Trends in Cardiovascular Medicine 26 (2016) 260–265

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/tcm

Gene scanning and heart attack risk

*Andreas S. Barth, MD, and Gordon F. Tomaselli, MD**

Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, MD

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 multi-dimensional 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.

Key words: 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 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) (Supplementary 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

The authors have indicated that there are no conflicts of interest.

*Corresponding author. Tel.: +1 410 955 2774; fax: +1 410 502 2096.
E-mail addresses: gtomase1@jhmi.edu, andreas.s.barth@gmail.com (G.F. Tomaselli).

http://dx.doi.org/10.1016/j.tcm.2015.07.003
1050-1738/

(Supplementary Table 1). Most of these loci have small effect sizes with Odds Ratios (ORs) in the range of 1.1–1.3. 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, Kathiresan 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. Fig. 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 (Fig. 1),

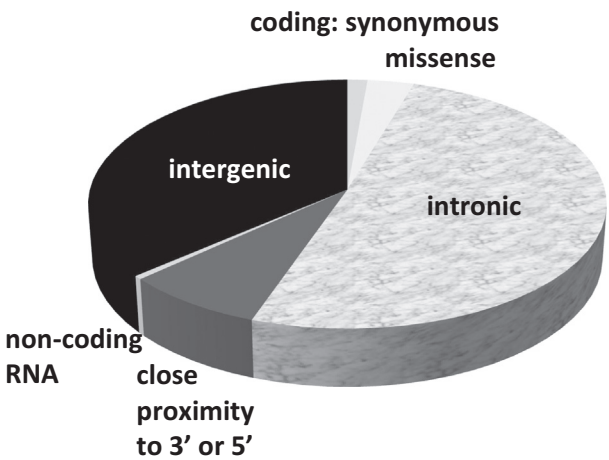

Fig. 1 – Relationship of known SNPs associated with coronary artery disease and myocardial infarction to nearby genes.

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 (Fig. 2 and Supplementary Fig. 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 (Fig. 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 et al. [11] 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*. 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 comprise the human genome, genome-wide association studies using a 500,000 SNP array will, on average, result in a genetic marker every 6000 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 non-coding RNAs. Of interest, a 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 (Fig. 3). The only pathway which was significantly enriched in genic and intergenic SNPs was cholesterol transport/metabolism (Fig. 3).