



Published in final edited form as:

Trends Cardiovasc Med. 2016 April ; 26(3): 260–265. doi:10.1016/j.tcm.2015.07.003.

Gene Scanning and Heart Attack Risk

Andreas S Barth, MD, PhD and Gordon F Tomaselli, MD[‡]

Department of Medicine, Division of Cardiology, Johns Hopkins University, Baltimore, Maryland, USA

Abstract

Coronary heart disease remains the leading cause of death in the Western World. The advent of microarray and next-generation sequencing technologies has generated multidimensional data sets allowing for new pathophysiological insights into this complex disease. To date, genome-wide association studies (GWAS) have identified 152 associated loci and 320 candidate genes contributing to the genetic risk of coronary artery disease (CAD) and acute myocardial infarction (AMI). The majority of single nucleotide polymorphisms (SNPs) mediate their risk by unknown mechanisms. A functional analysis based on Gene Ontology and KEGG pathways of candidate genes that are associated with CAD/AMI-SNPs showed the strongest evidence for genes regulating cholesterol metabolism. Additional clusters were significantly enriched for pathways which play prominent roles during AMI and the development of atherosclerotic plaques in vascular tissue, including focal adhesion/extracellular matrix interaction, TGF- β signaling, apoptosis, regulation of vascular smooth muscle contraction, angiogenesis, calcium ion binding, and transcription factors. A systems genetics approach which incorporates genetic risk with gene expression data, metabolomic data, and protein biochemistry into genome-wide network studies holds promise to elucidate the complex interplay between genetic risk and environmental factors for coronary artery disease.

Keywords

Coronary artery disease; Gene; Genome-wide association studies; Functional genomics

Genetic Risk of Coronary Artery Disease: From Monogenic Diseases to GWAS and Beyond

In monogenic disorders, including autosomal dominant hypercholesterolemia, caused by mutations either in LDL receptor [1, 2], apolipoprotein B (*APOB*) [3], or proprotein convertase subtilisin kexin type 9 genes (*PCSK9*) [4], a single mutation confers a very high

[‡]To whom correspondence should be addressed: Gordon F. Tomaselli, Division of Cardiology, Johns Hopkins University, 720 Rutland Avenue, Ross Bldg. 844, Baltimore, MD 21205. gtomase1@jhmi.edu. Tel (410) 955-2774/FAX (410) 502-2096.

Conflicts of Interest: None.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

risk of premature coronary artery disease (CAD) and acute myocardial infarction (AMI). Studies of these very rare monogenic, inherited disorders have proven invaluable in identifying genes that play key roles in the disease process. However, in the vast majority of cases, CAD and AMI are polygenic disorders which are additionally influenced by environmental factors. In 2007, two independent groups demonstrated the first genetic risk variant for CAD/AMI located on the short arm of chromosome 9, referred to as 9p21. This locus has now been replicated in 10 genome-wide association studies (GWAS) (Supplemental Table 1). Homozygotes and heterozygotes have a 50% and 25% increased risk for premature CAD, respectively. In recent years, GWAS have shed light on many of the specific genetic risk alleles for CAD/AMI. The National Human Genome Research Institute (NHGRI) Catalog of Published GWAS provides a publicly available, manually curated collection of published GWAS assaying at least 100,000 single nucleotide polymorphisms (SNPs, occurring at a frequency of greater than 1% in the general population) and all SNP-trait associations with $p < 1 \times 10^{-5}$ [5]. At the current time, 29 GWAS studies are included that identified more than 150 genomic loci associated with CAD and AMI in different populations (Supplemental Table 1). Most of these loci have small effect sizes with Odds Ratios (ORs) in the 1.1 – 1.3 range. In each case, despite the often large numbers of loci identified, only a small proportion of the phenotypic variance is explained: it has been estimated that the 152 known CAD-associated variants explain <10.6% of the genetic variation across the population [6]. While more common SNPs are often only associated with small increases in risk, the recent availability of whole-exome sequencing has enabled identification of rare variants (minor allele frequency <1%), which often have larger effects than common variants. For instance, Kathirasan et al., found rare *LDLR* and *APOA5* alleles conferring a 4.2 and 2.2-fold increased risk of myocardial infarction, respectively [7].

In addition to SNPs and rare variants, the heritability could partly be also explained by structural variations. These represent insertions, deletions, duplications, copy-number variants (CNVs), inversions and translocations, which typically affect DNA between 1 kilobases to several megabases in length and are mostly found in non-coding regions of the genome. Yet, tests of common (>1% allele frequency) and rare CNVs failed to identify associations with risk of AMI [8].

An additional mechanism of heritability that is not caused by changes in the DNA sequence involves alteration in the epigenome, including aberrant DNA methylation and histone modifications. Recent studies suggest that epigenome-wide changes are associated with CAD occurrence in men. Specifically, *COL14A1* and *MMP9* DNA methylation levels were associated with CAD and age of onset of CAD [9].

Functional Analysis of CAD-associated SNPs

Only a handful of SNPs are exonic and cause non-synonymous missense substitutions, thereby directly altering the amino acid protein sequence. Most of the genetic risk variants for CAD are located in DNA sequence that do not code for protein. Figure 1 shows that genomic location of 214 SNPs associated with CAD, highlighting that more than 85% of SNPs are located either in introns or intergenic regions. Given the non-coding nature of most SNPs associated with CAD/AMI (Figure 1), translating results from GWAS studies into

biological function has proven challenging. A functional analysis based on evidence from gene co-expression, protein-protein interaction networks, experimental evidence and text-mining showed strong evidence for a central cluster consisting of genes/proteins regulating cholesterol metabolism (Figure 2 and Supplemental Figure 1).

In addition to cholesterol metabolism, functional analysis based on Gene Ontology and KEGG pathways using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [10] indicated that SNPs associated with CAD and AMI are significantly enriched for genes involved in focal adhesion/extracellular matrix interaction, TGF- β signaling, apoptosis, regulation of vascular smooth muscle contraction, angiogenesis, and calcium ion binding (Figure 3), all of which have been implicated to play prominent roles during the development of an atherosclerotic plaque in vascular tissue. Of particular interest are gene classes for which the mechanistic link to the atherosclerotic disease process is not immediately obvious. For instance, gene loci involved in transcriptional processes like alternative splicing and RNA polymerase II transcription factor activity were overrepresented in CAD/AMI-SNPs. These transcriptional regulators could act in concert with aforementioned traditional risk factors like cholesterol metabolism, as, for instance, Cefalù and colleagues recently identified two splicing mutations affecting both the donor and the acceptor splice sites of the same intron of the *APOB* gene resulting in two truncated *APOB* fragments and the total absence of *APOB* [11]. Yet, for the majority of these loci, it is unclear whether they are merely a marker of atherosclerosis or whether they play a role in the pathogenesis of the disease.

Of note, given the 3 billion base pairs that comprises the human genome, genome wide association studies using a 500,000 SNP array will, on average, result in a genetic marker every 6,000 base pairs. Thus, rather than individual SNPs, GWAS studies identify haplotypes of SNPs associated with a higher risk for CAD/AMI. Hence, rather than exerting its effect on an adjacent exon of the same gene (via changing splice sites or effects on transcriptional efficiency), an intronic SNP could be in linkage disequilibrium with a second SNP that changes the function of a neighboring, but more remote gene. The intergenic location of the many of these common variants suggests that they mediate their increased risk indirectly by regulating other DNA sequences through intergenic transcription factor binding sites, enhancer regions, or noncoding RNAs. Of interest, one SNP identified in African-American individuals is located within a long non-coding RNA (LINC00333) [12]. However, previous analyses have shown that intergenic regions, despite harboring the largest fraction of trait/disease-associated SNPs blocks (TAS), were significantly depleted for TAS blocks [13]. This is consistent with the assumption that intergenic regions, although containing important regulatory sequences, have the smallest ratio of functional to total DNA. In line with this idea, we noted that many of the aforementioned pathways are significantly enriched in genic, but not in intergenic SNPs (Figure 3). The only pathway which was significantly enriched in genic and intergenic SNPs was cholesterol transport/metabolism (Figure 3).

Promises and Limitations of GWAS Data

GWAS data have played a prominent role in the discovery of novel candidate genes for CAD/AMI. As *a priori* knowledge of the biological pathways involved in the disease process is not required, GWAS are largely unbiased. Interpretation of GWAS data, however, is complicated by the fact that a genomic location rather than a specific gene is identified. As most SNPs are located in non-protein coding sequences which could influence the disease process by altering the expression of neighboring (or remote) genes, the biological role of a given SNP is usually not immediately obvious. Therefore, integration of additional data, including expression quantitative trait loci data, protein interaction, and experiments to identify epistatic interactions with modifier genes will be important to elucidate whether a specific locus plays a role in the disease process. Ultimately, it needs to be stressed that GWAS studies detect association and not causation and, thus, while a given SNP may be useful for risk prediction, it may not play a prominent role in the pathophysiology of the disease.

Additionally, for any particular trait, the cumulative effects of multiple SNPs only explains a small fraction of an individual's risk for the trait [6]. In order to account for the large number of SNPs and to minimize false positives, large study populations with hundreds or thousands of well-phenotyped subjects are required. Interpretation of the results also need to take into account different ethnic backgrounds, as results obtained in a population of European ancestry generally cannot be extrapolated to another, e.g. Asian or African-American populations.

One of the greatest promises in the field have been associated with the rise of high-throughput technologies, including next-generation sequencing (NGS). As NGS platforms become more widely available, they will help to address the question how rare variants contribute to the disease risk and help to pave the path to personalized medicine. Yet, large-scale exome and genome-wide sequencing is expected to yield numerous SNPs, missense mutations and variants of uncertain significance (VUS), found only in individual patients and families. The challenge of interpreting NGS data will be to define objective criteria for pathogenicity of individual variants. A much more sophisticated understanding of genetic and environmental modifiers will be necessary for the interpretation even of large effect size variants [14].

Transcriptomic Biomarkers of Acute and Recurrent Myocardial Infarction

Disease-associated SNPs and thus, the genetic risk, are stable throughout the course of a lifetime. However, the risk for a specific phenotype is influenced by complex gene–gene and gene–environment interactions which are missed by examining the genomic SNPs alone. Examination of the transcriptome has been proposed as a way to capture dynamic environmental influences and assess the risk for CAD/AMI. This approach is complicated by the fact that the transcriptome is highly dynamic and can exhibit significant changes within minutes [15]. Additionally, as causative genes are likely to be operational across several tissues [16], it is currently unclear which cell types would be most informative as biomarkers. Several blood cell subpopulations, including whole blood, monocytes, and

circulating endothelial cells have been proposed as targets for transcriptomic analyses. Given that these circulating cells play an active role in the pathogenesis of CAD [17, 18] and can be obtained by phlebotomy, blood-based gene expression signatures have the potential to facilitate risk assessment in CAD/AMI. Proof-of-concept studies in animal models have demonstrated that RNA blood biomarkers can serve as an appropriate surrogate for cardiovascular disease [19].

The first data on the predictive value of a blood-derived gene expression signature in CAD came from the CardioDx Corus[®]CAD signature. The GeneDx score, comprised of the peripheral blood cell expression levels of 23 genes, was initially developed to discriminate obstructive from non-obstructive CAD in >1,100 patients [20–22]. This score was later also found to be associated with the composite primary endpoint of major adverse cardiac events (MACE) and revascularization over 1 year follow-up [22]. To determine molecular signatures at the time of AMI associated with long-term outcomes, a recent whole-genome expression microarray analysis collected blood samples within 48 hours of an AMI. Patients were then followed for 18-months, comparing those with (n = 5) and without (n = 22) any recurrent myocardial ischemia [23]. Despite comparable clinical baseline characteristics, this analysis identified more than 550 differentially-expressed genes. Bioinformatic analysis of this differential gene-set for associated pathways revealed that recurrent AMI events are associated with modulation of cholesterol transport genes that include *ABCA1*, *CETP*, *APOA1*, and *LDLR* as well as a decreased expression of genes involved in the developmental epithelial-to-mesenchymal transition pathway [23]. The importance of cholesterol transport genes in the pathogenesis of CAD is reinforced in a second study, examining the differential transcript expression in monocytes of young men with premature CAD vs. those from controls matched for age, sex and smoking status without a family history of cardiovascular disease. By whole genome expression arrays, two (*ABCA1* and *ABGC4*) of only three differentially expressed genes in CAD cases vs. controls were again involved in cholesterol metabolism [24].

The value of peripheral blood gene expression signatures in defining the prognosis of CAD/AMI patients is also highlighted in a recent study, examining 338 patients with suspected or confirmed CAD undergoing cardiac catheterization [25]. When the patients were followed for a mean 2.4 years, a specific blood gene expression profile was associated with a significant risk of cardiovascular death. A significant overlap between this signature, related to inflammation and altered T-cell signaling, and gene expression in AMI was noted, suggesting that altered expression levels of both pro-inflammatory/pro-thrombotic and anti-inflammatory mediators could play a role in the disease process. Yet, given the small samples sizes of these studies, these results can only be regarded as preliminary until they are validated in larger cohort studies with the tens of thousands of participants needed to achieve a sufficiently high number of CAD/AMI events and to account for the inherent variability of clinical variables in human samples.

Is CAD Genetics ready for Prime-Time? Considerations for the Practicing Cardiologist

While a thorough history, including a family history, should be part of the initial medical evaluation, considerable uncertainty exists regarding the value of genetic testing. In the circumstance where several first-degree relatives are affected, a cardiovascular geneticist should be part of the medical team to identify possible Mendelian inheritance patterns. In cases with a high pre-test probability, i.e. when a Mendelian inheritance pattern of premature CAD is suggested, genetic testing may be justified. In most cases, however, modern genetic testing offers only little added value over traditional risk factors (family history of premature CAD, LDL – cholesterol level, smoking, diabetes mellitus, and hypertension). For instance, addition of a genetic risk score based on SNPs associated with CAD modestly improved discrimination and reclassification beyond traditional risk factors in white participants of the ARIC Study, however, no significant improvement was seen in the Rotterdam Study and Framingham Offspring Studies [26–28]. While several studies have confirmed the ability of a multilocus genetic risk score derived from validated markers to identify people at higher risk of CVD, these scores have not yet convincingly demonstrated clinical utility and improved patient outcomes [26–28] and randomized controlled trials to guide implementations of genetic testing are lacking. Thus, the 2013 Scientific Statement of the American Heart Association (AHA) on “Genetics and Genomics for the Prevention and Treatment of Cardiovascular Disease” state that the “*routine* genotyping of genetic markers in the primary prevention of cardiovascular disease (CVD) cannot be recommended at the present time”[29].

As CAD represents a polygenic disease with complex interactions at the genetic, epigenetic and environmental level, prospective and large collections of tissue samples in well-phenotyped populations which integrate different layers of genomic, epigenomic, transcriptomic, proteomic, and metabolomic data using a systems biology approach are necessary. Additionally, future studies are needed in order to prospectively examine whether the combination of traditional risk factors, common and rare gene variants identified by genome-wide association studies, peripheral blood gene expression and epigenetic data will translate to improved risk assessment and better patient outcomes, by identifying individuals at very high risk for adverse outcomes in whom aggressive lifestyle modification and therapeutic intervention should be considered.

Conclusions

Genome-wide association studies (GWAS) have identified 152 genomic loci with 320 potential candidate genes contributing to the genetic risk of CAD and AMI. A functional pathway analysis of candidate genes associated with CAD/AMI-SNPs showed the strongest evidence for genes regulating cholesterol metabolism, however, genes involved in focal adhesion/extracellular matrix interaction, TGF- β signaling, apoptosis, angiogenesis, and transcriptional processes were also enriched. The prominent role of altered cholesterol metabolism is also underscored by transcriptomic signatures associated with recurrent AMI events and poor long-term outcomes which include many cholesterol transport genes. Future

studies are needed to show that results obtained from GWAS and functional genomics translate into strategies for risk prediction and personalized interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. *Science*. 1986; 232:34–47. [PubMed: 3513311]
2. Brown MS, Goldstein JL. Receptor-mediated control of cholesterol metabolism. *Science*. 1976; 191:150–4. [PubMed: 174194]
3. Law A, Wallis SC, Powell LM, Pease RJ, Brunt H, Priestley LM, Knott TJ, Scott J, Altman DG, Miller GJ. Common DNA polymorphism within coding sequence of apolipoprotein B gene associated with altered lipid levels. *Lancet*. 1986; 1:1301–3. [PubMed: 2872432]
4. Abifadel M, Varret M, Rabès J-P, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D, Derré A, Villéger L, Farnier M, Beucler I, Bruckert E, Chambaz J, Chanu B, Lecerf J-M, Luc G, Moulin P, Weissenbach J, Prat A, Krempf M, Junien C, Seidah NG, Boileau C. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. 2003; 34:154–6. [PubMed: 12730697]
5. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014; 42(Database issue):D1001–6. [PubMed: 24316577]
6. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier J-B, Johansson A, Hall AS, Lee J-Y, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lytikäinen L-P, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013; 45:25–33. [PubMed: 23202125]
7. Do R, Stitzel NO, Won H-H, Jørgensen AB, Duga S, Angelica Merlini P, Kiezun A, Farrall M, Goel A, Zuk O, Guella I, Asselta R, Lange LA, Peloso GM, Auer PL, Girelli D, Martinelli N, Farlow DN, DePristo MA, Roberts R, Stewart AFR, Saleheen D, Danesh J, Epstein SE, Sivapalaratnam S, Kees Hovingh G, Kastelein JJ, Samani NJ, Schunkert H, Erdmann J, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2014; 518:102–6. [PubMed: 25487149]
8. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissono D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009; 41:334–41. [PubMed: 19198609]
9. Guay S-P, Brisson D, Mathieu P, Bossé Y, Gaudet D, Bouchard L. A study in familial hypercholesterolemia suggests reduced methylomic plasticity in men with coronary artery disease. *Epigenomics*. 2015; 7:17–34. [PubMed: 25687463]
10. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
11. Cefalù AB, Norata GD, Ghigliani DG, Noto D, Uboldi P, Garlaschelli K, Baragetti A, Spina R, Valenti V, Pederiva C, Riva E, Terracciano L, Zoja A, Grigore L, Averna MR, Catapano AL. Homozygous familial hypobetalipoproteinemia: Two novel mutations in the splicing sites of apolipoprotein B gene and review of the literature. *Atherosclerosis*. 2015; 239:209–17. [PubMed: 25618028]
12. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, Bennett F, Bowden DW, Chakravarti A, Dreisbach A, Farlow DN, Folsom AR, Fornage M, Forrester T, Fox E, Haiman CA,

- Hartiala J, Harris TB, Hazen SL, Heckbert SR, Henderson BE, Hirschhorn JN, Keating BJ, Kritchevsky SB, Larkin E, Li M, Rudock ME, McKenzie CA, Meigs JB, Meng YA, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet.* 2011; 7:e1001300. [PubMed: 21347282]
13. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362–7. [PubMed: 19474294]
 14. Arndt A-K, MacRae CA. Genetic testing in cardiovascular diseases. *Curr Opin Cardiol.* 2014; 29:235–240. [PubMed: 24717670]
 15. Barth AS, Kumordzie A, Frangakis C, Margulies KB, Cappola TP, Tomaselli GF. Reciprocal Transcriptional Regulation of Metabolic and Signaling Pathways Correlates with Disease Severity in Heart Failure. *Circ Cardiovasc Genet.* 2011; 4:475–483. [PubMed: 21828333]
 16. Hägg S, Skogsberg J, Lundström J, Noori P, Nilsson R, Zhong H, Maleki S, Shang M-M, Brinne B, Bradshaw M, Bajic VB, Samnegård A, Silveira A, Kaplan LM, Gigante B, Leander K, de Faire U, Rosfors S, Lockowandt U, Liska J, Konrad P, Takolander R, Franco-Cereceda A, Schadt EE, Ivert T, Hamsten A, Tegnér J, Björkegren J. Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. *PLoS Genet.* 2009; 5:e1000754. [PubMed: 19997623]
 17. Schmidt DE, Manca M, Hofer IE. Circulating endothelial cells in coronary artery disease and acute coronary syndrome. *Trends Cardiovasc Med.* 2015
 18. Ghattas A, Griffiths HR, Devitt A, Lip GYH, Shantsila E. Monocytes in coronary artery disease and atherosclerosis: where are we now? *J Am Coll Cardiol.* 2013; 62:1541–51. [PubMed: 23973684]
 19. Gerling IC, Ahokas RA, Kamalov G, Zhao W, Bhattacharya SK, Sun Y, Weber KT. Gene expression profiles of peripheral blood mononuclear cells reveal transcriptional signatures as novel biomarkers of cardiac remodeling in rats with aldosteronism and hypertensive heart disease. *JACC Heart Fail.* 2013; 1:469–76. [PubMed: 24622010]
 20. Wingrove JA, Daniels SE, Sehnert AJ, Tingley W, Elashoff MR, Rosenberg S, Buellesfeld L, Grube E, Newby LK, Ginsburg GS, Kraus WE. Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis. *Circ Cardiovasc Genet.* 2008; 1:31–8. [PubMed: 20031539]
 21. Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, Sager PT, Sehnert AJ, Yau M, Kraus WE, Newby LK, Schwartz RS, Voros S, Ellis SG, Tahirkheli N, Waksman R, McPherson J, Lansky A, Winn ME, Schork NJ, Topol EJ. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Ann Intern Med.* 2010; 153:425–34. [PubMed: 20921541]
 22. Rosenberg S, Elashoff MR, Lieu HD, Brown BO, Kraus WE, Schwartz RS, Voros S, Ellis SG, Waksman R, McPherson JA, Lansky AJ, Topol EJ. Whole blood gene expression testing for coronary artery disease in nondiabetic patients: major adverse cardiovascular events and interventions in the PREDICT trial. *J Cardiovasc Transl Res.* 2012; 5:366–74. [PubMed: 22396313]
 23. Suresh R, Li X, Chiriac A, Goel K, Terzic A, Perez-Terzic C, Nelson TJ. Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction. *J Mol Cell Cardiol.* 2014; 74:13–21. [PubMed: 24801707]
 24. Sivapalaratnam S, Basart H, Watkins NA, Maiwald S, Rendon A, Krishnan U, Sondermeijer BM, Creemers EE, Pinto-Sietsma SJ, Hovingh K, Ouwehand WH, Kastelein JJP, Goodall AH, Trip MD. Monocyte gene expression signature of patients with early onset coronary artery disease. *PLoS One.* 2012; 7:e32166. [PubMed: 22363809]
 25. Kim J, Ghasemzadeh N, Eapen DJ, Chung NC, Storey JD, Quyyumi AA, Gibson G. Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death. *Genome Med.* 2014; 6:40. [PubMed: 24971157]
 26. De Vries PS, Kavousi M, Ligthart S, Uitterlinden AG, Hofman A, Franco OH, Dehghan A. Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction

- of incident coronary heart disease: the Rotterdam Study. *Int J Epidemiol.* 2015; 44:682–688. [PubMed: 25953786]
27. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, Ridker PM. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA.* 2010; 303:631–7. [PubMed: 20159871]
28. Brautbar A, Pompeii LA, Dehghan A, Ngwa JS, Nambi V, Virani SS, Rivadeneira F, Uitterlinden AG, Hofman A, Witteman JCM, Pencina MJ, Folsom AR, Cupples LA, Ballantyne CM, Boerwinkle E. A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis.* 2012; 223:421–426. [PubMed: 22789513]
29. Ganesh SK, Arnett DK, Assimes TL, Assimes TL, Basson CT, Chakravarti A, Ellinor PT, Engler MB, Goldmuntz E, Herrington DM, Hershberger RE, Hong Y, Johnson JA, Kittner SJ, McDermott DA, Meschia JF, Mestroni L, O'Donnell CJ, Psaty BM, Vasani RS, Ruel M, Shen W-K, Terzic A, Waldman SA. Genetics and genomics for the prevention and treatment of cardiovascular disease: update: a scientific statement from the American Heart Association. *Circulation.* 2013; 128:2813–51. [PubMed: 24297835]
30. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013; 41(Database issue):D808–15. [PubMed: 23203871]

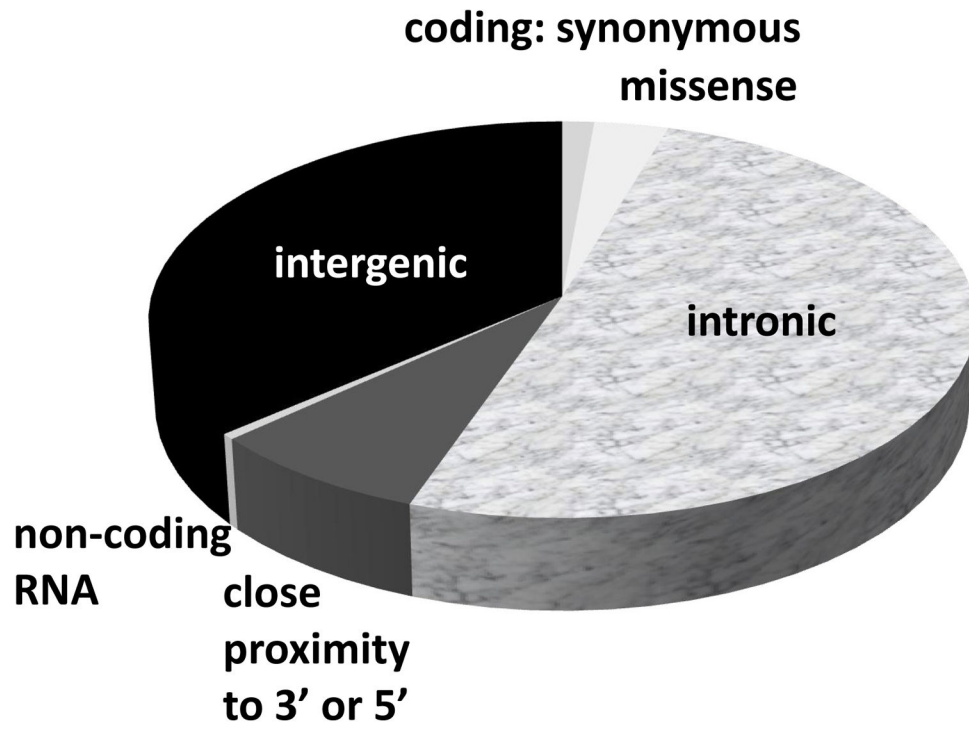


Figure 1. Relationship of 214 SNPs associated with coronary artery disease and myocardial infarction to nearby genes.

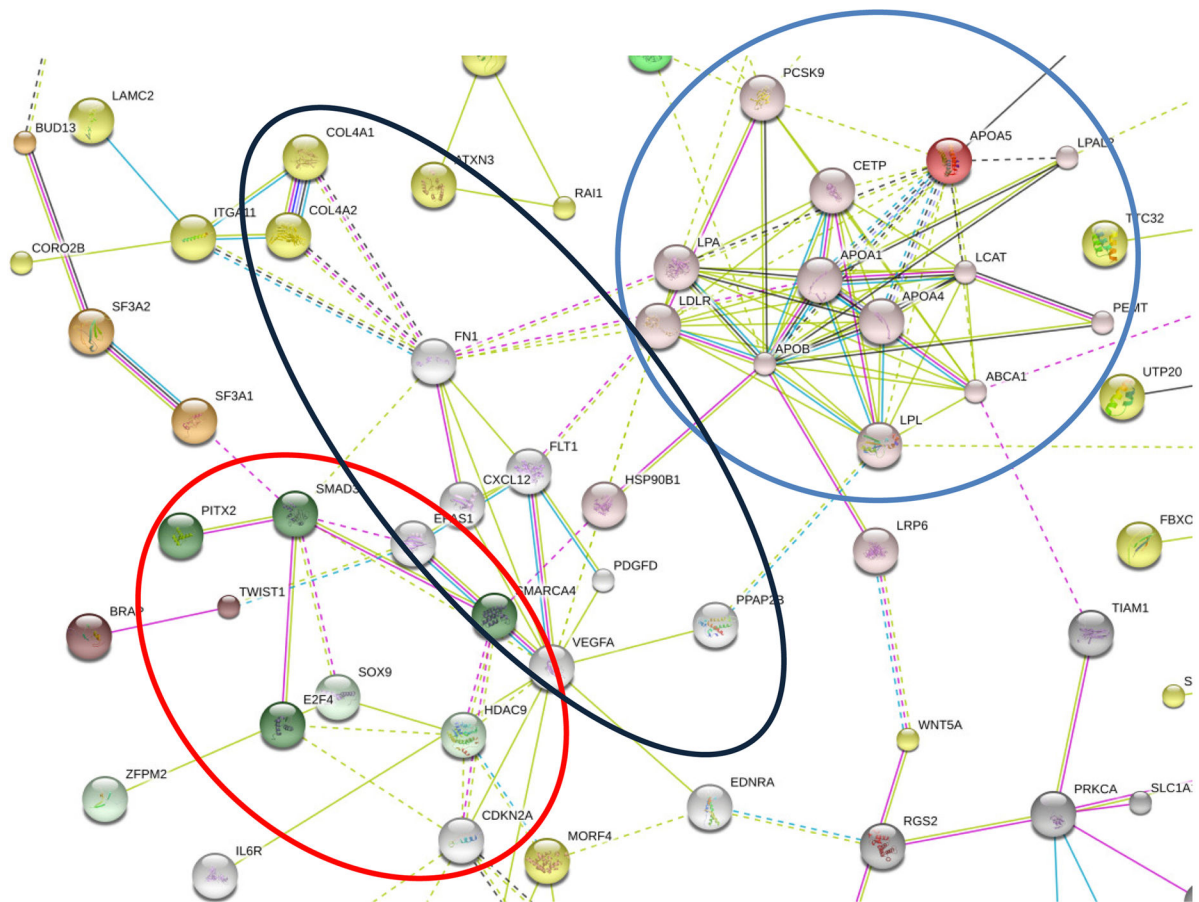


Figure 2.

Protein-protein interaction network of 320 genes associated with SNPs for CAD/AMI based on the STRING v9.1 database [30]. For better legibility, only proteins which display interactions with other proteins are shown (the full protein-protein network is shown in Supplemental Figure 1). There is strong evidence for a central cluster consisting of proteins regulating cholesterol metabolism (blue circle). Additional clusters include proteins involved in focal adhesion/extracellular matrix interaction (black circle: *FN1*, *FLT1*, *COL4A1*, *COL4A2*, *VEGFA*, *ITGA11*, *LAMC2*, *PDGFD*), and TGF- β signaling (red circle: *E2F4*, *CDKN2B*, *CDKN2A*, *SMAD3*, *PITX2*). Proteins interconnected by K-means clustering are colored. Edges, i.e. predicted functional links, consist of up to eight lines: one color for each type of evidence (STRING v9.1 integrates experimental, predicted and transferred interactions, together with interactions obtained through text mining).

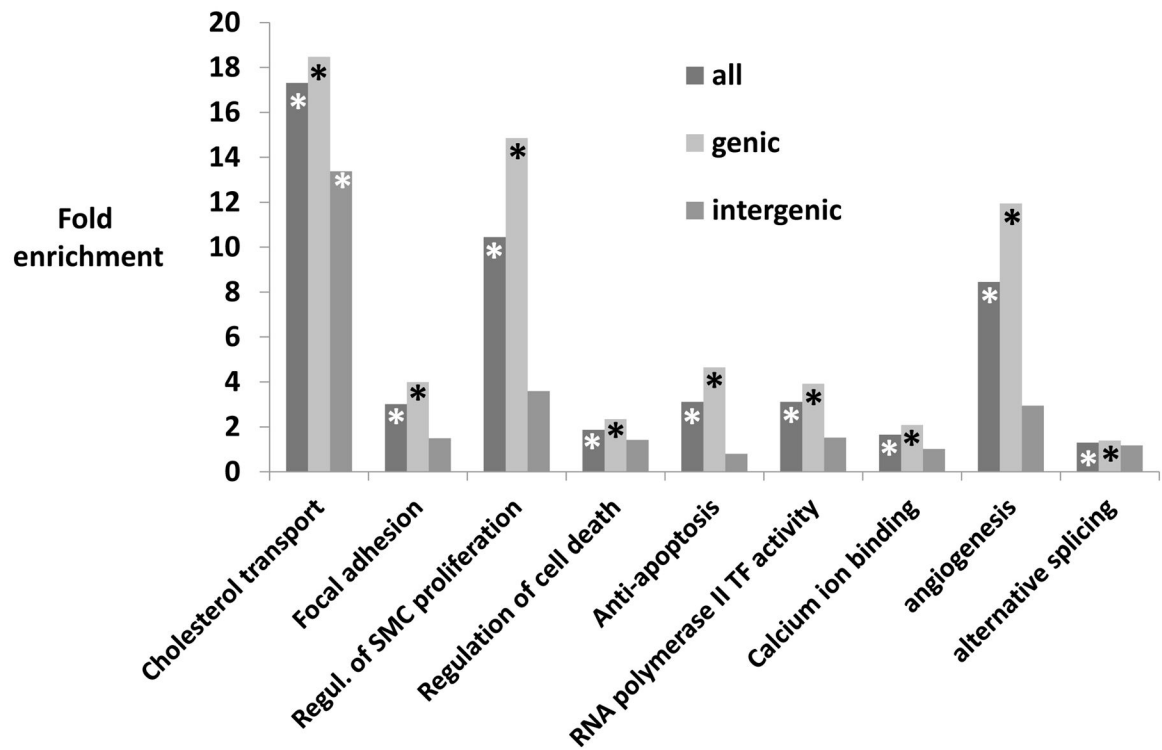


Figure 3.

Functional analysis with fold enrichment of genic vs. intergenic SNPs compared to expected rate against the whole genome. While cholesterol transport pathways were enriched in genic and intergenic regions, other pathways including focal adhesion, anti-apoptosis, regulation of smooth muscle (SMC) proliferation, angiogenesis, alternative splicing and RNA polymerase II transcription factor (TF) activity were enriched for SNPs in genic regions only. * indicates $p < 0.05$.