Family studies throughout the 1970s and 1980s documented the role of shared genetic factors in the familial aggregation of cardiovascular disease and its risk factors, including hypertension. These familial aggregation studies, however, do not identify and characterize the role of particular genes. Identification of the genes contributing to interindividual variation in disease risk may facilitate early identification of patients who are at elevated risk of cardiovascular disease before the onset of any clinical symptoms, development of more efficacious treatments by exploiting previously unidentified metabolic and physiological pathways, and the tailoring of particular treatments to patients who are most likely to respond on the basis of their genetic constitution.

Cardiovascular disease risk and risk factor levels are controlled by complex interactions among numerous metabolic and physiological systems, as well as demographic and lifestyle factors. Because so many systems are involved, variation in a large number of genes can potentially influence interindividual variation in disease risk, and the impact of any one gene is likely to be small to moderate in size. Before the current revolution in genomic analyses, studies identifying genes contributing to cardiovascular disease risk were of 2 basic types: studies of rare inborn errors of metabolism and association studies of a priori biologic candidate genes. The former have proved very useful for the identification of novel pathways, but the frequencies of these conditions are very rare, so their contribution to the prevalence of disease in the general population is minimal. The latter have proven useful in a few cases, such as the apolipoprotein E polymorphism with plasma cholesterol levels and risk of myocardial infarction, but have been plagued by lack of consistency. Modern genomic analyses have provided 2 additional pathways to identify genes that may be contributing to disease risk: gene expression profiles and genome-wide linkage analyses. Nothing further will be presented in this editorial on gene expression arrays except to state that they represent a complementary approach and that differentially expressed genes should be followed up for their role in interindividual variation in disease risk. Genome-wide linkage analyses, combined with the emerging sequence of the human genome, have become the most powerful tools for identifying genes contributing to the common chronic diseases, including cardiovascular disease.

In this issue of Circulation, Rice et al report the results of a genome-wide linkage analysis of blood pressure levels in 2 types of pedigrees: those selected for having obese family members and those not selected for any trait of interest. The study and its results are important for several reasons. First, they remind us that there is an enormous amount of activity using genome-wide linkage analyses to localize genes for cardiovascular disease risk factors, of which the work by Rice et al is a fine example. Fruits from these genome activities will provide the foundation for the identification of cardiovascular disease susceptibility loci for years to come. Second, the fact that there were no LOD (logarithm of odds) scores above 3 or 3.6, values that have been traditionally used as thresholds to indicate statistically significant linkage, underscoring that there are no genes with large effects influencing blood pressure levels in the general population. Instead, multiple significant or suggestive peaks were identified that signal the location of multiple genes each with small to moderate effects. It is this common variation with small to moderate effects that has the largest impact on risk in the general population. Any one study by itself typically does not include sufficient numbers of individuals to generate LOD scores of significant magnitude to localize genes with only moderate effects, thus necessitating the creation of large cooperative efforts (eg, www.bloodpressuregenetics.org). Third, many of the peaks identified by Rice et al had been reported in previous studies, while some novel peaks were also identified. This replication of findings is reassuring. Finally, and most importantly, it points out that genome-wide linkage analyses focus the search to find and identify the relevant susceptibility loci.

Genome-wide scans, such as that reported by Rice et al, yield chromosomal regions that show evidence for linkage with disease-related phenotypes. Typically, these linkage signals are detected with microsatellite polymorphisms that have no direct metabolic or physiological relationship with the phenotypes of interest. The challenge is to identify the genes contained in the broad chromosomal regions (20 to 30 cM of recombination or millions of DNA bases in length) that are responsible for the observed linkage results. The first step for following up genome scans is fine mapping to more precisely define the linked chromosomal region (the Figure). Fine mapping may begin with further linkage analyses using closely spaced markers but is limited by the low frequency of recombination events between any 2 closely spaced points in the genome. As a result, fine mapping must ultimately rely on simple biallelic single nucleotide polymorphisms (SNPs) and population association studies rather than meiotic recombination within families. If a known gene that is involved in relevant metabolic or physiological pathways has already been mapped close to a refined linkage peak, fine mapping...
will rapidly turn to detailed analyses of positional candidate genes. If no strong positional candidate genes are evident, fine mapping will rely on identification of SNPs throughout the linked interval for surveys by association/linkage analyses. This exercise may be greatly facilitated by public and private efforts to generate and make available a large numbers of SNPs distributed at high density throughout the genome. After SNPs that show associations in large population-based samples are identified, the investigator must identify which gene contains or is nearby the associated SNP or combination of SNPs. The process of identifying genes, although daunting only 1 or 2 years ago, is greatly facilitated by the burgeoning human DNA sequence data and sophisticated bioinformatics tools.

The second step is to identify the complete menu of DNA sequence variation within the identified genes through DNA resequencing or other methods (the Figure). Contrary to misunderstandings in the popular press and loose writing in the professional literature, a prototype sequence of the human genome does not exist for any individual. Rather, the sequence of the human genome is different among each and every one of us (except perhaps monozygotic twins), and this interindividual variation is a key contributor to differences in disease risk and their response to medical treatment among individuals, families, and populations. Like the gathering of the original sequence of the human genome, DNA resequencing will likely take place with existing technologies and has, in fact, already begun. Because of the possibility that variation within noncoding regions may influence the regulation of gene transcription and other functions, DNA resequencing should not be limited to only the protein-coding regions of the gene but rather should include 5' and 3' regulatory regions, as well as introns.

The third step is to identify those variable sites (or combinations of sites) identified by resequencing that are influencing the traits of interest. This is typically carried out by genotyping the variable sites in large population-based samples to identify sites that show association or linkage with the appropriate phenotypes. It is this third step that we are least prepared to competently accomplish because we lack an agreed-on conceptual and analytic framework to relate the large amount of DNA sequence variation that exists in modern human populations with interindividual phenotypic variation in a sample of moderate size, although promising developments are emerging. After polymorphic sites that show association or linkage with relevant phenotypes have been identified, the fourth step is to demonstrate functional effects on gene expression or protein function.

In general, these studies will require experimental cellular and animal models to directly measure the effects of naturally occurring DNA sequence variation. As is often the case, the apolipoprotein E polymorphism can serve as a paradigm for such functional studies. The functional laboratory, however, will not be limited to cellular and animal models; it must also include the population in which the ultimate impact on disease and the interactions with the environment will be elucidated.

The report by Rice et al1 and several other recent examples2,3 are early ripples in a tidal wave of studies to localize genes for cardiovascular disease and its risk factors. Sequencing of the human genome by both public and private efforts is rapidly progressing, as is the identification of sequence variation. It is therefore imperative that similar progress be made to place the onslaught of information in the context of improving the human condition. Both consumers and providers of health care should stop thinking of genetic risk profiles as distant science fiction and begin a constructive dialogue as to how they may benefit patient and public health. In addition, appropriate protections need to be implemented so that genetic risk profiles are not used to discriminate against individuals and obstruct access to appropriate health care. Because of its common prevalence in the population, availability of successful prevention and treatment regimens, and traditional leadership in contemporary biomedical research, the heart and vascular diseases are ideally suited to lead the effort for translating advances in genome research to the betterment of health care and human health.

References

Key Words: □ Editorials □ genetics □ genes □ mapping
Peeking Under the Peaks: Following Up Genome-Wide Linkage Analyses
Eric Boerwinkle, James E. Hixson and Craig L. Hanis

Circulation. 2000;102:1877-1878
doi: 10.1161/01.CIR.102.16.1877

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2000 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/102/16/1877

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation 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 is online at:
http://circ.ahajournals.org//subscriptions/