Candidate Gene Discovery in Cardiovascular Disease

Raimond L. Winslow, Zhong Gao

One of the most exciting new experimental technologies to emerge in recent years has been methods for obtaining genome-wide mRNA expression data using oligonucleotide and cDNA microarrays (for review, see Cook and Rosenzweig). Application of gene expression profiling was limited initially by cost, which in turn imposed severe constraints on the number of hybridizations that could be performed and thus the statistical significance of experimental results. This barrier has been reduced as array fabrication techniques have advanced and it is now common for studies to investigate differential gene expression at relatively large sample numbers, see for example, Singh et al. and Margulis et al. Improved statistical methods for inferring regulated genes as well as for grouping genes on the basis of related expression patterns have also been developed and refined.

Application of these methods to analysis of differential gene expression in cardiovascular disease typically reveals large numbers of genes with altered expression. These numbers can range from hundreds to thousands and they pose a new challenge—how best do we sift through these lists to identify those genes having relevance to disease development and progression? The most common approach is to perform a manual or semi-automated search of publicly available databases to evaluate gene function on the basis of annotation, and to then select as candidate genes those with function deemed most relevant to the disease process. While useful, such approaches are subjective and limited by the availability and accuracy of functional annotations. The study of Yagil et al presented in this issue of Circulation Research represents an alternative, experimentally-driven approach to the problem of identifying a restricted, biologically relevant set of candidate genes influencing the disease phenotype of hypertension.

In prior work, Yagil et al developed the Sabra hypertension prone (SBH/year) and resistant (SBN/year) rat model and used this animal model to perform QTL mapping of salt-induced hypertension susceptibility. They identified 2 QTLs (SS1a and SS1b) on chromosome 1 in males as well as an additional QTL (SS17) on chromosome 17 in females accounting for salt loading-induced variance of blood pressure. To provide functional evidence of the identified QTLs, Yagil et al constructed congenic strains of the salt-sensitivity QTLs SS1a and SS1b by replacing the QTLs in SBH/year with the chromosomal counterparts from SBN/year and found that blood pressure decreased, confirming that SS1a and SS1b are associated with salt susceptibility.

In this study, Yagil et al have gone a step beyond this approach to integrate gene expression analysis with prior information from QTL mapping to identify candidate genes responsible for salt-induced hypertension in rat. Gene expression profiling was performed using kidney tissue samples from four different populations consisting of male SBN/year and SBH/year rats fed with either a normal or salt-loaded diet. They first identified a total of 2470 genes that were differentially expressed across the 4 groups. Using 2-way analysis of variance and clustering methods, they then grouped these genes into 9 different clusters. Two of these clusters were associated with rat strain, 2 with the effect of salt-loading, and 2 with rat strain and salt-loading. The authors then focused on analysis of genes in the latter 2 clusters, identifying 7 genes that were cluster members, and with chromosomal location mapping within the boundaries of the SS1a and SS1b QTLs identified previously on chromosome 1. These 7 genes therefore have the properties that: (a) they are located within QTLs previously determined as being related to salt-sensitive hypertension; and (b) their expression is regulated by both rat strain and salt-loaded diet. Yagil et al propose these as novel candidate genes for hypertension.

This work provides a starting point for further development of data-driven methods for identifying small numbers of candidate genes likely to have biological relevance to a disease phenotype. Microarray technology is advancing rapidly and chip-based assays capable of genotyping many thousands of single oligonucleotide polymorphisms in large populations are now under development. Computational methods for performing genome-wide association analysis using data from these arrays, such as the exhaustive allelic transmission disequilibrium test of Lin et al, are being developed. When combined with data from expression profiling, such whole-genome association analyses have the potential to support high-throughput discovery of candidate coding and noncoding genomic regions regulating disease phenotype.

Yagil et al have submitted their hybridization data sets to the European Bioinformatics Institute, where they are available for download and subsequent analysis. Doing so is of vital importance to the research community, as it enables others to not only reproduce published research results, but to undertake additional data-mining studies to discover new biological knowledge. However, archiving of data should be viewed only as a necessary first step toward a more broad-
based approach to data dissemination, data analysis and quantitative computational modeling of cardiovascular function. Increasingly, this capability is being provided by “bio-grids”—distributed networks of computers and storage systems that can make both data, and data analysis methods and models available to geographically-dispersed user groups. A leading example of a biogrid is the Biomedical Informatics Research Network (BIRN). Supported by the National Institutes of Health, BIRN is currently a consortium of 19 universities and 26 research groups developing methods for the collection, dissemination and analysis of brain imagery relating to neurologic disorders. BIRN provides users access to data, computational resources and data analysis applications through web portals. With the large investment being made by the National Heart, Lung and Blood Institute into genomic, proteomic and clinical proteomics centers, it is time to follow the lead of the BIRN project and develop a cardiovascular research grid supporting the collection and dissemination of cardiovascular data, data analysis methods, and quantitative models of cardiovascular system function at biological levels of organization ranging from that of the gene to the transcriptome, proteome, cellular, tissue and organ level. Such a grid would provide a valuable resource to the national and international cardiovascular research community.

References


Key Words: microarrays gene expression
Candidate Gene Discovery in Cardiovascular Disease
Raimond L. Winslow and Zhong Gao

Circ Res. 2005;96:605-606
doi: 10.1161/01.RES.0000162161.71447.3c
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/96/6/605

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