

The Top Ten Advances of the Last Decade

BY ERIC JAKOBSSON, PhD

The Top Ten Challenges of the Next Decade

© PHOTOGRAPHER: KUTT NINEFUJ / AGENCY: DREAMSTIME.COM

The last ten years have seen huge leaps in biomedical computing. We now have new ways to integrate and understand vast quantities of data; the capacity for multi-scale biological modeling; and a vastly more networked world of researchers and their data.

The ten advances in biomedical computing listed here strike me as the most significant of the past decade. Others might compile a different list, but few would argue that these ten don't belong near the top.

I found it much harder to predict the hurdles and likely achievements of the coming decade, but I've given it a shot. The ten I've chosen are difficult to categorize. Some are true advances in science. Others are true advances in computing. And still others seem almost administrative—but are vital to the future of the field. Ten years from now, looking back, these might not be the most important developments of the decade, but I expect most of these will have been accomplished, or at least significantly advanced, by 2015.

What's exciting about this list is that while some of the advances could have been predicted or expected, others transformed medical care (eg., AIDS treatments) or the world (the advent of the web browser) in ways that could not have been anticipated. Let's hope the coming decade has a few of those in store as well.



Until this month, author Eric Jakobsson, PhD, was Director of the NIGMS Center for Bioinformatics and Computational Biology at the National Institutes of Health. He has now returned to his professorial position at the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

The Top Ten Advances

ADVANCES IN SCIENTIFIC DATA INTEGRATION:

1 SEQUENCE ALIGNMENT TOOLS. In 1997, the seminal paper was published that demonstrated the technology for aligning sequences of DNA's component parts even in situations where the evolutionary process has inserted or deleted strings of those nucleotides. This was the critical technological advance that permitted the universal search for genes in genomic sequences, and the search for and definition of motifs—i.e., segments of genes that are characteristic for catalytic function or regulation of the gene. A closely related advance was the development of Hidden Markov models, which enable a computer to infer the rules of nucleotide substitution from a

first draft comparison and refine the comparison process on the fly. Further extensions of these alignment tools proved to be critical in assembling the human genome from the fragments that are processed by sequencing machines. The “shotgun method” for genome reconstruction would have been impossible without these computational tools.

2 ENABLING SYSTEMS BIOMEDICINE. The first great high-throughput data challenge for cell and molecular biomedicine was the development of high-throughput sequencing. The current challenge is to move into the postgenomic era, and deal with the addition of massive gene expression and proteomic data in a manner that

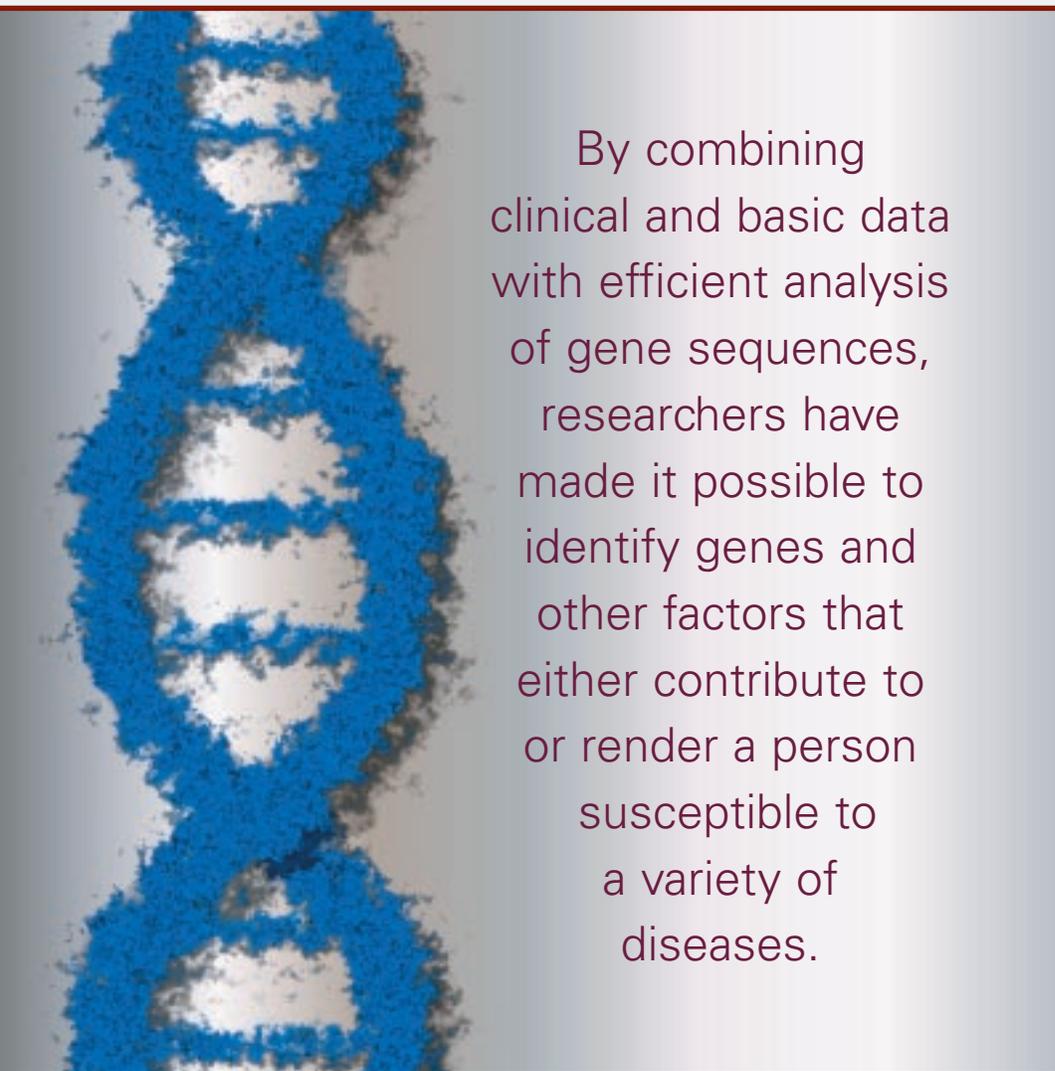
will permit the construction of comprehensive, interpretive and predictive models for networks, pathways, organelles, and whole cells and tissues. The foundations for meeting this challenge have been laid in the past decade in the form of software for analyzing gene expression and proteomic data, laboratory management systems for organizing these data into computable databases, software for inferring interaction networks from the contents of such databases, and software for modeling organelles and cells as either differential equations or linked stochastic processes.

3 IDENTIFICATION OF GENES FOR DISEASE SUSCEPTIBILITY. By combining clinical and basic data with efficient analysis of gene sequences, researchers have made it possible to identify genes and other factors that either contribute to or render a person susceptible to a variety of diseases. The result: improved efficacy of both diagnosis and treatment of diseases with a genetic component.

ADVANCES IN MULTI-SCALE BIOLOGICAL MODELING:

4 COMPUTATIONAL MODEL OF HIV INFECTION. Clinical researcher David Ho and mathematical/computational modeler Alan Perelson teamed up to use clinical and basic immunological data to understand the interaction between the HIV virus and the immune system during the apparently quiescent period between HIV infection and development of AIDS symptoms. The model, published in 1996, provided guidance in the development of multi-drug therapy that has led directly to the dramatic decline in AIDS mortality in Europe and the United States in the last five years.

5 TOMOGRAPHY. The ability to computationally reconstruct biological structures from various types of imaging—such as ultrasound and magnetic resonance—has dramatically enhanced both diagnosis and treatment for a variety of diseases by enabling detailed non-invasive visualization inside the body.



By combining clinical and basic data with efficient analysis of gene sequences, researchers have made it possible to identify genes and other factors that either contribute to or render a person susceptible to a variety of diseases.

© PHOTOGRAPHER: DIMITRIOS KESSARIS / AGENCY: DREAMSTIME.COM

6 COMPUTER-AIDED PROSTHETIC DESIGN. For a period of time culminating in the last decade, the design of prosthetics has been by computer-aided design systems. With limb prosthetics for example, designers routinely couple imaging of the residual limb with automated fabrication of the prosthesis. Mechanics and stability are based on computer modeling. This work is laying the foundation for a new generation of “smart” prosthetics that will modulate function based on electrical and/or chemical signals from the body.

7 MODELING ELECTRICAL BEHAVIOR OF NEURONS AND OTHER ELECTRICALLY EXCITABLE TISSUE. Within the past ten years it has become possible to accurately model the dynamic electrical behavior of nerve and other cells based on their structure, the distribution of ion channels and other transporters in their membranes, and the activity of relevant signaling systems. This and other analytical and modeling work at higher levels of organization lay the foundations for a deeper understanding of neural and other electrically excitable tissue (for example, cardiac muscle), for the integration of cellular neuroscience and cardiac science with higher level concepts and images, and for the application of cellular mechanisms of electrical excitability to engineered systems, such as artificial retinas to replace vision lost to macular degeneration or glaucoma.

ADVANCES IN THE NETWORKING OF SCIENCE:

8 DISSEMINATION OF BIOINFORMATICS TOOLS ON THE WEB. The hypertext Web browser is only 11 years old. It was introduced in 1993 in the form of NCSA Mosaic, a legacy that can still be seen if one pulls down the “Help” menu in Internet Explorer and clicks on “About Internet Explorer.” Within a few years, Web access to sequence analysis and other bioinformatics tools began to appear and become available to all researchers. This allows all



The hypertext Web browser is only 11 years old. Since its introduction, web access to sequence analysis and other bioinformatics tools has become invaluable to researchers.

researchers to augment experimental work in molecular and cellular biology with sequence analysis that provides insights into the particular mechanisms studied and the likely generality of those mechanisms among organisms and preparations other than those studied. These tools are also invaluable in hypothesis generation and designing future experiments.

9 TELEMEDICINE. High speed networks and high resolution image reconstruction now make it possible for specialists to share medical judgment re diagnosis and treatment remotely, providing in principle a many-fold expansion of the beneficial impact of specialized medical expertise.

10 ESTABLISHMENT OF COMPUTER NETWORKS FOR SURVEILLANCE OF DISEASE. We have not had a worldwide pandemic of respiratory infectious disease, killing tens of millions of people, since the 1918 flu pandemic. We may yet have one, but it has been made less likely and will be made less lethal by the establishment of world-wide computer networks for surveillance of infectious disease. These networks permit the identification of potential pandemics and effective early countermeasures such as quarantine and the development of vaccines. Beyond the issue of infectious disease, computerized surveillance has the ability to identify environmental factors influencing the spread of disease, and to perform multi-factoral analysis to assist in inferring causality.

The Top Ten Challenges

1 **IN SILICO SCREENING OF DRUG COMPOUNDS** would be a major achievement. The ability to predict the efficacy and side effects of lead compounds using computer modeling and simulation would also reduce the need for risky human testing and would save significant amounts of money and time that are now spent in the laboratory. The core of this improvement would be computation of affinities of lead compounds to targets and to effectors of potential side effects.

2 **PREDICTING FUNCTION FROM STRUCTURE OF COMPLEX MOLECULES** at an engineering level of precision is not now possible. Currently, molecular simulation and analysis methods capture the essence of the mechanism of biomolecular function, but do not predict that function with quantitative accuracy. Improved capability in this regard would lead to a pre-

cise understanding of the consequences of mutations and other biological variations, and the ability to design molecules for medical nanotechnology.

3 **PREDICTION OF PROTEIN STRUCTURE** to the point where all protein sequences have associated accurate structures would provide critically important information. At present there are many more known protein sequences than structures. By a combination of accelerated experimental structure determination and improved techniques for mining known structures to determine the rules for predicting unknown structures, it should be possible to fill in the gaps and assign a structure to every sequence. Success in this arena would advance the field of biomedicine in many and various ways.

4 **ACCURATE, EFFICIENT, AND COMPREHENSIVE DYNAMIC MODELS** of the

spread of infectious disease are needed to make sense of the extensive data that have been gathered on the spread of infectious disease and the consequences of various strategies of intervention. Such models will provide a basis for rational, informed, real-time decision-making in combating natural epidemics and bioterrorist attacks.

5 **INTELLIGENT SYSTEMS FOR MINING BIOMEDICAL LITERATURE** would provide better access to an abundance of information about the functioning of genes, gene products, and cells. At present there is no efficient and effective way to organize these data into computable databases from which accurate interpretive and predictive models can be constructed.

6 **COMPLETE ANNOTATION OF THE GENOMES** of selected model organisms so that we have either a known or putative function for all genes, would allow biomedicine to advance more rapidly. At present, so-called “complete” genomes are far from “complete.” We should select eukaryotic and prokaryotic model organisms for a focused attack on complete annotation, and use all experimental, bioinformatics, and data-mining tools on these organisms. A sequel to complete annotation should be complete elucidation of the metabolic, signaling, and homeostatic pathways and networks of these organisms.

In silico screening of drug compounds would be a major achievement.

7 **IMPROVED COMPUTERIZATION OF THE HEALTH-CARE DELIVERY SYSTEM** should be achieved in this decade. The relatively primitive information technology environment supporting the delivery of health care results in extra expense and avoidable error. There is a major need for a nationally interoperable



© PHOTOGRAPHER: ANDRZEJ TOKARSKI / AGENCY: DREAMSTIME.COM

ble system of medical records that will support transferable patient records, diagnosis and treatment based on integrating the patient record with relevant basic and clinical knowledge, and efficient patient monitoring. A logical consequence and extension of this computerization would be an information environment that would support the deployment of personalized medicine, where diagnosis and treatment would be optimally keyed to the patient's history and genotype.

8 MAKING SYSTEMS BIOLOGY A REALITY by integrating appropriate computational tools would provide a much-needed computational environment for information-based modeling of pathways, networks, cells, and tissues. At present many useful tools for systems biology have been created, but they are not integrated into computational environments that provide for automatic interaction of multiple programs and functionalities to address generally useful issues in biomedicine. The tools themselves also need improvement in their scope of applicability, computational efficiency, and ease of use.

9 TUNING BIOMEDICAL COMPUTING SOFTWARE TO COMPUTER HARDWARE will permit more efficient use of computing resources. Biomedicine uses substantial computing resources at all levels, from the desktop to high-end supercomputing centers. A large fraction of these resources are not efficiently used, as the hardware and software are not tuned to each other. For example, a computing system may have very high processing throughput but not a high bandwidth between the processor and massive data sets contained on hard drive. Thus, such a system is not well suited for a particular biomedical application that requires frequent and massive communication between the processor and the hard drive. Addressing this problem will allow research to advance more rapidly.

10 PROMOTING THE USE OF COMPUTATIONAL BIOLOGY TOOLS IN EDUCATION will help forestall a likely

Accurate, efficient, and comprehensive dynamic models of the spread of infectious disease will provide a basis for rational, informed, real-time decision-making in combating natural epidemics and bioterrorist attacks.



© PHOTOGRAPHER: MONIKA WISNIEWSKA / AGENCY: DREAMSTIME.COM

shortage of quantitatively competent researchers. The biomedical research enterprise is heading for a such a shortfall due to a convergence of two trends: increasing opportunity in other countries for international students who up to now have trained in the United States and made their careers here; and the relatively small numbers of American students with strong quantitative skills who are opting for careers in biomedical research.

In this context we recognize that the same developments that are making biomedical computing tools useful to experimental researchers also can make them the basis of compelling problem-solving educational environments for students. There is a strong need to adapt biomedical computing tools to education at all levels in order to capture their power to motivate youngsters to pursue biomedical research careers. □