The history of cardiovascular drug discovery is filled with delightful stories of serendipity. Had William Withering not first advanced his expertise in botany to court Helena Cook, a young woman who enjoyed painting flowers and later became his wife, he may not have been able to identify foxglove as a therapy for heart failure. Similarly, warfarin owes its discovery not to a search for therapeutic anticoagulants, but to studies of a hemorrhagic disease of cattle that devastated the American northern prairie farming community in the 1920s.

More recently, cardiovascular drugs have arisen less from chance and more from logical scientific approaches. The development of captopril, for example, depended on an understanding of the active site of the angiotensin-converting enzyme and logical chemical modifications of active site antagonists. However, advances in genomic technology over the past several years have transformed drug discovery. This article will provide an overview of one new approach to drug discovery that is often called Functional Genomics.

Although the progress of the Human Genome Project and privately-funded genome databases enabled the functional genomics approach, other powerful uses of the genome in drug discovery will be not be discussed in detail here. For example, single-base genomic differences between individuals called single nucleotide polymorphisms are common in candidate genes for cardiovascular diseases and can be used to identify genes associated with disease susceptibility in populations. Genes associated with disease susceptibility are obvious potential targets for novel therapies, although many of these genes may not be suitable targets for drugs. Single nucleotide polymorphisms may also be used to identify patients who may respond to specific therapies. Clinicians are aware that some patients benefit from a given drug while others do not; the science of pharmacogenetics aims to identify the genetic basis for this variation.

Thus, the functional genomics approach outlined here is only one benefit of human genomic information for finding new therapies. However, this process has important implications for drug discovery, particularly when combined with modern methods of identifying small-molecule drug candidates with robotic screening techniques. Our current pharmacological therapies are directed toward <500 human proteins, but various estimates of potential disease targets are in the thousands. Functional genomics strategies alone will not bring us to thousands of new therapies in the next decade, but they are already pointing us in new directions.

**Functional Genomics, Version 1.0**

**The Functional Genomics Process**

Although “functional genomics” has no single definition, it can be defined as a combined approach of biology, physiology, bioengineering, and all other available tools to assign physiological functions to the wealth of genomic sequence information. Functional genomics, therefore, is not simply a process toward novel drug discovery, but a general approach to assigning biological functions to genes with currently unknown roles in all organisms.

A traditional drug discovery approach begins with a pathophysiological or physiological process and then meticulously dissects the role of individual members of that process in the disease. This is typically a long and labor-intensive process; for example, we still do not know all of the downstream molecular pathways activated by angiotensin II. In contrast, in functional genomics, one begins with the genome and asks which genes or proteins may participate in the disease. This approach does not require a priori understanding of the molecular pathways causing the disease; in fact, the power of functional genomics is revealing roles of previously unsuspected pathways. The process can be described in 4 potential steps from the molecular target (gene or protein) to the lead (a chemical compound): target identification, target validation, lead identification, and lead optimization (Figure 1).

**Target Identification**

An initial step of functional genomics is filtering through the estimated 30 000 to 35 000 human genes to identify a much smaller set of genes that may participate in the disease. A target is a molecule (usually a protein but sometimes a nucleic acid) that may participate in a disease and also interact with a drug. An obvious approach to target identification is simply asking which targets are present or absent in diseased tissue relative to normal tissue. The technique of hybridizing RNA to DNA microarrays with known sequences has rapidly become the most common method of target identification. DNA microarrays typically have 10 000 unique sequences, and a single microarray experiment provides differential expression data on a significant proportion of Bioengineering and Environmental Health, Massachusetts Institute of Technology, Cambridge, Mass.

Correspondence to Richard T. Lee, MD, Partners Research Facility, Room 279, 65 Landsdown St, Cambridge, MA 02139.
(Circulation. 2001;104:1441-1446.)
© 2001 American Heart Association, Inc.

_Circulation_ is available at http://www.circulationaha.org
Figure 1. The functional genomics process, from identification of targets (genes or proteins) through optimization of leads (compounds) that inhibit or activate targets. Genomics and high-throughput screening methods allow rapid target identification and lead identification, respectively. In contrast, target validation and lead optimization are often laborious steps, and many compounds fail in the lead optimization phase. One strategy (chemical genomics) is to bypass the difficult target validation phase by screening targets for lead compounds very early and using the lead compounds to validate that the target participates in the disease.

Disease tissues and models

Target Validation: are these targets truly relevant to the disease?

Lead Identification: compound screening

Lead Optimization: search for optimal efficacy

GENOME

Target Identification: characterizing genes as candidates

Development Compound

Bypass biologic validation use chemical leads for validation

Comparison of genes expressed in diseased and normal human tissues is only one approach to target identification, and relying only on this approach can be misleading. Because tissues are multicellular, a more abundant cell type can mask genes overexpressed or underexpressed by an important cell type; for example, the heart is comprised primarily of cardiomyocytes by mass, but fibroblasts far outnumber cardiomyocytes in both normal and diseased myocardium. This may be one reason that differential gene expression experiments in hypertrophied hearts reveal numerous connective tissue genes that are most commonly associated with fibroblasts. Similarly, macrophages are relatively rare in the atheroma compared with smooth muscle cells, but macrophages may have crucial pharmacological targets for preventing atherosclerosis and improving plaque stability. Therefore, cellular models of disease (for example, activated versus resting endothelial cells or macrophages) are also commonly used for target identification.

Managing the deluge of genomic information has created the need for new approaches to data management and given rise to the field of bioinformatics. Bioinformatics is a cross-disciplinary approach of computer science, engineering, and molecular biology to data management and plays a key role in all genomic research. A single DNA microarray experiment can provide >100 pages of data in a traditional spreadsheet format; tools from bioinformatics provide logical analysis and integration of these large databases with preexisting data from prior human and animal experiments. Any single human gene target is likely to have no known function, but bioinformatic links with Drosophila or yeast gene databases can provide critical information. For example, the Down Syndrome Critical Region-1 gene encodes a protein that is expressed in cardiac hypertrophy, but the initial clue to its function as an endogenous inhibitor of the calcineurin-mediated cardiac hypertrophy pathway in mammals was its homology to a yeast protein known to inhibit calcineurin.

Target Validation

Target identification experiments can be performed very rapidly, typically in days or weeks, revealing targets associated with the disease. However, differential expression approaches can yield hundreds of potential targets, and the longer process of target validation requires distinguishing targets that are merely associated with a disease from targets that may truly participate in the pathophysiology of the disease. Target validation can employ all of the traditional tools for hypothesis testing in biology; however, the strategy must be designed not only for the disease model but also for practical considerations of the number of targets to validate. For example, if target identification yields only a handful of targets, proceeding immediately to gene deletion or transgenic overexpression experiments in mice is reasonable, particularly when mouse models of the human disease are already available.

Faster and less costly methods of target validation are desired to allow the target validation step to identify a few validated targets from among dozens or even hundreds of identified targets. Among 77,692 DNA sequences screened in C.C. Liew’s laboratory, 64 genes were potentially overexpressed in hypertrophied hearts compared with normal hearts; typical academic or industry laboratories do not have the resources to generate dozens of genetically engineered mouse lines for target validation. Furthermore, experiments with genetically engineered mice are powerful but not perfect methods of target validation. Deletion of a gene may be embryonically lethal, even though that gene participates in a disease postnatally. Deletion of a gene may also lead to unpredictable effects on other pathways, so that an apparent influence of a target gene on a disease may be due to a secondary effect of gene deletion.

Common methods of target validation when mouse gene deletion experiments are impractical include antisense oligonucleotides to inhibit expression of target genes, adenoviral gene transfer, and ribozymes, which are catalytically active RNA molecules that destroy target messenger RNA. With all of these methods, the goal is to mimic the effect of a potential drug with molecularly specific methods before the actual availability of that drug. In some circumstances, however, activation or overexpression rather than inhibition of a target may be beneficial, and finding a small molecule that causes overexpression or activation of the target may not be possible.

Lead Identification

Once sufficient target validation steps confirm that a gene and its protein are likely to be involved in a disease, chemical “leads” that modify that protein’s functions are identified. A
lead is an active new chemical entity, which by subsequent modification may be transformed into a clinically useful drug. Modern combinatorial chemistry and high-throughput screening can rapidly accelerate this process (Figure 2). Combinatorial chemistry, like functional genomics, has no clear definition, but can be considered the systematic, iterative chemical modification of “template” compounds leading to a larger number of diverse chemical structures. Using combinatorial chemistry, companies have established large chemical libraries or greatly expanded their libraries; chemical libraries typically have hundreds of thousands to millions of distinct compounds.

High-throughput screening takes advantage of the diversity of large chemical libraries to identify compounds that may interact with a target protein. Robotic technology, microfluidics, and miniaturized assays are integrated with automated compound storage and retrieval to screen the chemical library against the target. High-throughput screening technology is evolving quickly, and many companies screen >100,000 compounds per day.

How can this shotgun approach speed lead identification? Screening large chemical libraries against a new target protein does not require the type of structural insight that led to the development of captopril. Chemical leads from high-throughput screens can serve as additional templates for combinatorial chemistry or for insight into structural approaches to the target protein. Furthermore, a failed high-throughput screen can demonstrate at a relatively early stage that a target is a poor candidate for a small molecule approach. Many protein:protein interactions and protein:nucleic acid interactions occur over large surface areas that are difficult to block with small, orally-absorbed molecules.

Random screening of compounds in a chemical library against the endothelin receptor has recently led to a new class of endothelin antagonists. These chemical ligands served as leads that were modified to derive potent selective antagonists of the endothelin 1 type A and B receptors. Screening for inhibitors of the enzymatic activity of the cholesteryl ester transfer protein, which facilitates the transfer of cholesteryl esters from high density lipoprotein (HDL) to apolipoprotein B–containing lipoproteins, has yielded compounds that (after further structural modification) can raise HDL. Although high-throughput compound screening is a common approach in drug discovery, the ultimate benefit of random screening of very large chemical libraries generated through combinatorial chemistry remains unclear, and some investigators believe that more focused chemical libraries with compounds more likely to yield drugs are more useful.

Natural Products and Lead Identification

Enthusiasm for screening natural products has recently diminished, and many pharmaceutical companies no longer have natural product programs. Natural products typically contain dozens or even hundreds of bioactive chemicals, such that identifying, purifying, and synthesizing the active component can be a daunting task. However, 9 of the top 20 selling nonprotein drugs of 1999 were derived from natural product leads, including many currently used in cardiology (simvastatin, lovastatin, enalapril, pravastatin, atorvastatin, and cyclosporine).

Natural products will probably continue to be sources of chemical leads because only a small fraction of the world’s chemical biodiversity has been evaluated for biological activity, and the chemical novelty of known natural products far exceeds current synthetic chemistry. For example, tyrosine kinase receptors like the insulin receptor are challenging molecular targets for activation by small molecules, and the identification of a small molecule that could replace injected insulin has been a long-standing goal in drug discovery. Screening a library of 50,000 mixtures of synthetic and natural compounds led to the identification of a nonpeptide, orally absorbable molecule from a fungus found in the Congo; this molecule activates the insulin receptor selectively, and analogs of this compound reduce glucose when given orally to rodents with diabetes. This demonstrates the potential value of natural product chemical diversity.

Using Chemical Leads for Target Validation

One important potential of chemical libraries and high-throughput screening is the possibility of using chemical
leads themselves for target validation; this process is sometimes called “chemical genomics.” Compounds may be identified as capable of interacting with specific interesting proteins (such as kinases or receptors) even before that protein is known to be a target for a disease. After a set of potential targets has been identified, a potential chemical genomics target validation strategy directly tests the effects of these compounds in cell and animal models of the disease. Biological target validation methods such as gene transfer and genetically engineered mice are not only costly and slow, but one cannot directly predict the behavior of a compound that blocks or activates the target from these experiments. Chemical genomics provides the opportunity to move more quickly toward evaluating a potential chemical lead.

**Lead Optimization**

After the identification of a lead compound, further structural modifications of that lead can create a family of analogs. The initial lead is unlikely to have optimal efficacy and pharmacokinetics and minimal toxicity compared with some analogs, so dozens or even hundreds of analogs may be pursued. The compounds are then extensively examined in cell and animal studies to determine which is the best compound for clinical development (phase I human studies). Lead optimization is labor-intensive and risky but perhaps the most important process in drug development; many potential drugs fail at this stage.

**Functional Genomics, Version 2.0**

Biotechnology and pharmaceutical companies are currently exploiting the functional genomics process. However, even in these early stages, it is clear that major developments like proteomics and structural genomics will have dramatic impacts.

**Proteomics**

Biotechnology rapidly embraced functional genomics because sensitive and highly accurate differential gene expression methods became available. However, the therapeutic targets are not the genes themselves but the proteins encoded by the genes, and changes in gene expression do not always correspond to changes in protein quantity or activity. The concepts of identifying targets through differential expression of proteins are old, but recent technologies now allow rapid and accurate analyses of minute quantities of proteins and even their activities. Examining all of the proteins expressed in a given cell or tissue is one of the goals of “proteomics.” Proteomics, like many of these new terms, has no clear definition but can be loosely considered the characterization of all proteins. One aspect of the small number of genes in the human genome (≈35,000 compared with the anticipated 50,000 to 130,000) is a growing consensus that relatively more biological complexity lies within protein functions and interactions, and gene expression information alone cannot unravel that complexity.

---

**Figure 3. Proteomics.** A, Identification of differentially expressed proteins may be performed by 2D electrophoresis of samples and identification of differentially expressed proteins with image analysis. Proteins can then be rapidly sequenced with mass spectrometers, usually after digestion with an enzyme like trypsin to create smaller fragments. B, Proteins can be separated with affinity arrays. Chemistry-based chromatographic affinity arrays do not have the specificity of DNA microarrays, thus leading to the capture of a large number of proteins at a time. Proteins captured in this manner are analyzed by laser desorption time-of-flight mass spectrometry, allowing for discrimination of the various captured proteins by molecular weight. Biology-based arrays can use covalent coupling of a bait molecule (e.g., an antibody) to a preactivated spot surface, thus yielding a high specificity capture scheme. Figure provided by Ciphergen Biosystems, Inc, Fremont, Calif.
Proteomics techniques include protein arrays that allow high-throughput analysis of protein function and mass spectrometers that can identify proteins in only seconds to minutes (Figure 3). Proteomics will integrate into the functional genomics process, improving both the sensitivity and specificity of target identification. Proteomic analyses of human hearts with dilated cardiomyopathy have already identified >50 proteins that are differentially expressed compared with normal human hearts.

**Structural Genomics**

The 3D structure of a protein can provide critical insight into the function of that protein and improve the discovery process at all steps. Traditionally, because of the investment in time and money required to determine 3D protein structure, this has been reserved for late steps in evaluating a target or designing a lead compound. However, as genomic and proteomic technologies have advanced, so have innovations like multilength anomalous diffraction analysis, which have increased efficiency in structural biology. This has led to a worldwide movement that includes the National Institutes of Health, industries, and the Wellcome Trust in structural genomics—the systematic solution of most protein structures. Although there are probably >100,000 distinct human proteins (more than the ∼35,000 genes through alternative splicing and post-translational modification), there may only be 1000 to 6000 superfamilies of protein topologies, with ∼800 now known. If structural genomics programs can assign 3D structures to most of the genome, functions of genes and their proteins in a disease can be anticipated early in the drug discovery process.

**Cardiovascular Diseases and Functional Genomics**

Atherosclerosis is a multicular and temporally dynamic disease that has defied understanding on the basis of a single factor. In addition to molecular epidemiological approaches to identifying genes that determine atherosclerosis susceptibility, differential gene expression studies at different stages of human and animal atherosclerosis and in different regions of the vasculature will provide important insights. High-throughput screening and bioinformatics have already played a role in identifying a second angiotensin-converting enzyme gene (ACE2), whose protein product is not inhibited by lisinopril or captopril. ACE2 is expressed predominantly in the endothelium and in renal tubular epithelia, and it thus may be an important new cardiovascular target. Callow et al performed DNA microarray experiments in the livers of genetically engineered mice with decreased HDL plasma concentrations and found several novel genes involved in oxidative stress and sterol metabolism.

At the cellular level, differential gene expression studies have defined novel endothelial signaling pathways and chemotactic pathways that may participate in atherosclerosis. DNA microarrays have even been applied to gene expression in single endothelial cells, providing insight into cell-to-cell genomic variation. The identification of downstream molecules that regulate transforming growth factor- β signaling through genomic investigation of the effects of shear stress on endothelial cells illustrates both the power and limitations of functional genomics. Smad7, which inhibits the formation of transforming growth factor-β–dependent Smad2/Smad4 complexes, represents a potential new endothelial target for preventing atherosclerosis, but the subsequent steps between demonstrating function and developing a useful drug from this novel pathway remain formidable. Thus, although functional genomics reveals new pathways, it represents only the middle ground between basic biology and successful drug development.

Cardiac hypertrophy, remodeling, and the progression to failure represent premier arenas for genomic discovery. An examination of gene expression in failing versus nonfailing human hearts revealed 19 of 7000 genes that were differentially expressed, including genes involved in the cytoskeleton, protein turnover, and energetics. Friddle et al studied mice during both the induction and regression of cardiac hypertrophy; using microarrays, they identified 30 genes that were not previously associated with cardiac hypertrophy. Experimental myocardial infarction in animals has many features in common with human myocardial infarction. Stanton et al found >200 of 4000 genes were differentially expressed in microarray experiments after myocardial infarction in rats, and statistical analysis showed patterns of genes expressed in wound healing, cell signaling, and energetics. Sehl et al also used microarrays and rat myocardial infarction tissue to identify 14 novel mRNAs, although the majority of the 1075 genes that they studied did not change expression significantly after infarction. These examples of myocardial expression profiling demonstrate that myocardial remodeling draws on multiple interactive pathways that are largely undefined. Microarrays can also be used to evaluate the effects of a drug on remodeling. Jin et al randomized rats to receive captopril or no treatment 1 day after infarction. Thirty-seven genes were differentially expressed between rats with untreated infarction and controls, and captopril inhibited changes in 10 of these genes. This strategy has the potential to reveal novel targets in remodeling that are not currently affected by angiotensin-converting enzyme inhibitors, which are standard treatment after infarction.

**Waiting at the Bedside**

For the clinician, these biotechnologies may seem more like science fiction than science. In fact, functional genomics has not yet yielded a successful therapy. Why, then, all the excitement, and when will functional genomics deliver on its promise?

Historically, major technological innovations undergo a delay between the breakthrough and the practical delivery of the benefit to the public. This “technological lag” has typically been ∼2 decades in industrialized society and has been observed for the steam engine, electricity, the computer and, most recently, the Internet. It takes time to learn how to use technologies and apply them optimally and to turn them into reliable products. Although predicting the impact of the genome in general and functional genomics specifically leads us into uncharted waters, we may reasonably expect a lag of a decade or more before the maximum impact of functional
genomics is felt in clinical medicine, particularly given the necessity of careful clinical trials.

However, functional genomics actually began several years before the completion of the rough draft of the human genome in 2000. Tens of thousands of human DNA sequences were available in the late 1990s, and functional genomics drug discovery been underway for years. Hundreds of chemical leads identified through combinatorial chemistry, screening, and lead modification are in preclinical testing, and the first functional genomics compounds are now moving out of the laboratory and into human developmental testing. One can easily envision dozens of entirely new therapies over the next decade from this process reaching the bedside.

Acknowledgments

This work was supported by grants from the National Heart, Lung, and Blood Institute. I am grateful to the following scientists for their advice: John F. Thompson of Pfizer Global Research and Development; David S. Thorpe of Aventis Pharmaceuticals, Inc; Paul England of Aurora Biosciences; George Schreiner of Scios; John Babiak of Pharmacopeia; James N. Topper of Cor Therapeutics; and Steven Gygi and Chong-Chin Liew of Harvard Medical School.

References

Functional Genomics and Cardiovascular Drug Discovery
Richard T. Lee

Circulation. 2001;104:1441-1446
doi: 10.1161/hc3701.097176
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 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/104/12/1441

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/