Frontiers in Nephrology: Genomic Approaches to Understanding the Molecular Basis of Atherosclerosis

Euan A. Ashley, Joshua M. Spin, Raymond Tabibiazar, and Thomas Quertermous

Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California

ABSTRACT

Atherosclerosis is a complex multicellular disease that is responsible for pathology in various organ systems. The understanding of its initiation and progression has been enhanced in recent years by the application of high-throughput genomic tools such as the microarray. Increasing in genomic coverage, such tools allow a view of the disease unaffected by previous conjecture as to the primary signal of interest. New statistical tools and pathway modeling techniques have established definitively for the first time the central role of inflammation in this process. This article reviews the genomic literature relating to atherosclerosis from cell culture, animal models, and human tissues. In this comparison of these differing approaches, the available data are synthesized to reach a new understanding of the complex interplay between vascular wall and immune system components.


Atherosclerosis is a chronic disease characterized by abnormal accumulation of smooth muscle cells (SMC), lipid, and necrotic debris in the intima of the elastic arteries, which leads to narrowing, ischemia, and infarction of the tissue supplied. It is the primary cause of heart disease and stroke, diagnoses that present not only a significant health problem in the United States and worldwide but also a considerable economic challenge. Part of that challenge is that patients with atherosclerosis represent a heterogeneous group of individuals, with disease that progresses at different rates and in distinctly different patterns. Although epidemiologic risk factors have been defined at a population level, the mechanisms by which changes in these variables contribute to the primary disease process in the blood vessel wall are unclear. An expansion in our understanding of this disease at the genomic level will enable better approaches to the diagnosis of individuals with active vascular disease and the identification of pharmacologic therapies that are specific for the primary disease process.

It has been appreciated for some time that inherited genetic variation contributes significantly to the risk for atherosclerosis.1 Studies from the Swedish twin registry2 established relative hazards for death from coronary heart disease when one twin died of coronary heart disease before the age of 55 yr. For male twins, the hazard ratio was 8.1 for monozygotic and 3.8 for dizygotic twins. In female twins, the effect was even more dramatic: 15 for monozygotic versus 2.6 for dizygotic twins. Despite the weight of this evidence, however, insights regarding the genetic programs that are responsible for these effects remain elusive.

Both genetic and genomic approaches have been used to assess nonenvironmental determinants of atherosclerosis. While candidate gene approaches have yielded some insight both into human population genetics and molecular pathways, the advent of highly parallel techniques such as the microarray provides a paradigm change in our approach. Whereas previous studies relied on earlier data or the special interest of an investigator to determine the focus end point of a given study, newer techniques allow assay of a representation of the entire genome.

Despite the sophistication of these tools, the disease process itself provides further challenges. Atherosclerosis is a multicellular, multisystem disease with local and circulating components. Cell culture models of vascular wall and immune cells offer the most control and allow assessment of the contribution and functioning of individual cell types. Animal models provide the opportunity to control the multicellular environment and assess the evolution of the disease but suffer from manifesting differences in critical ways from human disease. Human studies are maximally relevant but least controlled and provide only limited access to late-stage tissue.

We argue in this article that the new tools of high-throughput genomics offer an opportunity to change the paradigm in atherosclerotic disease research. We believe that combining these techniques with cellular, animal, and human disease models gives us the greatest opportunity...
to understand the true nature of the disease.

HIGHLY PARALLEL GENOMIC ASSAYS

Although radioactive membrane-spotted cDNA “macroarrays” existed in the 1980s, the idea of multiplexing hybridization experiments to assess simultaneously the expression levels of many genes was borne out of Brown’s3 Stanford laboratory in the mid-1990s. In these early experiments, almost 50 cDNA were robotically placed on one glass slide and a dual dye fluorescence system was used for readout from hybridization experiments using micrograms of RNA. In the past 10 yr, there has been an explosion in technological advancement as private-sector competition has led to exponential increases in the capacity of arrays, the sensitivity of fluorescence detection, the ability to amplify small amounts of RNA, and the ability to deal with the resulting data. Such has been the progress that in 2004, several companies launched “whole genome” expression arrays (albeit that estimates regarding the true number of human genes were shrinking during this period). Meanwhile, high-throughput genotyping of genomic DNA lagged behind somewhat, plagued by the low molar concentration of genomic DNA and difficulties with primer-dimer formation in multiplex PCR; however, driven by the desire to describe comprehensively common human genetic variation and the haplotype structure of the human genome (Hapmap project), solutions such as the Molecular Inversion Probe and Goldengate assays were developed. Both use extension ligation reactions followed by PCR to avoid issues of primer-dimerization. The advance resulted in the first whole-genome genotyping platforms in approximately 2005. Because the number of predicted common (>1%) single-nucleotide polymorphisms (SNP) is greater than 2 million, these platforms achieve their whole-genome label by placing one “tag” SNP for each haplotype on a single slide (approximately 0.5-m probes).

The pace of technological innovation continues in the expression array space. Newer exon-specific arrays allow splice variants to be quantified, and tiling arrays can identify novel transcriptional elements, intronic and antisense transcription. Meanwhile, genomic DNA can be resequenced using tiling technologies and transcriptional control investigated. The latter is achieved by focusing on nonexonic regions, DNA sequences modified or bound by regulatory proteins, and histone modification using chromatin immunoprecipitation on chip technology. Experiments looking at DNA methylation show that genomic hypomethylation occurs during atherogenesis in human, mouse, and rabbit lesions, both correlating with heightened transcriptional activity and increasing with lesion severity.4,5 In addition, histone hyperacetylation (leading to global increases in transcriptional activation) is present in atherosclerotic lesions and in activated vascular SMC.6

EVOLUTION OF BIOSTATISTICS

As technological advancement has allowed many more genes to be assayed simultaneously, biostatistics evolved to deal with these challenges. Initial approaches were extensions of those used in low n number molecular biology experiments. Genes whose expression had changed by a given “fold” were reported as most important, but this inherently flawed technique took no account of the principles of variation and probability. In addition, investigators learned that the largest values were most likely to be error while biologically significant changes could be “significant” even at low fold changes below the “noise” cutoff for some of the earlier experiments.7

The natural evolution was then to apply traditional statistics such as the t statistic or F statistic to the large data sets. Controlling the type I error (false-positive rate) appropriately is more challenging when the dependent variables number 20,000 genes or 500,000 SNP. P < 0.05 would ensure that only one of 20 findings was “significant” by chance alone, but when dependent variable numbers are large, a correction factor needs to be applied. This has traditionally been that recommended by Bonferroni.8,9 This sets the α value for the entire set of n comparisons equal to α by making the α value for each comparison equal to α/n; however, this control of the family wise error rate turns out to be overconservative when more than a single-digit number of variables are expected to be significant. Because of this, researchers using expression profiling tools turned toward the idea of controlling false discovery, first described by Benjamini et al.10 In fact, the original method described in 1995 also turns out to be overly restrictive, and several variations have been suggested.11 Most prominent among them have been those of Tibshirani and colleagues,12–15 who described a permutation-based bootstrap method for controlling false discovery and extended the idea to include the “q” value—a measure of significance assigned to an individual feature and stated in terms of false-discovery rate. Whereas a false-positive rate (P value) speaks to the frequency with which features that are null are called significant (taking no account of the characteristics of other elements tested), the false-discovery rate is stated in reference to features that are called significant and gives the percentage of features that turn out in fact to be null. It is interesting that the genetics literature that evolved from linkage analysis where very few loci would be expected to be significant has tended to favor a conservative “false positive” approach, whereas the expression profiling literature (where researchers believe hundreds of genes may be changed in a relevant way) has favored false discovery.

Defining an appropriate cutoff for significance of a set of genes in a microarray experiment is, however, only the beginning. Early on, researchers found themselves challenged with long lists of genes and quickly developed methods for making more biologic sense of them. One early technique, which was rejuvenated by Eisen et al.,16 was hierarchical clustering. In this technique, one of a variety of distance...
measures is used to identify groups of genes with similar exression. These are then displayed diagrammatically in a tree-based format. Other cluster algorithms may in fact perform better than classical hierarchical clustering, and these include k-means clustering and self-organizing maps, both of which begin with a predetermined number of cluster centroids.17

Whereas clusters of genes imply similar expression, use of the Gene Ontology consortium’s classification system has been a widely used alternative for grouping genes based on similar known functions. In a common version of this technique, a “significant” group of genes derived from difference analysis is tested against the rest of the genes for the over-abundance of certain categories (ontologies) of biologic processes, cellular components, and molecular functions, usually using a variation of Fisher exact test. This data reduction allows functional signals to be derived from long lists of genes. Although ontologies can be overlapping, overspecific, and beholden to previous investigation, this represents a powerful method to understand the principal biologic signals of a set of a differentially expressed genes.

Applying more sophisticated structure to gene lists is the goal of pathway analysis. There are multiple forms of this, but a common one uses a variation of the Fisher statistic to test for overabundance of members of biologic pathways in a given list of significant genes. Both turnkey proprietary (e.g., Ingenuity, Pathway Studio) and publicly available pathway lists (Biocarta, Kyoto Encyclopedia of Genes and Genomes) exist and can be used for this purpose. A further layer of complexity can be added via physical protein interaction databases such as BIND (biomolecular interaction network database) or language-parsing algorithms that derive relationships from genes that are “connected” by sentences in the literature. Such techniques allow a sliding scale of confidence with which certain genes can be stated to be functionally related. Scale-free network theory leads to the natural prediction that the more connected genes may be functionally more relevant, more likely to be rate limiting, and, perhaps, ideal targets for therapeutic intervention.

APPLICATION TO Atherosclerosis

Atherosclerosis is a multicellular disease process that is initiated by the accumulation of LDL particles in the blood vessel wall. As these particles become oxidized, endothelial cells (EC) are stimulated to upregulate adhesion molecules and to secrete chemokines, which attract circulating monocytes and T cells to migrate to the intima, where the former mature into activated macrophages. Activated macrophages themselves upregulate scavenger receptors that allow them to ingest oxidized LDL and mature into fat-laden “foam” cells. In early-stage atherosclerosis, the “fatty streak” is made up predominantly of these cells. As the streak enlarges into a plaque, a fibrous cap forms when SMC from the media migrate to the intima and de-differentiate to secrete fibrous matrix. Later, proteolytic enzymes secreted by inflammatory cells can weaken this fibrous cap, leading to erosion or rupture and exposing “tissue factor,” a potent coagulant, to the blood. This rupture process and the resulting thrombosis are the central pathophysiologic mechanism of myocardial infarction. While the fibrinolytic system attempts to lyse the clot, this process can be overwhelmed, leading to complete vascular occlusion and downstream tissue and cell death.

Traditional techniques have allowed description of this disease process in cellular terms. Still little is known, however, regarding the molecular phenotype of differentiating SMC, the role of the endothelium, and genetic determinants of the susceptibility of individuals to these processes. In an attempt to understand this process, researchers have used multiple atherosclerosis models and cell systems (Table 1). Here, we summarize the principal approaches.

IN VITRO TRANSCRIPTIONAL PROFILING OF VASCULAR WALL COMPONENTS

Although in vivo genomic studies can provide a snapshot of vessel wall gene expression, the inherently complex and uncontrolled environment may make drawing biologic conclusions about process difficult. In vitro microarray studies have begun to examine the genetic pathways that are involved in the response of SMC and EC to inflammatory and growth-mediating factors, as well as SMC differentiation. The primary advantages and disadvantages of these approaches both arise from the same basis: They use simplified models that therefore show clear responses and distinctions unburdened by the complex milieu of the vessel wall (Table 1).

EC

Vascular EC are responsible for transmitting and mediating the inflammatory injury signals that initiate lesion formation.18 Many studies have used microarrays to examine EC in the context of tissue classification; response to cytokines, growth factors, and physical stress; and morphogenesis models.19–28 Furthermore, a combination of database mining and microarray analysis has been used to identify EC genes. In one study, EC from four different vascular sources were compared with five non-EC types and yielded 64 significant pan-endothelial markers.29 These included well-characterized markers of endothelium (e.g., CD31, VE cadherin, multimerin, von Willebrand factor) as well as genes that are known to be preferentially expressed in endothelium (e.g., melanoma adhesion receptor, endothelin-1, plasminogen activator inhibitor type 1). In addition, a number of previously uncharacterized expressed tag sequence with conserved protein motifs that may have important roles in vascular wall biology were identified.

One preferentially expressed endothelial gene identified in this study, matrix Gla, serves as an example of the way in which microarray analysis can point to fruitful areas of atherosclerosis research.

www.jasn.org SPECIAL ARTICLE
### Table 1. Key studies of the genomics of atherosclerosis

<table>
<thead>
<tr>
<th>Model/Cell System</th>
<th>Intervention</th>
<th>Array Platform</th>
<th>Prominent Signals</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cells</td>
<td></td>
<td></td>
<td>Differentiation, development, chromatin remodeling, inhibin β</td>
<td>Spin et al.</td>
</tr>
<tr>
<td>murine A404/P19 cells</td>
<td>Differentiation to SM</td>
<td>Agilent Mouse Development oligo array</td>
<td></td>
<td></td>
</tr>
<tr>
<td>human saphenous vein and coronary artery EC</td>
<td>Comparison of baseline differences, difference with stimulation by oxLDL, cytokines</td>
<td>Agilent Human 1A oligo array</td>
<td>Regulation of cell growth, oxodreductase activity and stress, immune and anti-inflammatory responses, fibrinolysis, thrombogenesis</td>
<td>Deng et al.</td>
</tr>
<tr>
<td>human saphenous vein and coronary artery SMC</td>
<td>Comparison of baseline differences, difference with stimulation by oxLDL, growth factors</td>
<td>Agilent Human 1A oligo array</td>
<td>α-Chemokines, proinflammatory cytokines, apoptosis, inflammation, lipid biosynthesis, some β-chemokines, metalloproteinase inhibitors</td>
<td>Deng et al.</td>
</tr>
<tr>
<td>rat aortic SMC</td>
<td>Sodium butyrate</td>
<td>Atlas cDNA array</td>
<td>Cell growth, differentiation, stress response, vascular function</td>
<td>Rangana et al.</td>
</tr>
<tr>
<td>human mononcytic THP-1 cells</td>
<td>PMA for differentiation, stimulation with oxLDL, acLDL, LDL</td>
<td>Sanger Centre Hver1.20.1 DNA array and Clontech antibody array</td>
<td>Early response genes (transcription factors), upregulated early; cell proliferation, migration, inflammation, and lipid metabolism activated late</td>
<td>Tuomisto et al.</td>
</tr>
<tr>
<td>Animal models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>murine aortic tissue</td>
<td>ApoE, genetic background, high-fat diet, disease, diet, age</td>
<td>Agilent oligo array</td>
<td>Inflammatory response, TH1, TH2, wound healing, ossification, proteo- and peptidolysis, apoptosis, nitric oxide-mediated signal transduction, cell adhesion and migration</td>
<td>Tabibiazar et al.</td>
</tr>
<tr>
<td>murine aortic tissue</td>
<td>ApoE−/−, C57Bl6, genetic background</td>
<td>Agilent oligo array</td>
<td>Growth, differentiation, inflammation, catecholamine synthesis, phosphatase activity, peroxisome function, IGF activity</td>
<td>Tabibiazar et al.</td>
</tr>
<tr>
<td>murine aortic tissue</td>
<td>ApoE−/−, C57Bl6</td>
<td>Affymetrix MG-U74Av2</td>
<td>Lipid and lipoprotein metabolism, defense response, inflammation, nuclear organization and biogenesis, morphogenesis</td>
<td>Karra et al.</td>
</tr>
<tr>
<td>Human tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>human aortas (transplant donors)</td>
<td>Risk factors</td>
<td>U95av2 Affymetrix 10k genes (63k probes)</td>
<td>Prediction gene set for disease severity (includes ApoE, osteopontin, icr1, also capg, gm2, mmp9, ccr2)</td>
<td>Seo et al.</td>
</tr>
<tr>
<td>human arteries (various)</td>
<td>Risk factors</td>
<td>Sanger 10-k cDNA</td>
<td>HMG-CoA overexpression in macrophage-rich lesions</td>
<td>Tuomisto et al.</td>
</tr>
<tr>
<td>human arteries (various)</td>
<td>Risk factors</td>
<td>Incyte 18-k cDNA</td>
<td>Known (ApoE, CD68, TIMP, and phospholipase D) and unknown genes (JAK-1, VEGF receptor-2)</td>
<td>Hiltunen et al.</td>
</tr>
<tr>
<td>human coronary arteries (explanted hearts)</td>
<td>Risk factors</td>
<td>Agilent-Stanford oligonucleotide vascular array (22 k)</td>
<td>SM dedifferentiation, inflammation</td>
<td>King et al.</td>
</tr>
<tr>
<td>human coronary atherectomy specimens</td>
<td>Risk factors</td>
<td>Agilent-Stanford oligonucleotide vascular array (22 k)</td>
<td>Inflammation greater in de novo lesions</td>
<td>Ashley et al.</td>
</tr>
<tr>
<td>human carotid plaques, transplant donors (media)</td>
<td>Risk factors</td>
<td>21k cDNA array</td>
<td>Loss of RGSS from fibrous cap cells</td>
<td>Adams et al.</td>
</tr>
</tbody>
</table>

HMG-CoA, 3-hydroxy-3-methylglutaryl CoA; JAK-1, Janus kinase 1; oxLDL, oxidized LDL; TIMP, tissue inhibitor of metalloproteinase; VEGF, vascular endothelial growth factor; PMA, phorbol 12-myristate 13-acetate; AcLDL, acetylated low density lipoprotein.
Vessel calcification is a well-established element of progressive atherosclerosis. Matrix Gla protein (MGP) is a mineral-binding extracellular matrix protein that is known to be synthesized in a vitamin K–dependent manner in normal vascular SMC. MGP transcription is down-regulated in calcified atherosclerotic plaques in humans but upregulated at adjacent sites. Its expression is regulated in vascular SMC by a G-protein–coupled cation-sensing mechanism, and it is thought to play a role in inhibiting tissue calcification. Indeed, MGP knock-out mice show diffuse, massive calcification of the arteries and die within 8 wk after birth from aortic rupture. A limited study of MGP polymorphisms in two populations of patients (with and without myocardial infarction; with and without femoral atherosclerosis) identified two alleles (A-7 and Ala 83) that showed weak association with the disease. Another study identified a distinct variant of MGP with weak association with the disease. Despite this, the strongest transcriptional pathway in the diseased vessel wall remains SMC de-differentiation.

A concept not always well appreciated in atherogenesis research is the significant variation in gene expression and phenotype of SMC from different vascular beds. This likely results both from their varying embryonic origins and from their different hemodynamic environments. Also, primary SMC in culture rapidly de-differentiate after a few passages, with very significant drops in SMC marker gene expression, including SM α-actin, transgelin (SM22α), and SM myosin heavy chain. Despite these limitations, a few studies have examined cultured vascular SMC with microarrays or serial analysis of gene expression. In parallel to the endothelial experiments described, molecular signatures that are capable of distinguishing cultured human coronary and saphenous vein SMC were recently identified. These studies included stimulation–response using oxidized LDL (an atherogenic factor) and PDGF. The SMC subtypes exhibited different gene expression profiles at baseline and responded very distinctively to treatment. The data revealed that arterial SMC expressed higher levels of α-chemokine and lipid biosynthesis genes at baseline, whereas larger amounts of transcripts for β-chemokines, metalloproteinase inhibitors, IGF-binding proteins, and many atheroprotective genes were found in venous SMC. A lack of reliable in vitro models of SMC differentiation has long hindered detailed research, but in recent years, a few have arisen. The A404 clonal cell line, derived from P19 mouse embryonal carcinoma cells transfected with a SM α-actin promoter/puromycin-N-acetyltransferase, differentiates toward the SMC lineage in the presence of all-trans retinoic acid as demonstrated by upregulation of SM-MHC and other SMC marker genes, and allows selection for cells adopting an SMC fate through a differentiation–specific drug marker.

Comprehensive transcriptional profiling of the SMC differentiation process using the A404 model revealed thousands of genes that undergo differential regulation. Sequential patterns of gene expression were identified, providing complex insights into the regulation of the SMC differentiation process. Changes in genes associated with specific biologic ontology-based pathways were evaluated, and temporal trends were identified for functional pathways. An understanding of the differentiation process suggests approaches to alter SMC phenotypic modulation in vivo. These data serve as a platform from which studies seeking to prevent and treat atherosclerosis are being launched.

Macrophages
Monocytes respond to vascular injury and chemokines by migrating into regions of atherogenesis, differentiating into macrophages, and contributing to the complex atherosclerotic plaque, partly by ingesting oxidized lipoproteins and becoming foam cells. They also elaborate inflammatory factors that contribute to matrix degradation and plaque rupture. In vitro microarray studies of macrophages targeting atherogenesis typically have focused on their transcriptional response to cholesterol loading or other disease-related stimuli (e.g., Chlamydia pneumoniae, IL-13, fibrin D-dimer, free hemoglobin). Complex regulation seems to accompany foam cell formation and function. Early differentiation from monocyte to macrophage is associated with upregulation of transcription factors, whereas later differentiation and lipid loading ac-
tive genes related to cell survival, lipid metabolism, proliferation, migration, inflammation, and matrix remodeling\textsuperscript{[66,68]}; however, human macrophages also respond to oxidized LDL loading by activating antioxidative systems (glutathione and thioredoxin).\textsuperscript{[47]}

Array studies of THP-1–derived macrophages show that oxidized LDL may also modify the manner in which macrophages react to an inflammatory stimulus (LPS), causing substantial upregulation of a specific subset of chemo/cytokines and signal transduction genes. Nuclear receptors, including retinoid X receptor, and peroxisomal proliferator–activated receptor-\(\gamma\) seem to mediate this effect.\textsuperscript{[59]}

A panoply of stimuli may be present within the atherosclerotic plaque, and interest in their impact has prompted several \textit{in vitro} high-throughput macrophage gene expression studies. For example, the role of \textit{C. pneumoniae} in atherogenesis is widely debated. Differential-display reverse transcription–PCR was used to evaluate macrophages treated with \textit{C. pneumoniae} and/or LDL and found differential regulation of genes related to cell proliferation, apoptosis, ubiquitination, and immunomodulation.\textsuperscript{[52]}

Heme toxicity is also thought to contribute to atherosclerotic inflammation, and an array study found that human macrophages (crucial for hemoglobin clearance) activate an antioxidative/anti-inflammatory pathway in response to heme via the CD163 scavenger receptor\textsuperscript{[64]}; however, D-dimer, a marker of cardiovascular risk and atherosclerotic severity, upregulated inflammatory and adhesion molecule genes in macrophages.\textsuperscript{[53]}

Last, lipoprotein and endothelial lipase are associated with lesion macrophages. Suppression of lipoprotein lipase or endothelial lipase in THP-1 macrophages led to substantial decreases in inflammatory genes while increasing lipoprotein receptor genes. Both effects were enhanced by incubation with oxidized LDL.\textsuperscript{[55]}

### MOUSE MODELS

Mouse models allow us to take one step closer to human disease in the study of the molecular basis of atherosclerosis. They offer ease of access to the vessel wall during the progression of the disease process; assessment of interactions between the environment and the genetic factors that contribute to disease progression; and control of interacting variables such as genetic background, gender, diet, and medical therapy. There are also disadvantages. Some key processes are known to be different between mice and humans, for example, lipid transport, complement activation, and acute-phase protein reactants (\textit{e.g.}, C-reactive protein). Furthermore, most mouse models involve extreme elevations in cholesterol that generally do not exhibit distal coronary disease or evidence of thrombotic events that cause myocardial infarction (there is no reliable mouse model of plaque rupture).\textsuperscript{[56–58]}

Despite these concerns, the flexibility of genetically manipulated mouse models of atherosclerosis has made them a compelling tool for investigation into the genomics of atherosclerosis. The principal models are the apolipoprotein E (ApoE) knockout and the LDL receptor knockout mice used either as a single model of plaque rupture.\textsuperscript{[56–58]}

Despite these concerns, the flexibility of genetically manipulated mouse models of atherosclerosis has made them a compelling tool for investigation into the genomics of atherosclerosis. The principal models are the apolipoprotein E (ApoE) knockout and the LDL receptor knockout mice used either as a single model of plaque rupture.\textsuperscript{[56–58]}

The genetic background of the mouse contributes significantly to development of atherosclerosis.\textsuperscript{[59–61]}

On high-fat, high-cholesterol diets, several strains of inbred mice, such as C57Bl/6, not only exhibit increases in their cholesterol levels but also develop several layers of foam cells reminiscent of early atherosclerotic lesions. In contrast, other strains, such as C3H/HeJ mice, do not develop such lesions. In a recent study\textsuperscript{[62]} comparing the genome-wide transcriptional profile of aortas from C3H/HeJ and C57Bl/6 mice, differences in gene expression were identified at baseline as well as during normal aging and longitudinal exposure to high-fat diet.

Studies relating to the gene expression of models of atherosclerosis were recently reviewed\textsuperscript{[63]} with discussion of critical issues in method, analysis, and interpretation of the data. Some of these studies focused on identifying individual genes, whereas others included analysis of underlying biologic pathways.\textsuperscript{[64–67]}

In a comprehensive recent study of genes and pathways linked to development of atherosclerosis in the mouse aorta, we used several mouse genetic models with differing propensity to develop atherosclerosis, two different diets, and a longitudinal experimental design.\textsuperscript{[52,56]}

The goal of this complex design was to eliminate genes that are differentially regulated in response to aging or exposure to high-fat diet, focusing on the primary disease process in the ApoE-deficient model. Analysis of the genes demonstrated prominence of inflammatory pathways. Subanalysis of the inflammatory response pathway revealed genes that are characteristic of the macrophage lineage as well as both the TH-1 and TH-2 T cell populations. Other biologic processes and molecular functions were noted to be significant, including pathways such as wound healing, ossification, proteo- and peptidolysis, apoptosis, nitric oxide–mediated signal transduction, cell adhesion and migration, and scavenger receptor activity. Several pathways that have not previously been appreciated to have a role in atherosclerosis, such as carbohydrate metabolism, complement activation, calcium ion homeostasis, collagen catabolism, glycosyl bonds and hydrolase activity, taurine transporter activity, and heparin activity, were also identified through these studies.

Time-course analysis of atherosclerosis using mouse models offers insights that are not possible using human tissues, which are generally collected at one moment in time. Two groups have assessed murine models in a time-dependent manner. Karra \textit{et al.}\textsuperscript{[68]} identified

---

**mouse models**

Mouse models allow us to take one step closer to human disease in the study of the molecular basis of atherosclerosis.
genes in different transition periods from early- to late-stage atherosclerosis in ApoE−/− and C56Bl6 mice. At disease initiation, genes relating to metabolic processes, such as lipid or lipoprotein metabolism, were overrepresented, whereas during early disease progression, defense response and inflammation genes were prominent. In a similar analysis, we identified disease-related pathways that were expressed very early in the disease process and then reduced in expression as the disease progressed: Oxidoreductase activity, fatty acid and glucose metabolism, and mitochondrial-related biologic activities. Other pathways were expressed early in the disease process and maintained at a relatively high expression level throughout the course of the experimental period: Cytokine and chemokine activity, cell cycle, and lipid metabolism. Some biologic pathways seemed to be most active at later stage of the disease process, such as extracellular matrix metabolism, metalloproteinase activity, and solute transport. Not surprising, the key pathways regulating plaque development, including growth factor, cytokine, and cell adhesion activity, increased in their expression level as the atherosclerosis developed. In this analysis, inflammatory genes and pathways were represented in almost all of the patterns.

**HUMAN STUDIES**

Although human studies offer obvious advantages in the direct study of human disease processes, there are also limitations. The clearest is the lack of availability of tissue with early-stage disease. Human tissue for study is typically derived from surgical specimens such as carotid endarterectomy or from hearts explanted at the time of heart transplantation. Less invasive techniques, such as atherectomy, are little used in the coronary circulation, but some data exist from these sources. Because tissue availability is limited, some researchers have used the circulating transcriptome to gain insight into disease patterns such as collateralization.69

Although approaches and array platforms differ significantly among researchers, it is reassuring that overall similar findings were observed by different laboratories. Hiltunen et al.70 analyzed gene expression in normal arteries and in advanced atherosclerotic lesions using an 18-k cDNA array. Arterial samples were derived from abdominal aorta, iliac arteries, femoral arteries, and one coronary artery specimen. These investigators identified a panel of 17 genes that are already known to be mediators of atherosclerosis, including ApoE, CD68, tissue inhibitor of metalloproteinase, and phospholipase D, then went on to describe 75 differentially expressed genes that had not previously been associated with atherosclerosis. The latter included a subgroup of genes involved in cell signaling and proliferation, including Janus kinase 1 and vascular endothelial growth factor receptor-2.

Aorta was also the primary focus of Seo et al.,71 who analyzed samples with varying degrees of atherosclerosis to find gene expression signatures that could predict severity of disease. Their technique identified a set of 207 genes that discriminated severity of disease with 94% accuracy. The prediction set included genes such as ApoE, osteopontin, and oxidized LDL receptor 1. Ontologies identified included cell-cycle regulation and inflammatory response. Novel genes included ganglioneuroactivator protein and chemokine receptor-like 2.

Our laboratory has also focused on gene expression changes with disease severity. We focused primarily on coronary artery disease, examining gene expression in 51 coronary artery segments isolated from the explanted hearts of 22 cardiac transplant patients.38 Histologic grading of vascular segments was followed by hybridization to a customized 22-K oligonucleotide microarray, and assessment of gene expression according to histologic grade of disease revealed that loss of differentiated SMC gene expression was the primary expression signature of disease progression in atherosclerosis. We also provided insight into diabetic coronary artery disease, reporting an overabundance of immune and inflammatory signals in diabetic arteries, which extended across the severity spectrum. We have also had a significant interest in seeing significant genes within their biologic context. In this article, we presented a pathways analysis using both curated and newly derived pathways. The latter were networks based on connectivity, determined by language parsing of the published literature, and network ranking, determined by the significance of differentially regulated genes in each network. In doing this, we identified highly connected “nexus” genes that we speculate are attractive candidates for therapeutic targeting and follow-up studies. This follow-up validation is an often-neglected aspect to these studies but can be provided proximately by confirming the presence of expected findings. An example is that the most up-regulated gene in the diabetes group was IGF-1, whereas the most significant “network” was that of the insulin receptor.58 These confirmations with respect to known data provide extra confidence in the novel findings.

Patient follow-up and earlier stage disease were benefits of investigating coronary atherectomy in addition to assessing the unique pathophysiology of in-stent restenosis (ISR).72 We analyzed 89 samples of de novo atherosclerosis (n = 55) or ISR of a bare metal stent (n = 34) with histology and a dual-dye 22,000 oligonucleotide microarray. As expected, we found significantly greater cellularity and significantly fewer inflammatory infiltrates and lipid pools in the ISR group. Gene ontology analysis demonstrated the prominence of cell proliferation programs in ISR and inflammation/immune programs in de novo restenosis, the latter a finding that may reflect the later nature of ISR disease. We also applied our network analysis described, which combines semantic mining of the published literature with the expression signature of ISR to reveal gene expression modules suggested as candidates for selective inhibition of restenotic disease. Specifically, we identified the procollagen type 1 α2 gene and the ADAM17/TNF-α–converting enzyme gene as particularly important. In
this study, we went further to test our contention that our method was capable of identifying successful targets of therapy by comparing mean significance scores for networks generated from subsets of the published literature containing the term “sirolimus” or “paclitaxel.” All analyses revealed higher mean values for sirolimus, suggesting that this agent has a broader suppressive action against ISR than paclitaxel and recapitulating the results of clinical trials of existing agents.

One significant challenge that carries across both animal and human studies is the multicellular nature of atherosclerosis. Investigators will often carry out in situ hybridization or immunohistochemistry to identify the cellular origin of a particular transcript. Although this is successful and offers the further advantage of localizing the protein, it is time-consuming and hard to multiplex. To address this, investigators have focused on certain areas of disease macroscopically or, more recently used laser capture microdissection to pick off single cell types from histologic slides.

Adams et al. focused on identifying the unique expression program of the fibrous cap, because this structure is so important for plaque stability. They assessed gene expression of carotid artery media, nonatherosclerotic adjacent intima, fibrous cap of advanced atherosclerotic plaques, and whole advanced plaque using 21-k cDNA arrays. They found that both intima and cap expressed novel genes that had not previously been appreciated to be involved in vascular disease. Having a catalog of the genes and pathways that are active in the disease process will aid whole-genome efforts aimed at identifying allelic variation that is associated with clinically active human vascular disease: Knowing that a gene is expressed in the diseased vessel wall provides a previous probability that sets it apart from other genes, allowing more insightful analysis of massive data sets provided by whole-genome scanning of large case and control cohorts. In addition, although investigation of gene expression of vascular tissue is not a realistic option in patients who are suspected of having active vascular disease, the investigation of atherosclerosis-related genes in circulating mononuclear cells may provide critical information on the activity of the vascular inflammatory disease. Furthermore, genomic data can provide insights regarding disease-related circulating biomarkers that reflect activity of the disease. Such areas are under active investigation and should allow the development of risk markers that are more specific and informative than the currently available surrogates.

CONCLUSIONS

Investigation into the gene expression signatures of atherosclerosis has provided incremental insight into a disease the understanding of which has changed dramatically in the past 10 yr. These comprehensive data sets provide for the first time a view of the process that was not affected by the requisite previous conjecture. They provide a framework for more detailed mechanistic molecular studies in particular through the identification of genes and encoded proteins that had not previously been appreciated to be involved in vascular disease. Having a catalog of the genes and pathways that are active in the disease process will aid whole-genome efforts aimed at identifying allelic variation that is associated with clinically active human vascular disease: Knowing that a gene is expressed in the diseased vessel wall provides a previous probability that sets it apart from other genes, allowing more insightful analysis of massive data sets provided by whole-genome scanning of large case and control cohorts. In addition, although investigation of gene expression of vascular tissue is not a realistic option in patients who are suspected of having active vascular disease, the investigation of atherosclerosis-related genes in circulating mononuclear cells may provide critical information on the activity of the vascular inflammatory disease. Furthermore, genomic data can provide insights regarding disease-related circulating biomarkers that reflect activity of the disease. Such areas are under active investigation and should allow the development of risk markers that are more specific and informative than the currently available surrogates.

DISCLOSURES

None.

REFERENCES

51. Lim WC, Chow VT: Gene expression profiles of U937 human macrophages exposed to Chlamyphila pneumoniae and/or low density lipoprotein in five study models using differential display and real-time RT-PCR. Biochimie 88: 367–377, 2006
59. Paigen B: Genetics of responsiveness to Chlamyphila pneumoniae and/or low density lipoprotein in five study models using differential display and real-time RT-PCR. Biochimie 88: 367–377, 2006