Genomic Medicine and the Future of Health Care

Chris Sander

Genomic technologies and computational advances are leading to an information revolution in biology and medicine. Simulations of molecular processes in cells and predictions of drug effects in humans will advance pharmaceutical research and speed up clinical trials. Computational prognostics and diagnostics that combine clinical data with genotyping and molecular profiling may soon cause fundamental changes in the practice of health care.

As we enter the 21st century, we are participants in a historic transition in science. A few years ago, only bits and pieces of the information stores of life, the genomes, were known in detail. After dramatic advances in molecular biology and technology, the first complete sequence of a human genome will soon be available. Information processing on computers and a new kind of biological information science are crucial in this transition. The impact on biology, medicine, and health care will be enormous.

Patient Scenario in the Future

Let us look at how an imaginary patient will benefit from this revolution. Shortly after a person is born, her genotype is recorded at her physician's office, and the information is transmitted to a secure database. Here, genotype means the presence or absence of specific variations in genes known to be relevant for assessing disease susceptibility and predicting responses to known drug types. Assisted by a decision support system, her physician may prescribe a personal immunization and screening schedule or recommend specific preventive measures. The genotyping information is complemented throughout her life by a screening program based on biomolecular profiling. At any point, screening may lead to recommendations about life-style or nutrition, or to detection of early stages of a disease. Refined diagnosis and choice of personalized therapy follow, which take into account her genotype and patient history and details of her molecular health profile.

Personalized therapy is supported by an expanded spectrum of drugs developed to target particular disease subtypes on a particular genetic background. Molecular profiling is used to monitor the progress of the disease,

Millennium Pharmaceuticals and Millennium Predictive Medicine, 640 Memorial Drive, Cambridge, MA 02139, USA, and MIT Center for Genome Research, One Kendall Square, Building 300, Cambridge, MA, 02139, USA.

and therapy may be adjusted flexibly. This scenario is most likely to apply to life-threatening diseases and to those for which disease disposition and response to therapy are known to vary considerably between individuals, such as cancer and heart and brain disease. Overall, the primary goal of personalized medicine should be to increase the quality of life first, and life-span second. But how will this kind of health care be achieved? And why is information the key to such functional genomics?

Information is the key because life at the molecular level can be understood as a process in which information is copied from generation to generation, expressed by producing biomolecules, protected by compartments and repair mechanisms, and adapted by a balanced process of mutation and selection. Decoding the genome—describing the connection between gene sequences and macroscopic life phenomena—is thus fundamentally a problem of describing and modeling biological information processes. In practice, this implies the generation, processing, and analysis of large data sets. The outcome will be a quantitative and predictive understanding of life processes, from molecular detail to macroscopic phenotype, that is a new predictive biology.

Toward Computing Gene Function

The technologies that underlie the generation of these information-rich data sets are extensions of molecular biology by genomics, robotics, and miniaturization. These include DNA chips (1, 2), mass spectrometry of proteins (3), and large-scale scans of protein-protein interactions (4–6). Applied to yeast cells, a single DNA chip experiment will now yield about 6000 data points per experiment, one for each gene. Soon, close to 120,000 data points per experiment will be collected from microarrays representing most human genes. Compared to the gel-based molecular biology of only a few years ago, which produced about 10 data points per experiment, data flow has increased by an impressive four orders of magnitude. How do computational tools cope with these data?

Fortunately, the volume of data as such is easily manageable. With current technology, a robot performing 40 DNA chip experiments per day, with 25,000 genes per chip (7), produces only 1 terabyte of raw image data per year, and this can be reduced to a few gigabytes by recording only a single expression value per gene per experiment. This volume of data is small compared to what is routinely processed and archived in science (in astronomy or particle physics, for example), commerce (in credit card transactions or Internet search engines), or intelligence agencies (satellite images). Even large-scale genotyping of human patients will not lead to unmanageable amounts of data in the next few years. Assuming that there are about 1000 clinically relevant genotypic markers per person, then genotyping one billion people would result in about 10 terabytes of data, an amount that would fit on a mere 1000 DVD optical disks.

The tough computational challenges resulting from large-scale genomic experiments lie in the specificity and complexity of the biological processes. How does one find the needle in the haystack: the gene(s) directly involved in disease or the single drug target that may lead to a cure? How does one perform computations involving biological function?

A set of expression profiling experiments in yeast (8) has been designed to reveal which genes are involved in cell cycle control and how their expression is regulated. Similarly, sets of marker genes permitting the classification of particular tumor cell lines have been sought by analyzing the gene expression patterns of a panel of cancer cell cultures (9, 10). In such investigations, the principal challenge is the interpretation of these patterns in terms of the underlying biological effect.

To perform any computational analysis of the biological function of a large number of genes, one needs to expand the concept of gene function. Each type of experiment leads to its own notion of gene function, from biochemistry ("protein phosphatase") to cell biology ("cell division control gene") and genetics ("radiation resistance protein"). Soon, the function of a gene can be described by its expression profile in a large number of controlled experiments. Comparative computation can then be performed on gene function in ways not previously possible. Answers to questions such as "which gene is most similar in function

to gene A?" or "which past experiment is most similar to experiment X with respect to involvement of a specific gene set Y?" become possible through definitions of appropriate similarity measures in gene and experiment space. Efforts to construct archival databases of gene expression experiments to facilitate such predictive computations are under way (*11*).

Toward E-Cell Simulation

Arguably the largest impact of genomic technologies on biological research will come from the emerging ability to simulate cells and organisms on the computer. The goal is to simulate the causal and temporal behavior of a cell as a network of genes and gene products and to simulate the behavior of the organism as a network of cells. Quantitative and predictive simulations have the potential of reducing or replacing experimental effort. Precedents from other areas of science and engineering abound; for example, car crash experiments have now largely been replaced by computer simulations that optimize materials and design for maximum safety. But biological simulations will be fundamentally different from those in physics and engineering. Knowledge of the historically evolved specificity of genetic information and the resulting individuality of proteins and functional RNAs is essential.

Work on full-cell simulations has started. For example, the "e-cell" project in Japan (*12*) reports the simulation of a minimal set of metabolic pathways in a cell that takes up glucose and excretes lactate. Other simulations are being attempted in this rapidly developing field (*13–15*). Early applications of e-cell models will probably come from simulations addressing questions such as "what are the qualitative consequences of inhibiting the function of gene X under conditions Y?" (*16*).
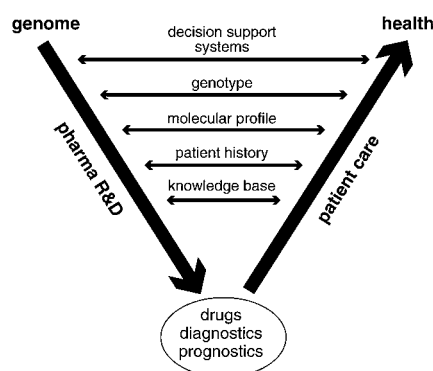

Fig. 1. Genomic and computational technologies have an increasing impact on pharmaceutical R&D (arrow at left). The same technologies will be adapted to directly serve patients (arrow at right), leading to personalized, information-driven medical care.

Toward the Perfect Drug Candidate

What changes will we see in the process of drug discovery and development? Currently, the failure rate in the transition of preclinical drug candidates to approved drugs is unacceptably high, with enormous attendant costs. Large savings would come from early detection of undesirable drug properties. The difficulty lies in the complexity and multiplicity of the desirable properties. Beyond the specific binding of the drug to its target, these include a compound's behavior in absorption and distribution in the body, the way the drug is metabolized and excreted, and the avoidance of negative side effects.

A combination of rich cellular data, genomic profiling, and computational prediction may provide a way out. For example, the effect of known toxic compounds can be assessed by measuring the genomic expression profile in cell cultures and accumulating a set of characteristic profiles as a background information base. New compounds can then be filtered out if their expression profile classifies them as potentially toxic. The advantage of such methods lies in the much lower cost of cell culture tests as compared to tests in animals and clinical trials. Extrapolations from laboratory measurements using databases and computational predictions are being attempted, for example, in drug absorption studies (*17*). Information from functional genomics experiments will be crucial for the predictive elimination of unpromising drug candidates.

Today's clinical trials are expensive and time-consuming. To accelerate the assessment of clinical outcomes using genomic technologies, a detailed and accurate link between molecular profiles and clinical outcomes is required. Patient progress can be assessed by detailed measurements of thousands of molecular indicators from bodily fluids or biopsies, such as RNA expression, protein expression, protein modification, or concentration of metabolites. Computational processing and reference to information and knowledge bases about organismic and disease processes would allow conclusions about the likely results of therapy to be reached much faster than with classical macroscopic indicators of clinical outcomes.

Imagine the benefit to the development of new therapies if drugs entering clinical trials are almost ensured to be well tolerated in the body and to have the desired effect. Or imagine relatively short clinical trials, confirmatory final tests to guarantee that drugs and diagnostics are safe and effective.

Toward Personalized Medicine

Genomics-based molecular profiling and related technologies may have a direct and early impact on the delivery of health care to patients long before clinical trials have been transformed and genomics-based drugs have come to market (Fig. 1). There are several reasons. First, the regulatory approval process for predictive and diagnostic techniques is shorter than that for drugs. Second, people are increasingly interested in information regarding their state of health, and such information can be made widely accessible by means of the Internet. Third, low-throughput genotyping for genetic markers (as for cystic fibrosis) and profiling for disease markers (such as prostate-specific antigen) are already in use. Applications of the new technologies to patient care are thus likely to be developed in parallel with pharmaceutical development.

These changes in health care practice are likely to trigger changes in socioeconomic relations. Strict regulations must ensure that genotypic information and molecular profiles are collected for medical purposes only and remain the exclusive property of the patient. For use in a knowledge base, genotypic and clinical information about patients will have to be made anonymous, using secure protocols. Certain routine examinations will perhaps no longer be done at the physician's office. The acquisition of medical expertise in software systems and knowledge bases may change the role of health care professionals in fundamental ways. There may also be dramatic shifts in the economics of health care, with details that are hard to predict.

Although it will take painfully long years for the wave of novel "genomic" drugs to come to market, it may not be long before patients feel concrete improvements in the quality of life—as soon as prognostic genotyping and diagnostic molecular profiling are used in routine medical practice.

References and Notes
1. P. O. Brown and D. Botstein, *Nature Genet.* **21**, 33 (1999).
2. M. Schena *et al.*, *Trends Biotechnol.* **16**, 301 (1998).
3. J. R. Yates, *Trends Genet.* **16**, 5 (2000).
4. A. R. Mendelsohn and R. Brent, *Science* **284**, 1948 (1999).
5. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
6. A. Abbot, *Nature* **402**, 219 (1999).
7. C. Muir and G. Kirk, personal communication.
8. P. T. Spellman, *Mol. Biol. Cell* **9**, 3273 (1998).
9. J. N. Weinstein *et al.*, *Nature Genet.* **23**, 81 (1999).
10. T. R. Golub *et al.*, *Science* **286**, 531 (1999).
11. A. Brazma, personal communication (see also www.ebi.ac.uk/arrayexpress).
12. M. Tomita *et al.*, *Bioinformatics* **15**, 72 (1999).
13. J. Schaff and L. M. Loew, *Pac. Symp. Biocomput.* **1999**, 228 (1999).
14. A. Arkin, J. Ross, H. H. McAdams, *Genetics* **149**, 1633 (1998).
15. R. F. Service, *Science* **284**, 80 (1999).
16. R. Brent, *Cell* **100**, 169 (2000).
17. D. E. Clark and S. D. Pickett, *Drug Discovery Today* **5**, 49 (2000).
18. The author is grateful for comments by J. Ahouse, R. Brent, M. Kauffman, G. Kirk, C. Muir, M. Pavia, and C. Reich.