predicted effects on protein structure or function, implying, as supported by our computational studies, local or long-distance effects on the expression of genes via a number of mechanisms or on their proximal regulators including microRNAs or long noncoding RNAs.

Our program has included studies on the *TRIB1* locus for plasma triglycerides and CAD^{21,60–62} and has a particular focus on the functional interrogation of novel GWAS loci for CAD that converge on VSMC and macrophage biology including an lncRNA, *CDKN2BAS* at the 9p21.3 locus,¹⁹ *SMAD3*,⁶³ *TGFB1*, *MFBG8*,⁶⁴ and a locus encompassing the *COL4A1/4A2/CARS2*²² genes that harbors 7 independent signals for CAD.¹⁶

Interrogation of Noncoding Polymorphisms

The GWAS-identified SNPs are rarely causal but rather in LD with a neighboring or even distal causal polymorphism. The standard approach is to first visualize the various LD blocks in the closest gene using Haploview software and linkage data from the 1000 Genomes Browser. We then seek possible causative SNP(s) by interrogating large-scale consortium data (ENCODE [Encyclopedia of DNA Elements], Roadmap Epigenomics) for chromatin marks identified by chromatin immunoprecipitation sequencing to denote histone modifications associated with promoters, primed and active enhancer regions, transcribed genes, and transcription factor binding. This is supplemented with information in other publically available data bases (RegulomeDB). Evidence for a *cis* or *trans* effect on gene expression (genotype tissue expression, SCAN [SNP and CNV Annotation Database]) is particularly informative where eQTL data are available on relevant tissues including liver and artery (STARNET [Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task]). Drs U. Hedin and L. Folkersen (BiKE [Biobank of Karolinska Endarterectomies] and ASAP [Advanced Study of Aortic Pathology]) have provided us with eQTL data in normal and diseased arterial wall.¹⁰ Finally, we use extensive laboratory analysis, for example, enhancer and promoter assays, allele-specific chromatin immunoprecipitation, allele-specific expression analyses, *CRISPRCas9* gene editing,⁶⁵ in iPS cells and cell types relevant to atherosclerosis to identify and molecularly characterize the functional variant underlying the GWAS signal (Figure 4).

As an example of this approach, our GWAS meta-analysis identified intronic SNP in *SMAD3*, rs56062135C>T with the minor allele (T) associating with protection from CAD.¹⁶ This locus is of interest because *SMAD3* is a key contributor to the TGF β pathway signaling. We then sought to identify causal CAD-associated SNPs at the *SMAD3* locus and unravel mechanisms underlying the association with CAD. Although there are other weaker CAD signals at the *SMAD3* locus, in our conditional and joint association analysis, only rs56062135 remained significant. This SNP lies within a haplotype block, containing 5 other intronic SNPs. Genetic and epigenetic fine mapping highlighted an SNP rs17293632C>T that was in tight LD with the GWAS tag SNP in intron 1 of *SMAD3*. RegulomeDB and ENCODE data indicated functional effects including enhancer histone marks, DNase hypersensitivity sites, alteration of transcription factor binding site, and chromatin immunoprecipitation sequencing data demonstrating binding of multiple transcription factors. In contrast, the GWAS tag SNP had no data to suggest functional effects. In a series of experiments, we showed that the sequence encompassing rs17293632 acts as a strong enhancer in human arterial smooth muscle cells. The common allele (C) preserves an AP-1 (activator protein-1) site and enhancer function, whereas the protective (T) allele disrupts the AP-1 site and significantly reduces enhancer activity ($P<0.001$). Pharmacological inhibition of AP-1 activity upstream demonstrated that this allele-specific enhancer effect is AP-1 dependent ($P<0.001$). Chromatin immunoprecipitation experiments revealed binding of several AP-1 component proteins with preferential binding to the C allele. We then showed that rs17293632 is an eQTL for *SMAD3* in blood and atherosclerotic plaque with reduced expression of *SMAD3* in carriers of the protective allele. Finally, siRNA knock-down of *SMAD3* in human arterial smooth muscle cells increased cell viability.⁶³ Consistently, Iyer et al⁶⁶ recently demonstrated that *SMAD3* inhibits proliferation and facilitates differentiation of HCASMCs (human coronary artery smooth muscle cells) to promote atherosclerosis.

Missense Variants

A small number of GWAS-identified CAD risk SNPs are coding variants, including rs11556924 in *ZC3HC1*, resulting in an Arg363His substitution in NIPA (nuclear interaction partner of ALK). In functional studies, we demonstrated that the ensuing cellular phenotype of the protective variant is one of increased NIPA expression and nuclear mobility and lower rates of cell growth.⁶⁷

APPLICATIONS OF GENOMIC DATA IN CLINICAL MEDICINE AND DRUG DEVELOPMENT

There are 2 potential applications of GWAS findings in clinical cardiovascular medicine. These are MR analyses to

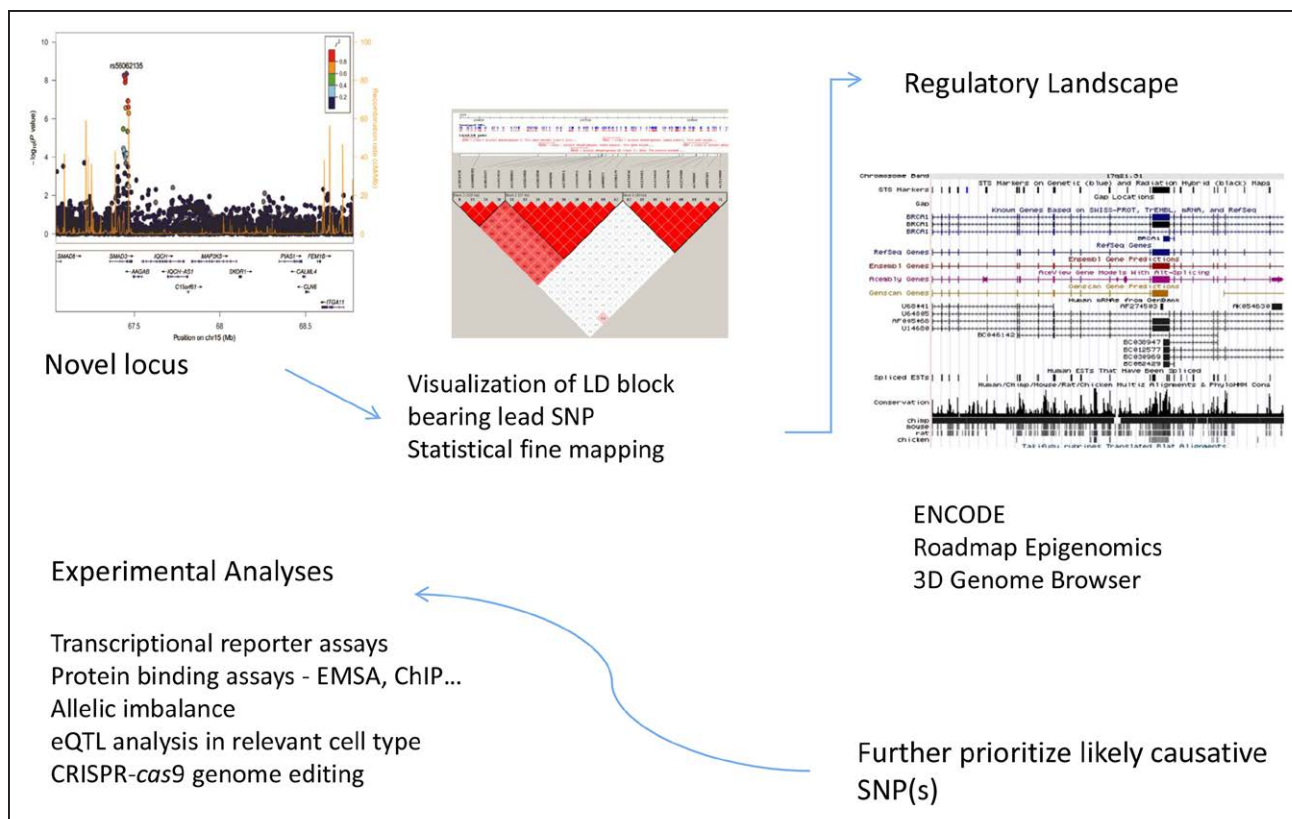


Figure 4. From locus to function.

A, Typical Manhattan plot of a significant locus for coronary artery disease identified in a meta-genome-wide association studies (GWAS). **B**, Visualization of the various linkage disequilibrium (LD) blocks in the region using Haploview and 1000 Genomes LDL (low-density lipoprotein) data. **C**, UCSC genome browser annotation and ENCODE (Encyclopedia of DNA Elements) data for the region bearing the lead single-nucleotide polymorphisms (SNPs), including chromatin regulatory features, histone modifications, and predicted transcription factor binding from chromatin immunoprecipitation (ChIP) sequencing data. **D**, Interrogation of publicly available databases can provide further information relevant to function. **E**, Finally, extensive laboratory analyses are necessary to identify and molecularly characterize the functional variant underlying the GWAS signal. 3-D indicates 3-dimensional; CRISPR-CAS9, clustered regularly interspaced short palindromic repeats/clustered regularly interspaced short palindromic repeat-associated 9; EMSA, electro-mobility shift assay; eQTL, expression quantitative trait loci; and UCSC, University of California, Santa Cruz.

identify causal biomarkers and the application of a PRS to assess future CAD risk relevant to the earlier application of preventive therapies in otherwise healthy individuals.

Mendelian Randomization

Clinical and epidemiological studies have identified a large number of biomarkers that are robustly associated with CAD. However, many of these are not causal risk factors because of confounding and reverse causation. MR is based on a simple tenet. If a biomarker has a causal association with a disease, the genetic determinants of the biomarker will also associate with disease risk.⁶⁸ An important requirement of MR analysis is lack of pleiotropy, meaning that the genetic variant(s) in question influences only the biomarker of interest. This is not always the case for plasma lipid traits, where pleiotropic effects are evident for several genes including *CETP*, *LPL*, and *APOA5*.

Using MR, there have been major advances in the demonstration of a causal association between

plasma levels of lipoprotein (a), LDL-C, and plasma triglycerides (as a marker of remnant cholesterol) and risk of CAD. Elevated levels of LDL-C are causally associated with CAD risk, as evident in individuals with monogenic familial hypercholesterolemia, most often due to a mutation in the *LDLR* gene. MR studies demonstrate that genetic variants at the *PCSK9*, *NPC1L1*, and *HMGCR* loci uniquely associate with plasma concentrations of LDL-C and are predictive of CAD risk. Thus, these data provide robust genetic evidence for the value of specific pharmaceutical treatments, not only statins but also ezetimibe and PCSK9 inhibitors.

In contrast, although CRP (C-reactive protein) and Lp-PLA2 (lipoprotein-associated phospholipase A2) are robust CAD risk markers,⁶⁹ they have not been shown to be causal.^{70,71} MR analyses would have predicted the negative outcome of large randomized controlled trials targeting Lp-PLA2.⁷² A closer inspection of MR data before drug design and initiation of clinical trials could mitigate the high costs of drug development.

Utility of a Polygenic Genetic Risk Score for Risk Assessment and Treatment Decisions

Although the effect of the identified susceptibility variants for CAD is individually small, their effects are independent and additive. These can be incorporated into a PRS consisting of the number of risk alleles adjusted for their individual effect size. In 2010, we demonstrated the ability of 12 robustly associated CAD loci to augment the ability of CAD risk prediction beyond 9p21.3.⁷³ In 2011, as part of the CARDIoGRAM study,⁵ we explored a weighted score based on the 23 validated CAD risk variants and observed a >3-fold difference in CAD risk between the top and bottom deciles of the risk scores. The availability of larger GWAS data sets, improved computational methods, and large-scale biobank cohorts for testing and replication has now made it possible to create and validate PRS for CAD using ≤ 6.6 million SNPs across the genome⁷⁴ (Figure 5).

Using this approach, Khera et al demonstrated that among patients with an MI before the age of 55 y, 17% have a high PRS for CAD as compared with 2% harboring a mutation in the *LDLR*, *APOB*, or *PCSK9* genes. Furthermore, they showed that a high PRS for CAD confers a relative risk of 4- to 5-fold, similar to that of familial hypercholesterolemia.⁷⁵ In a separate analysis, the UK Biobank CardioMetabolic Consortium CHD Working Group created a meta-GRS (genetic risk score) consisting of 1.7 million genetic variants. Their study demonstrated that those in the top 20% of the meta-GRS distribution had a hazard ratio for CAD of 4.2 compared with those in the bottom 20%. Importantly, the corresponding hazard ratio was 2.8 among individuals

on lipid-lowering or antihypertensive medications, reinforcing the mitigating effects of preventive therapies.³⁷ Thus, a PRS can provide important information on future CAD risk beyond traditional risk factors and in the future may have both clinical and economic value in identifying those persons who will benefit most from the early application of various disease-deferring therapies, including statin treatment.

FUTURE PERSPECTIVES

The last decade of GWAS has provided a broader understanding of genetic risk for CAD, complementing the earlier candidate gene approach. Although traditional risk factors remain important, application of these data using a systems genetics approach, has pointed to substantial roles for genes and pathways relevant to vessel wall biology and immune function. However, the published GWAS data provide insight rather than mechanism. Moving forward, a major priority is to develop and apply high-throughput methodology to understand at the molecular and cellular levels the function of each of the novel loci, the majority of which are in noncoding regions of the genome. This is a tedious process as outlined above. Bioinformatic analysis of available epigenomic, eQTL, mQTL (metabolite QTL), and miQTL data and traditional molecular approaches at the cellular level are a first step. CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats/clustered regularly interspaced short palindromic repeat-associated 9) and other techniques allow creation of genetically modified human iPSC that can then be differentiated into a relevant cell type for further study. A promising approach,

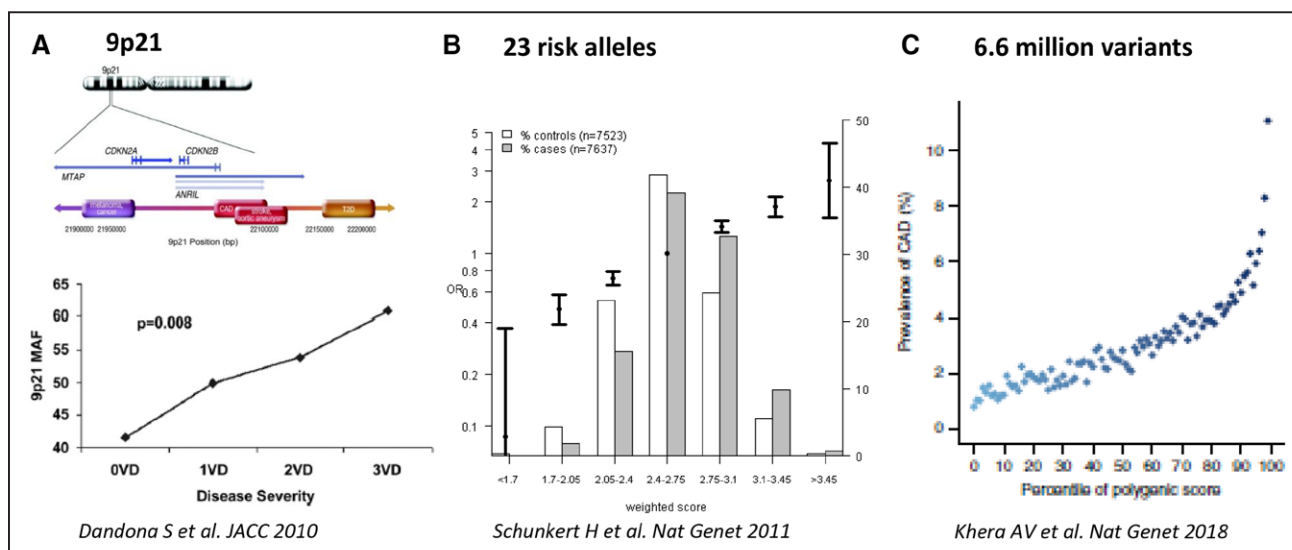


Figure 5. Predictive value of a polygenic risk score for coronary artery disease (CAD).

A, In an angiographic study, we demonstrated that the number of copies of the 9p21 risk allele strongly predicted atherosclerosis burden.⁶ **B**, In an early analysis from CARDIoGRAM (Coronary Artery Disease Genome-Wide Replication and Meta-Analysis), using a weighted score based on the 23 validated CAD risk variants, we demonstrated a >3-fold difference in CAD risk between the top and bottom deciles of the risk scores.⁵ **C**, Using ≤ 6.6 million single-nucleotide polymorphisms across the genome, Khera et al⁷⁴ demonstrated that a polygenic risk score for CAD confers a progressive increase in CAD risk ≤ 10 -fold for individuals in the top 1 percentile. OR indicates odds ratio; and MAF, minor allele frequency.

albeit with many steps and acknowledged pitfalls.⁷⁶ Finally, animal models have been the backbone of atherosclerosis research, but studies in mice are limited by the small effect size of GWAS-identified genetic variants crossed into an *Ldlr*^{-/-} or *ApoE*^{-/-} background. Furthermore, deletion of an entire CAD risk locus, as shown for 9p21,⁷⁷ may not result in an expected phenotype.

Beyond function, these data have been of clinical utility for interrogation of biomarker causality and drug discovery, through MR, albeit with a number of caveats. Polygenetic risk scores for a host of human diseases can be created at birth, at a genotyping/computational cost of <\$100. However, questions remain as to how these data could be best communicated to the patient to encourage lifestyle changes, early screening, or medication adherence without causing undue psychological distress. Another important limitation is that the published PRSs apply to individuals of European ancestry. Equitable application of these clinical tools will require expanded GWAS in all other ethnic groups.

ARTICLE INFORMATION

Received June 15, 2019; accepted August 6, 2019.

Affiliation

From the Division of Cardiology, Atherogenomics Laboratory, Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, ON, Canada.

Sources of Funding

This study was funded by a Canadian Institutes of Health Research Foundation grant.

Disclosures

None.

REFERENCES

- McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488–1491. doi: 10.1126/science.1142447
- Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491–1493. doi: 10.1126/science.1142842
- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, et al; WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443–453. doi: 10.1056/NEJMoa072366
- Grånsbo K, Almgren P, Sjögren M, Smith JG, Engström G, Hedblad B, Melander O. Chromosome 9p21 genetic variation explains 13% of cardiovascular disease incidence but does not improve risk prediction. *J Intern Med*. 2013;274:233–240. doi: 10.1111/joim.12063
- Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, et al; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338. doi: 10.1038/ng.784
- Dandona S, Stewart AF, Chen L, Williams K, So D, O'Brien E, Glover C, Lemay M, Assogba O, Vo L, et al. Gene dosage of the common variant 9p21 predicts severity of coronary artery disease. *J Am Coll Cardiol*. 2010;56:479–486. doi: 10.1016/j.jacc.2009.10.092
- Chan K, Patel RS, Newcombe P, Nelson CP, Qasim A, Epstein SE, Burnett S, Vaccarino VL, Zafari AM, Shah SH, et al. Association between the chromosome 9p21 locus and angiographic coronary artery disease burden: a collaborative meta-analysis. *J Am Coll Cardiol*. 2013;61:957–970. doi: 10.1016/j.jacc.2012.10.051
- Ye S, Willeit J, Kronenberg F, Xu Q, Kiechl S. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based, prospective study. *J Am Coll Cardiol*. 2008;52:378–384. doi: 10.1016/j.jacc.2007.11.087
- Smith JG, Melander O, Lökvist H, Hedblad B, Engström G, Nilsson P, Carlson J, Berglund G, Norrving B, Lindgren A. Common genetic variants on chromosome 9p21 confers risk of ischemic stroke: a large-scale genetic association study. *Circ Cardiovasc Genet*. 2009;2:159–164. doi: 10.1161/CIRCGENETICS.108.835173
- Cluett C, Mc Dermott MM, Guralnik J, Ferrucci L, Bandinelli S, Miljkovic I, Zmuda JM, Li R, Tranah G, Harris T, et al. The 9p21 myocardial infarction risk allele increases risk of peripheral artery disease in older people. *Circ Cardiovasc Genet*. 2009;2:347–353. doi: 10.1161/CIRCGENETICS.108.825935
- Helgadottir A, Thorleifsson G, Magnusson KP, Grétarsdottir S, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankensteijn JD, Ronkainen A, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet*. 2008;40:217–224. doi: 10.1038/ng.72
- Patel RS, Schmidt AF, Tragante V, McCubrey RO, Holmes MV, Howe LJ, Direk K, Åkerblom A, Leander K, Virani SS, et al; CARDIoGRAMplusC4D Consortium. Association of chromosome 9p21 with subsequent coronary heart disease events. *Circ Genom Precis Med*. 2019;12:e002471. doi: 10.1161/CIRCGEN.119.002471
- Vilne B, Schunkert H. Integrating genes affecting coronary artery disease in functional networks by multi-OMICs approach. *Front Cardiovasc Med*. 2018;5:89. doi: 10.3389/fcvm.2018.00089
- Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, Zeng L, Ntalla I, Lai FY, Hopewell JC, et al; EPIC-CVD Consortium; CARDIoGRAMplusC4D; UK Biobank CardioMetabolic Consortium CHD Working Group. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49:1385–1391. doi: 10.1038/ng.3913
- van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122:433–443. doi: 10.1161/CIRCRESAHA.117.312086
- Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–1130. doi: 10.1038/ng.3396
- Pen A, Yam Y, Chen L, Dennie C, McPherson R, Chow BJ. Discordance between Framingham risk score and atherosclerotic plaque burden. *Eur Heart J*. 2013;34:1075–1082. doi: 10.1093/eurheartj/ehs473
- Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010;466:714–719. doi: 10.1038/nature09266
- Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, Buerki C, McLean BW, Cook RC, Parker JS, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol*. 2009;29:1671–1677. doi: 10.1161/ATVBAHA.109.189522
- Lo Sardo V, Chubukov P, Ferguson W, Kumar A, Teng EL, Duran M, Zhang L, Cost G, Engler AJ, Urnov F, et al. Unveiling the role of the most impactful cardiovascular risk locus through haplotype editing. *Cell*. 2018;175:1796–1810.e20. doi: 10.1016/j.cell.2018.11.014
- Douvrin A, Soubeyrand S, Naing T, Martinuk A, Nikpay M, Williams A, Buick J, Yauk C, McPherson R. Functional analysis of the TRIB1 associated locus linked to plasma triglycerides and coronary artery disease. *J Am Heart Assoc*. 2014;3:e000884. doi: 10.1161/JAHA.114.000884
- Turner AW, Nikpay M, Silva A, Lau P, Martinuk A, Linseman TA, Soubeyrand S, McPherson R. Functional interaction between COL4A1/COL4A2 and SMAD3 risk loci for coronary artery disease. *Atherosclerosis*. 2015;242:543–552. doi: 10.1016/j.atherosclerosis.2015.08.008
- Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015;16:197–212. doi: 10.1038/nrg3891
- Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*. 2013;368:20120362. doi: 10.1098/rstb.2012.0362
- Culverhouse RC, Suarez BK, Beckmann L, Chen P, Chen YS, Chiu YF, Chang-Claude J, Dempfle A, Hein R, Kazma R, et al. Gene by environment interactions. *Genet Epidemiol*. 2007;31(suppl 1):S68–S74. doi: 10.1002/gepi.20282

26. Cole CB, Nikpay M, McPherson R. Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol*. 2015;26:133–138. doi: 10.1097/MOL.0000000000000160
27. Saleheen D, Zhao W, Young R, Nelson CP, Ho W, Ferguson JF, Rasheed A, Ou K, Nurnberg ST, Bauer RC, et al. Loss of cardioprotective effects at the ADAMTS7 locus as a result of gene-smoking interactions. *Circulation*. 2017;135:2336–2353. doi: 10.1161/CIRCULATIONAHA.116.022069
28. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic risk score for coronary heart disease modifies the elevated risk by cigarette smoking for disease incidence. *Circ Genom Precis Med*. 2018;11:e001856. doi: 10.1161/CIRCGEN.117.001856
29. Cole CB, Nikpay M, Lau P, Stewart AF, Davies RW, Wells GA, Dent R, McPherson R. Adiposity significantly modifies genetic risk for dyslipidemia. *J Lipid Res*. 2014;55:2416–2422. doi: 10.1194/jlr.P052522
30. Jia P, Wang L, Fanous AH, Pato CN, Edwards TL, Zhao Z; International Schizophrenia Consortium. Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLoS Comput Biol*. 2012;8:e1002587. doi: 10.1371/journal.pcbi.1002587
31. Ghosh S, Vivar J, Nelson CP, Willenborg C, Segrè AV, Mäkinen VP, Nikpay M, Erdmann J, Blankenbregt S, O'Donnell C, et al. Systems genetics analysis of genome-wide association study reveals novel associations between key biological processes and coronary artery disease. *Arterioscler Thromb Vasc Biol*. 2015;35:1712–1722. doi: 10.1161/ATVBAHA.115.305513
32. Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res*. 2010;38:W90–W95. doi: 10.1093/nar/gkq324
33. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*. 2009;37:D619–D622. doi: 10.1093/nar/gkn863
34. Björkregren JLM, Kovacic JC, Dudley JT, Schadt EE. Genome-wide significant loci: how important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders. *J Am Coll Cardiol*. 2015;65:830–845. doi: 10.1016/j.jacc.2014.12.033
35. Schmidt EF, Strittmatter SM. The CRMP family of proteins and their role in Sema3A signaling. *Adv Exp Med Biol*. 2007;600:1–11. doi: 10.1007/978-0-387-70956-7_1
36. Mayhew AJ, Meyre D. Assessing the heritability of complex traits in humans: methodological challenges and opportunities. *Curr Genomics*. 2017;18:332–340. doi: 10.2174/1389202918666170307161450
37. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, Kaptoge S, Brozynska M, Wang T, et al; UK Biobank CardioMetabolic Consortium CHD Working Group. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol*. 2018;72:1883–1893. doi: 10.1016/j.jacc.2018.07.079
38. Nikpay M, Stewart AFR, McPherson R. Partitioning the heritability of coronary artery disease highlights the importance of immune-mediated processes and epigenetic sites associated with transcriptional activity. *Cardiovasc Res*. 2017;113:973–983. doi: 10.1093/cvr/cvx019
39. Mäkinen VP, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, Huan T, Segrè AV, Ghosh S, Vivar J, et al; Coronary Artery Disease Genome-Wide Replication and Meta-Analysis (CARDIoGRAM) Consortium. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet*. 2014;10:e1004502. doi: 10.1371/journal.pgen.1004502
40. Sarwar N, Butterworth AS, Freitag DF, Gregson J, Willeit P, Gorman DN, Gao P, Saleheen D, Rendon A, Nelson CP, et al. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet*. 2012;379:1205–1213.
41. Ridker PM. From C-reactive protein to interleukin-6 to interleukin-1: moving upstream to identify novel targets for atheroprotection. *Circ Res*. 2016;118:145–156. doi: 10.1161/CIRCRESAHA.115.306656
42. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, Fonseca F, Nicolau J, Koenig W, Anker SD, et al; CANTOS Trial Group. Anti-inflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med*. 2017;377:1119–1131. doi: 10.1056/NEJMoa1707914
43. Nikpay M, Beehler K, Valsesia A, Hager J, Harper ME, Dent R, McPherson R. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of Cardiometabolic phenotypes. *Cardiovasc Res*. 2019;115:1629–1645. doi: 10.1093/cvr/cvz030
44. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, Horenstein RB, Post W, McLenithan JC, Bielak LF, Peyser PA, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science*. 2008;322:1702–1705. doi: 10.1126/science.1161524
45. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobraira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, et al; Centers for Mendelian Genomics. The genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet*. 2015;97:199–215. doi: 10.1016/j.ajhg.2015.06.009
46. Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med*. 2010;363:2220–2227. doi: 10.1056/NEJMoa1002926
47. Dewey FE, Gusarova V, Dunbar RL, O'Dushlaine C, Schurmann C, Gottesman O, McCarthy S, Van Hout CV, Bruse S, Dansky HM, et al. Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease. *N Engl J Med*. 2017;377:211–221. doi: 10.1056/NEJMoa1612790
48. Graham MJ, Lee RG, Brandt TA, Tai LJ, Fu W, Peralta R, Yu R, Hurh E, Paz E, McEvoy BW, et al. Cardiovascular and metabolic effects of ANGPTL3 antisense oligonucleotides. *N Engl J Med*. 2017;377:222–232. doi: 10.1056/NEJMoa1701329
49. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, O'Dushlaine C, Van Hout CV, Staples J, Gonzaga-Jauregui C, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science*. 2016;354:aaf6814.
50. Dewey FE, Murray MF, Overton JD, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
51. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291. doi: 10.1038/nature19057
52. Do R, Stitzel NO, Won HH, Jørgensen AB, Duga S, Angelica Merlini P, Kiezun A, Farrall M, Goel A, Zuk O, et al; NHLBI Exome Sequencing Project. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2015;518:102–106. doi: 10.1038/nature13917
53. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitzel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, et al; TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22–31. doi: 10.1056/NEJMoa1307095
54. Stitzel NO, Won HH, Morrison AC, Peloso GM, Do R, Lange LA, Fontanillas P, Gupta N, Duga S, Goel A, et al; Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med*. 2014;371:2072–2082. doi: 10.1056/NEJMoa1405386
55. Sarwar N, Sandhu MS, Ricketts SL, Butterworth AS, Di Angelantonio E, Boehkholdt SM, Ouweland W, Watkins H, Samani NJ, Saleheen D, et al. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet*. 2010;375:1634–1639.
56. Nordestgaard BG, Varbo A. Triglycerides and cardiovascular disease. *Lancet*. 2014;384:626–635. doi: 10.1016/S0140-6736(14)61177-6
57. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, Darius H, Lewis BS, Ophuis TO, Jukema JW, et al; IMPROVE-IT Investigators. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372:2387–2397. doi: 10.1056/NEJMoa1410489
58. McPherson R, Hegele RA. Ezetimibe: rescued by randomization (clinical and mendelian). *Arterioscler Thromb Vasc Biol*. 2015;35:e13–e15. doi: 10.1161/ATVBAHA.114.305012
59. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, McConkey M, Gupta N, Gabriel S, Ardissino D, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N Engl J Med*. 2017;377:111–121. doi: 10.1056/NEJMoa1701719
60. Soubeyrand S, Martinuk A, Lau P, McPherson R. TRIB1 is regulated post-transcriptionally by proteasomal and non-proteasomal pathways. *PLoS One*. 2016;11:e0152346. doi: 10.1371/journal.pone.0152346
61. Soubeyrand S, Martinuk A, Naing T, Lau P, McPherson R. Role of Tribbles Pseudokinase 1 (TRIB1) in human hepatocyte metabolism. *Biochim Biophys Acta*. 2016;1862:223–232. doi: 10.1016/j.bbadis.2015.12.003
62. Soubeyrand S, Naing T, Martinuk A, McPherson R. ERK1/2 regulates hepatocyte Trib1 in response to mitochondrial dysfunction. *Biochim Biophys Acta*. 2013;1833:3405–3414. doi: 10.1016/j.bbamcr.2013.10.001
63. Turner AW, Martinuk A, Silva A, Lau P, Nikpay M, Eriksson P, Folkersen L, Perisic L, Hedin U, Soubeyrand S, et al. Functional analysis of a novel genome-wide association study signal in SMAD3 that confers protection from

- coronary artery disease. *Arterioscler Thromb Vasc Biol.* 2016;36:972–983. doi: 10.1161/ATVBAHA.116.307294
64. Soubeyrand S, Nikpay M, Turner A, Dang AT, Herfkens M, Lau P, McPherson R. Regulation of MFG8 by the intergenic coronary artery disease locus on 15q26.1. *Atherosclerosis.* 2019;284:11–17. doi: 10.1016/j.atherosclerosis.2019.02.012
 65. Doetschman T, Georgieva T. Gene editing with CRISPR/Cas9 RNA-directed nuclease. *Circ Res.* 2017;120:876–894. doi: 10.1161/CIRCRESAHA.116.309727
 66. Iyer D, Zhao Q, Wirka R, Naravane A, Nguyen T, Liu B, Nagao M, Cheng P, Miller CL, Kim JB, et al. Coronary artery disease genes SMAD3 and TCF21 promote opposing interactive genetic programs that regulate smooth muscle cell differentiation and disease risk. *PLoS Genet.* 2018;14:e1007681. doi: 10.1371/journal.pgen.1007681
 67. Linseman T, Soubeyrand S, Martinuk A, Nikpay M, Lau P, McPherson R. Functional validation of a common nonsynonymous coding variant in ZC3HC1 associated with protection from coronary artery disease. *Circ Cardiovasc Genet.* 2017;10:e001498.
 68. Evans DM, Davey Smith G. Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu Rev Genomics Hum Genet.* 2015;16:327–350. doi: 10.1146/annurev-genom-090314-050016
 69. Thompson A, Gao P, Orfei L, Watson S, Di Angelantonio E, Kaptoge S, Ballantyne C, Cannon CP, Criqui M, Cushman M, et al; Lp-PLA(2) Studies Collaboration. Lipoprotein-associated phospholipase A(2) and risk of coronary disease, stroke, and mortality: collaborative analysis of 32 prospective studies. *Lancet.* 2010;375:1536–1544. doi: 10.1016/S0140-6736(10)60319-4
 70. Casas JP, Ninio E, Panayiotou A, Palmen J, Cooper JA, Ricketts SL, Sofat R, Nicolaidis AN, Corsetti JP, Fowkes FG, et al. PLA2G7 genotype, lipoprotein-associated phospholipase A2 activity, and coronary heart disease risk in 10 494 cases and 15 624 controls of European Ancestry. *Circulation.* 2010;121:2284–2293. doi: 10.1161/CIRCULATIONAHA.109.923383
 71. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Silleesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med.* 2008;359:1897–1908. doi: 10.1056/NEJMoa0707402
 72. Nicholls SJ, Kastelein JJ, Schwartz GG, Bash D, Rosenson RS, Cavender MA, Brennan DM, Koenig W, Jukema JW, Nambi V, et al; VISTA-16 Investigators. Varespladib and cardiovascular events in patients with an acute coronary syndrome: the VISTA-16 randomized clinical trial. *JAMA.* 2014;311:252–262. doi: 10.1001/jama.2013.282836
 73. Davies RW, Dandona S, Stewart AF, Chen L, Ellis SG, Tang WH, Hazen SL, Roberts R, McPherson R, Wells GA. Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. *Circ Cardiovasc Genet.* 2010;3:468–474. doi: 10.1161/CIRCGENETICS.110.946269
 74. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50:1219–1224. doi: 10.1038/s41588-018-0183-z
 75. Khera AV, Chaffin M, Zekavat SM, Collins RL, Roselli C, Natarajan P, Lichtman JH, D'Onofrio G, Mathera J, Dreyer R, et al. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation.* 2019;139:1593–1602. doi: 10.1161/CIRCULATIONAHA.118.035658
 76. Lin J, Musunuru K. From genotype to phenotype: a primer on the functional follow-up of genome-wide association studies in cardiovascular disease. *Circ Genom Precis Med.* 2018;11:e001946.
 77. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, Blow MJ, Cohen JC, Rubin EM, Pennacchio LA. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature.* 2010;464:409–412. doi: 10.1038/nature08801