D

evelopments in the fields of genomics, transcriptomics, and proteomics have modified the definition of complex diseases in recent years. To have increasingly precise protein, genomic, and gene expression signatures would further define disease, advance treatment, and enhance the field of personalized medicine. In the area of cardiovascular disease, these specific areas have not followed parallel courses; a fact primarily driven by technology. Although proteomics holds the promise of large-scale identification of individual proteins from biological samples, the current technology is hampered by significant limitations, including the cost of truly high-throughput analysis and difficulty in the identification of low-abundant proteins. At the other end of the spectrum, human genome-wide association studies have analyzed 500,000 single-nucleotide polymorphisms in the human genome in large numbers of subjects and genetically redefined many common diseases and risk factors. Genome-wide association studies have also enabled investigation of the genetic contribution to specific heritable phenotypes and biomarkers.

Large gene expression profiling studies have not yet been completed in epidemiological cohorts focused on risk factors for atherothrombotic disease but have been reported in large oncological trials. These trials have been shown to predict oncological prognosis and classify precancerous disease states. Expression patterns from peripheral blood cells demonstrate B-cell differentiation that predicts specific stages of lymphoma. Although there are more limited data in vascular disease, gene expression from leukocytes in patients with sickle cell disease is consistent with increased oxidation and inflammation. Aortic samples have been analyzed for gene expression from patients with abdominal aortic aneurysm and arterial occlusive disease. Gene expression profiling has also been shown to predict cardiomyopathy etiology in heart failure. The major limitations of previous cardiovascular gene expression studies are their small size and the inherent problem of obtaining the tissue needed for analysis.

In this issue of Circulation: Cardiovascular Genomics, Wingrove and colleagues present data from patients who underwent coronary catheterization and compare a whole-genome microarray from peripheral blood mononuclear cells in case patients (n=27) and control subjects (n=14) without angiographically significant coronary artery disease. Using microarray, they report that >500 genes had more than a 1.3-fold change, with the vast majority being upregulated. Using this information and selecting additional genes reported in the literature, they confirmed changes in gene expression in 2 independent cohorts by reverse transcription–polymerase chain reaction. After multivariable analysis that included some risk factors and a few medications, 14 genes remained significantly associated with cardiovascular disease in 1 cohort and 11 in the second set.

In this study, specific genes were reported as upregulated in patients with angiographically defined significant coronary artery disease in all 3 patient groups. Several genes were confirmed by reverse transcription–polymerase chain reaction and may be related to cardiovascular disease. These included 5-lipoxygenase, which generates leukotrienes and has been associated with atherosclerosis, and, conversely, glutathione-S-transferase, which catalyzes the conversion of 5-lipoxygenase products. The clinical relevance of these observations remains to be defined.

Observations such as those reported by Wingrove and colleagues in this issue of Circulation: Cardiovascular Genetics suggest promise for an evolving technology; however, as illustrated by this study, there are many limitations that need to be overcome and issues that must be examined to optimize the utility and reliability of the transcriptomic information. The first is the general concern about the lack of hypothesis-driven experimentation. However, in the present study, genes were chosen for the confirmation sets with an eye toward biological plausibility. In addition, microarray studies by nature are not hypothesis driven. This leads to the second commonly cited concern with high-throughput methodologies, the problem with multiple testing and the false-discovery rate. As with many of the smaller single-nucleotide polymorphism studies, one must be concerned that the information reported in these early gene expression studies may not be confirmed in much larger data sets that will be reported in the near future.

There are also crucial technical considerations that are often ignored but may drive the validity of large gene expression studies in the future. Some stem from the distinct differences between genomic and transcriptomics data. With genomic data, a specific variant is present or not; however, with transcriptomic data, levels will also vary between individuals and in different disease states. There is currently
no consensus on how those data should be expressed, particularly in larger sets. Unresolved is the issue of whether the use of complementary DNA arrays versus quantitative reverse transcription–polymerase chain reaction provides more reliable determination of messenger RNA levels. Arrays are often normalized to commercially available “normals,” which requires a substantial leap in faith as to what one is controlling for. Quantitative reverse transcription–polymerase chain reaction relies on housekeeping genes, but these genes themselves may be influenced by disease states. A limitation of the present study is that the housekeeping genes chosen (RPL28 and Pro1853) were not validated; only 2 were used, and they have not been shown to be stable in patients with cardiovascular disease.

As discussed, studies that use blood to study gene expression should be larger and more powerful in the future; however, the source of RNA may also drive the results. Whole-blood isolation is simpler than leukocyte isolation and thus is widely used. The assumption of many has been that any 1 source of RNA from the blood is equivalent to any other to detect clinically meaningful differences. This is likely not a valid assumption. It has been shown that whole-blood methods do not provide the same data as isolated leukocytes. Whole-blood sources are significantly less reliable for detecting changes in gene expression in response to stimulation (only one third the number of genes associated with change were detected compared with the number detected with isolated leukocytes), and when whole-blood methods are used, the most predominant messenger RNAs present are residual messenger RNAs associated with erythrocytes or reticulocytes. When the 2 methods are compared, the primary determinant of gene expression is the source of RNA, not the disease state. In addition, a recent large epidemiological study that used isolated peripheral blood lymphocytes demonstrated that common gene variants influence transcript levels and that many are heritable. The source of RNA from the study presented in the current issue of the journal is mixed. In 1 confirmation set, isolated leukocytes were used, and in the second, whole blood was used. Although 11 of 14 genes were found to have similar increases in expression, this may have been because of the specific limited genes chosen and the small size of the data sets.

It has become clear over the past few years that information obtained from gene expression data is vital to discovering new mechanisms for disease and has clearly aided diagnosis and treatment in the field of oncology. Although less is known in relation to cardiovascular disease, its inherent complexity makes the vast amount of data available from transcriptomics methods both exciting and daunting with regard to the goals of diagnosis and treatment. Blood samples are readily available from patients and hold the promise of utility in the understanding, treatment, and diagnosis of vascular and thrombotic diseases; however, much remains to be clarified merely to effectively utilize these methods. It is crucial that our knowledge base not be diluted by underpowered studies, lack of diligence in addressing technical and scientific details, and failure to use appropriate statistical analysis. However, if these issues are addressed, transcriptomic data can be linked with phenotypic and genomic information to provide invaluable biological information that will address the biological complexity of cardiovascular disease at a new and intricate level.

Sources of Funding
Dr Freedman’s research is supported by the National Institutes of Health/National Heart, Lung, and Blood Institute.

Disclosures
None.

References
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661–678.


**Key Words:** Editorials ■ blood cells ■ cardiovascular diseases ■ genes ■ genetics
Blood and Cardiovascular Disease: The Promise and Limitations of Gene Expression Analysis
Kahraman Tanriverdi and Jane E. Freedman

doi: 10.1161/CIRCGENETICS.108.814186

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circgenetics.ahajournals.org/content/1/1/7

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org/subscriptions/