EDITORIAL COMMENT

Microarrays Coming of Age in Cardiovascular Medicine

Standards, Predictions, and Biology*

Geoffrey S. Ginsburg, MD, PhD, FACC,†‡
David Seo, MD,†‡ Camille Frazier, MD‡
Durham, North Carolina

High-density microarray, or so-called “gene chip,” experiments, wherein the levels of expression of the ~25,000 human genes is measured using RNA isolated from a tissue or cell, is now commonplace and uniquely adaptable to a broad array of medical and biologic questions. For example, the application of microarray analyses to cancer has provided phenotypic detail never before appreciated using traditional methods of analysis by illuminating patterns of gene expression or “gene expression signatures.” These signatures have in turn provided novel insights into molecular mechanisms of disease and laid the groundwork for the development of predictors for future health risks. In 1999, in one of the first demonstrations of the use of gene expression microarrays to dissect the molecular architecture of leukemia, investigators at MIT identified a signature of 50 genes out of 6,817 surveyed that could distinguish between acute myeloid leukemia and acute lymphoblastic leukemia (1). Since then, unique gene expression profiles have classified hematologic malignancies (2,3) as well as solid tumors (4) and have followed suit, and there are now some tantalizing examples of the benefits of this technology in cardiovascular medicine that illustrate the opportunity to uncover subtle distinctions in biological states. Such distinctions may be important for understanding the complexity of disease and developing novel diagnostics and therapeutics (9,10).

Cardiomyopathy is often a diagnosis in search of an etiology. Histologic classification is frequently nonspecific, but if the etiology could be determined, important prognostic information and therapeutic options might be more clearly defined (11). Access to myocardial tissue from endomyocardial biopsy and tissue cores at the time of left ventricular assist device placement or during transplantation, has provided an opportunity heretofore never available to subclassify and categorize this constellation of diseases molecularly using genome-wide expression analysis. Several studies have examined the use of microarray technologies to classify failing versus nonfailing hearts (12,13), dilated versus ischemic cardiomyopathy (14), and dilated versus hypertrophic cardiomyopathy (15). One notable study compared myocardial tissue from patients with end-stage disease with that from newly diagnosed cardiomyopathy to develop a classifier that correctly predicted the etiology of disease in the tissue with about 90% accuracy (14). In most of these studies, however, the numbers of patients were small and there were significant differences in methods for data analysis, in the regions of the myocardium sampled, and in the handling of the tissues at the time of harvest. Nonetheless, large lists of deregulated genes were found, and various pathways involving deregulated genes were found, and various pathways involving sarcomeric and cytoskeletal proteins, transcriptional regulation of cell signaling, mitochondrial metabolism, and inflammation appeared to be prominent. This leaves open the question as to which of these genes truly reflect the biology of the underlying myopathy and which reflect differences in clinical variables such as disease duration, age, gender, ethnicity, and medications at the time of tissue harvest of biopsy. Or, one might ask, do these changes in gene expression result from systematic experimental factors such as the choice of microarray platform, the methods of tissue handling, or the mechanical or oxidative damage sustained during sample collection, processing, and/or storage? Finally, only one of these studies verified or validated the results in a second unrelated population to assess for type I error or model overfitting. Therefore, the exact meaning of the gene expression signatures obtained in most of the studies is unclear.

In many cases investigators fail to recognize the extreme complexity of the disease tissue under study as well as the complexity of the microarray experiments themselves. These issues have led to the development of standards for microarray experimental information (16). Moreover, they emphasize the point that a research design that accounts for potential systematic errors is the most important component of gene expression-based investigations. Given the likelihood that gene expression analyses will increase in cardiovascular medicine, it is imperative that the data represent the biology of the disease and not noise or artifacts induced by experimental design. “Expression noise” (i.e., the inherent gene expression variation that does not correlate with the biology being studied and is introduced by or during the experimental procedure) has recently been the subject of debate and discussion among avid users of gene expression microarrays (17). Expression noise can obscure informative patterns of gene expression (resulting in false negative findings), but because most computational approaches to microarray analysis are focused on finding true associations, false positive findings are usually of greater concern, even in the presence of profound noise. Indeed, as the experimental protocols used to generate RNA from myocardial or vascu-
lar tissues for arrays may also be significantly confounded by ischemia, hypoxia, and hemodynamic stressors introduced during the tissue procurement and isolation, how can we be sure that the resulting data reflect the ischemic response in vivo versus the ischemic response ex vivo?

In this issue of the Journal, Barth et al. (18) have addressed the issue of confounders in microarray experiments probing end-stage cardiomyopathy. In addition they have explored the biologic significance of the genes identified by array experiments using gene ontologies to describe putative mechanisms underlying the different clinically classified forms of heart failure. Using 2 datasets from 2 different gene expression analysis platforms (cDNA spotted arrays and oligonucleotide arrays) in their own laboratory, Barth et al. (18) accessed a large number of samples from nonfailing and dilated cardiomyopathy and purposefully selected different regions of the ventricle (septum and left ventricular subendocardium) to keep their data as discrete and clean as possible. They also analyzed datasets from 2 additional laboratories available through the Gene Expression Omnibus microarray database (available through the National Center for Biotechnology Information). Thus, a large number of samples was available that allowed them to identify transcriptional signatures that could distinguish between dilated cardiomyopathy, ischemic cardiomyopathy, and nonfailing hearts. They were able to develop training sets of classifiers that had >90% accuracy in 3 of the 4 datasets. They derived a 27-gene classifier from one of these and used it to categorize the etiology of cardiomyopathy in each of the 3 remaining sets, for which also >90% accuracy was achieved. Notably, this signature contains brain natriuretic peptide (BNP) and BNP-related genes, sarcomeric structural proteins, and genes involved in the cell cycle, cellular proliferation, and apoptosis, consistent with published reports of biologic markers and molecular pathways known to be deregulated in cardiomyopathy. However, individual BNP-related genes were unable to classify the etiology of the myopathy in the absence of the complete set of genes, thereby illustrating the power and utility of this approach. In addition, a robust gene signature of immune response was seen across all the datasets that deserves further exploration. But because the hearts under study were end-stage, whether these mechanisms or pathways are relevant to the earlier stages of disease and its progression remains unclear. And although these observations are not necessarily profound, they deserve recognition for the meticulous attention to detail that went into the authors’ experimental design, as illustrated by the reproducibility of their work in the multiple datasets from different sources. This approach should serve as a guidepost for subsequent gene expression-based studies in cardiovascular disease.

So where do we go from here? As Barth et al. (18) point out, there is a need to explore tissues other than the myocardium for potential diagnostic precision in heart failure for all the reasons outlined in the preceding as well as for practical issues of tissue accessibility. Gene expression profiling from blood RNA may hold promise in this regard, because it has already been demonstrated that individual proinflammatory cytokine RNA transcripts from blood monocytes can discriminate between different forms of cardiomyopathy (19). Moreover, microarray analyses of RNA from blood have led to a noninvasive test that can classify heart transplant patients into low- and high-risk categories for cardiac rejection before clinical symptoms or biopsy evidence appear (20). At the end of the day, whether the RNA is from diseased tissue or from a blood surrogate, if the right standards are in place, gene expression profiling holds tremendous promise for classifying clinical phenotypes, developing prognostic predictors and, most importantly, providing novel unbiased insights into the mechanisms underlying heart disease. Barth et al. (18) have made an excellent start and set a standard for others to follow in cardiovascular genomic medicine.

Reprint requests and correspondence: Dr. Geoffrey S. Ginsburg, Director, Center for Genomic Medicine, Duke Institute for Genome Sciences & Policy, Box 3382, 101 Science Drive, Durham, North Carolina 27708. E-mail: geoffrey.ginsburg@duke.edu.

REFERENCES


