Functional Genomics: It's All How You Read It

Philip Hieter and Mark Boguski

"Functional genomics" is a term that has taken root in the scientific community. What exactly do people mean when they refer to functional genomics? An informal poll of colleagues indicates that the term is widely used, but has many different interpretations. There is even some sentiment that the term is unnecessary and that it does nothing more than refer to biological research as a whole. Perusal of the several hundred functional genomics websites that have sprung up over the last 12 months clearly demonstrates that interpretations of the term are diverse and highlights the substantial degree of "hype" that is being used to promote the functional genomics approach, with little data to support it. Nevertheless, the concept of functional genomics has arrived and it is stimulating the creation of new ideas and approaches to understanding biological mechanisms in the context of knowledge of whole genome structure.

To fully understand functional genomics, we must examine its roots. The term "genome" itself is more than 75 years old and refers to an organism's complete set of genes and chromosomes. The term "genomics" was coined rather recently (in 1986) by Thomas Roderick to describe the scientific discipline of mapping, sequencing, and analyzing genomes and to provide a name for the new journal *Genomics* (*1*). The term has become universally accepted over the past decade. Genomics is now undergoing, however, a transition or expansion from the mapping and sequencing of genomes (the original stated goals of the Human Genome Project) to an emphasis on genome function. To reflect this shift, genome analysis may now be divided into "structural genomics" and "functional genomics." Structural genomics represents an initial phase of genome analysis and has a clear end point—the construction of high-resolution genetic, physical, and transcript maps of an organism. The ultimate physical map of an organism is its complete DNA sequence.

Functional genomics represents a new phase of genome analysis. It provides a fertile ground for (and will require) creative thinking in developing innovative technologies that make use of the vast resource of structural genomics information. Specifically, functional genomics refers to the development and application of global (genome-wide or system-wide) experimental approaches to assess gene function by making use of the information and reagents provided by structural genomics. It is characterized by high throughput or large-scale experimental methodologies combined with statistical and computational analysis of the results. The fundamental strategy in a functional genomics approach is to expand the scope of biological investigation from studying single genes or proteins to studying all genes or proteins at once in a systematic fashion. Computational biology will perform a critical and expanding role in this area: whereas structural genomics has been characterized by data management, functional genomics will be characterized by mining the data sets for particularly valuable information. Functional genomics promises to rapidly narrow the gap between sequence and function and to yield new insights into the behavior of biological systems.

Several recent studies fall under the operational definition of functional genomics. The recent completion (*2*) of the genome sequence of the budding yeast *Saccharomyces cerevisiae* (in other words, completion of the structural genomics phase) has provided the raw material to begin exploring the potential power of functional genomics approaches. An international consortium of yeast biologists is systematically constructing a comprehensive set of yeast strains, each of which will be deleted for one of the roughly 6000 predicted genes (*3*). Individual yeast open reading frames (ORFs) are being systematically replaced by oligonucleotide "bar codes," which can be used in a polymerase chain reaction (PCR) or DNA microarray assay for revealing those strains that survive under particular conditions (*4*). This reference collection will be made publicly available as soon as it is finished and will provide yeast researchers specializing in the study of a particular cellular process or class of genes the opportunity to devise assays or genetic screens utilizing the strain set. Three recently devised methods for obtaining genome-wide mRNA expression data, oligonucleotide "chips" (*5*), SAGE (*6*) and DNA microarrays (*7*), are particularly powerful in the context of knowing the entire genome sequence (and thus all genes) (*8*). The report by De Risi *et al.* in this issue (*9*) provides a powerful example of the way in which the DNA microarray methodology can provide a global view of changes in gene expression patterns in response to physiological shifts or manipulation of transcriptional regulators. The SAGE method, in the context of nearly comprehensive expressed sequence tag (EST) data, has also been elegantly applied to analysis of genes differentially expressed in human cancer (*10*).

In addition, knowledge of the yeast genome sequence has made feasible the systematic analysis of protein-protein interactions for all 6000 yeast proteins by means of the two-hybrid method (*11*). Analysis of all 18 million pair-wise combinations is under way. Furthermore, partial protein sequences from high-resolution, two-dimensional gels and electrospray mass spectrometry of protein complexes can be used to unambiguously assign peptides to specific gene sequences in the context of the whole genome sequence (*12, 13*).

As Peter Goodfellow has said (*14*), the central belief embedded in functional genomics is that the complete sequence of the genomes of many organisms, including humans, will change the way we do biology. Daniel Tosteson, dean of the Harvard Medical School, described the situation more explicitly: "In the past we have had functions in search of sequences. In the future, pathology and physiology will become 'functionators' for the sequences" (*15*). Traditional disciplines are already adopting a genome-scale viewpoint when it comes to approaching research problems. One example is the Cancer Genome Anatomy Project (CGAP), which seeks to foster infrastructure and new methodologies for cancer detection, diagnosis, prognosis, and therapy (*16*). Functional genomics will not only make traditional research approaches more productive and efficient, but will also supplement the detailed understanding of gene function provided by traditional approaches with a powerful new perspective on the holistic operation of biological systems. Functional genomics, however, will not replace the time-honored use of genetics, biochemistry, cell biology, and structural studies in gaining a detailed understanding of biological mechanisms. The extent to which any functional genomics approach actually defines the function of a particular protein (or set of proteins) will vary depending on the method and gene involved. In general, the functional information gained will provide a framework and a starting point for further analysis (*17, 18*), much like a primary genetic screen identifies can-

P. Hieter is at the Center for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, BC, V6T124 Canada. M. Boguski is at the National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA 20892.

didate genes that require extensive subsequent validation.

We are entering a phase in which we shall see more functional genomics data and hear less hype. Despite the unprecedented volume of data being generated from individual experiments, rigorous, reproducable design must be the watchword, so that emerging technologies can be fairly evaluated. The traditional format of scientific publication cannot reflect the scope and depth of data being produced. Summaries of results and conclusions in publications are certainly of interest, but are not very useful for subsequent analysis or utilization of the data by others and may not even be adequate for effective peer review. A key issue regarding the access to data from publicly funded, genome-scale, functional analyses must be addressed. A great legacy of the structural genomics era is the philosophy and practice of the public release of data that we hope will carry over to the functional genomics age. The timely submission of expression data, for example, in some standard format independent of specific technique, would lead to the most effective analysis and utilization of the results by the scientific community.

REFERENCES AND NOTES

1. V. McKusick, *Genomics*, in press.
2. A. Goffeau *et al.*, *Science* **274**, 546 (1996).
3. S. Oliver, *Trends Genet.* **12**, 241 (1996).
4. V. Smith, D. Botstein, P. O. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6479 (1995); D. D. Shoemaker *et al.*, *Nature Genet.* **14**, 450 (1996).
5. S. Fodor *et al.*, *Nature* **364**, 555 (1993).
6. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
7. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *ibid.*, p. 467.
8. V. E. Velculescu *et al.*, *Cell* **88**, 243 (1997).
9. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
10. L. Zhang *et al.*, *ibid.* **276**, 1268 (1997).
11. S. Fields and O. Song, *Nature* **340**, 245 (1989).
12. J. Yates, J. Eng, A. McCormack, D. Schieltz, *Anal. Chem.* **67**, 1426 (1995).
13. M. Wilm *et al.*, *Nature* **379**, 466 (1996).
14. P. Goodfellow, *Nature Genet.* **16**, 209 (1997).
15. Symposium on "Genomics and Gene Therapy: Meaning for the Future of Science and Medicine," Harvard Institute of Human Genetics, Cambridge, MA, 26 March 1997.
16. http://www.ncbi.n1m.nih.gov/ncicgap/
17. G. Miklos, L. Gabor, G. M. Rubin, *Cell* **86**, 521 (1996).
18. S. Fields, *Nature Genet.* **5**, 325 (1997).

19. We thank J. Roskams, S. Fields, and D. Bassett for critical reading of the manuscript and helpful input.

Predictive Genetic Testing: From Basic Research to Clinical Practice

Neil A. Holtzman,* Patricia D. Murphy,† Michael S. Watson, Patricia A. Barr‡

Announcements of discoveries of disease-related genes often suggest that tests to predict people at risk of future disease will soon be available (*1*). Few regulatory barriers stand in the way (Table 1A). If a commercial or academic clinical laboratory wants to offer a genetic test service (whereby it receives specimens, analyzes them, and reports results), it must register under the Clinical Laboratory Improvement Amendments (CLIA) and receive certification from the Health Care Financing Administration, which is the federal agency primarily responsible for the administration of CLIA (*2*). The process is not expensive and causes no delays in offering tests. For an organization that wants to market kits that independent laboratories, health care providers, or consumers can use to perform the test, the process is longer and more complex (*3*). In this case, the organization must first notify the Food and Drug Administration (FDA). If the test kit is not substantially equivalent to others already on the market, the FDA will require the organization to go through a premarket approval process during which it must collect data under an Institutional Review Board (IRB)–approved protocol to demonstrate that the test is clinically valid for the use intended by the manufacturer. Clinical validity includes determining test sensitivity and the predictive value of a positive test result (PVP) (*4*). Few genetic tests emerging from genome discoveries are being marketed as kits today, primarily because of the complexities of the assays and the interpretations. Under CLIA, there are no requirements for demonstrating clinical validity, but in the certification process, each laboratory may be called on to provide data on the analytic validity of its tests (*5*).

If the discovery of a disease-related gene provided sufficient information on a test's validity and other aspects of its effectiveness, this regulatory environment might be adequate. Seldom is this the case. First, the data collected on research subjects may not be representative of the findings in others at risk of the disease. Second, additional questions regarding the benefits and risks of testing, which are unlikely to be part of the original research, need to be considered. Under current regulations, the acquisition of sufficient data to warrant the transition of predictive genetic testing into health care cannot be ensured. More data must be collected in an investigative stage, during which results may be given to subjects (through their providers) if they have consented to participate and receive results. Before consenting, subjects must be informed of the questions the study is designed to answer and the potential risks and benefits.

The Task Force on Genetic Testing, was convened by the National Institutes of Health–U.S. Department of Energy (NIH-DOE) Working Group on Ethical, Legal, and Social Implications of Human Genome Research to review the state-of-the-art of genetic testing in the United States and to make recommendations when necessary to ensure (i) development of safe and effective genetic tests, (ii) their performance in laboratories of assured quality, (iii) their appropriate use by health care providers and consumers, and (iv) the continued delivery of tests for rare diseases. The Task Force, representing a wide array of stakeholders, has just issued its final report, concluding that, for the most part, genetic testing for Mendelian disorders in the United States has developed successfully, providing options for avoiding, preventing, and treating inherited disorders (*6*). However, problems arise in attaining each of the goals. Below we will consider the steps needed to establish the safety and effectiveness of a genetic test before it is incorporated into health care and the relevant Task Force recommendation (Table 1B). In its report, the Task Force makes recommendations on the last three goals as well. The focus of the Task Force on potential problems in no way is intended to detract from the benefits of genetic testing. Its overriding goal is to recommend policies that will reduce the likelihood of damaging effects so the benefits of testing can be fully realized.

N. A. Holtzman is with the Genetics and Public Policy Studies, Johns Hopkins Medical Institutions, 550 North Broadway, Suite 511, Baltimore, MD 21205, USA. P. D. Murphy is in the Department of Obstetrics and Gynecology, Albany Medical Center, Albany, NY 12208, USA. M. S. Watson is in the Departments of Pediatrics and Genetics, Washington University School of Medicine, St. Louis Children's Hospital, One Children's Place, St. Louis, MO 63110, USA. P. A. Barr is with Barr, Sternberg, and Moss, PC, 507 Main Street, Bennington, VT 05201, USA. N. A. Holtzman was the chair, M. S. Watson was the co-chair, and P. A. Barr and P. D. Murphy were members of the Task Force on Genetic Testing.

*To whom correspondence should be addressed.
†Consultant to Oncormed, a company that provides genetic tests.
‡Member of the board of the National Breast Cancer Coalition.