Published in final edited form as:
Curr Atheroscler Rep. 2012 August ; 14(4): 295–299. doi:10.1007/s11883-012-0245-0.

The Enigma of Genetics Etiology of Atherosclerosis in the Post-GWAS Era

A.J. Marian, M.D.
Center for Cardiovascular Genetics, Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center and Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX, 77030

Abstract

Coronary atherosclerosis is a complex heritable trait with an enigmatic genetic etiology. Genome-wide association studies (GWAS) have successfully led to identification of over 100 different loci for susceptibility to coronary atherosclerosis. Most identified single nucleotide polymorphisms (SNP)s and genes have not been previously implicated in the pathogenesis of atherosclerosis and hence, have modest biological plausibility. The novel discoveries, however, might provide the opportunity for identification of new pathways and consequently novel preventive and therapeutic targets. A notable outcome of GWAS is relatively modest effect sizes of the associated SNPs. Collectively, the identified SNPs account for a relatively small fraction of heritability of coronary atherosclerosis, which raises the question of "missing heritability". Because GWAS test the common disease – comment variant hypothesis, a plausible explanation might be the presence of uncommon and rare variants in the genome that are untested in GWAS but that might exert large effect sizes on the risk of atherosclerosis. The latter, however, remains an empiric question pending validation through experimentation. Alternative mechanisms, such as transgenerational epigenetics including microRNAs, might in part account for the heritability of coronary atherosclerosis. Collectively, the recent findings are indicative of the etiological complexity of coronary atherosclerosis. Hence, it is expected that genetic etiology of coronary atherosclerosis will remain enigmatic in the foreseeable future.

Keywords

Atherosclerosis; Coronary artery disease; Genetics; GWAS; Polymorphism

Introduction

Atherosclerosis, whether involving coronary or other vascular systems, is a complex phenotype and like any other complex trait, is in part genetically programmed. The initial evidence for a genetic etiology of atherosclerosis was based on the demonstration of familial aggregation and heritability estimates [1, 2], which stimulated the search for identifying the risk alleles. The candidate gene approach through case-control allelic association studies has implicated a number of DNA sequence variants (DSVs) in susceptibility to coronary atherosclerosis and myocardial infarction. However, the results have been inconsistent and subject to a high rate of spurious association. The advent of genome-wide association

Address for Correspondence: AJ Marian, M.D., Center for Cardiovascular Genetics, 6770 Bertner Street, Suite C900A, Houston, TX 77030, Phone: 713 500 2350, Fax: 713 500 2320, Ali.J.Marian@uth.tmc.edu.

Disclosure
A.J. Marian receives part of his salary in the form of grants from the NIH and some travel expenses are paid for by NIH study sections.

studies (GWAS) ushered in an unbiased genome-wide approach that raised considerable hope in identifying the genetic determinants of susceptibility to atherosclerosis. GWAS has become the state-of-the-art approach for genetic studies of complex disease and has been applied almost to all complex phenotypes successfully to identify the risk alleles. As of February 23, 2012, 1179 GWAS studies reporting 5876 phenotype-associated single nucleotide polymorphisms (SNPs) have been listed in the catalog of published GWAS (http://www.genome.gov/gwastudies/). Likewise, 16 GWAS studies of "coronary heart disease", 4 of "myocardial infarction" and 2 of "coronary artery calcification" have been catalogued (www.genome.gov/gwastudies/). GWAS of "coronary heart disease" collectively have led to identification of 80 susceptibility genes (clusters) and about 40 intergenic regions. Despite the plethora of GWAS, it has become evident that the results of GWAS are unlikely to fully explain the genetic basis of coronary atherosclerosis and explain only a small fraction of its heritability. Thus, it appears that a significant portion of heritability of coronary atherosclerosis remains to be discovered. In this review, we provide an overview of the results of GWAS of "coronary heart disease" and discuss other genetic and genomic factors that might account for the "missing heritability" of coronary atherosclerosis.

Genetic Etiology of Coronary Atherosclerosis Identified by GWAS

GWAS is an array-based genotyping approach designed to type the study population – typically cases and controls – for a very large number of known common variants, as defined by a minor allele frequency of > 0.05) [3•]. Thus, GWAS tests the common disease–common variant hypothesis (CD-CV). In less than a decade since its initial introduction, over 100 loci have been associated with "coronary heart disease", the term commonly used in GWAS of coronary atherosclerosis, even though it likely entails multiple related but partially independent phenotypes and is not sufficiently specific. As would be expected and to some extent gratifying, the findings show association of SNPs in several genes coding for the established risk factors and "coronary heart disease" (Table 1). However, the majority of the genes identified through the GWAS have not been previously implicated in the pathogenesis of atherosclerosis, and hence, in this aspects are novel genes (Table 2). Identification of the novel genes is indeed the strength of the GWAS, as it is devoid of *a priori* hypothesis for the candidacy of a gene and is rather an unbiased survey of the genome. In addition, a high threshold for the statistical significance is often adopted in order to account for the opportunity for multiple hypothesis testing and reduce the chance of spurious associations. Moreover, the GWAS findings are typically replicated in independent study populations to strengthen the findings, as otherwise considered provisional. Thus, considering these cautionary steps built into the design of GWAS, all novel genes identified through this approach have the potential to shed new light into the molecular mechanisms of atherosclerosis.

The list of novel genes for atherosclerosis identified through GWAS is intriguing (Table 2). Most if not all have met *a priori* set criteria for statistical significance and most have been replicated in independent populations. Several genes, such as *CXC12* [4], are biologically plausible candidates that might regulate pathways implicated in atherosclerosis. The majority, however, suffer from low biological plausibility for influencing susceptibility to coronary atherosclerosis. Nevertheless, biological plausibility is based on *a prior* knowledge and hence, a modest biological plausibility does not necessarily negate a potential causal role in atherosclerosis. These novel findings raise several possibilities, among them the possibility that a large number of pathways responsible for coronary atherosclerosis remain to be discovered. It is, however, perplexing to fathom SNPs in a gene that has been implicated in cancer or myopathies predisposes to coronary atherosclerosis, knowing that such phenotypes are not known to be associated with atherosclerosis. Nevertheless, phenotypic plasticity is not an uncommon feature of genes as many might exhibit

remarkable phenotypic plasticity, as best exemplified by *LMNA*, which encodes nuclear envelope protein lamin A/C. Mutations in *LMNA* are known to be responsible for over a dozen distinct phenotypes [5]. A plausible alternative possibility is linkage disequilibrium (LD) between alleles identified through GWAS and the true risk alleles. The size of LD varies in different genomic regions and in different ethnic backgrounds [6]. It typically encompasses blocks of 10 to 25 kbp in size but could extend into much larger blocks and on occasion involve several million base pairs containing multiple genes [6–8]. Thus, it is likely that the SNPs identified through GWAS are not the true susceptibility alleles but in LD with them. One also has to consider the possibility of spurious association in same cases, in spite of robust statistical association and even replication. Collectively, the results of GWAS provide the initial clues necessary for further refined genetic studies to identify the true risk alleles.

A notable feature common to genetic studies of complex traits by the GWAS approach is the relatively modest effect sizes of the phenotype-associated alleles. This is best illustrated for quantitative traits such as plasma high-density lipoprotein cholesterol (HDL-C) level or systemic arterial blood pressure. SNPs identified through GWAS shift plasma HDL-C level by 1 mg/dL or less and those associated with hypertension shift systolic or diastolic blood pressure by 1 mm Hg or less [9, 10, 11•]. Likewise, for phenotypes such as “coronary heart disease” SNPs identified through GWAS impart modest risk, typically a relative risk in the range of 1.1 to 1.2 [12]. The relatively modest effect sizes of the alleles on complex phenotypes mitigate the clinical impact of identification of the susceptibility alleles for genetic-based risk stratification or prognostication and with some exceptions, in individualization of pharmacological and non-pharmacological therapy.

The main, and probably the only, strength of the results of GWAS is in providing novel mechanistic insights into the pathogenesis of the complex traits. Two notable loci discovered through GWAS are the 9p21 and 1p13 loci, which are probably the best-characterized GWAS loci for coronary heart disease, other than the lipid-related genes. The 9p21 locus is a gene-desert locus that has been linked to coronary atherosclerosis in multiple studies [13–15]. The locus contains *CDKN2B-AS1*, also known as *ANRIL*, which codes for a long non-coding RNA (lnRNA). Immediately adjacent to the refined locus are *CDKN2A* and *CDKN2B*, which code for cell cycle regulators and tumor suppressor proteins CDKN2A and CDKN2B [16]. Reduced expression levels of *Cdkn2a* and *Cdkn2b* are associated with enhanced proliferation and reduced senescence of smooth muscle cells [17•]. The region also contains a number of enhancers including an enhancer for STAT1, which is a signal transducer for interferons and cytokines [18••]. The 9p21 risk alleles disrupt the binding site for STAT1, leading to enhanced expression of CDKN2B-AS1 in response to inflammation [18••]. Thus, it is plausible that the 9p21 locus imparts its effects on susceptibility to atherosclerosis through an enhanced expression of CDKN2B-AS1 to inflammation, suppression of expression of CDKN2B and ensuing increased proliferation of smooth muscle cells in the vessel wall [17•, 18••].

The precise mechanism(s) by which the 1p13 locus influences the risk of coronary atherosclerosis is uncertain but seems to relate to the LDL-C uptake in the liver, as discussed by Strong and Rader in this issue of the *Current Atherosclerosis Reports*. The minor allele of the risk SNP creates a C/EBP-α binding site on the *SORT1* promoter, which leads to increased expression level of SORT1 [19••]. Forced expression of Sort1 is associated with reduced plasma LDL-C levels. In contrast, suppression of its expression is associated with increased plasma LDL-C levels [19••]. The findings suggest enhancing expression of SORT1 could serve as a potential therapeutic strategy for hypercholesterolemia. In contrast, however, another study reported an opposite effect upon over-expression of Sort1, which was associated with increased hepatic release of lipoproteins and plasma LDL level [20].

The contrasting results of these two studies hint to the complexity of identifying the molecular mechanisms in the model organisms to link the GWAS findings to the pathogenesis of atherosclerosis in humans.

Unaccounted Genetic Etiologies of Coronary Atherosclerosis ("missing heritability")

The findings of the GWAS do not seem to resolve the enigma of the genetic etiology of coronary atherosclerosis. The common alleles, as tested in the GWAS, account for a small fraction of heritability of complex phenotypes including coronary atherosclerosis. The tagger SNPs, which represent the common haplotypes, might not adequately cover the uncommon (MAF of 0.01 to 0.05) or rare (< 0.01) variants, which might predispose to atherosclerosis. The unaccounted genetic determinants of coronary atherosclerosis along with technological advances in the whole genome and whole exome-sequencing approaches have raised considerable interest in testing the rare variant–common disease (RV-CD) hypothesis, which posits that multiple uncommon and rare susceptibility alleles are in part responsible for susceptibility for complex phenotypes. In accord with the large effect size of the causal rare variants in single gene disorders, it is generally expected that the uncommon and rare variants to exert larger effects on complex phenotypes than the common alleles. However, the RV-CD is largely an empiric question that remains to be tested through large-scale sequencing in carefully selected study populations (reviewed in [21]). The "missing heritability", however, might have several other components, in addition to yet-to-be identified common and rare alleles. It might reflect in part an over-estimation of the heritable component of etiology of coronary atherosclerosis, genotype-genotype interactions (epistasis), and heritable epigenetic factors including microRNAs (reviewed in [22]). Thus, finding the "missing heritability" of complex phenotypes, such as atherosclerosis might require discovering all DSVs in the genome, whether common or rare; discovering epigenetic determinants and deciphering the intertwined, often non-linear and stochastic interactions among the constituents that contribute to the pathogenesis of atherosclerosis.

Conclusions

Genotyping of over 230,000 cases with "coronary heart disease" and near 300,000 control individuals for > 500,000 common SNPs over the past 7 years has led to identification of over 100 susceptibility loci for coronary atherosclerosis. The vast majority of the loci identified through GWAS encompass genes not previously implicated in the pathogenesis of atherosclerosis. Despite the enthusiasm for the discoveries, it remains to be seen whether and how soon the findings of GWAS will lead to discovering novel pathways in the pathogenesis of atherosclerosis and hence, new preventive and therapeutic targets. The clinical impact of mechanistic discoveries would likely be independent of the effect sizes of the susceptibility alleles detected in GWAS on atherosclerosis. Although the latter is modest and clinically negligible, the former could be as formidable as the effects of statins. Notwithstanding the significance of the mechanistic discoveries, the enigma of genetic etiologies of coronary atherosclerosis is expected to remain unexplained in the foreseeable future.

Acknowledgments

This work is supported in part by grants from NHLBI (R01-088498 and R34HL105563); NIA (R21 AG038597-01), Burroughs Wellcome Award in Translational Research (#1005907), TexGen Fund from Greater Houston Community Foundation

References

1. Marenberg ME, Risch N, Berkman LF, et al. Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med. 1994; 330:1041–1046. [PubMed: 8127331]
2. Mayer B, Erdmann J, Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. Clinical research in cardiology: official journal of the German Cardiac Society. 2007; 96:1–7. [PubMed: 17021678]
•3. Marian AJ, Belmont J. Strategic approaches to unraveling genetic causes of cardiovascular diseases. Circulation Research. 2011; 108:1252–1269. This review article will bring the readers up-to-date with the current approaches to genetic studies of complex diseases, such as atherosclerosis. [PubMed: 21566222]
4. Farouk SS, Rader DJ, Reilly MP, Mehta NN. CXCL12: a new player in coronary disease identified through human genetics. Trends in cardiovascular medicine. 2010; 20:204–209. [PubMed: 22137643]
5. Worman HJ, Bonne G. "Laminopathies": A wide spectrum of human diseases. Experimental Cell Research. 2007; 313:2121–2133. [PubMed: 17467691]
6. Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. Science. 2005; 307:1072–1079. [PubMed: 15718463]
7. Dawson E, Abecasis GR, Bumpstead S, et al. A first-generation linkage disequilibrium map of human chromosome 22. Nature. 2002; 418:544–548. [PubMed: 12110843]
8. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet. 2002; 3:299–309. [PubMed: 11967554]
9. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009; 41:666–676. [PubMed: 19430483]
10. Pirruccello J, Kathiresan S. Genetics of lipid disorders. Curr Opin Cardiol. 2010; 25:238–242. [PubMed: 20224388]
•11. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011; 478:103–109. This is a genome-wide association study that describes identification of multiple loci associated with systemic arterial hypertension and shows that the effect of each locus is very small. [PubMed: 21909115]
12. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. PLoS genetics. 2011; 7:e1002260. [PubMed: 21966275]
13. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–1491. [PubMed: 17478681]
14. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007; 316:1491–1493. [PubMed: 17478679]
15. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007; 357:443–453. [PubMed: 17634449]
16. Collado M, Blasco MA, Serrano M. Cellular senescence in cancer and aging. Cell. 2007; 130:223–233. [PubMed: 17662938]
•17. Visel A, Zhu Y, May D, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature. 2010; 464:409–412. This is an interesting study that analyzes phenotypic consequences of deletion of the 9p21 analogous locus in mice and shows enhanced smooth muscle cell proliferation. [PubMed: 20173736]
••18. Harismendy O, Notani D, Song X, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature. 2011; 470:264–268. This study provides mechanistic data to explain association of the 9p21 locus with coronary atherosclerosis. The findings are notable for enhanced interferon gamma response of expression of an anti-sense RNA-CDKN2B-AS1 in the presence of the risk allele. The latter leads to suppression of cell cycle regulator CDKN2B and hence, anticipated increased proliferation of smooth muscle cells. [PubMed: 21307941]
••19. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010; 466:714–719. This study identifies SORT1, coding for Sortilin 1, as the gene associated with coronary atherosclerosis in the GWAS. The

authors perform a series of in vivo and in vitro mechanistic studies and discover that the risk alleles generate a binding site for C/EBP-α and increased expression of Sort1 is associated with lower plasma LDL-C level. [PubMed: 20686566]

••20. Kjolby M, Andersen OM, Breiderhoff T, et al. Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. Cell Metab. 2010; 12:213–223. The authors study the mechanism by which Sort1 influences plasma LDL-C level and report that over-expression of Sort1 is associated with increased hepatic release of lipoproteins and plasma LDL level, a finding that seems to be in contrast to the study by Musunuru et al. [19• •]. [PubMed: 20816088]

21. Marian AJ. Molecular genetic studies of complex phenotypes. Transl Res. 2012; 159:64–79. [PubMed: 22243791]

22. Marian AJ. Elements of missing heritability. Current Opinion in Cardiology. 2012:27.

Table 1

Lipid-related genes for “coronary heart disease” identified through genome-wide association studies

Gene symbol	Gene name	Function
ABCA1	ATP-binding cassette, sub-family A	Cholesterol efflux
APOA5	Apolipoprotein A-V	Plasma triglyceride and HDL
CETP	Cholesteryl ester transfer protein	Transfers cholesteryl esters between lipoproteins
LCAT	Lecithin-cholesterol acyltransferase	Esterifies cholesterol for transport
LDLR	Low-density lipoprotein receptor	LDL-C update
LIPA	Lipase A, lysosomal acid, cholesterol esterase	Catalyze the hydrolysis of cholesteryl esters and triglycerides
LPA	Lipoprotein, Lp(a)	A serine proteinase that inhibits PAI-1
PCSK9	Proprotein convertase subtilisin/kexin type 9	LDL-C homeostasis
SORT1	Sortilin 1	Regulates LDL-C

HDL—high-density lipoprotein; LDL-C—low-density lipoprotein cholesterol; PAI—plasminogen activator inhibitor.

Table 2

Non-lipid related susceptibility genes for “coronary heart disease” identified through genome-wide association studies

Gene symbol	Gene name	Function
ABO	ABO blood group	Glucosyltransferase
ADAMTS7	ADAM metallopeptidase with thrombospondin type 1 motif, 7	Zinc-dependent protease
ALDH2	Aldehyde dehydrogenase 2 family	Mitochondrial enzyme in the oxidative pathway
ANKS1A	Ankyrin repeat and sterile alpha motif domain containing 1A	A negative regulator of growth factors
CDKN2A (CDKN2B)	Cyclin-dependent kinases 2A and 2B	Cell cycle regulators
CDKN2B-AS1 (ANRIL)	CDNK2B antisense RNA 1	Antisense RNA 1 against CDKN2B
COL4A1 (COL4A2)	Collagen, type IV, alpha 1	Collagen
CSMD1	CUB and Sushi multiple domains 1	Regulator of complement pathway/ inflammation
CXCL12	Chemokine (C-X-C motif) ligand 12	Activate lymphocytes
CYP17A1 (CNNM2, NT5C2)	Cytochrome P450, family 17, subfamily A, polypeptide 1	Drug metabolism and cholesterol/steroid synthesis
DNM2	Dynamin 2	Endocytosis and cell motility
DOCK6	Dedicator of cytokinesis 6	Actin cytoskeletal reorganization
DTNB	Dystrobrevin, beta	Cytoskeletal protein
EDC4	Enhancer of mRNA decapping 4	mRNA processing
FNDC1	Fibronectin type III domain containing 1	G protein signaling
GCOM1	GRINL1A complex locus 1	Junction protein
HFE2	Hemochromatosis type 2	Iron metabolism
HHIPL1	HHIP-like 1	?
HLA, DRB-DQB	Major Histocompatibility Complex, class II, DQ beta1	Immune response in B lymphocytes, dendritic cells and macrophages
HSP90B1	Heat shock protein 90kDa beta (Grp94), member 1	Co-chaperone
IGSF5	Immunoglobulin superfamily, member 5	Cell-cell adhesion
KIAA1462	Uncharacterized protein	?
KLHL29	Kelch-like 29	?
LAMC2	Laminin, gamma 2	Non-collagen member of basement membranes
MIA3	Melanoma inhibitory activity family, member 3	Facilitates collagen transport
MRAS	Muscle RAS oncogene homolog	Signal transducer in cell growth and differentiation
MRPS6	Mitochondrial ribosomal protein S6	Ribosomal protein S6
MTAP	Methylthioadenosine phosphorylase	Polyamine metabolism
OPRM1	Opioid receptor, mu 1	Receptor for opioid peptides and analgesics
PDGFD	Platelet-derived growth factor D	Mitogenic factor
PHACTR1	Phosphatase and actin regulator 1	Regulates protein phosphatase 1
PLEKHO2	Pleckstrin homology domain containing, family O member 2	?

Gene symbol	**Gene name**	**Function**
PPAP2B	Phosphatidic acid phosphatase type 2B	Synthesis of glycerolipids and phospholipase D signal transduction
PPP1R3B	Protein phosphatase 1, regulatory subunit 3B	Regulates glycogen synthesis
RASD1 (SMCR3, PEMT)	RAS, dexamethasone-induced 1	Activator of G-protein signaling
RNF130	Ring finger protein 130	E3 ubiquitin-protein ligase
SH2B3	SH2B adaptor protein 3	Negative regulator of cytokine signaling
SLC12A9	Solute carrier family 12, member 9	Potassium/chloride transporters
SLC30A1	Solute carrier family 30, member 1	Zinc transporter
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	Regulate gene expression through chromatin modification
SMG6 (SRR)	Smg-6 homolog, nonsense mediated mRNA decay factor	Non-sense mediated mRNA decay
STK32B	serine/threonine kinase 32B	?
TCF21	Transcription factor 21	Transcription factors expressed in epithelial and mesenchymal cells
TCF7L2	Transcription factor 7-like 2	Transcription factor of Wnt signaling
TNIK	TRAF2 and NCK interacting kinase	Cytoskeletal regulation
UBE2Z (GIP, ATP5G1, SNF8)	Ubiquitin-conjugating enzyme E2Z	E2 ubiquitin ligase
WDR12	WD repeat domain 12	Maturation of the large ribosomal subunit.
ZC3HC1	Zinc finger, C3HC-type containing 1	?
ZFHX3	Zinc finger homeobox 3	Transcription factor regulating myogenic and neuronal differentiation
ZNF259	Zinc finger protein 259	Regulator of neuronal development