Genetic analysis of atherosclerosis: a research paradigm for the common chronic diseases

Eric Boerwinkle*, Darrell L. Ellsworth, D. Michael Hallman and Alan Biddinger

Human Genetics Center, University of Texas – Houston Health Science Center, PO Box 20334, Houston, TX 77225, USA

Received May 30, 1996

Rapid discoveries of novel and unexpected disease-associated genes for atherosclerotic coronary artery disease (CAD) are anticipated as genomic maps become more detailed and methods for mapping complex disease phenotypes become more refined. Although establishing association or linkage of a marker locus to a CAD susceptibility gene is an important first step, the long-term goal should be to define the underlying functional mutations and explore possible disease mechanisms, including the gene–environment interactions that culminate in clinically apparent disease. This review will define a contemporary research paradigm for study of the genetics of CAD and other common chronic diseases using the tools of modern molecular biology and human genetics.

INTRODUCTION

Great strides have been made in the prevention and treatment of atherosclerotic coronary artery disease (CAD), yet each year in the United States over half a million individuals die of this disease and far more suffer a first myocardial infarction (1). In all Westernized countries, CAD remains the number one contributor to morbidity and mortality in both men and women of all racial groups. Development and progression of atherosclerotic CAD is characterized by a steady accumulation of plaque constituents including inflammatory cells and molecules, complex lipids and lipoproteins, fibrin and smooth muscle. The majority of clinical disease results from mild to modest atherosclerotic lesions that abruptly progress to severe obstructions as a result of plaque rupture (2). Understanding the role of genes in CAD will improve our understanding of its etiology and facilitate early identification of individuals at increased risk of disease. Insight into the genetic basis of chronic disease etiology will have immediate impact by suggesting novel therapeutic approaches and aiding new drug discovery.

This review presents a research paradigm for the genetics of the common chronic diseases, particularly atherosclerotic CAD, and highlights the results of recent studies that demonstrate the utility of this paradigm in investigations of the genetic component of CAD. We will focus on research targeting common genetic variation that impacts on interindividual risk to CAD. This review is not intended to be comprehensive. It will not address recent developments in gene therapy for CAD, nor will it focus on rare mutations that may be catastrophic to individuals but have little impact in the population at large. Although CAD is the target of pioneering work in gene therapy, the field is in its infancy and success has been limited, though promising. Recent reviews and perspectives in gene therapy for CAD (3,4) and rare inborn errors of metabolism leading to CAD (5,6) have been published.

A RESEARCH PARADIGM

Figure 1 presents a cascade of current and future research activities aimed at defining the genetic etiology of CAD and assessing the utility of this information for preventing and treating disease. Although no one line of investigation is more important than another, they follow a logical order and are related to one another in a scientifically consistent manner. Genetic linkage and association analyses using hundreds of marker loci spanning the human genome define the location of CAD susceptibility genes. Once localized to a particular region, genes can be identified using knowledge of the metabolism and pathophysiology of CAD (a candidate gene approach) or by positional cloning strategies with the assistance of the developing human transcription map. The next step in this research paradigm, defining functional mutations within the localized gene, is one of the most intellectually challenging problems confronting human geneticists. Once candidate functional mutations have been identified, functional or causal relationships are most likely to be established using in vitro studies or animal models. Following identification of genes and functional mutations, the goal of the research shifts from gene/mutation discovery to defining the impact of this variation on CAD in the population at large. Large prospective epidemiologic studies are best suited for defining the ability of genetic information to predict disease risk. Gene therapy and gene pharmacological studies attempt to modify the expression of the gene and its effect on CAD risk. Clearly, CAD is the result of interactions among numerous environmental and genetic factors. Therefore, the effect of an allele measured in one environment may not be the same as its effect in another. Studies of genotype–environment interaction are necessary before implementing genetic information for treatment or prevention of CAD.

*To whom correspondence should be addressed
Localizing genes

Genetic linkage analysis is the most commonly applied method for localizing genes contributing to human disease to a chromosomal region. Increasingly refined and comprehensive maps of the human genome have accelerated progress in linkage analyses. The latest published version of the Genethon human linkage map contains 5264 short tandem repeat polymorphisms (average heterozygosity of 70%) spanning a sex-averaged distance of 3699 cM with an average interval of 1.6 cM between adjacent markers (7). Traditional methods of linkage analysis based on LOD scores require the mode of inheritance of a disease to be known \textit{a priori}, though alternatives have been proposed, such as combined linkage and segregation analysis (8). The mode of inheritance of CAD is complex and heterogeneous among families; CAD aggregates but does not segregate in pedigrees (9). Applying LOD scores to CAD may inflate type I and type II error rates for both linkage and exclusion. Therefore, robust ‘non-parametric’ methods of linkage analysis are preferred for complex diseases. Research strategies are available for both qualitative traits (e.g. CAD) using affected relative pairs (10), and quantitative traits (e.g. cholesterol levels) using entire pedigrees or pairs of relatives (11). These methods are all based on the principle that relatives sharing alleles identical by descent at a marker locus linked to the susceptibility gene will be more similar for the trait of interest.

Greenberg (12) compared linkage and association analyses for localizing disease genes when the genes under consideration increase the susceptibility to disease but are neither necessary nor sufficient to cause it. He concluded that if the relative risk of the disease given the susceptible genotype is small, then the chance of finding genetic linkage is correspondingly small. In this case, association analysis will be much more sensitive than linkage analysis. However, proper design of genetic association studies is difficult because heterogeneity and population substructure can produce misleading results. Recently, transmission disequilibrium tests (TDT) have been proposed to avoid some of the problems inherent in most linkage and association studies (13,14). In its simplest form, the TDT compares allele frequencies in cases with the frequencies of the non-transmitted alleles in their parents (Fig. 2). The TDT greatly reduces the likelihood that any allele frequency differences between cases and non-cases might be due to poorly chosen controls or unsuspected population substructure.

Two examples of linkage analyses for CAD risk factors are presented below; one used a candidate gene approach; the other, a genome-wide search. Jeunemaitre et al. (15) reported significant linkage between a highly polymorphic microsatellite marker near to the angiotensinogen gene and a locus influencing essential hypertension. Significant excess sharing of marker alleles was observed in a sample of 215 sibling pairs with essential hypertension. The relationship was more pronounced when the analysis was restricted to siblings with severe or early-onset hypertension. Implicating functional DNA variation in genes of the renin–angiotensin system is of particular importance because of its central role in maintenance of salt and fluid homeostasis and regulating blood pressure, and because of the role of the vascular renin–angiotensin system in maintaining vascular tone and regulating cell growth and migration (16).

There is currently only one published study describing the results of a genome-wide linkage analysis for a late-onset common chronic disease. Hanis et al. (17) used 500 highly polymorphic markers with an average distance between adjacent markers of <10 cM and robust methods of linkage analysis in a sample of 330 Mexican-American affected sibling pairs to localize susceptibility genes for non-insulin-dependent diabetes mellitus (NIDDM). NIDDM, hyperglycemia and insulin resistance are major risk factors for CAD in the Mexican-American population (18). Eleven markers were consistently and significantly related to NIDDM. One marker, D2S125, showed highly significant linkage to NIDDM and appears to be a major contributor to NIDDM susceptibility in this population. Surprisingly, a number of candidate genes showed no evidence of significant linkage to

---

**Figure 1.** Research paradigm for the genetic analysis of CAD.

**Figure 2.** Schematic representation of family-based case-control association studies. A, C, case alleles; B, D, control alleles.
NIDDM in this sample. Multipoint exclusion analyses indicated that 71% of the genome could be excluded as containing a locus having an effect large enough to increase the relative risk of disease by 1.6 in those individuals having the susceptible genotype. However, only 5% of the genome could be excluded as containing a locus having an effect large enough to increase the relative risk of disease by only 1.2. The work presented by these authors establishes a strong foundation for future molecular genetic studies of NIDDM and CAD in this population.

**Gene identification**

Phenomenal progress in the identification of monogenic disease genes has been achieved through positional cloning strategies that locate genes with only minimal information about their functions. However, alleles associated with increased susceptibility to multifactorial diseases are often common in the general population, and variation within any given gene may account for only a small portion of the total phenotypic variance underlying the affliction. Therefore, traditional methods of locating Mendelian disease genes may not be entirely feasible for genetically complex disorders. Locating a single gene within a chromosomal region rarely <2–5 Mb in length is severely hampered by the absence of chromosomal rearrangements or deletions that define the candidate region and by the often subtle nature of functional sequence variation which may be located outside the coding region (e.g. 19).

A positional candidate approach (20) involving linkage analysis of multiple affected families or sibling pairs to localize susceptibility genes to subchromosomal regions, followed by an intensive search for logical candidates within the interval, may prove to be an efficient strategy for locating quantitative trait genes. A proliferation of sequence data readily accessible through the GenBank/EMBL, Genome Data Base (GDB) and Human cDNA Data Base will greatly expedite the search for candidate susceptibility loci. Expressed sequence tags (ESTs) are powerful tools for identifying and characterizing genes contributing to cardiovascular disease by supplementing increasingly dense transcript maps (21) and identifying potential new candidate genes from tissue-specific libraries (22).

**Probable functional mutations**

Once potential candidates have been identified, an exhaustive search must be conducted for DNA variation within the candidate region. A variety of scanning techniques are available for the initial detection of unknown mutations (23). Most methods rely on the characteristic secondary structures, dependent on nucleotide sequence, formed by self-annealed single-stranded DNA (24,25). Following initial detection of variation, the precise nature of specific variants is usually characterized by DNA sequencing. Modern methods of high throughput DNA sequencing permit the search for DNA variation in linked regions to proceed by direct DNA sequencing (26) without first carrying out single-stranded conformation analysis. This direct sequencing method has the advantage of locating all variants in the sequenced region, and the nature of the nucleotide substitutions is immediately apparent.

A difficult task in the study of complex disorders is distinguishing functional variants from neutral polymorphisms. Brute-force DNA sequence analysis in subjects with and without disease will reveal an array of variants whose relationships to CAD are unknown and difficult to assess. Lack of a one-to-one relationship between genotype and phenotype in CAD means that some individuals who are at increased genetic risk will not develop clinically apparent disease, while other individuals will incur disease in the absence of increased genetic risk. Several approaches may be used to pinpoint the causative variant(s) or reduce the number of candidate variants requiring further investigation of expression and function. Multiple DNA variant association analysis (MVAA) (27) characterizes variants at all polymorphic sites throughout the candidate region as disease-associated (+) or non-associated (−). Linkage groups of variants (haplotypes) are inferred from observed genotype patterns, and disease-associated haplotype(s) are identified by comparing haplotype frequencies in case and control groups. The effects of individual variants or combinations of variants are then distinguished by comparing relative risks among haplotypes. Variant sites which are candidates for being functionality related to phenotypic differences are those that are associated with significant allelic effects after controlling for background haplotype variation. Templeton et al. (28) proposed a strategy based on phylogenetic analyses which will lead to detection of candidate functional mutations. The most likely evolutionary history of the present day observed haplotypes is first reconstructed. This cladogram is then used to define a set of nested analyses that detects associations between a phenotype and measured genotypes, and localizes the phenotypic effects to specific branches of the phylogeny. Therefore, this analysis identifies groups of individuals that are similar for silent marker DNA variation but whose haplotypes differ with respect to their phenotypic effects. Direct sequence analysis of these haplotypes should identify the potential functional sites.

**Experimental systems**

The appropriate in vitro system to use in characterizing functional mutations depends on the nature and location of variants within the gene. Candidate mutations in suspected regulatory regions (including the promoter and enhancer sequences) can be linked to reporter genes such as luciferase (29,30) and transfected into eukaryotic cells for assaying functional differences. Locations of amino acid substitutions that alter protein structure, the spatial position of variants relative to known functional sites and alternate conformations can be obtained from the Molecular Modelling Database (NCBI) and viewed using software programs such as RASMOl (31). However, direct assays of protein function are the preferred method for assessing the functional significance of common amino acid substitutions.

Animal models have significantly advanced our understanding of numerous single gene disorders in humans. However, dissecting contributing genetic factors and distinguishing between causation and correlation are more difficult for complex diseases (e.g. CAD) having multiple genetic and environmental components. Despite initial concerns regarding the biological relevance of ‘knockouts’ to human diseases with complex etiologies and the validity of extrapolating effects of single gene disruptions to multifactorial conditions, gene targeting is improving our ability to examine the often modest effects of complex disease genes independently of the effects of other loci or environmental factors that may influence susceptibility (32). Recent advances in transgenic technology have enabled researchers to examine the effects of altering the amount of gene
product by manipulating gene copy number (see reviews, see 33,34). For example, Smithies and Kim (35) have utilized
gap-repair gene targeting to introduce multiple functional copies of
the human angiotensinogen gene in mice resulting in
progressive and significant increases in the amount of
angiotensinogen in the plasma. The extent to which the results
from in vitro and animal model studies mimic the human
condition should always be questioned. However, the availability
of well-designed and appropriate experimental systems enables
the testing of hypotheses not possible in humans and will suggest
other hypotheses that can only be adequately addressed by direct
study of humans (36).

Gene-targeted and transgenic animals provide information on
the metabolic functions of candidate genes and the physiological
relationship between genetically determined alterations in gene
dosage and predisposition to disease. The feasibility of testing
human functional variants in animal systems will depend on
refining strains to eliminate phenotypic effects caused by
polymorphic differences among animals, and on gene
replacement technologies allowing intact disease-associated or
non-associated alleles to be introduced in defined copy number.

Developing animal models useful in revealing subtle differences
in the effects of naturally occurring functional variants will
greatly enhance efforts to identify genes associated with the
incidence and progression of complex diseases.

Disease prediction

Once functional mutations have been identified, one of their
greatest utilities is in primary prevention programs. There are
many reasons why measures of DNA variation may improve the
ability to predict disease risk beyond that provided by established
risk factors: (i) an individual’s genotype does not change
throughout life (barring somatic mutation); (ii) the genotype is not
influenced or changed by the disease process itself; (iii) DNA
variation can be measured more accurately than most
intermediate predictor traits; (iv) the intermediate physiological
and metabolic traits underlying disease may be unknown or
inaccessible to measurement; and (v)the ability of other traits
(e.g. weight) to predict disease may be genotype dependent.

Prospective studies of genetic effects on cardiovascular disease,
which will be necessary in developing accurate predictions of
genetic risk of CAD, have been relatively rare. The gene whose
effects on normal lipid variation and CAD risk are best understood
is apolipoprotein (apo) E. ApoE is a structural component of
several lipoprotein species and plays a major role in lipid
metabolism through cellular uptake of lipoprotein particles by
apoE-specific and apoB/E receptors on the liver and other tissues
(37). Human apoE is polymorphic with three alleles, ε2, ε3 and ε4
coding for three isoforms E2, E3 and E4, respectively.
The most extensively studied gene in this regard is again apoE.
Several studies have reported different responses to
lipid-lowering drugs among apoE types. In studying the response
of heterozygous familial hypercholesterolemia (FH) patients
to probucol, a cholesterol-lowering antioxidant, Nestruet et al. (47)
found that 11 subjects with at least one ε4 allele showed
significantly greater change than did 35 E3/3 homozygotes;
change did not differ among apoE types in 39 patients with
hypercholesterolemia not attributable to FH. There is other
evidence that drug dosage may influence the effect of apoE on
statin response. Ojala and colleagues (48) found that after 6 weeks
of lovastatin at 20 mg/day, non-FH hypercholesterolemic subjects
with apoE3/4 genotypes showed significantly smaller reductions
in plasma total and LDL cholesterol levels.

Genotype by environment interaction

The common chronic diseases, such as CAD, result from
interactions among numerous genetic and environmental factors,
but analyses of the putative interactions are infrequent. The term
‘interaction’ may be used in several different ways, and the
interested reader should consult the articles by Rothman et al.
(45) for a general discussion of the concept of interaction as it is
applied to the biological and biomedical sciences. The immediate
value of studying genetic interactions in complex diseases lies in
elucidating pathology and in guiding treatment. Studies of gene
effects on response to treatment represent one example of
gene–environment interaction. The efficacy of most cardio-
vascular disease risk factor treatments including blood pressure
and cholesterol-lowering medications is characterized by extensive
interindividual variability. Often the sample standard deviation of
a drug’s effect is at least as great as the average (e.g. 46).
Undoubtedly, a portion of the interindividual variation in response
to treatment is attributable to genetic factors. Knowledge about
the genes underlying differences in response to risk-lowering
treatments will enable physicians to match treatments to
individuals who are most likely to respond.
in low density lipoprotein (LDL)-cholesterol than those with apoE3/3 genotypes, while FH patients with apoE3/4 types showed significantly lower total cholesterol response after treatment at 40 mg/day. When doses in each patient group were doubled, the differences among apoE phenotypes disappeared. When FH and non-FH patients at 40 mg/day were pooled, no significant effects of apoE type on response were found.

HORIZONS

As health care costs continue to rise and the focus of medicine shifts to early detection of disease and primary prevention, methods for identifying asymptomatic individuals at risk for common diseases and the development of more efficacious interventions have become paramount. Recent developments in genetic maps, analytical methods and imaging technologies have provided the tools for unraveling the complexities of these disorders through examination of novel intermediate phenotypes. Measures of established CAD risk factors, including lipid metabolism, systolic blood pressure and lifestyle have been shown to be poor predictors of CAD deaths or coronary luminal narrowing (49,50). Recent developments in electron beam computed tomography and magnetic resonance imaging technology have opened new avenues for measuring structural and functional changes in arterial walls which precede the development of obstructive disease (51,52). Combined with the research paradigm and methods discussed in this review, these new technologies for ‘visualising’ the atherosclerotic plaque open new frontiers of research. These methods will permit direct study of the genetic and epigenetic factors contributing to unstable atherosclerosis that rapidly progresses into life-threatening events.

ACKNOWLEDGEMENTS

The authors would like to thank their many collaborators and friends in the Rochester Family Heart Study, The ARIC Study and The Starr County Health Study for years of support and encouragement. This work was supported by grants HL51021 and HL54481 from the National Heart, Lung and Blood Institute.

REFERENCES
