Abstract: Genome-wide association studies have provided a rich collection of ≈58 coronary artery disease (CAD) loci that suggest the existence of previously unsuspected new biology relevant to atherosclerosis. However, these studies only identify genomic loci associated with CAD, and many questions remain even after a genomic locus is definitively implicated, including the nature of the causal variant(s) and the causal gene(s), as well as the directionality of effect. There are several tools that can be used for investigation of the functional genomics of these loci, and progress has been made on a limited number of novel CAD loci. New biology regarding atherosclerosis and CAD will be learned through the functional genomics of these loci, and the hope is that at least some of these new pathways relevant to CAD pathogenesis will yield new therapeutic targets for the prevention and treatment of CAD. (Circ Res. 2016;118:586-606. DOI: 10.1161/CIRCRESAHA.115.306464.)

Key Words: atherosclerosis ■ coronary artery disease ■ functional ■ genome-wide ■ genomics
Efforts to perform unbiased discovery using the tools of human genetics to uncover novel pathways underlying complex diseases and traits have been pursued extensively over the last decade. In particular, genome-wide association studies (GWAS), reviewed elsewhere in this compendium, have capitalized on the millions of common single nucleotide polymorphisms (SNPs) to identify those SNPs that are genome-wide significantly associated with low-density lipoprotein (LDL) cholesterol, a known causal risk factor for CAD. The causal genes at these loci exert their effects through their expression in hepatocytes or enterocytes, consistent with their role in regulating LDL metabolism. Another locus, LPL, harbors the gene encoding the enzyme lipoprotein lipase, the most important regulator of triglyceride-rich lipoprotein metabolism. In addition, the APOA1/C3/A4/A5 locus associated with LDL-C is also associated with triglycerides and harbors 2 genes, APOC3 and APOA5, in which coding variants have been shown to be associated with both triglyceride levels and CAD. These and other observations have helped to confirm the causal role of triglyceride-rich lipoproteins in CAD. Another 5 loci are genome-wide significant for association with blood pressure, consistent with the causal role of elevated blood pressure in CAD. Also there is some overlap with GWAS studies for other vascular diseases, such as stroke. Interestingly, none of the 58 loci are associated with type 2 diabetes mellitus, raising interesting questions regarding the genetic overlap between type 2 diabetes mellitus and CAD and whether type 2 diabetes mellitus per se is causally related to CAD. Importantly, the majority of CAD GWAS loci is not associated with known risk factors for CAD (Table) and, thus, have the potential to provide novel insights into the biology and pathophysiology of CAD.

### Issues and Challenges for Functional Genomics of CAD GWAS Loci

Common variant GWAS studies only identify genomic loci associated with disease or trait. However, many questions remain even after a genomic locus is definitively implicated, including the nature of the causal variant(s) and the causal gene(s), as well as the directionality of effect. For the majority of the CAD GWAS loci, the answers to these questions are unknown, and after excluding the loci associated with lipids or blood pressure, virtually none of the remaining loci have answers to these fundamental questions. Identification of the causal variant is challenging because of linkage disequilibrium and the possibility that the variant(s) at a given locus with the lowest $P$ values for association with CAD may simply be proxies for the causal variant (see Figure 1). Furthermore, although the majority of variants with the lowest $P$ values fall in noncoding intergenic regions, they usually do not fall within a well-established *cis*-regulatory element, such as a known promoter, and thus challenge predictions of their impact on regulatory elements, like disruption of transcription factor binding or function of a long noncoding RNA (lncRNA). Below we discuss approaches to elucidating the causal variant at a GWAS locus.

Arguably the most important biological question to be addressed at each CAD GWAS locus is what the causal gene(s) at the locus are. By convention, GWAS loci are tabulated by the coding gene closest to the lead SNP with the lowest $P$ value. However, it is becoming clear that this approach does not always identify the causal gene. Because of chromatin looping that places regulatory enhancer elements in proximity to the promoters of genes that may be distant on the physical map, a causal variant may influence expression of distant genes (Figure 1). Furthermore, the causal gene at a GWAS locus need not necessarily be a protein-coding gene, but could be an lncRNA (eg, antisense ncRNA of the INK4 locus, an lncRNA at the 9p21 CAD locus), a microRNA, or some other transcribed or regulatory element. Finally, some loci may not have a single causal gene but in fact may be characterized by the coordinate regulation of several genes, potentially in different contributing cell types, that have additive effects on disease phenotype. Below we discuss in some detail the

<table>
<thead>
<tr>
<th>Nonstandard Abbreviations and Acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAV</td>
</tr>
<tr>
<td>Apo</td>
</tr>
<tr>
<td>ASE</td>
</tr>
<tr>
<td>ASO</td>
</tr>
<tr>
<td>CAD</td>
</tr>
<tr>
<td>Cas9</td>
</tr>
<tr>
<td>CESD</td>
</tr>
<tr>
<td>ChIP-Seq</td>
</tr>
<tr>
<td>CRISPR</td>
</tr>
<tr>
<td>eQTL</td>
</tr>
<tr>
<td>GWAS</td>
</tr>
<tr>
<td>HDR</td>
</tr>
<tr>
<td>hiPSC</td>
</tr>
<tr>
<td>LDL</td>
</tr>
<tr>
<td>LDL-C</td>
</tr>
<tr>
<td>lincRNA</td>
</tr>
<tr>
<td>IncRNA</td>
</tr>
<tr>
<td>LOF</td>
</tr>
<tr>
<td>MI</td>
</tr>
<tr>
<td>NHEJ</td>
</tr>
<tr>
<td>RISC</td>
</tr>
<tr>
<td>shRNA</td>
</tr>
<tr>
<td>siRNA</td>
</tr>
<tr>
<td>SNP</td>
</tr>
<tr>
<td>VLDL</td>
</tr>
</tbody>
</table>
Table. Overview of 58 Previously Reported Loci With Genome-Wide Significance for Coronary Artery Disease

<table>
<thead>
<tr>
<th>Reported Gene</th>
<th>First Reported Lead SNP</th>
<th>Chr</th>
<th>Location*</th>
<th>1000G P Value</th>
<th>New Lead SNP</th>
<th>Updated Location*</th>
<th>Updated 1000G P Value</th>
<th>Reported Phenotypes†</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCSK9</td>
<td>rs11206510</td>
<td>1</td>
<td>Intergenic</td>
<td>2.34E-08</td>
<td></td>
<td></td>
<td></td>
<td>MI, LDL-C</td>
</tr>
<tr>
<td>PPAR2B</td>
<td>rs17114036</td>
<td>1</td>
<td>PPAR2B</td>
<td>2.22E-13</td>
<td>rs9970807</td>
<td>PPAR2B</td>
<td>5.00E-14</td>
<td>MI</td>
</tr>
<tr>
<td>SORT1</td>
<td>rs646776</td>
<td>1</td>
<td>CELSR2</td>
<td>9.01E-19</td>
<td>rs7528419</td>
<td>CELSR2, PSRC</td>
<td>1.97E-23</td>
<td>MI, LDL-C; phospholipase A2</td>
</tr>
<tr>
<td>IL6R</td>
<td>rs4845625</td>
<td>1</td>
<td>IL6R</td>
<td>3.93E-08</td>
<td>rs6689306</td>
<td>IL6R</td>
<td>2.60E-09</td>
<td>MI</td>
</tr>
<tr>
<td>MIA3</td>
<td>rs17466857</td>
<td>1</td>
<td>TAF1A</td>
<td>4.18E-05</td>
<td>rs67180937</td>
<td>MIA3</td>
<td>1.01E-12</td>
<td>MI, Lp(a)</td>
</tr>
<tr>
<td>AK097927</td>
<td>rs16986953</td>
<td>2</td>
<td>Intergenic</td>
<td>1.45E-08</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>APOB</td>
<td>rs515135</td>
<td>2</td>
<td>Intergenic</td>
<td>3.09E-08</td>
<td>chr2:21378433:D</td>
<td>Intergenic</td>
<td>2.89E-08</td>
<td>MI, LDL-C</td>
</tr>
<tr>
<td>ABCG5-ABCG8</td>
<td>rs6544713</td>
<td>2</td>
<td>ABCG8</td>
<td>8.88E-07</td>
<td>chr2:44074126:D</td>
<td>ABCG8</td>
<td>2.60E-08</td>
<td>MI, LDL-C</td>
</tr>
<tr>
<td>VAMP5-VAMP8-GGCX</td>
<td>rs1561198</td>
<td>2</td>
<td>VAMP5, VAMP8</td>
<td>6.37E-10</td>
<td>rs7568458</td>
<td>GGCX, VAMP8</td>
<td>3.62E-10</td>
<td>MI</td>
</tr>
<tr>
<td>ZEB2-AC07493.1</td>
<td>rs2252641</td>
<td>2</td>
<td>TEX41</td>
<td>5.16E-04</td>
<td>rs17678683</td>
<td>LINC01412, ZEB2</td>
<td>3.00E-09</td>
<td>MI</td>
</tr>
<tr>
<td>WDR12</td>
<td>rs6725887</td>
<td>2</td>
<td>WDR12</td>
<td>9.51E-18</td>
<td>chr2:203828796:I</td>
<td>ALS2CR8</td>
<td>2.15E-18</td>
<td>MI</td>
</tr>
<tr>
<td>MRAS</td>
<td>rs9818870</td>
<td>3</td>
<td>MRAS</td>
<td>2.21E-06</td>
<td>chr3:138099161:I</td>
<td>MRAS</td>
<td>2.89E-09</td>
<td>MI</td>
</tr>
<tr>
<td>REST-NOA1</td>
<td>rs17087335</td>
<td>4</td>
<td>NOA1</td>
<td>4.6E-08</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>EDNRA</td>
<td>rs1878406</td>
<td>4</td>
<td>Intergenic</td>
<td>1.24E-06</td>
<td>rs4593108</td>
<td>Intergenic</td>
<td>8.82E-10</td>
<td>Plaque, stroke</td>
</tr>
<tr>
<td>GUCY1A3</td>
<td>rs7692387</td>
<td>4</td>
<td>GUCY1A3</td>
<td>7.35E-09</td>
<td>rs72689147</td>
<td>GUCY1A3</td>
<td>6.07E-09</td>
<td>MI</td>
</tr>
<tr>
<td>SLC22A4-SLC22A5</td>
<td>rs273909</td>
<td>5</td>
<td>SLC22A4, LOC555103</td>
<td>1.24E-04</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>ADTRP-C6orf105</td>
<td>rs6903956</td>
<td>6</td>
<td>ADTRP</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>PHACTR1</td>
<td>rs12526453</td>
<td>6</td>
<td>PHACTR1</td>
<td>2.14E-20</td>
<td>rs9349379</td>
<td>PHACTR1</td>
<td>1.81E-42</td>
<td>MI, calcification, artery dissection, migraine</td>
</tr>
<tr>
<td>ANKS1A</td>
<td>rs17609940</td>
<td>6</td>
<td>ANKS1A</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>KCNK5</td>
<td>rs10947789</td>
<td>6</td>
<td>KCNK5</td>
<td>1.63E-06</td>
<td>rs56336142</td>
<td>Intergenic</td>
<td>1.85E-08</td>
<td>MI</td>
</tr>
<tr>
<td>TCF21</td>
<td>rs12190287</td>
<td>6</td>
<td>TCF21, TARIID</td>
<td>1.07E-03</td>
<td>rs12200217</td>
<td>TARIID, LINC01312</td>
<td>1.98E-11</td>
<td>MI</td>
</tr>
<tr>
<td>SLC22A3-LPAL2-LPA</td>
<td>rs2048327</td>
<td>6</td>
<td>SLC22A3</td>
<td>2.46E-09</td>
<td>rs55730499</td>
<td>LPA</td>
<td>5.39E-39</td>
<td>MI</td>
</tr>
<tr>
<td>PLG</td>
<td>rs4252120</td>
<td>6</td>
<td>PLG</td>
<td>3.32E-03</td>
<td>rs4252185</td>
<td>PLG</td>
<td>1.64E-32</td>
<td>MI</td>
</tr>
<tr>
<td>HDAC9</td>
<td>rs2023938</td>
<td>7</td>
<td>HDAC9</td>
<td>1.36E-04</td>
<td>rs2107595</td>
<td>Intergenic</td>
<td>8.05E-11</td>
<td>Stroke</td>
</tr>
<tr>
<td>ZC3HC1</td>
<td>rs10953547</td>
<td>7</td>
<td>BCAP29</td>
<td>1.02E-05</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>NOS3</td>
<td>rs3918226</td>
<td>7</td>
<td>NOS3</td>
<td>1.7E-09</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>LPL</td>
<td>rs264</td>
<td>7</td>
<td>LPL</td>
<td>1.06E-05</td>
<td></td>
<td></td>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>TRIB1</td>
<td>rs2954029</td>
<td>8</td>
<td>RP11-136012.2</td>
<td>2.61E-06</td>
<td></td>
<td></td>
<td></td>
<td>MI, LDL-C, total cholesterol</td>
</tr>
<tr>
<td>9p21</td>
<td>rs3217992</td>
<td>8</td>
<td>CDKN2B, CDKN2B-AS1</td>
<td>1.03E-42</td>
<td>rs2891168</td>
<td>CDKN2B-AS1</td>
<td>2.29E-98</td>
<td>MI, venous thromboembolism</td>
</tr>
<tr>
<td>ABO</td>
<td>rs579459</td>
<td>9</td>
<td>ABO</td>
<td>1.14E-10</td>
<td>rs2519093</td>
<td>ABO</td>
<td>1.19E-11</td>
<td>MI</td>
</tr>
<tr>
<td>KIAA1462</td>
<td>rs2505083</td>
<td>9</td>
<td>KIAA1462</td>
<td>1.57E-10</td>
<td>rs2487928</td>
<td>KIAA1462</td>
<td>4.41E-11</td>
<td>MI</td>
</tr>
<tr>
<td>CXCL12</td>
<td>rs2047009</td>
<td>9</td>
<td>Intergenic</td>
<td>2.75E-11</td>
<td>rs1870634</td>
<td>Intergenic</td>
<td>5.55E-15</td>
<td>MI</td>
</tr>
<tr>
<td>LIPA</td>
<td>rs11203042</td>
<td>10</td>
<td>LIPA</td>
<td>1.22E-04</td>
<td>rs1412444</td>
<td>LIPA</td>
<td>5.15E-12</td>
<td>MI</td>
</tr>
<tr>
<td>CYP17A1-CNNM2-NT5C2</td>
<td>rs12413409</td>
<td>10</td>
<td>CNNM2</td>
<td>1.07E-07</td>
<td>rs11191146</td>
<td>PPN1P11</td>
<td>4.65E-09</td>
<td>MI, intracranial aneurysm</td>
</tr>
</tbody>
</table>

(Continued)
methodological approaches to solving these critical biological conundrums.

The directionality of effect at the locus is a critically important issue, particularly with regard to the question of whether the biology represented by that locus can be approached from a therapeutic targeting standpoint. For example, if the minor allele at a locus is associated with protection from CAD, it is essential to know whether the minor allele is associated with increased or decreased expression of the causal gene in the relevant cell type. Approaches such as expression quantitative trait loci (eQTL) and allele-specific expression (ASE) can be used to establish directionality of effect. However, the effects of many variants on differential gene expression are cell type–specific, and in most cases, we do not know with any confidence the relevant cell type for the genetic effect. Again, below we discuss the experimental approach to establishing directionality of effect given these challenges.

The Tools of Functional Genomics

There is large variety of experimental tools available to investigate the mechanism by which GWAS loci exert their effect on biological phenotype. They are tailored to the scale of interrogation, sampling on the genome-wide level, targeting a gene, a genomic region, or even a specific SNP of interest. Figure 2 gives an overview of the experimental techniques which will be discussed in detail later.
In contrast, several different histone modifications have been shown to sample distinct fractions of the human genome, such as promoters (Histone 3 Lysine 4 trimethylation), transcribed genes (Histone 3 Lysine 36 trimethylation), primed (Histone 3 Lysine 4 monomethylation), and active (Histone 3 Lysine 27 acetylation) enhancers. Chromatin immunoprecipitation with massively parallel sequencing (ChIP-Seq) is the method of choice to identify these chromatin marks. This technique can also be used to identify the binding events of specific transcription factors of interest.

Active enhancers are identified with the highest confidence when data sets are intersected, for example, by combining Histone 3 Lysine 27 acetylation abundance with Polymerase II–binding patterns from ChIP-seq, as well as Gro-Seq. The latter technique maps actively transcribed RNAs from enhancers through the isolation of newly synthesized RNA from nuclei incubated with bromouridine. Large-scale consortium data for histone marks and transcription factor binding is now publicly available from the ENCODE (Encyclopedia of DNA Elements) project (https://genome.ucsc.edu/ENCODE/dataMatrix/encodeChipMatrixHuman.html). The experimental data stems from in vitro cultured cells, many of them immortalized human cell lines. Of particular interest for the annotation of CAD GWAS loci are data generated from HepG2 cells (a hepatocellular carcinoma line that serves as a proxy for hepatocytes), endothelial HUVECs (human umbilical vein endothelial cells), CD14+ monocytes, and aortic smooth muscle cells. Similarly, the Roadmap Epigenomics Project (www.roadmapepigenomics.org/data/tables/adult) compiled a large portfolio of histone marks from human tissues, including data from liver, aorta, and primary CD14+ monocytes.

In addition to revealing transcription factor occupancy and histone modification at sites of interest, ChIP-Seq data can also be interrogated for allelic imbalance on an individual basis, if paired with corresponding genotype information. In this case, sequencing reads from ChIP-Seq experiments are assigned to either the maternal or paternal allele. A significant difference in coverage on either allele is indicative for sequence-dependent differential transcription factor binding or histone abundance in this region and provides direct evidence for the functional role of a polymorphic site.

A second step in elucidating the mechanism by which noncoding polymorphisms affect disease risk lies in the analysis of transcriptome data from disease-relevant cells and tissues. Whole-transcriptome RNA-Seq data provides useful information on the transcriptional repertoire of a cell of interest. No meaningful association of regulatory variants with disease risk genes is possible without the knowledge of the cell’s transcriptional machinery in single cells. Future applications may allow simultaneous derivation of both data sets from the same cell, which would greatly reduce the information lost because of sampling heterogeneous pools of cells at different cell cycle and developmental stages. This could be particularly useful for the characterization of the diverse cell populations within the atherosclerotic neointima.

An important consideration in functional transcriptomic analysis is that because both transcriptional regulation and gene expression are tissue-specific processes, particular attention should be paid to the suitability of cell type from which the data are generated. Early data sets have depended...
heavily on samples generated in cell culture for practical reasons, namely the accessibility of long-established cell lines, and the ability to generate homogeneous material, reproducibly, and in large scale. However, all cultured cells, whether they be immortalized or cancer cell lines, induced pluripotent stem cell–derived cells, or cultured primary cells, display to some extent rather an immature, precursor-like phenotype in comparison to the corresponding fully differentiated primary cell in vivo. Additionally, cancer cell lines, in particular, often carry genomic rearrangements, leading to artifacts that are not representative of their cell type of origin. On the other hand, primary ex vivo tissues, although displaying the most authentic transcriptional profiles, are often comprised of multiple cell types which complicates analysis and can mask subtle effects within data noise or because of numerically underrepresented cell types. The best approach when dealing with tissues of mixed cell types, such as coronary arteries, may be to focus on effects also observed in in vitro cultured pure cell populations.

Furthermore, primary tissues need to be extensively phenotyped as to their disease status to enable identification of changes within the transcriptome between the healthy and diseased state. Additionally, the exact tissue origin can play an important role. For example, aortic, coronary, or femoral artery transcriptomes, although often similar, can show distinct differences in their expression profiles, which may be based on their differing tissue environments or developmental origins. Differences in gene expression are well established along the different section of the aorta itself and follow embryological and hemodynamic patterns.

For meaningful correlation of ChIP-Seq and RNA-Seq data to identify SNP-to-gene interactions, data should be generated from the same cell type, if possible even from the same individual culture. In case of primary tissue, combined data sets from the same individual are particularly informative. The association of cis-regulatory regions with one or multiple nearby regulated targeted gene or genes is, however, problematic. Frequently used distance-based methods do not adequately reflect the true biology. More meaningful approaches to directly link SNP genotype with gene expression levels include eQTL and ASE studies.

eQTLs and ASE

QTL are polymorphic sites within a genome which show significant association with a quantitative trait, such as plasma...
lipid levels,7 carotid intima-media thickness,8 or gene expression levels (eQTL).46,47 eQTL studies combine genotype with gene expression level information—the latter usually assessed using expression microarrays or RNA-seq—and can detect both local cis-effects as well as distal trans-effects of regulatory elements on gene expression. eQTL data has been generated from multiple human cells and tissues.38–40 Of particular relevance is data derived from liver, which has proven especially successful for loci involved in dyslipidemia, as well as data from peripheral blood monocytes51 and in vitro cultured human aortic endothelial cells (ECs).52

eQTL studies require sampling from several hundreds of individuals to identify the majority of loci with statistical significance because interindividual noise arises from differences in genetic background, host factors, such as age or sex, as well as environmental factors like diet and lifestyle. The large sample sizes needed for eQTL analysis are particularly problematic when interrogating tissues that are rare or difficult to sample. Additional challenges, such as insufficient platform coverage and batch variation, stem from the use of microarrays to determine gene expression levels. These issues can be circumvented in large part by the use of RNA-Seq as a data source, which greatly increases statistical power, and is more compatible with meta-analysis.

eQTL studies have been successfully used to identify functional SNPs and directionality from GWAS studies for a variety of diseases because of changes in the expression of their downstream putative risk genes.31 Because CAD is a complex disease involving multiple tissues, recent studies have aimed at integrating expression data from several tissues for a more comprehensive annotation of CAD GWAS loci.54,55 However, the portfolio of tissues used included several human tissues not relevant to CAD, which may have introduced bias in the SNP-to-gene association process. Fortunately, the GTex consortium is generating publicly available RNA-Seq data from currently under represented CAD-relevant tissues, including coronary artery and aorta.

ASE has emerged as an alternative to eQTL analysis for linking genetic variation in cis-regulatory regions to gene expression.56–59 It is based on the identification of allelic imbalance, showing differences in gene expression levels between the 2 alleles in a single heterozygous individual. The within-person allelic analyses greatly reduces impact of interindividual variation from environmental and genetic trans effects, thus enhancing statistical power. ASE requires allele-specific transcriptomic data, such as stranded RNA-Seq data, but in contrast to eQTL analysis, a relatively small number of samples that carry the same heterozygous site of interest are needed. The information value of each sample is limited by the number of heterozygous sites of its genome. Additionally, the data sets that are generated need to be of high read coverage to ensure presence of multiple reads at interrogated sites of interest. With regards to CAD, to date, ASE has been applied to assess general principles of the regulation of gene expression in mouse liver.60,61

Additionally, Chromosome Conformation Capture can be used as a complementary technique to associate cis-regulatory regions with their target genes. This method captures the physical interaction between 2 genomic regions, such as enhancers with promoters, and has recently been applied to study human liver and aorta.62 Several different variations of the technique are in use, with Hi-C and 5C as genome-wide methods. A major obstacle is, however, their low resolution of tens to hundreds of kilobases. Targeted approaches, such as Capture-C63 and Capture Hi-C,64 that can interrogate hundreds of select loci simultaneously are displaying higher resolutions of down to 1 kilobase (kb), and single cell approaches are starting to emerge.65 The improved availability of tissues relevant to CAD and increasingly sensitive methods at hand to link association SNPs to downstream risk genes suggest that a comprehensive reannotation of all 58 known CAD association loci in disease-relevant cells and tissues using the latest methods may be worthwhile.

Noncoding RNAs

A substantial proportion of trait-associated SNPs identified by GWAS lie outside of protein coding regions and map to the noncoding intervals.66 However, the mechanistic relationship of trait-associated SNPs with the noncoding functional genome is poorly understood. Because protein-coding genes account for only a small proportion of the transcribed human genome, noncoding (nc) RNA are now emerging as alternative functional genomic elements underlying GWAS hits. Along with microRNAs, IncRNAs, ncRNAs defined as transcripts >200 nucleotides (nt) in length, are emerging as important regulators involved in cancer as well as in neurological, cardiovascular, developmental, and other human diseases,67 highlighting the need to investigate the possible contributions of variations in ncRNAs to human diseases.

Systematic analyses are emerging that evaluate the potential association of regulatory ncRNAs with complex traits. miRNAs comprise a class of short (20–24 nt) regulatory RNAs that modulate mRNA translation and turnover. A recent study leveraged GWAS meta-analysis in >188,000 individuals to identify 69 miRNAs located in genomic regions associated with abnormal blood lipid levels.68 The work identified 4 miRNAs (miR-128-1, miR-148a, miR-130b, and miR-301b) that are associated with LDL-C uptake and cholesterol efflux by possibly controlling the expression of the LDL receptor and the ATP-binding cassette transporter A1, respectively. miR-QTL studies using liver tissue from 424 morbidly obese individuals revealed an association of miR-128-1 and miR-148a expression with SNPs linked to abnormal human blood lipid levels, suggesting the relevance of these miRNAs identified by GWAS to human cardiometabolic disorders.69

A subset of IncRNAs, intergenic IncRNAs (lincRNAs), represents a rapidly evolving catalog of IncRNA species that does not overlap with exons of protein-coding genes.69 Several studies examined the implication of lincRNAs in complex diseases based on GWAS (LinPoly70 and LineSNP71). These studies represent initial efforts to integrate disease-associated SNPs and human lincRNAs, but both data sets do not include complete GWAS SNP data and focus on only a few thousand lineRNAs. A recent study identified 495,729 and 777,095 SNPs in >30,000 lincRNA transcripts in human and mouse, respectively. A large number of SNPs were predicted to impact the lincRNA secondary structure and modulate lincRNA–miRNA
interactions. By mapping these SNPs to GWAS results, 142 human lncRNA SNPs are GWAS tagSNPs and 197,827 lncRNA SNPs are within these GWAS linkage disequilibrium regions.72 Kumar et al examined the association of SNPs with expression of lincRNAs in human blood and identified 112 cis-regulated lincRNAs. A considerable number of the observed lincRNA cis-eQTLs had disease or trait associations,73 suggesting that intergenic GWAS-associated SNPs may act by modulating expression of specific lincRNAs.

Although functional roles of most lincRNAs remain elusive, mechanistic insights into distinct nuclear and cytoplasmic actions69,74,75 for a small number of well-studied lincRNAs strongly suggest that some lincRNAs play major regulatory roles in a variety of cellular processes, such as X chromosome inactivation,28 embryogenesis,77 cell pluripotency,78 cell development, and differentiation.79 Depending on their subcellular localization, lincRNAs can mediate gene expression through distinct mechanisms. In the nucleus, they are involved in co-transcriptional regulation, recruitment of proteins complexes to specific loci for cis or trans regulation of gene expression, or scaffolding of nuclear complexes.69,74 In the cytoplasm, lincRNAs can function as competitive endogenous RNAs that bind miRNAs and inhibit their activity, pair with mRNAs to trigger post-transcriptional regulation, or interact with target proteins to modulate their function.69,74 LincRNAs are increasingly implicated in human diseases, including cancer,80 neurological diseases,81 and cardiovascular disorders,82–88 and also modulate physiology and pathophysiology in cells relevant to cardiometabolic disease. For example, cardiac lincRNAs like Braveheart, CHRF, and Mhrt regulate cardiomyocyte differentiation and cardiac hypertrophy.82,85,86 LincRNAs in smooth muscle and ECs (MALAT, linc-p21) regulate proliferation.87,88 A few lincRNAs have been implicated in macrophage functions: lincRNA-Cox2 in mouse represses the basal expression of interferon-stimulated genes by partnering with the heterogeneous nuclear ribonucleoproteins hnRNPA/B and hnRNPA2/B1,89 and a human monocytic THP-1 lincRNA called tumor necrosis factor and hnRNPL-related immunoregulatory lincRNA regulates expression of tumor necrosis factor through its interactions with hnRNPL.90 Conserved adipose lincRNAs, such as Fmr1 and Blnc1,91–93 and species-specific lincRNAs, such as Inc-BATE1 and ADINR,94,95 regulate adipogenesis of white and brown adipocytes in mouse and human.

Several studies have tried to identify and validate the causal GWAS variants that regulate lincRNA expression and function. A well-known example on chromosome 9p21 that encompasses an antisense lincRNA, antisense ncRNA of the INK4 locus, has been significantly associated with susceptibility to coronary disease, as well as abdominal aortic and intracranial aneurysms.96 Some associated SNPs in this region have been shown to alter the transcription and processing of antisense ncRNA of the INK4 locus transcripts.97

To further uncover the effects of GWAS-associated genetic variants on ncRNAs function, future studies are needed to (1) define a comprehensive genome-wide set of human lncRNAs across all disease-relevant tissues. Because lncRNAs have lower and more tissue-specific expression patterns than mRNAs, this requires deeper RNA sequencing of multiple human cells and tissue than is currently available in GENCODE98 or the human bodymap99; (2) investigate ncRNA enrichment and association in targeted yet genome-wide approaches within catalogued GWAS and particularly whole genome sequencing projects as they emerge, for example the National Heart, Lung and Blood Institute’s Trans-Omics for Precision Medicine (TOPMed) Program (https://www.nhlbiwgs.org); (3) establish bioinformatics pipelines to prioritize trait-associated ncRNAs (systematically using synteny and conservation, tissue expression, eQTL, and ASE and ChIP-Seq promoter, enhancer and transcription factors marks at trait-associated lncRNAs); and (4) establish high-throughput pipelines for human-relevant functional follow-up of trait-associated ncRNAs in disease-relevant cell types and in vivo animal models.

Taken together, ncRNAs, such as miRNAs and lncRNAs, may be important for interpreting GWAS data and may in many cases act as the causal genomic element in contributing to human cardiometabolic diseases. The therapeutic tractability of potent and specific antisense technologies targeting single or multiple ncRNAs implicated in human cardiometabolic diseases may thus have important clinical ramifications for the treatment of these diseases.

### Genome Editing

The putative functional variants identified by fine mapping overlapping with regulatory marks, eQTL and ASE analysis, and other bioinformatic approaches require further experimental validation to establish causality. Genome engineering and human-induced pluripotent stem cells (hiPSCs), when combined, represent powerful tools to accelerate GWAS-driven functional validation of causal variants at trait-associated loci.

Genome engineering represents strategies and techniques developed in recent years for the targeted modification of the genetic information. The type II clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9), an RNA-guided nuclease, are based on a bacterial system that has been modified for genome engineering in mammalian cells.100,101 Because of its ease of adaptability and improved efficiency, CRISPR/Cas9 has rapidly become one of the most popular approaches for genome engineering.102 Consisting of Cas9 and a short guide RNA, it generates site-specific DNA breaks, which are repaired by either nonhomologous end joining (NHEJ), creating insertions or deletions at the site of the break, or homology-directed repair (HDR) by precise change of a genomic sequence using an exogenously introduced donor template.103 In addition to the disruption of genomic sequence through nucleases, the CRISPR-associated catalytically inactive Cas9 protein, termed dCas9, can be fused to repressor104 or activator domains,105–107 named inhibition of gene transcription using the CRISPR/Cas9 technology104 and activation of gene transcription using the CRISPR/Cas9 technology,104–106 respectively. Such modified CRISPR/dCas9-fusion proteins, together with guide RNA, can then be introduced to control the expression or activity of candidate gene or regulatory elements that harbor GWAS signals.107

Although the CRISPR/Cas9 system is highly efficient in human cell lines, gene editing in primary human cells is challenging. Human iPSCs have the potential to be differentiated...
to all adult cell types, including rare or inaccessible human cell populations, for reliable disease modeling. By generating clonal lines carrying desired genetic modification introduced by CRISPR/Cas9 and then differentiating to somatic cells relevant to atherosclerosis (eg, hepatocytes, macrophages, smooth muscle cells, ECs, cardiomyocytes), genome editing in hiPSCs provides a unique platform for functional validation of GWAS CAD loci across multiple disease-relevant cell types. Applications include (1) double strand break repaired by NHEJ pathway may introduce frame-shift mutation or early stop codon, which, if in a critical coding exon, likely causes nonsense-mediated decay of the mRNA and effectively eliminates gene function for loss of function (LOF) studies; (2) by HDR-mediated precise nucleotide alteration using a donor template, it is feasible to generate hiPSC lines in which the disease-associated SNP is the sole experimental variable, thereby investigating the causal role of genetic variants—for instance, the targeted alteration of specific transcription factor–binding site motifs in otherwise intact loci could reveal the functional contribution of transcription factor binding to the function of a regulatory element; (3) the unique multiplexing capabilities of the CRISPR/Cas9 system facilitate the deletion of a large stretch of genomic DNA, enabling the functional interrogation of noncoding regulatory elements and noncoding transcripts. In addition, because most disease-associated SNPs confer only modest risk, the relevance of multiple monallelic and biallelic combination can be addressed by multiplexing of CRISPR/Cas9 gene editing.

One example of such an interrogation investigates the role of an intronic variant (rs9349379) in the CAD/MI риск locus PHACTR1 prioritized by genetic fine-mapping and eQTL in human coronary arteries. The study using EC extracts first showed that alleles at rs9349379 are differentially bound by the transcription factors myocyte enhancer factor-2. The deletion of this myocyte enhancer factor-2–binding site using CRISPR/Cas9 in hiPSC and subsequent differentiation to ECs then revealed that heterozygous ECs carrying the deletion express 35% less PHACTR1 transcript.108

Despite being powerful, precise editing of human genomes in pluripotent stem cells by HDR of targeted nuclease-induced cleavage has been hindered by the low efficiency of HDR over NHEJ, making the screening of clones containing desired genotypes time-consuming and labor-intensive. However, great strides have been made to improve the efficiency of HDR. A Cas9D10A mutant functioning as a nickase yields similar strides have been made to improve the efficiency of HDR.

One example of such an interrogation investigates the role of an intronic variant (rs9349379) in the CAD/MI risk locus PHACTR1 prioritized by genetic fine-mapping and eQTL in human coronary arteries. The study using EC extracts first showed that alleles at rs9349379 are differentially bound by the transcription factors myocyte enhancer factor-2. The deletion of this myocyte enhancer factor-2–binding site using CRISPR/Cas9 in hiPSC and subsequent differentiation to ECs then revealed that heterozygous ECs carrying the deletion express 35% less PHACTR1 transcript.108

Despite being powerful, precise editing of human genomes in pluripotent stem cells by HDR of targeted nuclease-induced cleavage has been hindered by the low efficiency of HDR over NHEJ, making the screening of clones containing desired genotypes time-consuming and labor-intensive. However, great strides have been made to improve the efficiency of HDR. A Cas9D10A mutant functioning as a nickase yields similar effects. These approaches are, therefore, not a substitute for subsequent genetic validation of promising putative causal genes, but rather a way to quickly prioritize candidate genes for further study.

Future studies need to (1) further improve the efficiency and reduce the cost of hiPSC differentiation; (2) optimize differentiation protocols to produce mature cells phenotypically, functionally, and transcriptomically highly similar to primary somatic cells; (3) improve the efficiency of HDR-mediated precise nucleotide alteration over NHEJ; (4) apply inhibition of gene transcription using the CRISPR/Cas9 technology and activation of gene transcription using the CRISPR/Cas9 technology in hiPSC for dynamic and precise control of expression of individual transcripts in hiPSC and differentiated cells; (5) adapt conditional knockout for the assessment of gene function in different lineages of differentiation; (6) establish more advanced techniques to facilitate rapid screening of rare iPS clones carrying the desired genotypes and thoroughly evaluate potential off-target effects.

In summary, facile high efficiency genome editing coupled with hiPSC differentiation can pave the way for functional interrogation of GWAS variants and loci of complex non-Mendelian diseases, such as CAD, and can help delineate human genotype–phenotype relationship in human cellular disease models and, potentially, in genetically modified mice carrying mutations, reporter, or conditional alleles for in vivo modeling using CRISPR/Cas9-mediated genome engineering.116,117

**Somatic Gene Targeting (Small Interfering RNA, Antisense Oligonucleotide, and Adeno-Associated Virus)**

As mentioned earlier, transgenic mouse models that are genetically predisposed to develop atherosclerosis because of partial or complete loss of apolipoprotein E (ApoE) or Ldlr function or that have been modified to have plasma lipid profiles which more closely reflect human biology are invaluable in the functional study of candidate GWAS genes. Mouse lines with humanized lipid profiles include the Apobec knockout and the human ApoB(100) transgenic mouse models, which have been combined with haploinsufficient Ldlr deficiency in the LAhB-H mouse strain.118 Genetic ablation remains the gold standard for the characterization of gene function: combining conditional approaches and tissue-specific Cre drivers allow precise interrogation of the potential role of a gene to the phenotype of interest. On the other hand, even with the advent of facile CRISPR/Cas9 genome editing, the development of genetic models is laborious, expensive, and time-consuming.

Alternatives to genetic approaches include the use of small interfering RNA (siRNA) or antisense oligonucleotide (ASO) inhibitors. In addition, adeno-associated virus (AAV) platforms can be used for either overexpression or for permanent loss of function by expression of short hairpin RNAs (shRNA). Combining these approaches with the existing genetic atherosclerosis models is a potent way to accelerate GWAS functional analysis, but compared with more rigorous genetic approaches involves some compromises and limitations in interpretability. These approaches are, therefore, not a substitute for subsequent genetic validation of promising putative causal genes, but rather a way to quickly prioritize candidate genes for further study.
ASO and siRNA have in common that they are systemically delivered, modified nucleic acids that target the gene of interest through complementary base-pairing between their primary sequences and those of their target transcripts. Despite these similarities, the 2 forms differ in their mechanism of action. In the case of siRNA, the technology takes advantage of the RNA-direct RNA endonuclease activity of Argonaute2, the miRNA binding component of the RNA-interference silencing complex (RISC). The cytoplasmic RNAse III endonuclease, Dicer1, cleaves double-stranded or short-hairpin RNA and concomitantly loads one strand of the circa 22 basepair cleavage product into a binding cleft in Argonaute2. The solvent-exposed bases of the Ago-loaded RNA serve to target RISC to complementary sequences.\textsuperscript{119} Endogenous microRNAs in mammals have imperfect complementarity with their targets: downregulation occurs by a combination of miRNA destabilization because of the recruitment of decapping and deadenylation factors and to the inhibition of translational initiation.\textsuperscript{120} In contrast, artificial shRNA and siRNA systems take advantage of an evolutionary remnant activity of Argonaute2, which cleaves the paired target strand where perfect complementarity exists between itself and the loaded RNA.\textsuperscript{121} The interfering RNA serves only as a targeting component and is not cleaved itself. The 2 types of RNA interference most significantly differ in their entry points to the endogenous system: although exogenous siRNAs can be transfected directly in cell culture, the shRNA are supplied as transgenes (usually along with a reporter gene, such as enhanced Green Fluorescent Protein). As such, the shRNA genes must be transcribed in the nucleus and the hairpin RNAs exported to the cytoplasm, cleaved by Dicer and loaded into RISC. The advantage of the shRNA approach is that candidate inhibitory RNAs can be validated in cell culture and then readily adapted to viral or transgenic applications in vivo. However, the nuclear export protein, Exportin-5, which shuttles shRNA to the cytoplasm has been shown to be limiting in the biogenesis of shRNA, leading to initial limitations of this approach.\textsuperscript{122} Several subsequent innovations have alleviated these concerns, which were partly because of saturation of the miRNA biogenesis pathway, but also to off-target effects of the passenger strands of the shRNAs.\textsuperscript{123–125}

In contrast to shRNA, siRNA bypasses the requirement for nuclear transcription and export: on entry into the cytoplasm, the siRNA is rapidly loaded into RISC and interference begins. Because naked RNA has an extremely low half-life in plasma and because endocytosed RNA is targeted to the lysosome and degraded, a variety of strategies have been developed to evade these obstacles to in vivo use.\textsuperscript{126} In addition to its usefulness in basic research, RNA interference has already shown therapeutic potential in cardiovascular disease (CVD), for example, by targeting PCSK9.\textsuperscript{127}

ASOs are also short nucleic acids, but do not depend on the RISC complex for their action. In the context of functional analysis of GWAS hits, ASOs may be deployed in 3 ways: (1) targeted to a transcript of a protein-coding gene to interfere with the initiation of translation; (2) targeted to intron-containing genes to block splicing, and (3) targeted against microRNAs to block their inhibitory effects on target genes. In each of these cases, pairing of the ASO with its target physically precludes the interaction of the target RNA with another molecule, elongation initiation factors, splicing factors, or target mRNAs, respectively. In addition, the specialized class of ASOs termed gapners are designed with a central stretch of unmodified DNA nucleotides, which when base-paired to a complementary RNA target yields a heteroduplex that is recognized as a substrate by the ubiquitous intracellular ribonuclease, RNase H1. The resulting cleavage of the RNA strand of the heteroduplex by RNase H1 is analogous to that of siRNA, albeit by a completely different mechanism, and induces rapid turnover of the cleaved RNA. It should be noted that mipomersen, a therapeutic ASO targeting apoB, which is approved for the treatment of homozygous familial hypercholesterolemia, is based on a gapmer strategy.\textsuperscript{128}

The advantage of RNA-targeting strategies as an approach, irrespective of the precise mechanism, is that they are relatively straightforward in design and validation. There are, however, key limitations of this approach. Systemic delivery of nucleic acids has been shown to induce an inflammatory response—for any given inhibitor, there is a possibility of off-target effects of the artificial RNA, and the biodistribution of the oligonucleotides can be influenced slightly, but not tightly controlled. The inflammatory side effects have been largely mitigated by successive innovations in the chemistry of the synthetic nucleic acids used, and concerns over off-target effects can be addressed by the separate use of independent ASOs targeting a given gene. The limitations concerning the delivery and distribution are harder to address: the fact that most oligonucleotides end up by default in the liver (predominantly in hepatocytes) is less of a concern in the functional analysis of CVD GWAS candidate risk genes that are liver-expressed than it might be in other biological contexts. In addition, both siRNA and ASO approaches only permit gene knockdown but not upregulation.

AAV vectors provide a tractable system to perform the reciprocal experiment: overexpression of candidate causal genes for functional analysis in vivo.\textsuperscript{129} As noted earlier, they can also be adapted for corresponding loss of function experiments using shRNA expression cassettes. As the name suggests, AAV was identified as a coinfecting parvovirus with adenovirus. To date, AAV has no identified role in any human disease and does not replicate in the absence of adenovirus. These characteristics have made it an attractive platform for candidate gene therapy development, which has been a boon for parallel uses in basic science. AAV induces minimal immune response compared with other viral vectors commonly used for somatic gene expression (specifically adenovirus and lentivirus). An additional advantage is the availability of multiple serotypes with varying tropisms, which addresses a significant shortfall in the application of system siRNA/ASOs: combining serotype-limited tropism with tissue-restricted promoters allows a significantly nuanced expression of genes or shRNAs that target them. Furthermore, significant efforts have been made to improve and refine the naturally isolated serotypes by repeated rounds of in vivo selection and expansion, suggesting that even more tissue-selective versions will be available in the near future.\textsuperscript{130} A significant validation of this platform is that
AAV-driven expression of PCSK9 when coupled with a dietary stress (high fat diet) induces atherosclerosis in mice.\textsuperscript{131}

Hepatocyte-expressed genes are particularly amenable to study using AAV: AAV8 transduction rates of hepatocytes are high, and the relatively quiescent nature of the adult liver allows expression to be maintained for many months. One significant limitation is, however, that the packaging capacity of AAV is not large: \(\approx 4.8\ \text{kb}\) for AAV and \(\approx 2.4\ \text{kb}\) for its self-complementarity derivative. Although this precludes the use of AAV for some genes, the platform is nonetheless powerful and widely used. To date, clinical AAV gene therapy has been approved only in Europe and only for the treatment of familial lipoprotein lipase deficiency,\textsuperscript{132} but homozygous familial hypercholesterolemia is a CVD-relevant condition that is an excellent candidate for an AAV-based therapeutic.\textsuperscript{133}

**Examples of Functional Genomics at Selected CAD GWAS Loci**

Recent progress has been made in understanding the biology underlying some of the genes which have been implicated in risk for CAD. Being expressed in distinct cell types involved in atherosclerosis (Figure 3), these genes exert their effects in a cell type–specific manner, which determines their specific contributions to disease. We provide here 5 selected examples of CAD GWAS loci for which the tools described above have been variably used to identify the causal gene at the locus and probe the underlying biology linking the gene to CAD.

**SORT1 (Sortilin)**

A compelling and now widely replicated novel locus associated with plasma lipid traits is the chromosome 1p13 locus, which had the lowest \(P\) value of association in the Global Lipids Genetics Consortium study.\textsuperscript{19} Notably, this locus had been independently and genome-wide significantly associated with MI/CAD, suggesting that it is of high importance to human cardiovascular health.\textsuperscript{134–136} The locus harbors a high density of genes that might plausibly contribute to the phenotype, which necessitated thorough functional analysis. Fine-scale mapping of the locus refined the signal to a 6.1 kb genomic region containing 6 SNPs in high linkage disequilibrium. Cloning of this region into a luciferase reporter construct and the separate replacement of each SNP with the corresponding minor allele variant identified rs12740374 as the causal SNP. The mechanism by which it exerts its effect is because of the creation of a novel CCAAT/enhancer binding protein alpha (C/EBPa)–binding site, which was functionally validated by gel shift assays.\textsuperscript{136} Nonetheless, the causal gene remained ambiguous: \textit{SORT1} and \textit{PSRC1} both had strong eQTLs in liver, and a priori, neither gene could be eliminated as causal. To address this, the genes were separately overexpressed using the hepatocyte-tropic AAV8 system. \textit{SORT1} overexpression, but not that of \textit{PSRC1}, substantially decreased plasma LDL-C in a mouse model with a humanized lipid profile (LahB, as described earlier), identifying \textit{SORT1} as the causal gene.\textsuperscript{136}

Sortilin is a type I transmembrane multiligand receptor that is synthesized in the endoplasmatic reticulum as a propeptide and is further processed to an active, mature form in the Golgi. It localizes to both the Golgi and plasma membranes and facilitates trafficking of a variety of proteins bidirectionally between the Golgi lumen and the extracellular environment. Sortilin can also facilitate protein degradation by shuttling proteins from the Golgi to the lysosome through the endolysosome. Preliminary characterization of the role of sortilin in regulating very low–density lipoprotein (VLDL) secretion was performed through a series of \textit{Sort1} overexpression studies in hepatocytes and hepatocyte-like cell lines and in a variety of mouse models. \textit{Sort1} expression was shown both to decrease VLDL secretion rates and increase plasma LDL turnover, thereby reducing plasma cholesterol additively.\textsuperscript{137}

**Figure 3.** Coronary heart disease (CHD) genome-wide association studies (GWAS) risk genes are active in selective cell types involved in atherosclerosis. Coronary heart disease follow-up studies have demonstrated roles for LIPA, \textit{SORT1}, and \textit{TRIB1} as plasma lipid regulators in the liver, as well as in macrophages biology. Within the vessel wall, \textit{TCF21} is upregulated in dedifferentiated smooth muscle cells which migrate to the forming fibrous cap. \textit{Adams7} is also a regulator of smooth muscle migration but also a role in endothelial cells has been suggested.
Surface plasmon resonance demonstrated a high-affinity pH-dependent interaction between sortilin and apoB-containing lipoproteins, and mutants defective in their ability to traffic to the endolysosomal system were used to show that sortilin serves as a bona fide cell surface LDL receptor. Wild-type sortilin binds LDL at the cell surface in an LDL receptor–independent manner and delivers the LDL to the endolysosomal system for degradation.18,137

Based on the concordance of the human GWAS and mouse overexpression data, it was hypothesized that genetic knockout or knockdown of Sort1 would have the opposite effect, increasing plasma cholesterol and VLDL secretion. However, the reported effects of the genetic loss of sortilin on VLDL secretion have been contradictory and perplexing: loss of sortilin has been shown in different studies to result in either increased and decreased VLDL secretion.137–139 These discrepancies likely reflect the differences in models, methods of ablating sortilin function, lengths of time under diet-induced lipid overload, and technical approaches to measuring outcomes. Further complicating the story, Sort1−/− mice have recently been shown to be more insulin-sensitive but nonetheless were associated with decreased atherosclerosis. This effect was found to be attributable to the ability of sortilin to serve as a receptor of LDL: macrophages lacking sortilin to the endolysosomal system were used to show that sortilin binds LDL at the cell surface in an LDL receptor–independent manner and delivers the LDL to the endolysosomal system for degradation.136,137

The detailed molecular mechanisms by which sortilin influences insulin signaling, the sortilin protein is itself regulated by insulin signaling.141,142 Although these data suggest that sortilin influences insulin signaling, the sortilin protein is itself regulated by insulin signaling.141 Increased insulin sensitivity in extrahepatic tissues in Sort1−/− mice could be responsible for a decrease in free fatty acid flux to the liver, a major contributor to hepatic lipid accumulation and a driver of VLDL secretion during insulin resistance. The role of sortilin in the tissues other than the liver under these conditions is unknown and confounds the interpretation of the effect of sortilin knockout and knockdown in the liver on VLDL secretion. In addition, there may be other aspects of extrahepatic sortilin biology that could influence disease risk: in a mouse model with a humanized plasma lipid profile, whole body knockout Sort1−/− had no effect on plasma lipids, but nonetheless was associated with decreased atherosclerosis. This effect was found to be attributable to the ability of sortilin to serve as a receptor of LDL: macrophages lacking sortilin have reduced LDL uptake, which led to decreased foam cell formation.144

The detailed molecular mechanisms by which sortilin influences the complex processes of hepatic and plasma lipid metabolism, VLDL secretion, and MI/CVD risk have not yet been elucidated, but under conditions consistent with western lifestyle, it clearly impacts ApoB100 secretion, LDL clearance, and foam cell formation. Although the roles of sortilin in diverse CVD-relevant cell and tissue types have confounded the analysis of its biological function, this underscores the importance of the gene and likely explains its robustness as a GWAS signal.

**TRIB1 (Tribbles-1)**

GWAS have consistently associated variants at the 8q24 locus containing the gene TRIB1 with multiple human metabolic phenotypes. The TRIB1 locus was first implicated in plasma lipid metabolism by 2 papers published simultaneously that showed noncoding variation in the TRIB1 gene locus was associated with circulating triglyceride levels in humans. The landmark Global Lipids Genetics Consortium meta-analysis of >100,000 individuals further illustrated the importance of TRIB1 in lipid metabolism by associating the locus with not only triglycerides but also total cholesterol, high-density lipoprotein cholesterol, LDL-C, and CAD,19 making TRIB1 the only novel locus from these studies to be associated with all 4 lipid traits and CAD. The plasma lipid associations were replicated in a Global Lipids Genetics Consortium follow-up meta-analysis of =200,000 individuals,145 whereas the CAD association was confirmed in 2 recent GWAS from the CARDIoGRAMplusC4D consortium, in which researchers investigated the association of genome-wide sequence variation with atherosclerosis, regardless of plasma lipid phenotype.146 The significantly associated SNPs in all instances fall around 20 kb upstream of the TRIB1 gene, suggesting a role in the regulation of TRIB1 gene expression. One recent study showed that significantly associated SNPs in the TRIB1 locus alter the expression of a long noncoding RNA named TRIBAL (TRIB1 associated locus), although the role of TRIBAL in any disease pathology is currently unclear.147

The TRIB1 gene encodes a protein known as Tribbles-1, which was originally identified in a drosophila mutagenesis screen which revealed that the protein Trbl (the drosophila homolog of TRIB1) participates in oogenesis via promoting the proteasomal degradation of C/EBPα, which it may participate in CAD pathogenesis. Studies using AAV-mediated overexpression of mouse Trib1 (AAV_mTrib1) to investigate this association found that increasing levels of hepatic Trib1 decreased plasma total cholesterol, high-density lipoprotein cholesterol, LDL-C, and triglycerides levels in a dose-dependent manner.118 AAV-treated mice showed a decrease in hepatic lipogenic gene expression, and ex vivo studies of primary hepatocytes from those mice showed reduced cellular triglyceride production and secretion. Furthermore, Ldb8 mice treated with AAV_mTrib1 had decreased plasma ApoB levels, and HepG2 cells overexpressing TRIB1 had decreased ApoB secretion. These data suggest that TRIB1 can modulate VLDL secretion from the liver, presumably by affecting the level of triglycerides available for efficient VLDL assembly. More recent work from our group in a liver-specific knockout of Trib1 established that C/EBPα is the mechanistic link between TRIB1 and hepatic lipogenesis.148 Liver-specific knockout of Trib1 mice have increased hepatic triglyceride
content, lipogenic gene expression, and de novo lipogenesis. They also display increased hepatic C/EBPα protein, and this increase is both necessary and sufficient to drive the lipogenic phenotype. The liver-specific knockout of Trib1 mice also have increased plasma lipids; however, this seems to be a C/EBPα-independent effect and suggests that TRIB1 regulates plasma lipid metabolism via other mechanisms independent of lipogenesis. TRIB1 has also recently been shown via in vitro overexpression assays to interact with the transcription factor ChREBP, as well as SAP18, a component of the Sin3A-HDAC co-repressor complex. The role of these interactions in vivo and the extent to which they all participate in plasma lipid regulation remains to be determined.

The association of the TRIB1 gene locus with CAD is likely driven in large part by its putative regulation of VLDL secretion. However, it remains possible that other mechanistic links between the gene and CAD contribute to this genetic association. The aforementioned role of TRIB1 in macrophage polarization is one potential link because the M1/M2 association. The mentioned role of TRIB1 in macrophage polarization is one potential link because the M1/M2 status of macrophages in the lesion can contribute to plaque progression.

Human genetics have implicated the TRIB1 locus in a host of other human phenotypes, including levels of circulating adiponectin and liver enzymes—an association functionally confirmed by the liver-specific knockout of Trib1 mouse—as well as the onset of metabolic syndrome in humans. Each of these human traits could by themselves contribute to CAD, either directly or indirectly. Thus, it is possible that pleiotropic effects of TRIB1 contribute to CAD, and careful work in animal models of metabolic disease with TRIB1 tissue-specific deletion will be required to determine the specific contribution to disease burden by each specific function of TRIB1.

**LIPA**

Several GWAS studies have identified LIPA (lysosomal acid lipase) as a novel locus for CAD.41,163,164 Meta-analyses revealed that LIPA CAD risk alleles rs14124444T and rs2246833T (clustered in introns 2 and 3 in high linkage disequilibrium, r²=0.985) were associated with higher LIPA expression in monocytes but not in liver, nor did they alter plasma lipids. Fine mapping of the LIPA region by the CARDIoGRAM+C4D consortium failed to reveal additional variants with stronger signals than the original GWAS SNPs, and rs2246833 had the strongest CAD association (P=4.9×10⁻¹²). Both aforementioned SNPs show strong Histone 3 Lysine 27 acetylation enrichment and are in and near DNase I hypersensitivity site and TF-binding sites (ENCODE), suggesting possible regulatory roles.

LIPA encodes an enzyme called lysosomal acid lipase (LAL), which catalyzes the hydrolysis of cholesteryl ester and triglycerides in intracellular lysosomes after their internalization via receptor-mediated endocytosis of lipoprotein particles. Human LAL is encoded by the LIPA gene on chromosome 10q23.2–23.3 and is a 46-kDa glycoprotein. After undergoing cotranslational glycosylation in the endoplasmic reticulum and attachment of mannose-6-phosphate residues in the Golgi apparatus, LAL is targeted to prelysosomal compartments.

Before the GWAS discovery for CAD, LOF mutations in LIPA were identified as causes of rare lysosomal disorders. Wolman disease is an infantile-onset disorder with massive infiltration of cholesteryl ester/triglycerides-filled macrophages in multiple organs because of complete LIPA LOF. Cholesteryl ester storage disease (CESD) is a later-onset disorder with incomplete LIPA LOF mutations, resulting in hepatomegaly, hyperlipidemia, and premature atherosclerosis. The most common mutation seen in CESD patients is a splice junction mutation at exon 8 of LIPA, which leads to ≥3% to 5% of normally spliced LAL protein and similar low levels of LAL activity. Data of CESD fibroblasts suggest that LIPA deficiency leads to lysosomal cholesteryl ester accumulation that limits lysosome 1 FC release and cytosolic cholesterol esterification and impaired ATP-binding cassette transporter A1 (ABCA1)-mediated cholesterol efflux. In a phase 3 trial of enzyme replacement therapy in children and adults with CAD deficiency, recombinant human LAL Sebelipase Alfa resulted in a reduction in multiple disease-related hepatic and lipid abnormalities, with the long-term effects of Sebelipase Alfa on cardiovascular events undetermined.

Although recombinant human LAL enzyme replacement therapy is likely to exert protective effects against premature atherosclerosis in CESD patients, it is unclear what the effects of recombinant human LAL treatment would be in CAD patients without CAD deficiency. Indeed, our understanding of the role of LIPA in the progression of atherosclerosis is far from complete. eQTL studies have suggested that the GWAS risk alleles for CAD are associated with increased LIPA mRNA in monocytes. How the higher LIPA mRNA in monocytes relates to increased risk of CAD is unclear: whether the GWAS CAD variants associated with higher LIPA mRNA is correlated with higher CAD risk alleles or enzymatic activity, and if it is a true gain of function, mutation has yet to be determined. The GWAS CAD alleles are in linkage disequilibrium with a missense coding variant (rs1051338) in the signal peptide that may alter post-translational trafficking and secretion, so it is possible that the CAD signal marks a loss of normal LAL processing and function despite higher mRNA. Thus, whether increased or decreased monocyte–macrophage activity of LAL beyond the normal cellular response, in the general population and in CAD patients, is atherogenic or protective remains a completely open and controversial question.

A comprehensive understanding of the impact of LIPA on CAD pathogenesis relies on in vivo modeling. Lipa knockout mice display shortened life span, tissue cholesteryl ester accumulation, and hepatosplenomegaly. Furthermore, Lipa⁻/⁻ accelerates atherosclerosis in the ApoE⁻/⁻ hyperlipidemic mouse model. These findings recapitulate the pathological phenotypes of human LIPA LOF in CESD. Although systemic recombinant human LAL administration reduces hyperlipidemia and atherosclerosis in Ldlr⁻/⁻ mice, the atheroprotection was most likely attributable to the reduction in plasma lipids. Surprisingly, transgenic mice with whole body Lipa overexpression apparently have elevated plasma VLDL cholesterol and hepatocellular lipids on western diet. In context of the lack of association between GWAS CAD risk alleles and plasma lipid levels or liver LIPA expression, the whole body knockout or transgenic mice do not serve as an appropriate model of the CAD-associated locus identified in
GWAS studies. Indeed, because the CAD risk alleles are specifically associated with higher monocyte LIPA mRNA, it is now imperative to define the monocyte/macrophage-specific role of LIPA CAD risk alleles in vitro in macrophage function using isogenic hiPSC lines carrying risk or nonrisk alleles with subsequently differentiation to macrophages—or in vivo in the progression of atherosclerosis using monocyte/macrophage-specific gain of function of LIPA in mouse models of atherosclerosis.

**ADAMTS7**

The association of the A disintegrin and metalloproteinase with thrombospondin motifs-7 (ADAMTS7) locus with CAD risk has been identified and replicated through GWAS.\(^{1,4,18,184}\) This locus for coronary atherosclerosis was discovered in the PennCath cohort using angiographic CAD as the primary outcome.\(^{184}\) and subsequent studies have shown that ADAMTS7 also relates to MI.\(^{1,4,18}\) Its association is most robust for angiographic CAD, a marker of coronary atherosclerotic burden, suggesting that ADAMTS7 is likely to relate to clinical events through the development and progression of atherosclerosis. Recent findings in mouse vascular injury and atherosclerosis models\(^{185,186}\) are consistent with such an action in the humans. Genetic variation at the ADAMTS7 locus has no relationship with traditional risk factors or mechanistic biomarkers;\(^{4,18}\) hence, the directional impact of ADAMTS7 expression on CAD risk and the underlying biological mechanisms have been unclear. Functional studies suggest that ADAMTS7, a metalloproteinase expressed in vascular smooth muscle cells (VSMC) and ECs, is the probable causal proatherogenic gene at this locus.\(^{185-188}\) Briefly, the top CAD-risk SNPs at this locus are eQTLs for higher ADAMTS7 expression, whereas allelic variation at a nonsynonymous variant (rs3825807, Ser214Pro) in ADAMTS7 associates with reduced CAD risk and may impair ADAMTS7 function in VSMC.\(^{187}\) Our recent work demonstrated that deletion of *Adamts7* is atheroprotective in both Ldlr\(^{−/−}\) and ApoE\(^{−/−}\) mouse models.\(^{188}\) Thus, blockade of ADAMTS7 expression or inhibition of its function presents novel therapeutic opportunities for prevention and treatment of CAD.

Understanding the relationship between CAD risk alleles at the ADAMTS7 chr15q21.1 region and expression levels of ADAMTS7 in human-disease–relevant cells has not been straightforward. Interestingly, in available eQTL data sets with large sample sizes,\(^{189}\) the lead SNPs from the PennCath (rs1994016), CARDIoGRAM (rs3825807), and C4D (rs4380028) GWAS studies demonstrate a significant association with ADAMTS7 expression and match the directionality and causality of in vivo data using mouse model, with the CAD risk alleles being associated with higher ADAMTS7 expression. However, currently there are no large eQTL or RNA-Seq-based ASE data that provide adequate power to determine eQTL directionality in the most pertinent human vascular cells and tissues. Individual laboratories and the GTEx consortia\(^{189}\) are now generating data sets from a large enough sample pool of human vasculature to address this question. In the ENCODE Project,\(^{163}\) the National Institutes of Health Roadmap Epigenomics Mapping data and in our own ChIP-Seq experiments, the top CAD SNPs fall in ADAMTS7 5′ and 3′ regions that overlap regulatory elements in VSMC and aortic tissues and are, for instance, close to binding sites for TCF21, a VSMC transcription factor that regulates coronary development and is itself a GWAS locus for CHD.\(^{4,164,191}\) In unpublished data, several of these regions have been found to have enhancer activity in rat A7r5 VSMC, suggesting that this chr15q21.1 CHD SNPs may act on CHD by regulating human coronary arterial smooth muscle cell ADAMTS7 expression.

ADAMTS7 is a member of the ADAMTS family of secreted zinc metalloproteinases with characteristic protein domain composition including at least one thrombospondin type I repeat.\(^{192-195}\) The family of ADAMTS proteases degrades extracellular matrix, and several ADAMTS family members have been implicated in human diseases, including thrombotic thrombocytopenic purpura,\(^{196}\) Weill–Marchesani syndrome,\(^{197}\) and atherosclerosis.\(^{198}\) Unlike other metalloproteinases, ADAMTS family members demonstrate narrow substrate specificity because of their C-terminal exosites.\(^{192-195}\) Previous research on ADAMTS7 has mainly centered on its role in bone and cartilage growth because cartilage oligomeric matrix protein has been identified as a substrate.\(^{199}\) ADAMTS7 can regulate endochondral bone formation through interactions with cartilage oligomeric matrix protein. Cartilage oligomeric matrix protein is also expressed in VSMC and vasculature, and additional studies with viral-mediated overexpression and knockdown in vivo and in vitro suggests that ADAMTS7 might modulate VSMC phenotype switching and migration via interactions with cartilage oligomeric matrix protein.\(^{198}\)

Most domains in human and mouse *Adamts7* are highly conserved rendering the mouse as a useful model for actions in human disease. The first evidence that *Adamts7* deficiency (Adamts7\(^{−/−}\)) attenuates atherosclerosis in vivo, in both the ApoE\(^{−/−}\) and Ldlr\(^{−/−}\) mouse models, and *Adamts7*\(^{−/−}\) confers a specific loss of VSMC migration in response to inflammatory signals was recently published.\(^{196}\) It has been shown that *Adamts7*\(^{−/−}\) also reduces vascular response to mechanical injury.\(^{195,198}\) *Adamts7* gene expression was induced transiently in the mouse vasculature in response to stress, both in the wire injury model and in the atherosclerosis experiments, that tumor necrosis factor-a induces *Adamts7* expression in primary VSMCs, and that VSMC of *Adamts7*\(^{−/−}\) mice shows reduced tumor necrosis factor-a–induced migration.\(^{196}\) Immunostaining in human diseased coronary arteries reveals colocalization of ADAMTS7 with cells positive for VSMC markers, and immunofluorescence in human aortic smooth muscle cells shows subcellular localization with leading edges of migrating VSMCs. These data suggests that Adamts7 modulates VSMC phenotype and migration during inflammatory stress and mechanical injury and that *Adamts7* deficiency markedly reduces atherosclerotic lesions in hyperlipidemic mice.

Human eQTL interrogations reveal that common alleles that relate to lower CAD risk are also associated with reduced ADAMTS7 expression. This is consistent with rodent studies and supports a proatherogenic role of ADAMTS7 in humans. Because ADAMTS7 has narrow substrate specificity, it has promise as a potentially safe drug target. Thus, inhibition of ADAMTS7 is a potential novel therapeutic strategy for CAD.
in humans. However, several important questions still need to be addressed to accelerate clinical and therapeutic translation related to this locus.

**TCF21 (Pod1, Capsulin, Epicardin)**

The TCF21 gene codes for a basic helix–loop–helix transcription factor known to bind cis-regulatory elements as heterodimers with TCF3 or TCF12. It is expressed in the mesenchyme of developing organs, including the lung, kidney, gut, and heart, and constitutive Tcf21 knockout mice die at birth because of missing alveoli in their lungs. In addition, Tcf21 has been shown to play a role in vascular development: Tcf21 is highly expressed in the proepicardial organ, which contains progenitor cells of coronary artery smooth muscle and ECs and cardiac fibroblasts. It is necessary for epithelial-to-mesenchymal transition of epicardial fibroblasts and their subsequent migration into the cardiac interstitium.

The TCF21 gene locus has been linked to CAD risk by GWAS, which first reported rs12190287 as the association lead SNP. This polymorphism lies in the 3′UTR of 1 of the 2 TCF21 transcript variants. It has been shown to disrupt an AP-1-binding site inside an enhancer in VSMCs. Interestingly, this variant also alters a mir-224-binding site inside a TCF21 transcript variant, suggesting additional miRNA-dependent regulation of TCF21 on the post-transcriptional level. More recently, the 1000 Genomes-based CARDIoGRAMplusC4D GWAS meta-analysis reported rs12202017, which lies 3.7 kb upstream of the TCF21 gene within the TARID (TCF21 antisense RNA inducing demethylation) lncRNA locus as the lowest P value association SNP. This locus also harbors a separately reported CAD association signal in Han Chinese, with rs12524865 as the strongest association SNP in this study.

The potential role of TARID in the vasculature has to date not been interrogated and presents an attractive target for future study. The contribution of TCF21 to disease seems to be coronary artery–specific: a meta-analysis for shared susceptibility reported that the TCF21 association signal is confined to CAD and is not implicated in risk for ischemic stroke.

The role of the Tcf21 protein in CAD has recently been investigated in vivo using a lacZ reporter and a lineage tracing model in mouse. These studies show that Tcf21 expressing cells migrate into the forming atherosclerotic lesion and contribute to the fibrous cap. Further studies are needed to investigate how lack of TCF21 contributes to lesion size and composition. To identify TCF21 transcriptional target genes, ChIP-Seq studies have been performed in human coronary arterial smooth muscle cells. Enrichment analysis showed that TCF21-binding sites are enriched for CAD GWAS association SNPs. This suggests a role of TCF21 as regulator of genomic loci, such as ADAMTS7, conferring risk for atherosclerosis in coronary artery SMCs.

**Summary**

GWAS have provided a rich collection of CAD loci that suggest the existence of exciting new biology relevant to atherosclerosis that we never suspected but that require extensive functional follow-up studies. However, causality of a specific gene cannot be inferred solely based on proximity to a region of statistical association with disease. A thorough and meticulous annotation of the region with data generated in relevant cell types is paramount. Specific targeting of the SNP or region of association needs to be combined with the interrogation of changes in transcription of all genes in the haplotype block of interest. Far too often are candidate genes chosen for functional studies without solid evidence for their definitive association with the GWAS signal, and poorly characterized or long-distance genes are rarely followed-up on. Combinatorial effects of more than one gene within a locus, possibly across multiple tissues, could be one mechanism by which common, nonprotein-coding variation contributes to CAD. We may eventually be surprised by the complexity with which the implicated genomic regions modulate the CAD phenotype. As of now, the biology underlying most CAD GWAS loci remains unknown. The likelihood is high that fundamentally new biology regarding atherosclerosis and CAD will be learned through the interrogation of CAD GWAS loci. Furthermore, the hope remains that at least some of these new pathways relevant to CAD pathogenesis will yield new therapeutic targets for the prevention and treatment of CAD.

**Sources of Funding**

S.T.N. is supported by a Transatlantic Network of Excellence Award from Foundation Leducq. H.Z. is supported by the American Heart Association Postdoctoral Fellowship 15POST25620017.

**Disclosures**

None.

**References**


38. Innocenti F, Cooper GM, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary


168. Birney E, Stamatoyannopoulos JA, Dutta A, et al; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children’s Hospital Oakland Research Institute. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007;447:799–816. doi: 10.1038/nature05874.


From Loci to Biology: Functional Genomics of Genome-Wide Association for Coronary Disease
Sylvia T. Nurnberg, Hanrui Zhang, Nicholas J. Hand, Robert C. Bauer, Danish Saleheen, Muredach P. Reilly and Daniel J. Rader

doi: 10.1161/CIRCRESAHA.115.306464
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/118/4/586

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/