CARDIOVASCULAR GENOMIC MEDICINE

Gene Expression Analysis of Cardiovascular Diseases
Novel Insights Into Biology and Clinical Applications

David Seo, MD, Geoffrey S. Ginsburg, MD, PHD, Pascal J. Goldschmidt-Clermont, MD
Durham, North Carolina

Although the contribution of genetics to complex cardiovascular diseases such as atherosclerosis has been accepted for quite some time, full and detailed knowledge of the individual causative genes has been elusive. With the advent of genomic technologies and methods, the necessary tools are now available to begin pinpointing the genes that contribute to disease susceptibility and progression. One approach being applied extensively in candidate gene discovery is gene expression analysis of human and animal tissues using microarrays. The genes identified by these genomic studies provide valuable insight into disease biology and represent the initial steps toward the development of diagnostic tests and therapeutic strategies that will substantially improve human health. This paper highlights the progress that has been made in using gene expression analysis cardiovascular genomic research and the potential for applying these findings in clinical medicine.   (J Am Coll Cardiol 2006;48:227–35) © 2006 by the American College of Cardiology Foundation

On April 14, 2003, the International Human Genome Sequencing Consortium, led in the United States by the National Human Genome Research Institute (NHGRI) of the National Institutes of Health and the Department of Energy, announced completion of the Human Genome Project (1). This announcement heralded the start of a new era in medical science in which genomic data will markedly improve our ability to diagnose and treat heart disease, cancer, and a myriad of chronic diseases. Understanding the relevance of an individual's genomic profile or "molecular fingerprint" in a disease context now might help us to gain valuable insight into disease mechanisms and could be the foundation for tailoring health care to the individual patient on the basis of their predicted disease course and response to therapies. The opportunity also now exists to replace diagnostic tests that are currently invasive in nature with genomic tests that can be performed without detectable risk or significant stress to the patients (2,3). The objective of this article is to provide an overview of the state of genomic research in cardiovascular medicine, specifically with regard to the use of microarray technology and functional genomics.

BACKGROUND

With common cardiovascular diseases such as atherosclerosis, atrial fibrillation, and heart failure affecting millions of Americans and costing our health care system billions of dollars, the manner in which clinicians diagnose and treat disease must change (3–6). Considering disease progression as a series of distinct "states," the goal for clinicians is to stabilize patients within a disease state and reduce the probability that they will progress to the next, often worse state. Moreover, many of the current therapies for a given state have been derived from large prospective studies such as the PRISM (Platelet Receptor Inhibition in Ischemic Syndrome Management) and the CURE (Clopidogrel in Unstable Angina to Prevent Recurrent Events) trials. The subjects in the treatment and control groups in these studies are, through a process of randomization, ostensibly considered to be identical. And although there might be a statistically significant overall treatment effect, it is impossible to know whether a treatment regimen will be effective for a specific individual. As a case in point, the state-of-the-art treatment for acute coronary syndromes uses a cocktail of 5 to 8 drugs for every patient. At present, we cannot identify which medication or combination of medications will provide the optimal benefit-to-risk ratio for the individual patient. Thus, 2 patients in the same disease state that are otherwise identical with respect to measurable variables, such as demographics, laboratory values, and comorbidities, might have marked differences in clinical course with respect to tempo and intensity of disease development and therapeutic response (7–9).

The greatest limitation in optimizing therapeutic strategies for the individual patient is an incomplete understanding of disease biology and inability to differentiate individuals according to their relative disease risk. The inherent genetic makeup of individuals is likely to play a large role in determining clinical course by modulating response to risk factors such as smoking, cholesterol, and obesity. For example, the intrinsic capacity for arterial and cardiac repair might be genetically determined and play a dominant role relative to the eventual impact of cardiovascular injuries. By identifying the genes and gene variants that determine

---

Abbreviations and Acronyms

CARGO	=	Cardiac Allograft Rejection Gene Expression Observational study
DNA	=	deoxyribonucleic acid
ICM	=	ischemic cardiomyopathy
LVAD	=	left ventricular assist device
mRNA	=	messenger ribonucleic acid
NF	=	nonfailing (myocardial samples)
NICM	=	nonischemic cardiomyopathy
PMBC	=	peripheral blood mononuclear cell
QTL	=	quantitative trait locus
SNP	=	single nucleotide polymorphism

---

individual disease susceptibility, we might be able to design health care plans that not only identify patients in the preclinical stages of disease but also allow for individualized therapies that are most efficacious and least likely to cause side effects (3,7,8,10).

## IDENTIFYING CANDIDATE GENES THROUGH RNA PROFILING

In the pre-genomic era, the opportunity to identify a relevant set of causative genes for multigenic diseases such as atherosclerosis and cardiomyopathies was limited. Traditional genetic approaches were designed to find single loci or genes with the power to cause Mendelian cardiovascular disorders such as the prototypical familial hypercholesterolemia. These types of disorders can result from a single base change in the deoxyribonucleic acid (DNA) that leads to significant changes in protein abundance or function. However, in common "garden-variety" disorders, such as atherosclerosis, the artery-specific inflammatory process likely results from multiple gene variants that interact to modulate an individual's response to clinical risk factors such as smoking, serum lipids, and hypertension. These gene variants usually take the form of single base changes in the DNA or single nucleotide polymorphisms (SNPs) that cause only a modest to moderate change in the resulting translated protein (either concentration, function, or a combination of both). However, it is the concurrent presence of a number of SNPs within a specific mix of risk factors that determines an individual's susceptibility to disease development and progression (11–13). The small contribution of each SNP to the disease makes them difficult to detect by traditional genetic association studies.

The obvious critical challenge is to identify the genes that collectively contribute to multigenic disorders such as atherosclerosis and cardiomyopathies. With the advent of broad-based genomic technologies, one approach that has emerged as a real opportunity is to perform gene expression analysis of disease-relevant tissues to look for changes in the abundance of transcribed genes or messenger ribonucleic acids (mRNAs) that correlate with a particular disease state, clinical outcome, or therapeutic response. Although DNA defines a person's inherent genetic make up, it is the active transcription of the DNA to RNA that integrates an individual's dynamic interaction with the environment. Therefore, genomic studies provide us with the opportunity to identify genes whose relative RNA abundance changes under differing biological conditions, with the premise that such altered expression response highlights culprits for the disease process. There are numerous ways to measure RNA abundance, including Northern blotting and multiplex real time polymerase chain reaction. However, these techniques can quantitate expression levels on the order of 10 to 50 genes at a time.

More recently, genomic technologies such as DNA microarrays provide us with an opportunity to assay thousands of genes simultaneously, giving us an unprecedented opportunity to survey the genomic contribution to cardiovascular diseases (14–16). Microarrays refer to solid substrates of glass, plastic, or silicon that contain thousands of microscopic printed spots of DNA (Fig. 1). Each of the DNA spots consists of hundreds to thousands of identical short pieces of DNA, also called probes, that have been apposed to the solid material. Each probe has a unique sequence of nucleotide base pairs that is complementary and unique to a single gene. Most microarray experiments today use "gene chips" on which the entire complement of the human genome is represented. In a typical microarray experiment, RNA is extracted from a tissue or sample of interest. The mRNA, which is the fraction of the RNA population that represents the transcribed genes, is amplified, and fluorescent tags are attached to each molecule. The entire pool is then incubated with the microarray, and the tagged mRNA molecules hybridize to the probes containing the specific complementary sequence via base pairing. Complementary mRNAs bind to probes on the array during hybridization. The microarray is then placed in a scanner where a laser is directed onto each spot, causing the tags to fluoresce. The resulting fluorescent intensity is proportional to the number of tagged RNAs that have hybridized to the probes in each spot, and therefore the measured intensity represents the activity or expression level of that gene. Consequently, a genomic profile can be generated for a particular tissue at any stage of health or disease.

The utility of these gene signatures, generally speaking, is two-fold: 1) as an initial step in identifying disease-associated genes, or "candidate genes;" and 2) as a disease
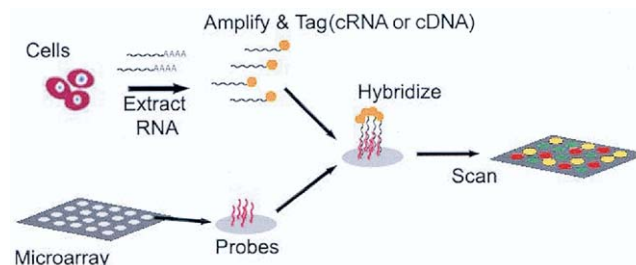
Figure 1. Schematic of a microarray assay. cDNA; complementary deoxyribonucleic acid; cRNA = complementary ribonucleic acid.

biomarker. In the first instance, the genes that comprise the genomic signature are by definition disease-associated and thus are the candidates in which to look for SNPs that are associated with disease susceptibility or a particular treatment outcome. By identifying a potential panel of disease SNPs, we could then screen a patient's DNA via a blood test for the presence of these gene variants to aid in diagnosis or planning treatments. Certainly these candidate genes might substantially improve our understanding of disease biology and subsequently lead to identification of potential targets for new treatment strategies (17). The other potential use of genomic information is to use the gene signatures directly as a disease biomarker. Gene expression data in the appropriate setting is an extraordinarily detailed patient phenotype that could be used to accurately stratify patient populations as to their disease risk or response to therapies. In both cases, genomic data represent a means to distinguish between patients who are otherwise identical by traditional clinical variables (14–16). Although this is still a work in progress in cardiovascular medicine, we can look to the field of oncology to see where the future lies.

CANCER GENOMICS AS A TEMPLATE FOR CARDIOVASCULAR MEDICINE

The clinical use of genomics has advanced most rapidly in the area of oncology, owing in part to the inherent advantage of being able to directly assay the tumors. The cancer literature provides us with some guidance on how gene expression data might be applied in the context of cardiovascular medicine. A predominant role for cancer genomics has been to risk-stratify patients that have been classified by traditional clinical factors with greater granularity (18). After the initial diagnosis and clinical staging, gene expression profiling of the tumor has been used to predict long-term disease recurrence and survival as well as possibly for planning treatment regimens (19–21). In the most concrete example, patients who are at high risk for breast cancer recurrence can be identified even at early stages of the disease. Two tests are commercially available in the U.S. for patients with early invasive breast cancer. Both tests are designed to complement the diagnostic armamentarium available to clinicians to select the most appropriate therapeutic strategy for patients with early-stage breast cancer. In this population, patients traditionally undergo surgical resection followed by radiation treatment and hormonal adjuvant therapy. These new genomic assays seek to identify individuals that are at significantly higher risk for recurrence, potentially warranting more aggressive treatment strategies with chemotherapy in spite of a rather good prognosis as indicated by traditional risk profiling (18). It is the integration of the molecular characteristics from the patients' tumors with clinical and demographic data that allows for this additional granularity in clinical decision-making.

Genomic research has progressed more slowly in the cardiovascular field, owing to the inability to access enough informative tissues. Perhaps this has been the reason that much more effort has gone into looking for disease SNPs, because these can be both studied and tested through the use of blood (DNA) samples. Still, within the past 3 years, landmark studies using human aortic, atrial, and ventricular tissues have improved our understanding of disease biology and are being translated to applications that might improve our ability to make clinical decisions. It has also become evident that blood cells can also provide usable genomic information for cardiovascular conditions. We now turn our attention to recent significant research in the area of cardiovascular genomics.

CARDIOVASCULAR GENOMICS

Atherosclerosis. Given the extraordinarily heavy health burden of atherosclerosis and its thromboembolic complications, this condition has been one of the first foci for applied genomic technologies. The availability (or relative lack) of appropriate human target tissues has until recently been a critical limiting factor. In a recent study, investigators performed gene expression analysis exclusively on human aortic tissues to identify genes that are predictive for the extent of atherosclerotic burden in the aorta (22). The tissues were collected at the time of organ harvest from heart and lung donors and preserved to maintain premortem RNA integrity. The inherent symmetry of disease development about the longitudinal midline of the thoracic aorta allowed a quantitative analysis of atherosclerosis on one-half of the specimen and performance of gene expression profiling on the other half. The seemingly palindromic character of the disease mosaic allowed for direct association gene expression profiles to disease burden phenotype. Gene expression data were then used to develop predictive models that accurately classified unknown aortas as having minimal or severe disease with over 90% accuracy. The predictive capability of these genes implies potential biological relevance. The genes that provided this predictive capability included not only many previously associated with atherosclerosis as well as an entirely novel set of genes that never before had been linked to cardiovascular disorders but also a large number of genes whose role in atherosclerosis had gone unrecognized.

Given the challenges in obtaining human arterial tissues for studies of atherosclerosis, animal models have been invaluable as tools for determining the mechanisms underlying disease progression and the influence of diet and other environmental factors on its expresssion. The apolipoprotein E knock out (apoE$^{-/-}$) is a well-established disease model of atherosclerosis that has been used for years with great relevance to human disease. This model allows the study of atherosclerotic inflammation while controlling for a number of important clinical variables such as age, gender, and diet on a genetically homogenous background. In a

recent study, gene expression analyses using aortas from apoE$^{-/-}$ mice of different ages and diets defined a genomic signature for disease burden that could differentiate between the different mouse atherosclerotic disease states (23). This genomic model was then tested in human disease by profiling genes in mRNA from atherectomy samples extracted from atherosclerotic lesions and from in-stent restenotic lesions. The findings were that the mouse genomic signature could clearly differentiate between atherosclerotic and restenotic lesions with a high degree of accuracy, thus validating in humans the genomic predictor discovered in the mouse. This is important, because it suggests that genomic data derived across species might be applicable to human disorders. A similar gene expression analysis of mouse atherosclerosis models showed similar findings with regard to genomic signatures indicative of different disease states and that genomic signatures from mouse tissues were homologous to those seen in human tissues (24). This study also went further to demonstrate the presence of a molecular repair signal in the mouse aortas that implicates the role of bone marrow-derived progenitor cells in the balance between vascular damage and restoration. This is one of the first instances in which chronic disease progression has been directly linked to changes in the body's ability to repair itself. Therefore, this study not only implicated a series of gene programs as disease progresses but also demonstrated that gene expression analysis can implicate specific disease mechanisms.

The results from these studies currently have no direct diagnostic utility, because they rely on the availability of disease tissues. However, the information contained in the gene expression signature (i.e., the genes making up the gene expression pattern) that differentiates one disease state from another serves as the basis for "candidate gene" identification for use in genetic association studies. Although whole genome scanning and association studies will likely supersede the candidate gene approach, candidate genes identified in a gene expression signature can also be used to facilitate whole genome scans by prioritizing targets within chromosomal regions. In contrast to the traditional method of identifying candidates through their established role in the biology of disease, gene expression represents a novel and nonbiased methodology. Thus genes identified by genomic studies can be used as the basis for genetic association studies, or these genes might identify new biology relevant to the disease process. A key challenge that remains is to identify which gene or genes are causative and functionally relevant.

The most clinically relevant application of genomic technologies in cardiovascular medicine would be "point-of-care" assays for diagnostics with a simple blood-based test, and there is now evidence to suggest that this might be a real possibility. The peripheral blood mononuclear cells (PMBCs) play a critical role in inflammatory pathways leading to atherosclerosis and constitute a major cellular component of atherosclerotic lesions and their repair. Therefore, gene expression profiling of PBMCs could

potentially yield a diagnostic signature for atherosclerosis. One of the earliest studies was performed in a small group of patients with and without significant coronary artery disease (four cases and three control subjects) (25). With RNA extracted from the PBMCs from these patients, investigators evaluated the relative abundance of over 10,000 transcripts. They found over 100 genes that were significantly up-regulated or down-regulated in the patients with coronary disease compared with the normal control subjects. Although this was a small study, it points to the feasibility of using gene expression profiling of blood to detect coronary heart disease. Another recent study also provides additional evidence that blood gene expression profiling might have utility in the detection of atherosclerosis (26). Investigators performed gene expression analysis of PBMCs to find genes that were differentially expressed in patients with carotid atherosclerosis. Gene expression signatures from patients with severe carotid atherosclerosis necessitating surgical intervention were compared with a control cohort without measurable disease. The investigators found a number of regulatory genes and transcription factors, including Finkel-Biskis-Jinkins osteosarcoma proto-oncogene, dual specificity phosphatase-1, and nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-alpha, that were significantly altered in patients with severe carotid disease. Given the sample sizes used, this study was meant to be an exploratory analysis that identified differentially expressed genes for improving biological insight into atherosclerosis with PBMCs as reporter cells. It has not been validated for use as a diagnostic or prognostic tool. However, it provides proof on concept that blood-based gene expression data seem to be informative for an individual's overall inflammatory state, with possibly more specific information relevant to carotid atherosclerosis. The identified genes might represent genes whose variants might contribute to pathologically enhanced inflammatory responses to endothelial damage from high serum cholesterol or hypertension. This is one of the first studies to show that assays of PBMCs have potential to be informative in a diagnostic context for presence or absence of atherosclerosis in humans.

One could envision that genomic profiling of PBMCs could serve as an adjunctive test to noninvasive functional studies to improve sensitivity and specificity for the presence or absence and severity of disease. An important consideration here is whether carotid atherosclerosis will be associated with gene expression profiles that are distinct from those associated with coronary, aortic, or lower extremity vascular disease.

Cardiac transplant. Significant progress has been made in the application of genomic information to predict the risk of acute rejection in patients who have undergone cardiac transplantation. In one study, researchers profiled PBMCs in a cohort of cardiac transplant patients and found a pattern of gene expression statistically associated with rejection (27). In this study, longitudinal clinical data and blood samples

were collected at the time of surveillance right heart biopsies. Periodic right heart biopsies are standard of care for assessing signs of acute rejection, often before its clinical manifestations. Analysis of the gene expression data from the PBMCs collected from transplant patients showed differentially expressed genes that predicted the development of acute rejection as defined by histological diagnosis and blood serum assays. In addition, in patients that experienced rejection, the majority of discriminatory genes returned to baseline levels of expression after treatment from the immune rejection episode. These results not only support the utility of gene expression for diagnosis but also the potential use of genomic data to measure response to treatment.

With a similar study design, another research group also found genes in PBMCs that are predictive for acute rejection in cardiac transplant patients (28,29). These findings have led to the development of a commercially available, blood-based test for detection of acute cardiac rejection. The investigation examined data from 600 transplant recipients in the multi-institutional CARGO (Cardiac Allograft Rejection Gene Expression Observational) study. The CARGO study investigators collected longitudinal clinical data and blood samples at the time of surveillance right heart biopsies. Analysis of the gene expression data from the PBMCs identified a core group of genes that were differentially expressed in patients at various stages of acute rejection. With this information, a blood-based test was developed that assigns risk for developing rejection within the next 3 months with a negative predictive value of over 90% (29,30). Therefore, purely on the basis of gene expression data from peripheral blood, this test might allow patients to forego the risk, albeit minimal, of surveillance right heart biopsies if the results are negative. These studies have also shown that the genomic data might be informative as to the level of immunosuppression required for individual patients, particularly as to whether they were receiving an effective dose of corticosteroids independent of the actual dose (31). For example, a patient receiving 40 mg of prednisone might have a gene expression profile that is more typical of a patient receiving an 80-mg dose, indicating that the gene expression profile might be a more accurate pharmacodynamic biomarker for dose finding than traditional methods. Although further data are needed to substantiate this finding, it represents an opportunity for the early application of pharmacogenomics to assist in drug titration.

One issue with these types of studies will be the sensitivity of the test for predicting clinical outcomes. The ability to thoroughly assess an accurate and reproducible phenotype is an absolute necessity for the use of gene expression to drive decision-making. In the CARGO study, the phenotype associated with gene expression was the histological grade of rejection determined by a panel of clinicians. If there is substantial variability in the grading of rejection in the myocardial samples, this will affect the consequent reliability of the assay for predicting rejection. As more applications of genomics to clinical practice are developed, the accuracy of the measure phenotype will be crucial. In any event, the gene expression-based assay for transplant rejection represents the first substantial and clinically relevant application of genomics in the field of cardiovascular medicine.

Electrophysiology—atrial fibrillation. The field of electrophysiology provides a tremendous opportunity for genomic data to guide clinical decision-making. New device-based therapeutic strategies have been developed that are of unmatched benefit to the patients who need them most but also come with substantial side effects, particularly if applied to patients who need them least (32–34). Furthermore, it has become increasingly evident that arrhythmias result from the combined effect of myocardial degeneration, aging, and to a smaller extent, "channelopathies," or genetic variations in ion channels (35–37). Each individual effect is influenced by genetic factors, which by themselves do not lead to abnormal rhythms. However, when the genetic factors are combined and interact, arrhythmias might develop. In this context, the first application of microarray technology has been to improve our understanding of the biology of arrhythmias, in particular atrial fibrillation. Two groups have recently published results of functional genomic studies using human tissues. Both studies used the right atrial appendages obtained at the time of coronary artery bypass surgery or valve repair surgery for their gene expression profiling assays. One study evaluated 1,800 known genes with DNA microarrays and found 50 genes that were differentially expressed in the atrial fibrillation samples relative to the control samples in normal sinus rhythm (38). Within the predictive gene expression signature was a preponderance of genes related to reactive oxygen species and enzymes that modulate response to oxidative stress. This led to speculation that oxidative damage, perhaps through the inflammatory pathways, could be one of the initiators of atrial fibrillation as well as a potentiator of the condition. This study used a relatively limited gene set by current standards; it was one of the first broad, nonbiased investigations of molecular changes in humans with atrial fibrillation. Another similar investigation also used atrial tissue samples taken at the time of cardiac surgery to find genes associated with the presence of chronic atrial fibrillation but assayed for a larger set of over 12,000 genes (39,40). In patients with atrial fibrillation, compared with individuals in normal sinus rhythm, the investigators found differential expression in genes related to remodeling, contractile proteins, and ion channel changes. In addition, changes in expression of genes related to metabolism were observed that might be indicative of higher energy needs in the fibrillation state. Each of these studies sought to improve biological insights into atrial fibrillation. As in the study of atherosclerosis, neither was meant to yield data that could be used directly in diagnostics. The prioritized genes can help to suggest possible mechanisms that might contribute to development of atrial fibrillation. Obviously, the ability to

study readily accessible tissues would be ideal for clinical genomics, but at this point it is difficult to say whether this will be possible. What is more likely will be the construction of clinical tests that assess for genetic polymorphisms that might indicate susceptibility for developing arrhythmias or response to pharmacologic therapies.

Hypertension. Although hypertension is extraordinarily important in cardiovascular medicine, genomic studies of humans have been difficult, given the lack of accessible tissues. Much of what we know about the molecular mechanisms contributing to hypertension has been developed with animal models. In the search for hypertension candidate genes with animal models, one technique that has been used extensively is the identification of quantitative trait loci (QTLs). A QTL is a portion of a chromosome that is highly associated with a trait, such as high blood pressure. One problem has been that a QTL might consist of hundreds of genes. The emergence of microarray technologies has provided a method for prioritizing the genes within these QTLs that might warrant further study. With a strategy known as "genomic convergence," some investigations have combined gene lists derived from QTL mapping and gene expression analysis of disease-relevant tissues to find the overlapping members (41,42). The genes identified at the interface between two disparate approaches might therefore represent particularly intriguing candidate genes. One study that illustrates this approach used the Sabra rat hypertension model of salt susceptibility. Independent QTL analysis was performed, showing two loci on chromosome 1 of the rat genome that was highly associated with the hypertensive phenotype (42). One locus spanned 43.1 centimorgans and has been shown to contain 2,933 known genes, whereas the other spanned 18 centimorgans with 1,102 genes. Gene expression profiles were generated from the rat kidneys, because it was felt that this organ was intimately involved in sodium balance. The researchers found seven genes that overlapped between the QTL and gene expression analyses. Although this is an animal study, it represents a powerful approach for studying human disease biology when informative tissues are not readily available. The polymorphisms from the genes identified by this study can be used in human genetic association studies.

Heart failure. Heart failure is the fastest-growing cardiovascular disease. Recently, new device therapies such as biventricular resynchronization and automatic defibrillators have expanded the armamentarium of strategies to treat patients with heart failure (43–46). Patients with heart failure also receive a complex medication regimen and have a heterogenous clinical course, just as with other common cardiovascular diseases. As such, genomic analyses might play a substantial role in improving diagnostics and health care planning. The availability of human tissues obtained at the time of transplant or placement of left ventricular assist devices (LVADs) has facilitated the use of microarray analyses to study cardiomyopathies. The first groups of

recent genomic studies have sought to differentiate between ischemic and nonischemic cardiomyopathies.

A recent study examined a number of myocardial samples to look for gene expression differences between end stage ischemic and nonischemic subtypes (47). They also attempted to develop the genomic profile for more intermediate clinical phenotypes by studying tissues obtained at the time of diagnostic biopsy and LVAD placement. This larger study examined nearly 50 myocardial samples, from two different institutions, obtained at the time of transplantation/ explantation, LVAD placement, or biopsy. Furthermore, the samples represented a range of clinical disease stages, from new diagnosis to end-stage disease requiring transplant. The researchers demonstrated the ability to differentiate between patients with ischemic and nonischemic cardiomyopathies among the end-stage patients with remarkable 100% accuracy. However, predictive accuracy for classifying samples as ischemic or nonischemic dropped substantially in the LVAD and newly diagnosed cohorts. Instructively, age, gender, and hemodynamic differences in the patient cohorts did not add power to any of the predictive models used in the study. Overall, this study demonstrates the potential clinical applications for classifying patients into cardiomyopathy subtypes by using genomic phenotypes. The genes that differentiate between the groups in this study might provide clues as to how treatment regimens could be tailored for specific subtypes. Finally, although not described specifically, the genes that differentiate between the different cardiomyopathy subtypes over time aid in targeting of therapies depending upon the stage of the natural history of the disease process as well as tracking response to therapies.

This same group of investigators also performed additional analyses with the same sample cohort and identified differentially expressed genes without assessing their predictive capability (48). The reason for the additional analyses was the authors' assertion that the candidate genes arising from the predictive models, although instructive, were not truly indicative of disease biology. Rather, they represent a stable expression profile for a particular disease setting. Although this is still an unresolved issue, they nevertheless provided a more detailed analysis of the gene expression differences between patients with ischemic and nonischemic cardiomyopathies. In 21 nonischemic cardiomyopathy (NICM) samples, 10 ischemic cardiomyopathy (ICM) samples, and 6 nonfailing (NF) myocardial samples, 257 genes were found that were significantly differentially expressed in NICM compared with NF hearts and 72 genes when comparing ICM with NF hearts. There were 41 genes that overlapped across these two comparisons and represented biological processes relevant to cell growth and signal transduction. Genes unique to NICM were related to metabolism, and those unique to ICM seemed to be related to catalytic activity. Angiotensin-converting enzyme-2 was up-regulated in NICM but not in ICM samples, pointing to a potentially different pharmacologic target in patients with these two entities. An important issue with genomic studies

of cardiomyopathy is that of phenotype used as the basis for the comparison. In one sense, the diagnosis of nonischemic or idiopathic dilated cardiomyopathy likely represents patients with a wide spectrum of unrelated conditions as well as patients with a mix of both ischemic and nonischemic features. In addition, such analyses performed thus far do not take into account the gene changes induced by treatment regimens. Furthermore, comparisons between end stage and normal tissues might yield results that are difficult to interpret and potentially less useful clinically. However, these studies do illustrate that gene expression profiling of myocardium or another tissue, such as blood, could be studied as a means to differentiate between clinical cardiomyopathy entities, each of which might warrant slightly different pharmacologic regimens.

CHALLENGES AND OPPORTUNITIES

There have been a number of substantial findings to date in cardiovascular genomic research with the development of a commercially available diagnostic test. However there are still some significant challenges if we are realize the full potential of genomics in cardiovascular medicine. Among these hurdles are issues surrounding the use of larger numbers of appropriate samples, validation of results, and societal and ethical concerns.

The issue of having enough appropriate samples in gene expression studies is a substantial one on many levels. On one level, many studies, including ones described here, have used small numbers of samples—perhaps a total of 20 to 40. Such limits relate to both the expense of the assays as well as the availability of sufficient sample numbers. Without larger sample numbers, the validity of analyses will remain controversial. Therefore, although the data from these studies are still useful for exploration and hypothesis generation, they require further substantiation before any clinical application of the results. The intrinsic genetic and socioenvironmental heterogeneity within human populations dictate that larger sample sizes, perhaps on the order of hundreds of samples, could make the results more clinically applicable. This might also allow researchers to fine tune their genomic analyses to detect nuances in disease states or clinical outcomes. For example, it is likely that there are distinct gene groups that determine atherosclerotic susceptibility in men versus women as well as between racial and ethnic groups. Sample size limitations prevent us from being able to study these important subpopulations. In the U.S., multi-institutional investigations such as the CARGO study are good examples of how the pooling of cohorts enable researchers to generate clinically applicable information (3,49–54). Internationally, there are a number of efforts underway to develop repositories of biological materials on a massive scale that can be used for genomic, genetic, metabolomic, and proteomic studies. One of the first projects of this kind was the North Health project of Sweden (55), which has a biorepository of over 85,000

unique individuals. Two recent national projects are the United Kingdom Biobank (56) and Biohealth Norway (57) that plan to collect genetic material and clinical information on up to 500,000 individuals.

Discussion of performing gene expression experiments on a significantly larger scale brings up the issue of costs. Larger samples sizes would allow for more robust results and larger, truly independent training and testing sets to test clinical utility. Depending on the microarray platform being used, the per-sample costs for materials (reagents, microarrays) and RNA processing (extract, label, hybridize) can range from $300 to $700. This does not account for the costs of accruing, collecting, and storing comprehensive, longitudinal clinical information. Certainly, development of core facilities can help to defray these costs. Still, genomic research will require a substantial budget when using sample sizes that are useful for developing clinically usable results. There are a number of technical advancements that might dramatically reduce costs, with the development of high throughput microarray platforms. But regarding this issue of sample numbers, we must make sure that we are using the appropriate samples. For example, in the context of cardiomyopathies, the samples easiest to obtain might come from individuals with end-stage disease, already receiving a multitude of medications. It might be difficult to decipher the information derived in this situation and might limit the applicability of the resulting information.

Another challenge facing genomic research is the process of moving beyond exploratory studies for candidate gene discovery toward improving understanding of disease biology. After robust statistical analysis, identified candidate genes can be used for diagnostic testing in a clinical setting. However, an understanding of disease mechanisms is needed if we are to truly learn how to improve health care strategies. Therefore, candidate genes must be assessed for functional relevance to patients and disease processes. In particular, it will be important to determine whether expression changes in the prioritized genes are through a causative role or rather the result of the disease process itself. This is critical for furthering our understanding of disease biology and subsequently improving our health care strategies. Use of knock out/in or transgenic animals will be vital to these types of studies. In addition, considering the time, expense, and technical challenges of such animal confirmatory studies, researchers will need to develop some more novel and straightforward methods, perhaps through the use of small molecules, aptamers, or small interfering RNAs (58–62).

Finally, aside from the scientific challenges, there are a number of public policy issues that must be addressed. A key issue will be to shift health care delivery to one of individualized health assessment that emphasizes prevention and prognosis as genetic and genomic information is translated to clinical practice. This will require modification in patient education as well as changes in medical education. Clear improvements in health outcomes and reductions in health care costs will need to be demonstrated to justify new and

Table 1. Landscape of Applications for Gene Expression Data in Clinical Practice

	Application Available Now	Potential or Emerging Applications Available Within Next 5 Yrs
Cardiovascular		
Atherosclerosis		Diagnostic test to assess atherosclerosis risk (23,24)
Heart failure/transplant	Early detection of acute rejection	Blood test for efficacy of immunosuppression (29)
Electrophysiology		Results from gene expression studies lead to new therapies (36–38)
Hypertension		Results from gene expression studies lead to new therapies (39,40)
Oncology		
Breast	Predict recurrence risk in early invasive breast cancer	Test predicting response to chemotherapy (64)
Lymphoma		Classification of lymphoma subtypes (65–68)
Unknown primary	Classify tumors of unknown primary to guide therapeutic decisions	
Neuroblastoma		Assess prognosis for children with neuroblastoma (69)
Infectious disease		
HIV	Procleix	Predict response of HIV and HCV antiretroviral therapies (70)
SARS	Rapid classification of SARS virus (71)	
Bacterial resistance		Predict response to antimicrobial therapies
Hepatitis C		Predict response to treatment (72)

HCV = hepatitis C virus; HIV = human immunodeficiency virus; SARS = severe acute respiratory syndrome.

potentially expensive genetic and genomic testing. Certainly as we move toward being able to predict disease susceptibility and treatment outcomes, legislation will likely be necessary to protect medical privacy and prevent potential barriers to insurance or employment (63,64).

Cardiovascular genomic research is still in its formative stages with regard to the translation of results of studies to clinical applications. However, examination of other medical fields demonstrates the potential of genomic medicine and where cardiovascular medicine will be, possibly within the next 5 years (Table 1) (65–71). Microarrays are already being used in oncology, infectious disease, and pharmacology to help clinicians better risk-stratify patients as well as predict therapeutic responses and guide clinical decision making. The availability of a genomic test to detect acute rejection in cardiac transplant patients represents the first concrete example of how we will start to use gene expression data in the mainstream clinical enterprise. We have already started to gain substantial new insights into common cardiovascular diseases. It is likely that diagnostic tests using genomic information will be available in the near term to improve prediction of acute coronary syndromes as well as the potential for development of malignant arrhythmias. Although significant challenges remain, cardiovascular genomic medicine promises to improve patient care and lower health care costs and will change the way we practice medicine.

Reprint requests and correspondence: Dr. David Seo, Duke University Medical Center, Box 3382, Durham, North Carolina 27710-0001. E-mail: david.seo@duke.edu.

REFERENCES

1. Spencer G. International Consortium Completes Human Genome Project. Bethesda, MD: National Institutes of Health, 2003.
2. Collins FS, Green ED, Guttmacher AE, Guyer MS, U.S. National Human Genome Research Institute. A vision for the future of genomics research. Nature 2003;422:835–47.
3. Bell J. Predicting disease using genomics. Nature 2004;429:453–6.
4. Khoury MJ, McCabe LL, McCabe ER. Population screening in the age of genomic medicine. N Engl J Med 2003;348:50–8.
5. Nabel EG. Cardiovascular disease. N Engl J Med 2003;349:60–72.
6. Ross JS, Ginsburg GS. Integration of molecular diagnostics with therapeutics: implications for drug discovery and patient care. Expert Rev Mol Diagn 2002;2:531–41.
7. Roses AD. Pharmacogenetics and the practice of medicine. Nature 2000;405:857.
8. Williams RSaG.-C PJ. The genetics of cardiovascular disease: from genotype to phenotype. Dialogues in Cardiovascular Medicine 2004; 9:3–19.
9. Guttmacher AE, Collins FS. Welcome to the genomic era. N Engl J Med 2003;349:996–8.
10. Guttmacher AE, Collins FS. Genomic medicine—a primer. N Engl J Med 2002;347:1512–20.
11. Sing CF, Stengard JH, Kardia SL. Genes, environment, and cardiovascular disease. Arterioscler Thromb Vasc Biol 2003;23:1190–6.
12. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56.
13. Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. Annu Rev Genomics Hum Genet 2004;5:189–218.
14. Cook SA, Rosenzweig A. DNA microarrays: implications for cardiovascular medicine. Circ Res 2002;91:559–64.
15. Goldsmith ZG, Dhanasekaran N. The microrevolution: applications and impacts of microarray technology on molecular biology and medicine (review). Int J Mol Med 2004;13:483–95.
16. Napoli C, Lerman LO, Sica V, Lerman A, Tajana G, de Nigris F. Microarray analysis: a novel research tool for cardiovascular scientists and physicians. Heart 2003;89:597–604.
17. King JY, Spin JM, Chen MM, et al. Pathway analysis of coronary atherosclerosis. Physiol Genomics 2005;23:103–18.
18. Wooster R, Weber BL. Breast and ovarian cancer. N Engl J Med 2003;348:2339–47.
19. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
20. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006;439: 353–7.
21. Berchuck A, Iversen ES, Lancaster JM, et al. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. Clin Cancer Res 2005;11:3686–96.
22. Seo D, Wang T, Dressman H, et al. Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 2004;24:1922–7.
23. Tabibiazar R, Wagner RA, Ashley EA, et al. Signature patterns of gene expression in mouse atherosclerosis and their correlation to human coronary disease. Physiol Genomics 2005;22:213–26.
24. Karra R, Vemullapalli S, Dong C, et al. Molecular evidence for arterial repair in atherosclerosis. Proc Natl Acad Sci U S A 2005; 102:16789–94.

25. Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. J Mol Cell Cardiol 2003;35:993–8.

26. Patino WD, Mian OY, Kang JG, et al. Circulating transcriptome reveals markers of atherosclerosis. Proc Natl Acad Sci U S A 2005;102:3423–8.

27. Horwitz PA, Tsai EJ, Putt ME, et al. Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. Circulation 2004;110:3815–21.

28. Deng MC, et al. Cardiac allograft monitoring using a novel clinical algorithm based on peripheral leukocyte gene expression profiling. Presented at: American Heart Association, Scientific Sessions; Orlando, FL: 2003.

29. Mehra MR. The emergence of genomic and proteomic biomarkers in heart transplantation. J Heart Lung Transplant 2005;24 Suppl 7:S213–8.

30. Eisen HJ, et al. Longitudinal monitoring of cardiac allograft recipients using peripheral blood gene expression profiling: a retrospective observational analysis of molecular testing in 19 case studies. Presented at: International Society of Heart and Lung Transplantation; Philadelphia, PA: 2005.

31. Starling RC, et al. The influence of corticosteroids on the alloimmune molecular signature for cardiac allograft rejection. Presented at: International Society of Heart and Lung Transplantation; Philadelphia, PA: 2005.

32. Moss AJ, Hall WJ, Cannom DS, et al. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. Multicenter Automatic Defibrillator Implantation Trial Investigators. N Engl J Med 1996;335:1933–40.

33. Moss AJ, Zareba W, Hall WJ, et al. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. N Engl J Med 2002;346:877–83.

34. DiMarco JP. Implantable cardioverter-defibrillators. N Engl J Med 2003;349:1836–47.

35. Marban E. Cardiac channelopathies. Nature 2002;415:213–8.

36. Nattel S. New ideas about atrial fibrillation 50 years on. Nature 2002;415:219–26.

37. Noble D. Modeling the heart—from genes to cells to the whole organ. Science 2002;295:1678–82.

38. Kim YH, Lim do S, Lee JH, et al. Gene expression profiling of oxidative stress on atrial fibrillation in humans. Exp Mol Med 2003;35:336–49.

39. Ohki R, Yamamoto K, Ueno S, et al. Gene expression profiling of human atrial myocardium with atrial fibrillation by DNA microarray analysis. Int J Cardiol 2005;102:233–8.

40. Ohki-Kaneda R, Ohashi J, Yamamoto K, et al. Cardiac function-related gene expression profiles in human atrial myocytes. Biochem Biophys Res Commun 2004;320:1328–36.

41. Danziger RS, You M, Akil H. Discovering the genetics of complex disorders through integration of genomic mapping and transcriptional profiling. Curr Hypertens Rev 2005;1:21–34.

42. Yagil C, Hubner N, Monti J, et al. Identification of hypertension-related genes through an integrated genomic-transcriptomic approach. Circ Res 2005;96:617–25.

43. Abraham WT. Cardiac resynchronization therapy for heart failure: biventricular pacing and beyond. Curr Opin Cardiol 2002;17:346–52.

44. Bristow MR, Saxon LA, Boehmer J, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med 2004;350:2140–50.

45. Cazeau S, Leclercq C, Lavergne T, et al. Effects of multisite biventricular pacing in patients with heart failure and intraventricular conduction delay. N Engl J Med 2001;344:873–80.

46. Cleland JG, Daubert JC, Erdmann E, et al. The effect of cardiac resynchronization on morbidity and mortality in heart failure. N Engl J Med 2005;352:1539–49.

47. Kittleson MM, Ye SQ, Irizarry RA, et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. Circulation 2004;110:3444–51.

48. Kittleson MM, Minhas KM, Irizarry RA, et al. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. Physiol Genomics 2005;21:299–307.

49. Boguski MS, McIntosh MW. Biomedical informatics for proteomics. Nature 2003;422:233–7.

50. Haga SB, Khoury MJ, Burke W. Genomic profiling to promote a healthy lifestyle: not ready for prime time. Nat Genet 2003;34:347.

51. Ibrahim SM, Gold R. Genomics, proteomics, metabolomics: what is in a word for multiple sclerosis? Curr Opin Neurol 2005;18:231–5.

52. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. Nature 2000;405:827–36.

53. Nicholson JK, Wilson ID. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. Nat Rev Drug Discov 2003;2:668–76.

54. Tyers M, Mann M. From genomics to proteomics. Nature 2003;422:193.

55. The Swedish National Biobank Program. Available at: http://www.biobanks.se. Accessed February 22, 2006.

56. UK Biobank. Available at: http://www.ukbiobank.ac.uk. Accessed February 22, 2006.

57. Norwegian Institute of Public Health. Biohealth Norway. Available at: http://www.fhi.no/eway/. Accessed February 22, 2006.

58. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science 2002;298:2345–9.

59. Kiss-Toth E, Qwarnstrom EE, Dower SK. Hunting for genes by functional screens. Cytokine Growth Factor Rev 2004;15:97–102.

60. Mousses S, Caplen NJ, Cornelison R, et al. RNAi microarray analysis in cultured mammalian cells. Genome Res 2003;13:2341–7.

61. Vanhecke D, Janitz M. Functional genomics using high-throughput RNA interference. Drug Discov Today 2005;10:205–12.

62. Wang S, Sim TB, Kim YS, Chang YT. Tools for target identification and validation. Curr Opin Chem Biol 2004;8:371–7.

63. Personalized Medicine Coalition. Science and Public Policy. Available at: http://www.personalizedmedicinecoalition.org. Accessed February 22, 2006.

64. Clayton EW. Ethical, legal, and social implications of genomic medicine. N Engl J Med 2003;349:562–9.

65. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.

66. Dave SS, Wright G, Tan B, et al. Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. N Engl J Med 2004;351:2159–69.

67. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002;346:1937–47.

68. Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med 2003;198:851–62.

69. Maris JM, Hii G, Gelfand CA, et al. Region-specific detection of neuroblastoma loss of heterozygosity at multiple loci simultaneously using a SNP-based tag-array platform. Genome Res 2005;15:1168–76.

70. Kozal M, Chiarella J. Development of a DNA microarray to detect HCV and HIV drug resistance. Antiviral Therapy 2004;9:S136.

71. Wong CW, Albert TJ, Vega VB, et al. Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. Genome Res 2004;14:398–405.

72. Chen L, Borozan I, Feld J, et al. Hepatic gene expression discriminates responders and nonresponders in treatment of chronic hepatitis C viral infection. Gastroenterology 2005;128:1437–44.